Estimating Industry Conduct in Differentiated Products Markets*

The Evolution of Pricing Behavior in the RTE Cereal Industry

Christian Michel†
Stefan Weiergraeber‡

April 5, 2018

Abstract

We estimate the evolution of competition in the ready-to-eat (RTE) cereal industry. To separately identify detailed patterns of industry conduct from unobserved marginal cost shocks, we construct novel instruments that interact data on rival firms' promotional activities with measures of products' relative isolation in the characteristics space. These data are readily available for many differentiated products industries; therefore our empirical strategy has broad applicability. We find strong evidence for partial coordination between national cereal manufacturers in the beginning of our sample period with a 13.1 percent median margin over those implied by multi-product Nash pricing. Manufacturers' price coordination intensifies even more following a horizontal merger in 1993, increasing the median margin to 23.9 percent. Toward the end of our sample period, industry-wide margins reduce sharply to a level that is consistent with a change in conduct to multi-product Nash pricing among all firms.

Keywords: Markups, Market Power, Conduct Estimation, Differentiated Products Markets
JEL Classification: L11, L41, C51

*We thank Daniel Ackerberg, Steven Berry, Pierre Dubois, Juan Carlos Escanciano, Rosa Ferrer, Jeremy Fox, Amit Gandhi, David Genesove, Jean-François Houde, Kyoo-il Kim, Massimo Motta, Volker Nocke, Helena Perrone, Barbara Rossi, Philipp Schmidt-Dengler, Johannes Schneider, Alex Shcherbakov, Sandro Shelegia, André Stenzel, Yuya Takahashi, Frank Verboven, Matt Weinberg, as well as seminar and conference participants for invaluable comments. Michel gratefully acknowledges financial support from the Spanish Ministry of Economy and Competitiveness Grant ECO2016-76998.

†Department of Economics and Business, Universitat Pompeu Fabra. Email: christian.michel@upf.edu.
‡Department of Economics, Indiana University. Email: sweiergr@iu.edu.

1 Introduction

One of the central questions in industrial organization is to what extent firms exert market power. Product differentiation, as one source of market power, can lead to positive markups even if firms compete effectively with each other. In many industries, there are concerns that a low intensity of competition further contributes to high industry markups. Empirically disentangling legitimate from anti-competitive sources of market power is thus an important task. This task, however, is very difficult because neither the intensity of competition nor marginal cost, which is another price determinant, are usually observed in the data.

A key identification problem in empirical industry models is thus to distinguish whether firms charge high prices because of anti-competitive behavior or because of high unobserved cost shocks. To separate the two channels, one needs to find suitable instruments that are correlated with markups but not with underlying cost shocks. Recently, Berry and Haile (2014) have shown that it is in principle possible to empirically discriminate between different oligopoly models by exploiting variation in market conditions.[1] In practice, however, many of the instruments based on this type of variation tend to be weak. The few studies that have focused on estimating *industry conduct*, as a measure of an industry's competitive intensity, have used alternative identification strategies, such as exploiting plausibly exogenous industry shocks.[2] Such identification strategies can already lead to important insights. However, they often require the researcher to focus on estimating the conduct of only a subset of firms and time periods, or to assume that the structure of conduct is invariant across time and firms.

In most cases, it is not clear a priori that the conduct in an industry follows such a pattern. The level of conduct might not only deviate from competition but also differ substantially over time and across firms. Not accounting for this heterogeneity can lead to inconsistent estimates of markups and marginal costs. Allowing for more flexible conduct specifications is thus likely to lead to more accurate predictions and more effective policy recommendations.

In this paper, we estimate detailed patterns of industry conduct that account for changes over time and heterogeneity across firms in the US RTE cereal industry. To do so, we employ a structural differentiated products demand model, and a flexible conduct parameter framework on the supply side. To separately identify industry conduct and manufacturers' marginal costs, we propose novel instruments that exploit products' relative proximity in

[1] Examples for this type of variation are the number of firms, the set of competing products, or functions of their characteristics.

[2] For example, Miller and Weinberg (2017) consider a joint-venture as an exogenous shock to estimate a parameter that reflects how the behavior between the two leading firms in the US beer industry deviates from Bertrand-Nash pricing once one of them participates in the joint-venture. Ciliberto and Williams (2014) leverage a special feature of airport gate leasing contracts to estimate conduct as a function of multi-market contact in the airline industry.

the characteristics space interacted with information on rival firms' promotional activities, which we explain in detail below. Our paper contributes to the literature by incorporating an identification strategy based on variation in the characteristics space to estimate rich patterns of industry conduct. We focus on estimating the behavior of firms in the industry over time, and in particular on whether the behavior changes following important industry events. Specifically, we estimate the levels of conduct between firms before and after the 1993 Post-Nabisco merger, and following a massive wholesale price reduction for most cereal brands in 1996. Being able to accurately measure and explain the effects of such events is of key interest for competition policy.

Our results indicate that there are indeed substantial changes in industry conduct over time. We find partially cooperative levels of conduct between firms in the beginning of our sample, followed by a further increase in cooperation after the horizontal merger. When allowing our conduct parameters to differ across firms, we find that the pricing behavior of the smaller firms during these periods is more cooperative than that of the two market leaders, Kellogg's and General Mills. Finally, our estimates are consistent with a drastic change in industry conduct towards fully competitive behavior three and a half years after the merger.[3]

Section 2 introduces our data and provides detailed information about the RTE cereal industry and important industry events. We use scanner data from the Dominick's Finer Food (DFF) database. The database includes detailed information on DFF's supermarket stores located in the Chicago metropolitan area. In addition to detailed store-specific data on quantities, retail prices and temporary promotions, one convenient aspect of our data is that it contains information on wholesale prices. We analyze a five and a half year span of data from 1991 until 1996. Our sample period includes several important events, most notably the Post-Nabisco merger in January 1993 and a period that began in April 1996 in which manufacturers greatly decreased wholesale prices, which the business press referred to as a *price war*. For brevity, we use this terminology for the remainder of this paper.

To motivate our structural model and our identification strategy, we conduct a series of reduced form pricing regressions. We find that following the horizontal merger, prices increased significantly for the merging firms, as well as for two other firms. However, three and a half years later, almost all of the manufacturers greatly reduced their wholesale prices within only a few weeks which translated into shelf-price reductions of up to 18 percent

[3]Our results also relate to a long and extensive discussion regarding the underlying sources of market power of national cereal manufacturers. For example, Schmalensee (1978) argues that price competition is suppressed although firms might still partially compete via advertising and product entry. In contrast, Nevo (2001) finds that markups in the industry can be best explained solely by product differentiation and profit-maximizing behavior of multi-product firms.

for consumers (Cotterill and Franklin, 1999). These descriptive statistics provide the first evidence that the interaction between manufacturers significantly changed over our sample period.

Typically, it is not possible to disentangle the different explanations for observing these pricing patterns using only reduced form regression methods. Therefore, we introduce a structural empirical model in Section 3. On the demand side, we use a discrete-choice model in the style of Berry *et al.* (1995) (henceforth, BLP) and Nevo (2001), allowing for detailed consumer heterogeneity. On the supply side, we use a flexible conduct-parameter framework that specifies the degree of cooperation by a matrix of parameters that capture the degree to which firms internalize their rivals' profits.

We consider our approach to have significant advantages compared to simply assuming a particular form of industry conduct, as is often done in the literature. For example, the most commonly used assumption in such models is multi-product Bertrand-Nash pricing. This form of conduct implies that a firm maximizes the total profits of its own product portfolio but fully competes with all rival firms' products. Such a specification limits the heterogeneity of markups over time and across firms by assumption. There is a growing interest in the heterogeneity of markups both in the macroeconomics and the trade literature which usually rely on estimating output elasticities using a production function approach; see for example, De Loecker and Eeckhout (2017). Our approach allows us to estimate whether there is markup heterogeneity within an industry that can be attributed to heterogeneity in industry conduct.

Section 4 explains the details of our identification strategy. We exploit variation in firm's markup incentives caused by temporary promotions for different products across markets. The intuition is the following. In many consumer product industries, promotions are agreed upon between a manufacturer and a retailer several months in advance. This is done for various reasons, for example, a sufficient supply of the product must be ensured and promotional brochures must be printed.[4] Therefore, rivals' promotional activities in a given time period should be exogenous to a specific product's contemporaneous transitory demand and supply shocks. Furthermore, the promotional activities of a rival's product will affect a product's own demand, and the more so the more consumers consider these products as substitutes. Therefore, firms have an incentive to adjust the markups of all their products accordingly. This is why we interact the number of rivals' promotions with the products' relative proximity in the characteristics space. The relative proximity feature is shared with the class of differentiation instruments recently proposed by Gandhi and Houde (2015). In contrast to

[4]In both Europe and the US, retailers often set up a planogram of promotions for an entire product category several months in advance.

their approach and to classic BLP instruments, we do not require product entry or exit to induce variation in the characteristics space. Instead, we exploit variation in products' promotional activities as shifters of firms' pricing and markup behavior. We conduct a series of weak identification tests for both our demand and our supply side estimations, and find that our proposed instruments indeed prove to be very powerful in identifying both consumers' price elasticities and manufacturers' industry conduct. The data required to construct our instruments are readily available for many consumer goods industries so that our empirical strategy has broad applicability.

Section 5 presents our main estimation results. We find strong evidence for partial coordination in the beginning of our sample period, and for an additional increase following the Post Nabisco merger in 1993. When we restrict the conduct parameters to be equal across firms, our pre-merger conduct estimate of 0.465 indicates that a firm values US-$ 1 of its rivals' profits as much as US-$ 0.465 of its own profits. Furthermore, our estimates reveal that pre-merger, price-cost margins are 13.1 percent higher than under multi-product Bertrand-Nash pricing. Following the merger, the estimated conduct increases to 0.814, implying a 23.9 percent median margin over Nash pricing. Such an increase is, for example, consistent with a merger further facilitating coordination across firms in an industry.[5] When allowing the conduct parameters to differ across firms, we find that the small firms' internalization is higher than that of the two largest firms (Kellogg's and General Mills). The overall median margins are slightly higher in this model, at 14.6 and 25.0 percent over multi-product Nash pricing pre- and post merger, respectively. Moreover, towards the end of our sample period, for both specifications, we estimate conduct parameters close to 0, which is consistent with multi-product Bertrand-Nash pricing.

We use our parameter estimates to conduct a series of counterfactual exercises in which we simulate how prices and consumer surplus would have evolved under different levels of industry conduct. First, if firms had competed via Bertrand-Nash pricing prior to the price war, consumer welfare would have increased by around US-$ 3.5 million per year for the markets in our data set. Depending on the model specification, the median wholesale prices would have been between 5.2 and 6.0 percent lower during the pre-merger period, and between 11.1 and 11.2 percent lower during the post-merger period. Second, if industry conduct had remained at the post-merger level during the price war, median wholesale prices during the actual price war period would have been between 8.0 and 10.3 percent higher.

[5] Throughout the paper, we use the term *coordination* to describe cooperative pricing behavior, in the sense that firms' internalize the effect of their pricing on rival firms' profits to various degrees. We use this term for conciseness, and do not suggest that our model parameters correspond to anti-competitive behavior in the sense of violating antitrust laws.

Our paper relates to several different strands in the literature. First, it relies on the theoretical literature on identification of industry conduct and other structural elements of demand and supply in differentiated products models. Berry and Haile (2014) illustrate the potential to distinguish different oligopoly models in differentiated products industries by exploiting variation in market conditions. We show one way of how their arguments can be applied to real-world industry data and propose specific instruments that we find to be powerful for identifying detailed industry patterns that are difficult to identify using established instruments.

Early work in the literature on industry conduct has mostly relied on estimating conjectural variations; see, for example Bresnahan (1982) and Lau (1982) for identification results when estimating conduct for the homogeneous good case. Corts (1999) critically discusses such approaches. He argues that the estimated parameters usually differ from the "as-if conduct parameters" and therefore do not necessarily reflect the economic parameters of interest. This critique is not applicable in our case because we estimate the supply-side parameters structurally. In a series of seminal papers, Nevo (1998) discusses the advantages and disadvantages of a direct conduct estimation compared to a non-nested menu approach. He argues that in practice, estimating detailed industry conduct directly when having only a single demand rotator is impossible, and 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). 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 conduct in the US car industry around a price war in 1955. For 1954 and 1956, his results indicate a collusive industry outcome and for 1955 indicate multi-product Nash pricing. Nevo (2001) estimates a detailed differentiated products demand model for the RTE cereal industry, and recovers marginal cost for a menu of pre-specified models. He subsequently compares the different cost estimates with accounting data to select the most plausible specification, which he finds to be multi-product Bertrand-Nash pricing.

There is a small but growing literature on the estimation of industry conduct in a structural conduct parameter framework. The two papers most closely related to ours are Miller and Weinberg (2017), and Ciliberto and Williams (2014).

Miller and Weinberg (2017) assess the effects of a joint-venture on industry pricing behavior in the beer industry. They focus on estimating a conduct parameter that measures the magnitude of mutual profit internalization between Anheuser-Busch InBev (ABI) and MillerCoors after the MillerCoors joint-venture. Their model assumes industry-wide Bertrand-Nash pricing before the joint-venture for all firms and throughout the sample period for all firms except ABI and MillerCoors. Their identification strategy exploits the joint-venture as an exogenous shock together with the assumption that ABI's marginal costs are not affected by the MillerCoors joint-venture. They find a positive profit internalization between ABI and MillerCoors following the joint-venture, indicating that it potentially facilitated price coordination. Instead of relying on the merger itself as an exogenous instrument, our identification considers variation in rival firms' promotional activities and information on the relative proximity of products in the characteristics space. This allows us to identify a richer pattern of industry conduct. For example, we are able to quantify changes in conduct over time and differences across firms without assuming a specific conduct in any time period. Ciliberto and Williams (2014) estimate industry conduct in the airline industry. Their focus is on modeling industry conduct as a function of the degree of multi-market contact between different airlines. They find that firms with a lower degree of multi-market contact cooperate less when setting ticket fares. The identification strategy relies on the probability of a certain route being served by an airline being correlated with the number of gates an airline operates at an airport, and the number of gates not being easily adjustable in the short-term. Their model assumes a time-invariant and proportional relationship between the degree of cooperation between airlines and their level of multi-market contact.[6]

2 Data and Industry Overview

In this section, we describe our data and provide background information on the US RTE cereal industry. In addition, we conduct a series of reduced form regressions to guide our structural model and motivate the construction of our instruments.

2.1 Data Sources

Our main data consist of scanner data from the DFF database. The database includes information on DFF supermarkets located in the Chicago metropolitan area and weekly information on product prices, quantities sold, temporary promotions, and 1990 census data on demographic variables for each store area. For our analysis, we use data from 58 DFF stores and focus on 26 brands from the 6 different nationwide manufacturers present in the industry from February 1991 until October 1996. All of the products are offered throughout

[6] Although our paper focuses on estimating industry conduct for general industry settings, it is further related to the ex-post analysis of mergers. Crawford *et al.* (2017) analyze the welfare effects of vertical integration in the US 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. Michel (2017) analyzes the internalization of horizontally merging firms' pricing externalities in a structural model, and finds a relatively fast internalization for the first two years following the 1993 Post-Nabisco RTE cereal merger. Moreover, there is a growing literature focusing on the impacts of horizontal mergers on consumer surplus and industry prices; see, for example, Ashenfelter *et al.* (2013), and Björnerstedt and Verboven (2016).

the whole sample period and at all stores. There is no persistent entry of new products with a significant market share during our sample period. Therefore, we do not include these products. The database also includes data on in-store promotions, which DFF temporarily offers for different products. We explain this aspect in detail in the next subsection.

We complement the DFF data with input price data from the Thomson Reuters Datastream database and from the website www.indexmundi.com. The data include prices on commodities needed for the production of cereals such as sugar and various grains, and data on energy, electricity, and labor costs. Finally, we collect nutrition facts from the website www.nutritiondata.self.com and information on the different production and processing techniques for the different cereals. Throughout our analysis, we use deflated prices using a regional consumer price index.

We define a single unit of cereal as a 1 OZ serving of a specific brand. The total overall market size is defined as one serving per capita per weekday times the mean store-specific number of total customers.[7]

We are primarily interested in the interactions among the manufacturing firms. Observing a wholesale price measure rather than only the retail price allows for more precise inference regarding the manufacturing firms' marginal costs and markups. Specifically, we observe the retailer's average acquisition costs for each product at a given time. This variable reflects the inventory-weighted average of the percentage of the retail price that was paid to the producer. From this variable we 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; see Chevalier *et al.* (2003) for a discussion of this variable.[8] For our estimation, the data are aggregated at the monthly level. Consequently, DFF's inventory stocking at low wholesale prices for later dates should have a negligible effect on our wholesale price measure.

2.2 Industry Overview

The RTE cereal industry has been studied extensively; see for example Schmalensee (1978), Scherer (1979), and Nevo (2000b). At the end of our sample period, the industry had annual

[7] On average, our market size definition is very close to the specification of Meza and Sudhir (2010). We find the empirical results to be robust to using a time-variant market size specifications, and to changing the market size by factors $\frac{1}{3}$, $\frac{1}{2}$, 2, and 3, respectively. The implied elasticities from the main model are relatively close to those from studies using regional level data from the same industry, as for example in Nevo (2001). The demand results for the alternative market specifications are available upon request. We treat our market size number as exogenous to the RTE cereal prices because cereals only amount to a relatively small fraction of supermarket purchases for most consumers.

[8] DFF uses the following formula to calculate 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 reflects the weighted profit share for each product in a period, minus 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.

revenues of about US-$ 9 billion, which implies that almost 3 billion pounds of cereals were sold.

RTE cereals differ with respect to their observed and unobserved product characteristics, such as sugar and fiber content or package design. In the beginning of our sample period, the industry comprises 6 large nationwide manufacturers: Kellogg's, General Mills, Post, Nabisco, Quaker Oats, and Ralston Purina. It is common to classify the cereals into different groups, such as adult, family, and kids cereals. Kellogg's, which is 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 are strongest in the adult segment.

Table 1: Market Share Evolution

	GMI	KEL	POS	NAB	QUA	RAL
1991	32.4	46.1	7.9	3.1	7.2	3.4
1992	30.0	46.3	10.1	3.9	6.6	3.1
1993	28.9	47.0	11.6	0.0	8.9	3.6
1994	25.8	48.3	12.3	0.0	10.4	3.3
1995	31.9	43.8	14.4	0.0	6.8	3.1
1996	27.5	48.1	13.5	0.0	8.1	2.7

Notes: The table summarizes the firm-specific volume-based market shares (in percent) across all stores in our data set for each year. From 1993 onwards, Post's market shares include those of Nabisco. GMI stands for General Mills, KEL for Kellogg's, POS for Post, NAB for Nabisco, QUA for Quaker, and RAL for Ralston.

The products also differ in the type of main cereal grain and type of processing. The main types of cereal grains are corn, wheat, rice, and oats. The main production processes are flaking, puffing, shredding, and baking. We analyze the industry in a very mature state when no significant technological innovations occurred; therefore, we judge it safe to assume that production processes are constant over time.

On the retail level, RTE cereal products are primarily distributed via supermarkets. According to Nevo (2000b), more than 200 brands are available to consumers during the time span we analyze; however, the majority of of sales can be attributed to the 25 most popular brands. Table 1 summarizes the evolution of manufacturer market shares over our sample period. Although market shares vary over the different years in our sample, the industry structure is relatively stable. The two largest firms alone, i.e., General Mills and Kellogg's, cover around 75% of the market. The remainder of the market is split among the substantially smaller firms (Post, Nabisco, Quaker, and Ralston). Note that we do not include private label products explicitly in our analysis. This is because our main focus is on estimating the competitive interactions between nationwide manufacturers. We discuss how our model controls for potential changes in the popularity of private label products in Section 4.

An important feature of many consumer products industries is the prevalence and importance of temporary promotions. In our data, we observe four different forms of promotions: bonus buy, coupon, general and price reduction. With regard to their effects on prices and consumer demand, the promotions in our sample can be classified into two different categories. First, general and explicit price reduction promotions result in lower retail prices for all consumers and usually, also in a lower wholesale price. To obtain coupon and bonus buy price reductions, consumers typically must exert extra effort, and these promotions on average result in a lower price reduction than the other promotions. The different promotion types are often accompanied by measures that increase consumers' awareness of a product, for example, by being included in a retailer's advertising brochure or because of better shelf space or additional in-store promotion signs. Both direct price effects and increased product exposure typically increase the demand for products "on sale" and tend to temporarily decrease demand for rival products.

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. According to Rubinfeld (2000), the main concern of the antitrust authority regarding this merger was the strong substitutability in the adult cereal segment between Post's Grape Nuts cereal and Nabisco's Shredded Wheat, which would give the merging firms a non-trivial incentive to increase prices unilaterally. The merger did not lead to any product entry or exit or any changes to existing products. In fact, Nabisco cereals were even sold under the same brand names and in a packaging very similar to before the merger. Therefore, we abstract from product repositioning, as, for example, analyzed in Sweeting (2010), and treat the set of products as exogenous.

In April 1996, Post decreased the wholesale prices for its products nationwide by up to 20%, thereby also increasing its market share. This was followed by significant price cuts a few weeks later by the market leader Kellogg's and then by General Mills and Quaker. Cotterill and Franklin (1999) report an average decrease in the wholesale price of 9.66% across all products in the industry between April and October 1996, and an average 7.5% decrease in the retail price. These numbers suggest a systematic change in industry pricing for most products during this period.[9] One of the contributions of this paper is that we structurally estimate how much of the change in industry behavior is due to a breakdown of

[9] In March 1995, two US congressmen started a public campaign to reduce cereal prices, which received relatively high media attention; this campaign was revived one year later right before the start of the substantial wholesale price cuts (Cotterill and Franklin, 1999). Although negative publicity and political pressure might be potential reasons for the price cuts, we remain agnostic about any causes for the price war.

coordinated pricing rather than potential shifts in demand or marginal costs.

2.3 Reduced Form Analysis

To investigate whether our data supports anecdotal industry evidence and to guide our structural model, we run a series of reduced form regressions. In particular, we are interested in whether prices systematically changed following the merger and during the price war, and whether and how the promotions of rival brands affect manufacturers' pricing decisions. Figure 1 illustrates the evolution of wholesale and retail prices averaged over all stores during our sample period for some important brands.

We analyze the determinants of both wholesale and retail prices by estimating a series of OLS-regressions. The level of observation is a product-store-month combination resulting in a sample size of 96,512 observations. Our dependent variables are $\log(p_{it}^w)$ and $\log(p_{it}^r)$, i.e., the logged wholesale and retail prices of brand i for store-market combination t, respectively.

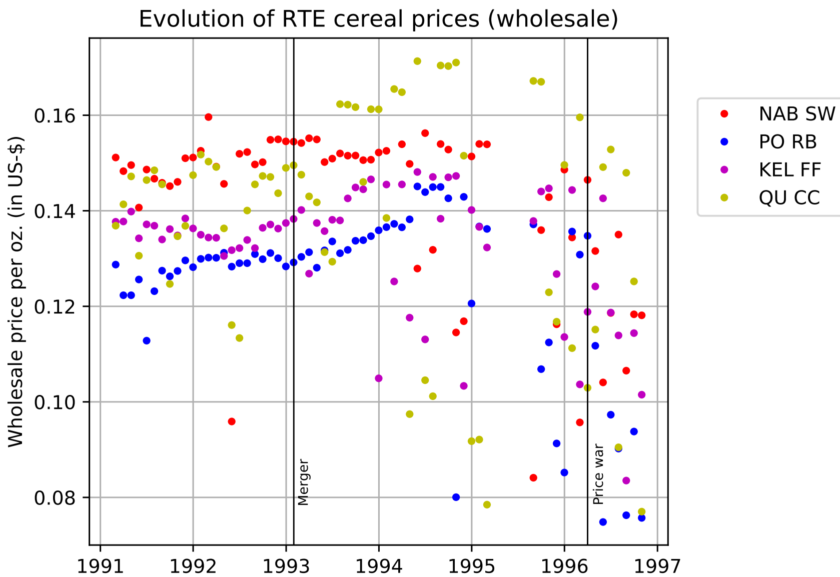
The large data set allows us to control for a wide variety of fixed effects such as brand and store fixed effects. Moreover, we include total market sales to control for overall industry shocks. For both dependent variables, i.e., wholesale prices and retail prices, our key regressors of interest are dummy variables for the post-merger and the alleged price war period, and variables summarizing a brand's own and rival firms' promotional activities in a specific store and month.

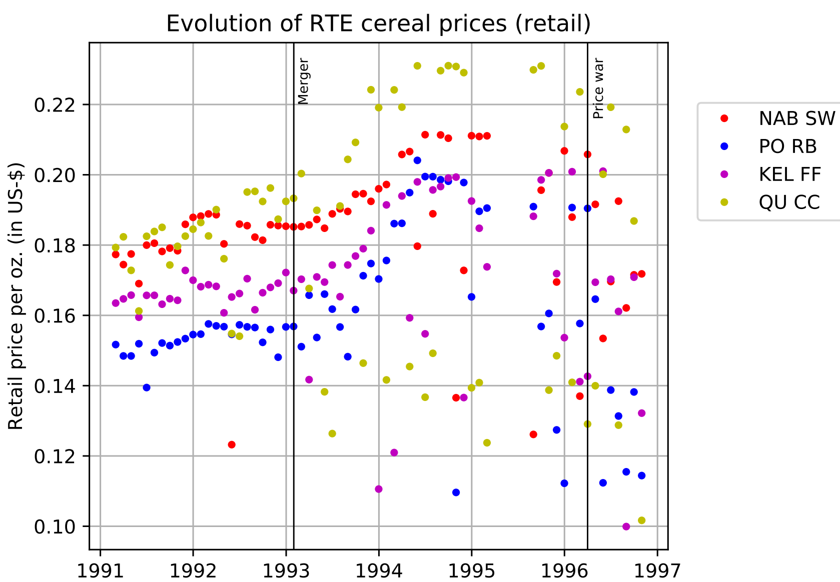
In the baseline specification, we interact a post-merger indicator only with a dummy for the merging firms and a dummy for the non-merging firms, allowing the merging firms (Post and Nabisco) to react differently to the merger than the non-merging firms. In more detailed specifications, we use the post-merger indicator interacted with the dummies for every post-merger firm (KEL, RAL, QUA, GMI and POSTNAB).

To motivate our identification strategy for industry conduct, we include several measures of the firm's own and rival brands' promotional activities as additional regressors. *Promo (own brand)* captures the number of promotions (general sales or price reduction-based sales) conducted for a given brand in a given market. *Promo (same firm)* indicates the number of promotions a firm conducted for its products other than brand i in a given store and month. *Promo (rival firm)* captures the number of promotions conducted by all rival firms. Because the reaction to rivals' promotions is likely to be affected by the prevailing industry conduct, we allow the effect of rivals' promotions to differ in the pre-merger, post-merger, and the price war periods.

We conjecture that general and price reduction sales are typically more visible and appeal to a broader range of consumers than bonus buy and coupon promotions. Since the latter are

Figure 1: Evolution of RTE cereal prices





Notes: The two figures display the evolution of the average wholesale and retail prices, respectively, across all stores over time for selected brands. The brands are Nabisco/Post Shredded Wheat, Post Raisin Bran, Kellogg's Frosted Flakes, and Quaker Cap'n Crunch.

usually more complicated promotions that have more restrictions and often require consumers to exert extra effort, we suspect their effects to be different from general sales and potentially much weaker. Therefore, in our baseline specification, we construct promotion regressors based only on general and price reduction promotions. In the second set of regressions, we include a measure summarizing bonus buy and coupon promotions conducted by the firm's own brand, the firm's other brands, and rival firms' brands. Appendix A provides additional results and presents the estimation equations for the reduced form regressions.

Table 2 summarizes the regression results when the dependent variable is the logged wholesale price. In the period following the Post-Nabisco merger, the merging firms increase their wholesale prices by 6% on average. Looking at the non-merging firms' reaction to the Post-Nabisco merger in detail (Column 2), we find that Kellogg's and Ralston increase their wholesale prices by almost as much as the merging firms, while Quaker and General Mills slightly decrease their prices. Furthermore, during the price war period, the wholesale prices drop substantially for all firms by almost 10% on average.

Not surprisingly, general and price reduction promotions result in a strong decrease in wholesale prices. An additional promotion on average decreases the wholesale price by 11%. In contrast, the cross-effects of promoting brands owned by the same firm are very small but positive (0.07%), i.e., a brand's wholesale price increases slightly when other products owned by the same firm are on promotion. When including regressors that capture the intensity of bonus buy promotions (Columns 3-4), our initial conjecture is confirmed, i.e., bonus buy promotions are associated with a substantially smaller (roughly 2%) reduction in wholesale prices.

Analyzing the effects of the promotions conducted by rival firms over time indicates important changes in how firms react to each other. In the early periods of our sample (i.e., pre-merger), the rival firms' promotions and wholesale prices for a given brand have a small but positive correlation both for general and bonus buy promotions. Following the merger, the effect becomes negative but remains very weak for general promotions, while the effect for bonus buy promotions remains positive and becomes stronger. During the price war period which starts approximately three and a half years after the merger, both general and bonus buy promotions have a substantial negative effect on rivals' wholesale prices.

This pattern is consistent with significant changes in industry conduct over time. In a collusive industry, firms internalize each others' profits. Therefore, rivals' promotions are not guaranteed to result in complementary price cuts by a firm's brands. In contrast, in a competitive environment, prices should be strategic complements: Rivals' promotions increase competitive pressure and should go hand in hand with price cuts for a firm's own brands.

Overall, our reduced form regressions provide the first evidence that promotional measures indeed capture relevant shifters of manufacturers' markups and can therefore constitute strong instruments for industry conduct.

Table 2: Reduced form analysis: Wholesale prices

	(1) Baseline	(2) Firm detailed	(3) Baseline w/ BB	(4) Firm detailed w/ BB
Promo (own brand)	-0.1099***	-0.1093***	-0.1116***	-0.1111***
	(0.0012)	(0.0012)	(0.0012)	(0.0012)
Promo (same firm)	0.0007*	0.0009**	0.0021***	0.0022***
	(0.0003)	(0.0002)	(0.0003)	(0.0003)
Promo pre-merger (rivals)	0.0029***	0.0024***	0.0026***	0.0023***
	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Promo post-merger (rivals)	-0.0019***	-0.0017***	-0.0010*	-0.0008
	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Promo price war (rivals)	-0.0071***	-0.0082***	-0.0054***	-0.0068***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Post-merger non-merging	0.0182***		0.0124**	
	(0.0022)		(0.0046)	
Post-merger POSTNAB	0.0609***	0.0584***	0.0511***	0.0436***
	(0.0023)	(0.0023)	(0.0049)	(0.0049)
Price war period	-0.0983***	-0.0936***	-0.0869***	-0.0854***
	(0.0015)	(0.0015)	(0.0017)	(0.0017)
Post-merger KEL		0.0515***		0.0413***
		(0.0021)		(0.0041)
Post-merger RAL		0.0501***		0.0376***
		(0.0027)		(0.0054)
Post-merger QUA		-0.0176***		-0.0273***
		(0.0026)		(0.0052)
Post-merger GMI		-0.0284***		-0.0394***
		(0.0025)		(0.0048)
BB (own brand)			-0.0217***	-0.0220***
			(0.0009)	(0.0009)
BB (same firm)			0.0051***	0.0045***
			(0.0002)	(0.0002)
BB pre-merger (rivals)			0.0009***	0.0008***
			(0.0001)	(0.0001)
BB post-merger (rivals)			0.0014***	0.0018***
			(0.0003)	(0.0003)
BB price war (rivals)			-0.0053***	-0.0040***
			(0.0003)	(0.0003)
Observations	96512	96512	96512	96512
R-square	0.76	0.76	0.77	0.77

Notes: All estimations include brand- and store- fixed effecs. Columns (2) and (4) allow for post-merger reaction to differ across firms. Promo (same firm) describes the number of promotions of other products in a market that belong to the same firm. Promo (rivals) describes the number of promotions of other products in a market that do not belong to the same firm. Pre-, post, and PW stand for pre-merger, post-merger, and price war period, respectively. BB stands for bonusbuy and coupon promotions, while all promo variables without BB reflect general / price reduction promotions.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9 in Appendix A summarizes the results for the retail price regressions using the same specifications as for wholesale prices. The results are mostly consistent with our wholesale price regressions. Prices increase substantially following the merger. When investigating the post-merger reaction in more detail, the same pattern as for wholesale prices emerges: The retail prices increase for two non-merging firms (Kellog's and Ralston), but remain constant for General Mills and Quaker. As expected, both wholesale and retail prices react strongly to promotions and decrease significantly during the price war period.

3 Empirical Model

The reduced form analysis presented in the previous section yields several important insights regarding the evolution of the RTE cereal industry in the 1990s. Most importantly, on average there is a significant price increase following the Post-Nabisco merger, which is followed by a dramatic reduction in wholesale prices three and a half years later. There are several potential reasons for observing this pattern. For example, consumers' preferences and willingness-to-pay may have shifted, resulting in changes in market power due to product differentiation. Alternatively, production costs may have changed over time. In addition, there may have been changes in industry conduct. Generally, it is extremely difficult to disentangle these explanations by using only reduced form regressions. To gain much more detailed insights about the different channels, we develop a structural model of the RTE cereal industry.

3.1 Demand Model

On the demand side, we estimate a random coefficients logit model with a specification that is similar to those in Nevo (2001) and Goldberg and Hellerstein (2013). One key advantage of this model is that it allows for very flexible substitution patterns. An accurate estimation of own- and cross-price elasticities is crucial in our model since they are the most important determinants of a firm's pricing first-order conditions. Consequently, using incorrect or weakly identified demand estimates is likely to result in confounded estimates of both marginal costs and industry conduct.

There are J brands available in each market. We denote the number of markets, defined as a store-month combination, by T. Each market consists of a continuum of individual consumers. Individual i's indirect utility from consuming product j in market t is given by

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

where x_j denotes a K-dimensional vector of brand j's observable characteristics (including several layers of fixed effects), p^r_{jt} denotes the retail price of product j in market t, and ξ_{jt} is a brand-market specific quality shock that is unobservable to the researcher but observable to and equally valued by all consumers. As in Nevo (2001), the inclusion of fixed effects allows us to decompose ξ_{jt} into a persistent part that is absorbed into the fixed effects and an idiosyncratic component so that $\xi = \bar{\xi} + \Delta\xi$. Therefore, only the latter is treated as our structural demand error for forming moment conditions. Finally, ϵ_{ijt} is an iid error term that is type I extreme value distributed.

The coefficients β_i and α_i are individual-specific. They depend on the mean valuations, a vector of i's demographic variables, D_i, and their associated parameter coefficients Φ that measure how preferences vary with demographics; therefore,[10]

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Phi D_i. \tag{2}$$

Consumers who do not purchase any cereal product in a period choose the outside good. The indirect utility of consuming the outside good can be written as $u_{i0t} = \xi_0 + \phi_0 D_i + \epsilon_{i0t}$. Because only differences in utility are identified in discrete-choice models, we normalize ξ_0 to zero.

The vector of demand parameters θ_D consists of a linear part $\theta_1 = (\alpha, \beta)$, that affects each consumer identically, and a nonlinear (random coefficients) part $\theta_2 = vec(\Phi)$. Analogously, the indirect utility of consuming a product can be decomposed into a mean utility part δ_{jt} and a mean-zero random component $\mu_{ijt} + \epsilon_{ijt}$ capturing heterogeneity from demographics and unobserved taste shocks. The decomposed indirect utility can be expressed as $u_{ijt} = \delta_{jt}(x_j, p^r_{jt}, \xi_{jt}, \theta_1) + \mu_{ijt}(x_j, p^r_{jt}, D_i; \theta_2)$; with

$$\delta_{jt} = x_j\beta + \alpha p^r_{jt} + \xi_{jt}, \tag{3}$$

$$\mu_{ijt} = [p^r_{jt}, x_j]' * \Phi D_i, \tag{4}$$

where $[p^r_{jt}, x_j]$ is a $(K + 1) \times 1$ vector of observable product characteristics.

Consumers buy either one unit of a single brand or take the outside good. They choose the option that yields the highest utility. The model's market share predictions are obtained via integrating over all the shock distributions

$$s_{jt}(x_{.t}, p^r_{.t}, \delta_{.t}, \theta_2) = \int_{A_{jt}} dP^*_\epsilon(\epsilon) dP^*_D(D), \tag{5}$$

where $A_{jt}(x_{.t}, p^r_{.t}, \delta_{.t}, \theta_2) = \{(D_i, \epsilon_{it}) | u_{ijt} \geq u_{ilt} \forall l \in \{0, .., J\}\}$ denotes the set of consumers'

[10] In extensive robustness checks, we also experimented with persistent preference heterogeneity in the form of classical normally distributed preference shocks. While these models led to elasticities that were similar to our baseline demand specifications, the standard errors increased; therefore we opted for a model with only demographic interactions, as in Goldberg and Hellerstein (2013) and Miller and Weinberg (2017).

shock realizations for which j yields the highest utility.

As discussed in Section 2.2, temporary product- and store-specific promotions are important determinants of consumers' cereal choices through both direct price effects and indirect awareness effects that increase the attractiveness of products "on sale". Our model captures direct price reductions in the observed retail price p^r_{jt}. We incorporate the indirect effects of promotions (e.g., retailers' advertising brochure, better shelf space or in-store promotion signs) by including the number of promotions for a given brand in a given market as an additional product characteristic in the utility function.

We abstract from dynamic consumer behavior for several reasons. In principle, our supply model and identification strategy can be combined with a dynamic demand model in the style of Hendel and Nevo (2006). However, dynamic models that allow for detailed high-dimensional heterogeneity are extremely computational intensive. A dynamic model would therefore have to heavily compromise in this dimension. In our application, we judge accounting for detailed consumer heterogeneity to be more important for estimating consumers' substitution patterns than dynamic storage behavior. Most importantly, we use data at the month level for which dynamic behavior is arguably much less relevant than for weekly data.

3.2 Supply Model

The J brands in the industry are produced by $R \leq J$ firms. Each brand is produced by only one firm, but each firm can produce multiple brands. We model marginal costs as a linear function of a battery of fixed effects, observable cost factors w_{jt}, and a brand-market specific cost shock ω_{jt} that is unobserved by the researcher but known to the firms, so that

$$mc_{jt} = \underbrace{w_{jt}\gamma}_{\tilde{mc}_{jt}} + \omega_{jt}, \tag{6}$$

where γ is a vector of marginal cost parameters to be estimated, and $\tilde{mc}_{jt}$ denotes the part of the marginal cost that is attributed to observable cost shifters.[11] As for the unobserved demand shocks, we decompose the unobserved cost shock ω_{jt} in a systematic component that is absorbed into the fixed effects and an idiosyncratic innovation such that $\omega = \bar{\omega} + \Delta\omega$. Our baseline cost specification implies that the production processes do not change over time, resulting in time-invariant marginal cost functions. For our application, we judge this to be a reasonable assumption since no technical innovations or relocations of production facilities occurred during our sample period. For a discussion on the types of time-varying

[11] For simplicity of notation, we omit index w for wholesale marginal costs.

cost functions our model can accommodate, for example, because of synergies following the merger, see Section 4.

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. We assume linear wholesale prices that are not contingent on the overall quantity sold in a period. As in Goldberg and Hellerstein (2013), we assume a model of double-marginalization in which manufacturers set their wholesale prices anticipating that the retailer takes these prices as given and optimally adjusts its retail prices.[12]

Focusing on data from a single retailer (DFF) allows us to observe detailed wholesale price data. The downside of this approach is that all substitution to different retailers is captured by the market share of the outside good. Given that cereals typically constitute only a small fraction of overall grocery expenses, we judge this channel to be much less important than the substitutability of different products within the same store. Slade (1995) finds that 90% of consumers do not compare the prices of different retailers on a week-to-week basis. Therefore, we do not expect that excluding other retailers will have a significant effect on our estimation results.

Henceforth, we denote a manufacturing firm simply as a firm. λ_{ijt} represents the degree to which brand i takes into account brand j's profits when setting its wholesale prices in market t. All λ_{ijt} can be arranged in an *internalization matrix* Λ_t. Consequently, Λ_t generalizes the ownership matrix of zeros and ones in classical BLP-models. We follow the literature (Miller and Weinberg, 2017; Ciliberto and Williams, 2014) in treating the elements of Λ as structural parameters. Black *et al.* (2004) and Sullivan (2017) illustrate how these parameters can be translated into the parameters of an underlying repeated game in which firms maximize their own discounted lifetime profits. Each λ_{ijt} is normalized to lie between 0 and 1, where 0 implies no internalization of firm j's profits by firm i, and 1 implies full internalization.[13]

[12]Testing of different forms of vertical price setting behavior, Sudhir (2001) finds evidence for such sequential price-setting behavior between manufacturers and retailers in the industry. Also see Villas-Boas (2007) for a framework to test for different forms of vertical relations using retail and input prices. Our model implies that each manufacturer sets store-specific wholesale prices. In principle, it is straightforward to estimate our model under various alternative assumptions, for example, under the assumption that firms set the same wholesale price for all stores or all stores within a pricing zone. We opted for a model of store-specific wholesale prices to capture that in reality, manufacturer-retailer contracts are often very high-dimensional. For example, they may specify additional payments for delivery or shelf-space allocations, which are likely to vary across stores.

[13]Conceptually, our model can also accommodate either $\lambda > 1$ or $\lambda < 0$. Negative internalization parameters would imply that a firm derives a positive utility from "ruining" another firm. In the cereal industry, there is no evidence of such behavior. $\lambda > 1$ implies that a firm values its rivals' profits more than its own, which does not seem reasonable in our application. See Appendix F for the normalization details and for how this affects the computation of standard errors.

The manufacturer's objective function for product j in market t can be written as

$$\Pi_{jt} = (p^w_{jt} - mc_{jt}) s_{jt} M_t + \sum_{k \neq j} \lambda_{jkt} (p^w_{kt} - mc_{kt}) s_{kt} M_t, \tag{7}$$

where s_{jt} denotes the market share of brand j as defined in Equation (5), M_t denotes the market size, and p^w_{jt} denotes the wholesale price per unit of brand j in market t. Following the literature, we assume that marginal costs are common knowledge among firms but unobserved by the researcher. Therefore, they must be backed out via the model's first-order conditions. The first-order condition for product j with respect to its own price can be written as

$$s_{jt} + \sum_{k=1}^{J} \lambda_{jkt} (p^w_{kt} - mc_{kt}) \frac{\partial s_{kt}}{\partial p^w_{jt}} = 0. \tag{8}$$

Define $\Omega_{jkt} \equiv -\lambda_{jkt} * \frac{\partial s_{kt}}{\partial p^w_{jt}}$, which combines information on consumers' price elasticities and firms' internalization behavior, and let Ω_t be the stacked version of Ω_{jkt} with j in the rows and k in the columns. Given the demand parameters θ_D, the vector of manufacturers' marginal costs of production for all products in market t, $mc_{.t}$, conditional on the ownership matrix Λ^t, is

$$mc_{.t}(\theta_D, \Lambda_t, p^r_{.t}, p^w_{.t}, x_{.t}) = p^w_{.t} - \Omega_t^{-1}\big(\theta_D, \Lambda_t, p^r_{.t}(p^w_{.t}), x_{.t}\big) s_{.t}\big(\theta_D, p^r_{.t}(p^w_{.t}), x_{.t}\big). \tag{9}$$

Rearranging and plugging in the marginal cost function from Equation (6) allows us to write the vector of structural cost shocks for all products in market t, $\omega_{.t}$, as a function of the model parameters and observed data, so that

$$\omega_{.t}(\theta_D, \gamma, \Lambda_t) = p^w_{.t} - \tilde{mc}_{.t}(\gamma, w_{.t}) - \Omega_t^{-1}\big(\theta_D, \Lambda_t, p^r_{.t}(p^w_{.t}), x_{.t}\big) s_{.t}\big(\theta_D, p^r_{.t}(p^w_{.t}), x_{.t}\big). \tag{10}$$

This structural cost shock forms the basis for our moment conditions to estimate the supply parameters.

In most BLP-style models, Λ is fully assumed. One of the key contributions of this paper is to flexibly estimate industry conduct as captured by the parameters within Λ. In principle, our empirical strategy is general enough to treat Λ non-parametrically, i.e., let its parameters vary freely across local markets, time, and products. However, a fully flexible conduct matrix for a specific market t consists of J^2 parameters. To keep the estimation tractable, we restrict the structure of Λ in an economically reasonable way.

Throughout the paper, we assume that the underlying conduct 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. For the cereal industry, we judge this to be a reasonable restriction, that allows us to focus in detail on variation in conduct over time and across firms. Moreover, we assume that each firm internalizes all products of a rival firm equally, so that our internalization parameters are not product- but firm-specific.

One of our primary goals is to quantify the evolution of conduct over time, in particular over three different periods: the pre-merger period (1990-1992), the post-merger period (January 1993 - April 1996), and the price war period (after April 1996). Throughout, we estimate conduct parameters that change across but are constant within periods. We employ the standard assumption that after the merger, merging firms fully internalize the profits of the other division. In our baseline specification, we assume that all firms internalize all rivals' profits to the same degree. In a more detailed specification, we allow different firms to internalize differently.

Example for Λ: industry with three firms For illustrational purposes, assume that there are 3 single-product firms. If each firm equally internalizes its pricing externalities on every rival, the pre-merger the conduct matrix is given by

$$\Lambda^{Pre} = \begin{pmatrix} 1 & \lambda^{Pre} & \lambda^{Pre} \\ \lambda^{Pre} & 1 & \lambda^{Pre} \\ \lambda^{Pre} & \lambda^{Pre} & 1 \end{pmatrix}.$$

If firm 1 and 2 merge, the conduct matrix post-merger changes to

$$\Lambda^{Post} = \begin{pmatrix} 1 & 1 & \lambda^{Post} \\ 1 & 1 & \lambda^{Post} \\ \lambda^{Post} & \lambda^{Post} & 1 \end{pmatrix}.$$

This matrix reflects that the merging firms fully internalize their profits post-merger. Moreover, this specification allows for non-merging firms to change their behavior as well. For example, if the merger resulted in increased industry-wide price coordination, then we expect λ^{Post} to be higher than λ^{Pre}. Finally, during the price war period, the conduct matrix evolves to

$$\Lambda^{PW} = \begin{pmatrix} 1 & 1 & \lambda^{PW} \\ 1 & 1 & \lambda^{PW} \\ \lambda^{PW} & \lambda^{PW} & 1 \end{pmatrix}.$$

If the price war leads firms to price competitively, then we expect λ^{PW} to be very close to zero.

Because we observe a horizontal merger in our sample, it is worth discussing how potential cost synergies could affect our estimation results. Note that we do not use the ownership change as an instrument, so that the occurrence of synergies would in principle not pose a problem. Our identification strategy would lead to biased estimates only if our instruments are correlated with the structural (transitory) cost shock $\Delta\omega$. This would be the case if there are synergies in the unobservable cost shock that are systematically related to our (promotion and relative proximity based) instruments for industry conduct. For example, our instruments would be invalid if following the merger, Post and Nabisco have systematically lower transitory cost shocks; rival firms anticipate these future shocks and therefore systematically change their promotional activities. We are not aware of any industry evidence for this kind of shift in manufacturers' strategies after the merger, nor do we find any support for such behavior in our data.

We have not found evidence suggesting that the Post-Nabisco merger caused significant marginal cost synergies. Moreover, cost synergy considerations have not been of significant importance during the merger case.[14] In addition, merger-related savings in fixed costs have no effect on firms' pricing because fixed costs do not affect the first-order conditions. An example for such savings are costs for administrative staff or rent for office space. Similarly, savings in financing costs due to a larger firm size should not affect the marginal costs of production in the short run.

We explicitly rule out synergies due to the increased bargaining power of the merged firm with suppliers of inputs. Because the production facilities of the different firms are geographically separated, the need to use different suppliers of wheat, sugar, and energy seems reasonable. In addition, there are no factory closures within the first five years of the merger. Nabisco's main production facility in Naperville, Illinois, continues to produce the same products after the merger as before. The merging firms' products also use different production technologies. Post's products primarily require flaking and baking processes, while Nabisco's products mainly rely on shredding.

Therefore, we treat the merging firms' cost functions as constant over time in our baseline specification. As a robustness check and to address potential remaining concerns about merger-related synergies, we re-estimate our supply models by including a post merger-

[14]See Rubinfeld (2000) for a detailed description of the arguments brought forward in the merger case. Synergies are not mentioned as an argument in favor of the merger but rather the discussion focused heavily on the consumers' substitution patterns between different cereals, which we estimate in detail. A potential non-synergy rationale for the merger was a reduction in debt for Nabisco's former parent company, RJR Nabisco. After the 1988 leveraged buyout of RJR Nabisco, which at this time was the largest leveraged buyout of all time, the ownership group accumulated substantial debt. Divesting different branches of the company such as the RTE cereal branch was thus a strategy to reduce the overall debt level.

merging firm dummy in the marginal cost function and evaluate the sensitivity of our results.

4 Identification & Estimation

In this section, we describe which variation in the data identifies consumer demand, manufacturers' marginal costs and industry conduct. Furthermore, we describe how we construct our instruments and the estimation algorithm.

4.1 Identification of Supply Parameters

The primary difficulty on the supply side is to separately identify the level of industry conduct from unobserved marginal cost shocks. Finding good instruments to identify industry conduct is complicated by two factors. First, many instruments used in practice turn out to be weak. For example, the classical BLP moment conditions, which are based on aggregate functions of rival products' characteristics, can often identify cost parameters reasonably well but they are rarely able to strongly identify conduct parameters. Second, in many applications, one does not observe variation in the set of products offered which makes many instruments collinear with brand fixed effects. In the following, we propose a novel set of instruments that rely only on standard market-level data and do not suffer from these issues.

To address the problem of weak instruments, we construct measures of products' relative isolation in the characteristics space. Gandhi and Houde (2015) illustrate that differentiation instruments, which exploit products' relative isolation, perform well in identifying heterogeneous consumer preferences. For our application, we find that instruments that are based on similar proximity measures are also very powerful for identifying industry conduct. More specifically, we construct several variables that capture how similar the characteristics of two products are to each other, for example, with respect to their sugar or fiber content.

To overcome the problem of a constant product space, we interact our isolation measures with information on the promotional activities of rival firms. Intuitively, our instruments count the number of promotions by rival firms in a given market but only consider those rival products that are "close enough" according to our relative proximity measures described above.

Typically, one can compute several proximity measures, and one often observes several types of promotions. This allows us to construct multiple instruments for industry conduct. For example, if we compute 3 different proximity measures and observe 2 types of promotions, we can rely on 6 different instruments. Appendix B.1 provides the details on how we specify our instruments.

In order for our instruments to be valid, they must satisfy two conditions. First, they must be correlated with the endogenous regressor. When estimating firm conduct, we effectively need to instrument firms' markups. How many and which rival products are on promotion affects the competitive pressure exerted on a product. When any substitute product of j is on sale, consumers become more likely to choose this product instead of j compared to when there is no promotion. The firm owning product j should consider this when setting prices and markups of its brands.

Second, the instruments must be exogenous to the structural error used to construct the moment conditions. Clearly, promotions are chosen by firms and are therefore endogenous. However, in many industries, including ours, decisions between retailers and manufacturers regarding whether a promotion for a particular product-store combination occurs in period t are made in advance, i.e., at the latest in $t-1$. Generally, these decisions are unlikely to be reversed due to operational and logistic issues; for example, advertising brochures have to be printed, and a higher supply than usual has to be delivered to the different stores.[15]

Note that since we include brand, store, and seasonal fixed effects in the marginal cost function, the structural error terms capture only transitory shocks. Therefore, it seems very plausible that the structural demand and supply errors for product j at time t are uncorrelated with other brands' sale periods that are decided in $t-1$ at the latest.

The key restriction we make is that while firms decide in period $t-1$ or before whether a promotion occurs in period t, they do not simultaneously determine the wholesale price. While one could in principle relax this timing assumption, the essential requirement for our instruments to work is that the promotion patterns are fixed before the wholesale prices are set.[16] In all other regards, we can be agnostic regarding the reason why the retailers and manufacturers agree to place products on promotion.[17] Thus, rival firms' promotional periods should affect firm j's pricing but should not be correlated with j's structural cost shocks.

To verify that our supply side instruments are indeed very powerful for identifying the conduct parameters, we conduct a series of weak IV and weak identification tests. We report

[15] This is a pattern observed in many consumer products industries in many different countries. In some countries, it is even known several months in advance at which retailer which brands will be on promotion, and this is common knowledge across the different manufacturers.

[16] A subtle additional requirement is that after the promotion pattern for period t is determined, but before the wholesale prices for period t are set, product-specific (demand or supply) shocks occur that lead two firms with identical promotion patterns today to charge different wholesale prices in the next period. This assumption is similar to common assumptions in the literature on production function estimation; see, for example, the extensive discussion in Ackerberg *et al.* (2015).

[17] There is a theoretical literature on why promotion occur in the first place; see, for example, Lal and Matutes (1994) for using promotions as loss-leadership and Varian (1980) and Villas-Boas (1995) for a price-discrimination rationale regarding different consumers.

these results in Appendix B.2.

Several advantages of our identification strategy are noteworthy. First, our instruments do not require the availability of exogenous industry shocks, such as a merger. Second, they do not rely on variation in the set of products offered or changes in products' physical characteristics. Finally, the information necessary to construct our instruments is available in many data sets used in empirical industrial organization or quantitative marketing; therefore, our empirical strategy can be easily applied to many consumer products industries.

4.2 Identification of Demand Parameters

Conceptually, our demand model does not differ significantly from most of those used in the literature. Our primary concern is that the estimated conduct parameters will depend crucially on the estimated demand elasticities. This results in two challenges for our demand model. First, we require realistic and flexible substitution patterns. Therefore, simple logit models are unlikely to describe the full picture. Second, a recent and growing strand of the literature has highlighted that many demand instruments commonly used in BLP-type models are weak. Weak instruments are likely to result in very imprecise and sensitive estimates of substitution patterns. Because reliable price elasticities and substitution patterns are the key inputs from the demand side to our supply model, it is extremely important to use strong instruments.

Our key instruments for identifying heterogeneity in consumer preferences are identical to the conduct instruments described above. We use the number of promotions of other products interacted with the relative proximity in the characteristics space.

As additional instruments for retail prices in the demand equation, we exploit input price variation over time interacted with 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. Because of the absence of major variation in the production processes, for example, due to firms' relocating their production facilities, and because our data covers only one metropolitan area, the relation between observed cost shifters, such as input prices, and retail prices can be opaque and statistically weak. Therefore, we exploit data on wholesale prices, as, for example, proposed by Chintagunta *et al.* (2003). We use predicted instead of actual wholesale prices to account for the possibility that a manufacturer's wholesale price could be correlated with a transitory demand shock. We calculate predicted wholesale prices as the fitted values from a linear regression of observed wholesale prices on a wide variety of fixed effects and observed demand and cost characteristics. We explain how we construct this regression in detail in the next subsection.

To ensure the power of our demand instruments, we conduct extensive checks to ensure that our model does not suffer from weak identification. We run a battery of first-stage F-tests and rank deficiency tests of the first-stage based on ideas in Cragg and Donald (1993) and Kleibergen and Paap (2006). The details and results are presented in Appendix B.2.

4.3 Estimation Algorithm

We estimate our model using the generalized method of moments (GMM) similarly to the seminal work by BLP and the subsequent literature. We estimate demand and supply parameters in two steps. Given our large data set, we judge the gain in efficiency from estimating both parts jointly to be less important than the gains in computational speed from estimating demand and supply separately.

Demand estimation For a given guess of the non-linear demand parameters, we solve the BLP contraction mapping to back out the mean utility levels δ for each brand, store and month to match the model's predicted market shares to the observed data. Then, we compute the structural demand shocks ξ for a given value of the linear demand parameters. Finally, we interact the implied demand shocks with a set of suitable demand instruments Z_D. Based on the identification arguments from the previous section we choose Z_D such that at the true demand parameter values θ_{D0}, the transitory demand shock $\Delta\xi(\theta_{D0})$ is uncorrelated with Z_D. The moment conditions for the demand model can be written as

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

In our main specification, Z_D contains the following variables. First, we include brand dummies and month-of-the-year (henceforth: month-year) dummies to capture potential seasonal effects in cereal demand, and the total number of a brand's promotions in a given market.[18] In addition, X and Z_D contain a linear-quadratic time trend that controls for long-term industry trends, such as potential changes in the popularity of private label products, that are part of the outside good in our model. Second, our main specification includes predicted wholesale prices as brand-specific cost shifters to identify the price coefficient. In our linear hedonic wholesale price regression, we use the following regressors: Brand dummies; month-year dummies; store fixed effects; a time trend; input prices for wheat, corn, sugar, rice, oats, electricity and gasoline; and the number of a brand's own and rival

[18] Since in our application products' characteristics do not change across markets, we follow Nevo (2001) and do not include exogenous product characteristics x in the estimation directly. Instead, we back out mean preferences for each time-invariant product characteristic by regressing the estimated brand fixed effects on these characteristics.

firms' promotions. Similar to Goldberg and Hellerstein (2013), we interact our cost shifters with brand dummies to form our instruments. Third, Z_D includes instruments based on promotional periods interacted with the relative proximity in the characteristics space. In particular, we use both the number of general promotions and bonus buy promotions and interact them with the relative proximity of the two products with respect to sugar content, fiber content, and sogginess. A detailed description of how these instruments are computed is provided in Appendix B.1. We regard these instruments as the most important for identifying the demographic interaction coefficients.

Our GMM estimate for the demand parameters minimizes the following objective function

$$\hat{\theta}_D = \arg\min_{\theta} \Delta\xi(\theta)' Z_D \hat{W}_D^{-1} Z_D' \Delta\xi(\theta), \tag{12}$$

where $\hat{W}_D^{-1}$ is an estimate of the efficient weighting matrix

$$W_D^{-1} = E[Z_D' \Delta\xi(\theta_{D0}) \Delta\xi(\theta_{D0})' Z_D]^{-1}$$

based on parameter estimates obtained from a first-stage estimation with 2SLS weighting matrix $E[Z_D' Z_D]^{-1}$. As proposed by Nevo (2001), we profile out all linear parameters contained in δ so that we have to optimize numerically only over the nonlinear (random) coefficients.

Supply estimation For the estimation of the marginal cost parameters γ and the conduct parameters λ, we generalize the algorithm by BLP to allow for the profit internalization matrix Ω to be estimated rather than assumed.

For a given parameter guess for the supply side parameters $\theta_S = (\gamma, \lambda)$, we solve the stacked first-order conditions, given by Equation (8), for the unobserved cost shock $\Delta\omega$ for each brand, store and month

$$\Delta\omega(\theta_S, \hat{\theta}_D) = p^w - \tilde{mc}(\theta_S) - \Omega^{-1}(\theta_S, \hat{\theta}_D) s.$$

Similar to our demand estimation, we exploit orthogonality conditions between the (transitory) structural cost term $\Delta\omega$ and a set of instruments Z_S. The moment conditions of the supply model can be written as

$$E[Z_S' \Delta\omega(\theta_{S0}, \hat{\theta}_D)] = 0. \tag{13}$$

Our supply side instruments consist of the following variables. First, they include brand dummies, month-year dummies, and store fixed effects to control for persistent cost differences, for example, due to different delivery costs for different locations or seasonal effects. Second, we include exogenous cost shifters. In our main model, we include only the electricity price in the Midwest region to avoid quasi-collinearity problems between different commodity prices. Finally, we include products' relative distance in the characteristics space interacted with rivals' promotion intensity. While the first two sets of moments identify the parameters of the marginal cost function, the last one identifies the conduct matrix. The objective function of our supply side estimation is given by

$$\hat{\theta}_S = \arg\min_{\theta_S} \Delta\omega(\theta_S, \hat{\theta}_D) Z_S' \hat{W}_S^{-1} Z_S' \Delta\omega(\theta_S, \hat{\theta}_D), \tag{14}$$

where $\hat{W}_S$ is an estimate of the asymptotically efficient weighting matrix based on parameters obtained from the first stage estimation using the 2SLS weighting matrix. As for the demand estimation, we profile out the linear parameters contained in the marginal cost function and search nonlinearly only for the conduct parameters.

5 Results

5.1 Demand Estimates

Table 3 displays the estimation results for our main demand specification. We include mean parameters for a constant, price, sogginess, sugar content, fiber content, and the total number of a brand's promotions in a given market. Furthermore, we interact consumer demographics with observed product characteristics. Specifically, we interact a dummy for households with small children (less than 10 years old) with the preference for sugar and a consumer's income with preferences for price and fiber content.

All of our demand coefficients are precisely estimated and highly significant, and the signs of the estimates for mean preferences seem reasonable. The price coefficient is highly negative, and ceteris paribus, consumers prefer cereals with higher sugar and less fiber content. Our estimated price-income coefficient is positive and significant indicating that high-income consumers are less price-sensitive. Households with small children have a stronger preference for cereals with a higher sugar content, which is consistent with popular kids' cereals having higher sugar content. Finally, the demand for fiber in cereal is negatively correlated with income, potentially because high-income consumers prefer to consume fiber from other food sources.

We experimented extensively with alternative demand specifications that include addi-

Table 3: Demand Estimates: Main Specification

	Mean	Children	Income
Constant	-5.2591^{***}		
	(0.0226)		
Price	-18.0342^{***}		1.4218^{***}
	(1.4124)		(0.2358)
Sogginess	0.3516^{***}		
	(0.0042)		
Sugar	2.0950^{***}	6.2271^{***}	
	(0.0337)	(0.9572)	
Fiber	-2.4320^{***}		-0.2305^{***}
	(0.0193)		(0.0516)
Promotion	0.5059^{***}		
	(0.0651)		

Notes: The estimation includes product- and month-year fixed effects and a linear-quadratic time trend. Standard errors are in parentheses. Number of observations: 96512.

tional demographic interactions and normally distributed random coefficients. The results are qualitatively similar. In particular, the implied price elasticities, which are the most important output of our demand model, are very similar to those of our main specification. However, larger demand models generally resulted in higher standard errors for some of the additional parameters, especially for the normally distributed random coefficients.

Table 13 and Table 14 show the median price elasticities over all markets for our main demand specification. The own-price elasticities are highly negative for all products. Moreover, our estimated substitution patterns exhibit significant variation across brands. The median cross-price elasticities are all positive, which is consistent with products being imperfect substitutes. Our estimates reveal that the cross-price elasticities are particularly high between Kellogg's' or General Mills' products, between the signature products of these two firms, and for kids cereals in general. Overall, our estimated substitution patterns are relatively similar to those of previous demand studies on the cereal industry using similar models but different instruments, such as Nevo (2001).

In general, we judge our model to be economically meaningful and to have a good fit with the observed data. The distribution of implied marginal costs based on the estimated demand elasticities seems reasonable. For example, under hypothetical Bertrand-Nash pricing our model predicts negative marginal costs for only 5 of our 96,512 observations. We further illustrate that the effect of the structural error terms ξ is small and not systematic. A series of figures in Appendix D shows that in general, when setting the ξ-errors to zero, our model predictions are close to the observed data on several levels. For example, our graphs suggest that when we predict aggregate sales for the whole Chicago area or for specific stores or market shares of individual brands, our prediction error is modest and non-systematic.

5.2 Supply Estimates

On the supply side, we focus on two different specifications. In our "small" model, we estimate 3 conduct parameters that reflect the level of conduct in each period, i.e., one parameter pre-merger, one post-merger, and one for the price war period. For this model specification we impose symmetry across all firms, such that each firm internalizes every rival's profit to the same degree. In our "large" model we let the conduct vary across firms. For each period (pre-merger, post-merger, and price-war), we estimate two distinct parameters capturing the potentially different internalization behavior of the two largest firms, Kellogg's and General Mills, and the smaller firms, i.e., Post, Nabisco, Ralston, and Quaker. Consequently, our large model estimates 6 conduct parameters. This specification allows us to capture the fact that industry leaders might have very different incentives to internalize rival firms' profits than smaller competitors. However, we remain agnostic about which type of firm cooperates more.

For our small model, there is significant internalization between firms pre-merger, with an estimate of 0.465. Intuitively, this parameter indicates that a firm values US-$ 1 profit of a rival firm as much as US-$ 0.465 of its own profits. This parameter further increases to 0.814 following the merger. Consistent with our descriptive evidence, in the price-war period, the industry conduct drastically decreases, with an estimated conduct parameter that is very close to zero. While the pre- and post-merger parameters are highly significant, the conduct parameter in the price war period is not significantly different from zero.

When allowing for heterogeneity in the internalization behavior of different firms, we find considerable differences across firms. Pre-merger, the small firms already internalize rivals' profits substantially more than the large firms (General Mills and Kellogg's). Following the merger, the degree of internalization increases for all firms. It remains higher for small firms, whose post-merger behavior is consistent with full industry profit internalization. During the price war period, the estimated parameters revert to close to 0 for all firms, indicating behavior that is consistent with Bertrand-Nash price competition during the price war. These results are broadly consistent with the descriptive and reduced form evidence presented in Section 2.

Table 16 presents the estimation results when accounting for a synergy dummy in the merging firms' cost function following the merger. The estimates are very similar to those of the baseline specification that abstracts from cost synergies.[19]

Table 4: Conduct Estimates: Model Comparison

	Small Model			Large Model		
	Pre-merger	Post-merger	Price War	Pre-merger	Post-merger	Price War
All Firms	0.4650***	0.8141***	0.0000			
	(0.0201)	(0.0298)	(0.0007)			
Large Firms				0.3639***	0.7577***	0.0170
				(0.1029)	(0.0495)	(0.0846)
Small Firms				0.7875***	0.9999***	0.0110
				(0.0948)	(0.0014)	(0.0597)

Notes: The table entries reflect the conduct estimates for both the small and the large conduct specification. Standard errors are in parentheses, and account for two-step estimation. Number of observations: 96512.

Table 5 summarizes the marginal cost estimates for our two conduct specifications, and under the assumption of multi-product Nash pricing. In the cost function, we account for product-fixed effects, month-year (seasonal) dummies, store fixed effects, a time trend, and electricity prices as a proxy for aggregate production costs. In a robustness check, we further incorporate a merging firms-post-merger dummy to allow for potential synergies. The time trend and price of electricity have a positive sign but are insignificant. The insignificant time trend is consistent with production processes being relatively constant over time. We experimented with several additional cost shifters, for example commodity (wheat, corn, rice, oats, and sugar) spot prices interacted with a brand's specific grain content. These specifications yielded very similar results. However, we did not obtain significant coefficients on the additional variables, most likely because of little variation in commodity prices, or because of potential commodity price hedging by manufacturers which makes observed spot prices only weak proxies for manufacturers' economic marginal cost shifters. In contrast, most of our fixed effects are highly significant and robust, explaining a large portion (62%) of the variation in marginal costs across brands, stores, and time.

For the large specification, the signs and absolute magnitudes of the parameters are similar to the small specification. The cost parameters do not change considerably when including a synergy dummy for the merging firms following the merger, as shown in Table 17. The results indicate a very small and insignificant decrease in the merging firms' marginal cost.

A very important and policy-relevant issue is to determine to what extent industry conduct translates into the markups of individual products. Table 6 compares product-specific median

[19]When interpreting our results with respect to policy recommendations, two caveats should be noted. First, we do not suggest that such estimates necessarily provide evidence that cereal manufacturers violated antitrust laws. Second, we do not claim that the merger or the price war actually caused the shifts in industry conduct.

Table 5: Marginal Cost Estimates: Model Comparison

	Small model	Large model	MP-Nash
Time Trend	0.0015	0.0026	
	(0.0039)	(0.0018)	
Elec. Price	0.0005	0.0008	
	(0.0068)	(0.0025)	
Median MC	0.085	0.084	0.099
Mean MC	0.082	0.081	0.095
St. Dev. MC	0.027	0.028	0.028
Median PCM	0.447	0.453	0.380
Mean PCM	0.447	0.453	0.392
St. Dev. PCM	0.094	0.095	0.080
Med. PCM pre	0.433	0.440	0.383
Med. PCM post	0.466	0.471	0.376
Med. PCM pw	0.389	0.391	0.389

Notes: The table entries reflect the marginal cost estimates from our small conduct model, the large conduct model, and under Bertrand-Nash pricing. For time-specific median price-cost margins, pre refers to pre-merger period, post to post-merger period, and pw to price war period. All estimations include brand fixed effects, store fixed effects and month-year fixed effects. Standard errors are in parentheses, and account for two-step estimation. Number of observations: 96512.

(across markets) price-cost margins for both of our conduct specifications. In addition, we compute the implied margins under multi-product Nash pricing which we use as a competitive benchmark.[20]

Both specifications lead to markups that are considerably higher than those implied by Bertrand-Nash pricing before the price war period. General Mills and Kellogg's products have higher markups in the small specification than in the large specification, while the opposite is true for the smaller firms' products. Again, this is likely to occur because General Mills and Kellogg's internalize their pricing externalities less in the large model than in the small model, while the opposite is true for the small firms.

The median marginal costs implied by our models are US-$ .099 per serving under multi-product Nash pricing, and US-$ .085 for both the small and large conduct specification. This implies that 17.6 percent of the median markups can be attributed to cooperative industry behavior in the small specification and 19.2 percent in the large specification averaged over the whole time span. In the small specification, the median margins increase from 13.1 percent in the pre-merger period to 23.9 percent in the post-merger period. The numbers are

[20] Recall that multi-product Nash pricing implies that each firm maximizes the profits of its own product portfolio, and all of the markups can be attributed to product differentiation rather than to cooperative behavior.

Table 6: Brand-specific PCM

	Small model	Large model	MP-Nash
NAB Shred Wheat	0.32	0.34	0.28
PO Raisin Bran	0.39	0.42	0.32
PO Grape Nuts	0.46	0.48	0.39
PO Honey Comb	0.45	0.48	0.36
GM RaisinNutBran	0.38	0.37	0.32
GM ApplCin Cheer	0.47	0.46	0.38
GM Wheaties	0.38	0.37	0.33
GM Cheerios	0.34	0.34	0.30
GM HonNut Cheer	0.41	0.41	0.35
GM Luck Charms	0.42	0.41	0.35
GM Tot CoFlakes	0.43	0.42	0.38
GM Trix	0.43	0.42	0.36
KE Froot Loops	0.47	0.46	0.41
KE Special K	0.45	0.45	0.41
KE Frost Flakes	0.57	0.56	0.50
KE Corn Pops	0.52	0.51	0.46
KE Raisin Bran	0.42	0.42	0.38
KE Corn Flakes	0.64	0.63	0.59
KE Honey Smacks	0.51	0.50	0.43
KE Crispix	0.48	0.48	0.44
KE Rice Krispies	0.53	0.52	0.48
RAL Chex	0.40	0.43	0.34
RAL Wheat Chex	0.36	0.39	0.30
RA Rice Chex	0.46	0.48	0.39
QU Quaker Oats	0.49	0.53	0.40
QU Capn Crunch	0.50	0.55	0.40

Notes: The table entries reflect the brand-specific median (across markets) price-cost margins for both the small and large model specification, and for multi-product Bertrand-Nash pricing.

slightly higher for the large model, with an implied increase from 14.6 percent pre-merger to 25.0 percent in the post-merger period.

Table 15 displays the median price-cost margins decomposed for both different brands and different time periods. Consistent with the model parameter estimates, most products experience considerable changes in markups over time.

Several figures in Appendix D illustrate that our supply estimates have a good fit with the observed data. Similarly to our demand model, we compare observed wholesale prices to the predictions from our model when setting all structural cost errors ω to zero. The graphs suggest that both wholesale price averaged at the store-level and brand-specific wholesale prices are explained well, and prediction errors seem reasonably small and non-systematic.

6 Counterfactual Simulations

In this section, we use our estimated parameters from the structural model to simulate how two different changes in the underlying industry conduct would affect consumer surplus and manufacturers' pricing behavior.

First, we examine how these two measures would change if the firms were competing a la multi-product Bertrand-Nash pricing before the price war. Table 7 shows the associated results. As a measure of consumer surplus, we estimate the compensating variation, i.e., the dollar value for which consumers would have been equally well off in both the observed industry state and the counterfactual simulation. Both the small and large conduct model

Table 7: Counterfactual Simulation 1: Change to Multi-product Nash Pricing

	Small Model		Large Model	
	Pre-merger	Post-merger	Pre-merger	Post-merger
Δ consumer surplus (in US-$ mio.)	5.0	12.8	5.2	13.1
Δ consumer surplus (in %)	11.3	19.6	11.8	20.2
Δprice All Firms (in %)	−6.0	−11.0	−5.5	−11.3
Δprice GM (in %)	−5.5	−10.9	−4.4	−10.2
Δprice RAL (in %)	−6.9	−12.1	−11.5	−14.9
Δprice KEL (in %)	−5.3	−9.5	−4.3	−9.1
Δprice POSNAB (in %)	−7.8	−11.7	−12.6	−14.4
Δprice QUA (in %)	−9.9	−18.2	−16.5	−22.1

Notes: The table entries reflect the results from the counterfactual simulations for both the small and the large conduct specification. The simulations compute the changes in consumer surplus and wholesale prices before the price war period when all firms play according to multi-product Nash pricing instead of the estimated conduct.

yield relatively similar results. We find that if firms had played according to multi-product Bertrand-Nash pricing instead of our estimated conduct, before the price war, consumers would have been between 17.8 and 18.1 million dollars better off. These numbers translate into roughly 3.45 million dollars of savings per year for consumers who purchase cereals at DFF stores in the Chicago area, which we believe is a plausible estimate. Consistent with the changes in industry conduct over time, the counterfactual simulations indicate that under multi-product Nash pricing, the median wholesale prices across all firms would have been between 5.5 and 6.0 percent lower in the pre-merger period and between 11.0 and 11.3 percent lower in the post merger period. When considering firm-specific prices, we find that Kellogg's would have had the lowest predicted wholesale price decrease, while the decrease would be significantly larges for Post-Nabisco, Quaker, and Ralston.

Table 8: Counterfactual Simulation 2: Change when there is no price war

	Small Model No price war	Large Model No price war
Δ consumer surplus (in US-$ mio.)	-1.9	-1.4
Δ consumer surplus (in %)	-16.6	-12.5
Δprice All Firms (in %)	10.3	8.0
Δprice GM (in %)	10.7	9.5
Δprice RAL (in %)	11.3	10.6
Δprice KEL (in %)	8.1	3.9
Δprice POSNAB (in %)	11.7	11.7
Δprice QUA (in %)	18.6	17.7

Notes: The table entries reflect the results from the counterfactual simulations for both the small and the large conduct specification. The simulations compute the changes in consumer surplus and wholesale prices when all firms continue playing according to the before price-war conduct instead of changing their conduct strategy in the price-war period.

Second, we examine how consumer welfare and pricing would have changed if the price war had never occurred, i.e., if the conduct had remained the same as before the price war period. Table 8 shows the associated results. We find that without the price war, consumers would have been between 1.4 and 1.9 million dollars worse off during this period, which spans the last 6 months of our sample. The median wholesale price responses of each manufacturer differ between the small and large conduct models. While for the small model, the predicted price responses are relatively homogeneous, for the large model, the wholesale price responses are more heterogeneous and generally higher for small firms than for large firms. This is fully consistent with the conduct parameters for small firms in the large model, which are higher in the pre-price war periods than those for large firms.

7 Conclusion

In this paper, we estimate the evolution of competition in the RTE cereal industry using a structural model of demand and supply. Our empirical strategy is flexible enough to accommodate detailed patterns of industry conduct; in particular, we allow levels of conduct to vary both across time and firms.

To overcome the identification problem of separating marginal costs from industry conduct, we construct novel instruments that interact measures of products' relative isolation in the characteristics space with data on rival firms' temporary and market-specific promotional activities. Intuitively, our identification of the supply parameters is based on the idea that a firm's markups react much more strongly to the promotions of a competing product that is close in the characteristics space than to those of a more distant product; and this relationship should be stronger the more competitive the industry is.

Our empirical strategy has several attractive features that allow it to be applied to many other industries. First, it does not rely on exogenous industry shocks, such as ownership changes, to identify industry conduct. Second, our instruments can be used even if there is no product entry or exit during the sample period. Third, the required data are available in many standard data sets for a broad range of consumer goods industries. Finally, a series of weak identification tests indicates that our instruments indeed are very powerful for identifying flexible patterns of industry conduct in contrast to many commonly used BLP-style instruments.

We use our model to shed new light on two important industry events during the 1990s: first, the Post-Nabisco merger and second, a period of large wholesale price cuts in 1996. Our estimation results suggest that in the beginning of our sample period in 1991, the industry was characterized by moderate levels of price coordination which increased significantly for all firms after the Post-Nabisco merger in January 1993. For both time periods, our model predicts price-cost margins that are significantly higher than those implied by multi-product Nash pricing. In particular, the median manufacturer margins over multi-product Bertrand-Nash pricing in these periods are 13.1 and 23.9 percent, respectively. These numbers suggest that a significant percentage of the median manufacturer markups can be attributed to cooperative industry behavior during the first half of the 1990s. Our conduct estimates for the last 6 months of our sample period are consistent with a shift in firms' behavior towards multi-product Nash pricing. Our results thus suggest that while product differentiation alone can explain the largest portion of the price-cost margins, for a long time cooperative behavior further increased markups even more. Moreover, we find that such behavior was more prevalent for small firms than for large firms.

Recently, there has been an increased interest in the evolution of markups over time from a macroeconomic perspective. De Loecker and Eeckhout (2017) document a substantial increase in markups from 1980 onwards for the US economy by using a production function approach. They attribute this pattern mainly to a sharp increase in the markups of already high-markup firms within the different industries. Our approach can be seen as complementary to this literature. By focusing on estimating competitive interactions between firms within an industry, one can gain detailed insights into the extent to which potentially heterogeneous conduct and differentiated consumer preferences can explain firms' markups.

Our model can be readily applied to estimate supply side patterns in many important industries because many standard data sets contain the information required for our estimation strategy. Comparing estimated conduct levels across industries can lead to a better understanding of the determinants of anti-competitive firm behavior, which is still a relatively open question with important implications for competition policy.

References

Ackerberg, D. A., Caves, K. and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, **83** (6), 2411–2451.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Appel, M., Labarre, R. and Radulovic, D. (2004). On accelerated random search. *SIAM Journal on Optimization*, **14** (3), 708–731.

Ashenfelter, O. C., Hosken, D. S. and Weinberg, M. C. (2013). The price effects of a large merger of manufacturers: A case study of Maytag-Whirlpool. *American Economic Journal: Economic Policy*, **5** (1), 239–261.

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

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

Björnerstedt, J. and Verboven, F. (2016). Does merger simulation work? evidence from the Swedish analgesics market. *American Economic Journal: Applied Economics*, **8** (3), 125–164.

Black, M., Crawford, G., Lu, S. and White, H. (2004). A virtual stakes approach to measuring competition in product markets. *mimeo*.

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.

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. (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.

Cragg, J. G. and Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, **9** (02), 222–240.

Crawford, G., Lee, R., Whinston, M. and Yurukoglu, A. (2017). The welfare effects of vertical integration in multichannel televistion. *Econometrica, forthcoming*.

De Loecker, J. and Eeckhout, J. (2017). The rise of market power and the macroeconomic implications. *NBER Working Paper 23687*.

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

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

Gasmi, F., Laffont, J. J. and Vuong, Q. (1992). Econometric analysis of collusive behavior in 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.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, **133** (1), 97–126.

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.

Lal, R. and Matutes, C. (1994). Retail pricing and advertising strategies. *Journal of Business*, pp. 345–370.

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

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

Michel, C. (2017). Estimation of within-firm pricing coordination in differentiated product industries. *mimeo*.

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

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

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

— (2000b). 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.

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.

Scherer, F. M. (1979). The welfare economics of product variety: an application to the ready-to-eat cereals industry. *The Journal of Industrial Economics*, pp. 113–134.

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

Slade, M. E. (1995). Product rivalry with multiple strategic weapons: An analysis of price and advertising competition. *Journal of Economics & Management Strategy*, **4** (3), 445–476.

Stock, J. H., Wright, J. H. and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, **20** (4), 518–529.

Sudhir, K. (2001). Structural analysis of manufacturer pricing in the presence of a strategic retailer. *Marketing Science*, **20** (3), 244–264.

Sullivan, C. (2017). *Three Essays on Product Collusion*. Ph.D. thesis, University of Michigan.

Sweeting, A. (2010). The effects of mergers on product positioning: evidence from the music radio industry. *The RAND Journal of Economics*, **41** (2), 372–397.

Train, K. (2000). Halton sequences for mixed logit. *mimeo*.

Varian, H. R. (1980). A model of sales. *The American Economic Review*, **70** (4), 651–659.

Villas-Boas, J. M. (1995). Models of competitive price promotions: Some empirical evidence from the coffee and saltine crackers markets. *Journal of Economics & Management Strategy*, **4** (1), 85–107.

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

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

Wright, J. H. (2003). Detecting lack of identification in GMM. *Econometric Theory*, **19** (02), 322–330.

A Reduced Form Estimations and Additional Details

A.1 Reduced Form Estimation Equations

In this section, we provide the estimation equation for our baseline reduced form regressions of the log wholesale price $\log(p^w_{ist})$ in Table 2. $PromoG$ denotes the number of general promotions for a given product in a given market. The superscript indicates which promotions are included. *own* considers only promotions for the same product, *firm* denotes the promotions of all other products owned by the same firm, and *rival* captures the number of rival firms' products' promotions. We also allow for different rival effects pre-merger (pre), post-merger before the price war ($post$), and during the price-war (pw). $\mathbb{1}^{POSNAB}_{Post}$, $\mathbb{1}^{nomerge}_{Post}$, and $\mathbb{1}_{PW}$ represent dummy variables for the merging firms post-merger, the non-merging firms post-merger, and a price war dummy for all firms, respectively. Furthermore, $Sales\_Tot$ indicates the total quantity of cereals sold in a given store and month. Finally, κ_i and κ_s denote brand and store fixed effects, respectively. Our baseline wholesale price equation can thus be written as

$$\begin{aligned}\log(p^w_{ist}) =& \beta_1 PromoG^{own}_{ist} + \beta_2 PromoG^{firm}_{ist} + \beta_3 PromoG^{riv,pre}_{ist} + \beta_4 PromoG^{riv,post}_{ist}\\ &+ \beta_5 PromoG^{riv,pw}_{ist} + \beta_6 \mathbb{1}^{POSNAB}_{Post} + \beta_7 \mathbb{1}^{nomerge}_{Post} + \beta_8 \mathbb{1}_{PW}\\ &+ \beta_9 Sales\_Total_{st} + \kappa_i + \kappa_s + \epsilon_{ist}\end{aligned}$$

where i, s, and t denote brands, stores, and months respectively. The second column in Table 2 and Table 9 substitutes the post-merger non-merging dummy with firm-specific post-merger dummies. The last two models add the *bonus buy* promotion variables, $PromoB$, in addition to the general promotion variables. The same models are estimated using the log retail price, $\log(p^w_{ist})$, as a dependent variable; see Table 9 for the results.

A.2 Market Share Generation

For our estimations, we include all package sizes between 10 and 32 ounces for the different products in our sample, and calculate aggregated quantities and the average price per ounce for each product. We obtain market shares for the inside goods by dividing aggregate quantities by our measure of the market size described in the main text. The remainder, i.e., 1 minus the sum of inside market shares per market, yields the market share of the outside

Table 9: Reduced form analysis: Retail prices

	(1) Baseline	(2) Firm detailed	(3) Baseline w/ BB	(4) Firm detailed w/ BB
Promo (own brand)	-0.1311***	-0.1316***	-0.1369***	-0.1365***
	(0.0015)	(0.0015)	(0.0015)	(0.0015)
Promo (same firm)	0.0070***	0.0063***	0.0052***	0.0056***
	(0.0003)	(0.0003)	(0.0002)	(0.0003)
Promo pre-merger (rivals)	-0.0007	0.0050***	0.0049***	0.0047***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Promo post-merger (rivals)	0.0061***	0.0030***	0.0022***	0.0022***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Promo price war (rivals)	0.0037***	0.0046***	0.0044***	0.0033***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Price war period	-0.0342***	-0.0163***	-0.0239***	-0.0228***
	(0.0011)	(0.0011)	(0.0012)	(0.0012)
Post-merger KEL		0.0690***		0.0775***
		(0.0019)		(0.0025)
Post-merger RAL		0.0874***		0.1000***
		(0.0024)		(0.0033)
Post-merger QUA		0.0027		0.0203***
		(0.0027)		(0.0036)
Post-merger GMI		-0.0013		0.0024
		(0.0025)		(0.0032)
Post-merger POSTNAB		0.1012***	0.1167***	0.1136***
		(0.0023)	(0.0029)	(0.0030)
BB (own brand)			-0.0340***	-0.0343***
			(0.0015)	(0.0015)
BB (same firm)			0.0010***	0.0003
			(0.0003)	(0.0003)
BB pre-merger (rivals)			0.0006***	0.0006***
			(0.0001)	(0.0001)
BB post-merger (rivals)			-0.0005*	-0.0002
			(0.0002)	(0.0002)
BB price war (rivals)			0.0034***	0.0045***
			(0.0003)	(0.0003)
Post-merger non-merging			0.0521***	
			(0.0027)	
Observations	96512	96512	96512	96512
R-square	0.75	0.76	0.76	0.77

Notes: All estimations include brand- and store- fixed effecs. Columns (2) and (4) allow for post-merger reaction to differ across firms. Promo (same firm) describes the number of promotions of other products in a market that belong to the same firm. Promo (rivals) describes the number of promotions of other products in a market that do not belong to the same firm. Pre-, post, and PW stand for pre-merger, post-merger, and price war period, respectively. BB stands for bonusbuy and coupon promotions, while all promo variables without BB reflect general / price reduction promotions.

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

good. We exclude five weeks in 1995 from our sample because of a substantial amount of missing data in the DFF database during these weeks.

B Construction of Instruments and Testing for Weak Identification

B.1 Details of Computation of Instruments

Our instruments for industry conduct consist of two parts. First, we construct proximity measures that describe how close each pair of products is in the characteristics space. Second, we compute a market-specific measure of the intensity of each product's rivals' promotional activities.

The first component is motivated by the fact that the effect of a rival product's promotions on another product's demand should strongly depend on the proximity of the two products in the characteristics space. A close rival product going on sale will exert much more competitive pressure than a very distant product on sale. For example, demand for Post's Raisin Bran should be affected much more by promotions of Kellogg's Raisin Bran than by promotions for Quaker Oats. For instance, the degree of closeness can be described by whether the difference between two products in terms of a specific characteristic, for example, sugar content, is in the first, second, third etc. decile of all differences in terms of that characteristic.[21]

More specifically, define $d_{ij}^x \equiv x_i - x_j, i \neq j$ as the difference in the product characteristic x between products i and j. Let $C^x = \{c_1^x, .., c_v^x\}$ denote v equally spaced percentiles of the entire distribution of characteristics differences d_{ij}^x with respect to product characteristic x.

The second part of our instruments is motivated by the fact that the pattern of promotions shifts the competitive pressure in a market. Denote by $PROMO_{it}^G$ and $PROMO_{it}^{BB}$ the number of general promotions and bonus buy promotions respectively for product i in market t. We compute our instruments by interacting these promotion measures with our proximity measures for the corresponding product-pairs. For a continuous product characteristic x, our instruments $z_{jt}^{x,k,w}$ can then be computed as

$$z_{jt}^{x,k,w} = \sum_{i \neq j}^{J_t \setminus \mathcal{F}(j)} \mathbb{1}\left(d_{ij,t}^x < c_k^x\right) \cdot PROMO_{it}^w,$$

where k indicates the percentile of closeness, w denotes the type of promotions, i.e., either *general* or *bonus buy*, and $\mathcal{F}(j)$ is the product portfolio of the firm owning brand j. This results in $v*2$ instruments per continuous product characteristic x. Intuitively, our instruments count the number of different types of promotions conducted by rivals in a given market but

[21] Similar ideas underlie the construction of *differentiation instruments* to identify consumer substitution patterns in Gandhi and Houde (2015).

only consider promotions of rival products that are close according to our definition above. The analogous instruments for a binary product characteristic, for example, whether a cereal is soggy in milk or not, can be computed as

$$z_{jt}^{x,k,w} = \sum_{i \neq j}^{J_t \setminus \mathcal{F}(j)} \mathbb{1}(d_{ij,t}^{x} = k) \cdot PROMO_{it}^{w},$$

which leads to $2 * 2$ instruments per binary product characteristic.

In our application, we use sugar and fiber content as continuous characteristics, and the binary variable sogginess. For the sugar- and fiber-based instruments we use the 33.33 and 66.67 percentiles of the respective distribution of characteristic differences. Computing our instruments based on other closeness definitions, such as using deciles instead of terciles, resulted in very similar estimation results.

B.2 Weak Identification

In the following, we show that our instruments have power for identifying both demand and supply parameters. Compared to traditional first-stage diagnostics for linear IV regressions, testing for weak identification in our model is more complicated for several reasons. First, our models are highly non-linear and contain multiple endogenous regressors. Second, even if instruments and endogenous regressors are correlated enough to result in a decently large F-statistic (larger than 10), the instruments can still be weak enough to result in very sensitive estimates and high standard errors.

In order to overcome the first problem, we adapt a testing procedure recently proposed by Gandhi and Houde (2015) for demand models. The main idea is to linearize the nonlinear BLP-model around the estimated parameter values using a first-order Taylor expansion. After the model is linearized, one can employ generalizations of the well-known F-statistics to test for identification of single parameters. While traditional F-tests test the null hypothesis of complete non-identification of a single parameter, rank deficiency tests as developed by Cragg and Donald (1993) and Kleibergen and Paap (2006) can be adopted to test for alternative hypotheses, such as underidentification or weak identification of single parameters or the model as a whole.

General procedure In the following, we describe a general procedure to test for various degrees of lack of identification and weak instruments based on Gandhi and Houde (2015). To the best of our knowledge, this procedure has so far not been used to test for weak identification of conduct parameters.

The starting point is a first-order Taylor expansion of the structural error $\kappa(\theta)$ as a function of the parameters around the true parameter vector θ_0

$$\kappa_{jt}(\theta) = \kappa_{jt}(\theta_0) + \sum_{k=1}^{K}(\theta_k - \theta_{0k})\frac{\partial \kappa_{jt}(\theta_0)}{\partial \theta_k} + \upsilon_{jt} \tag{15}$$

$$= \kappa_{jt}(\theta_0) + J_{jt}(\theta_0)b + \upsilon_{jt} \tag{16}$$

where J denotes the Jacobian stacking all the partial derivatives with respect to each parameter θ_k, b stacks the differences $\theta_k - \theta_{k0}$ and υ are higher-order residuals. When taking conditional expectations of the above equation with respect to the proposed instruments Z, $\mathbb{E}(\kappa(\theta_0)|Z)$ disappears and when evaluated at $\theta = \theta_0$ the Jacobian term becomes zero.

In order to have strong identification, we require $\mathbb{E}(\kappa(\theta)|Z)$ to be large for $\theta \neq \theta_0$. Therefore, we test whether the Jacobian of the objective function reacts strongly to the instruments (analogous to an F-test in linear GMM). Note that this test can be applied equally well to both demand and supply models.[22] For a given model, we proceed in the following steps.

1. Estimate the model using a set of instruments $A(Z)$ to get the parameter estimates $\hat{\theta}$.

2. Compute the Jacobian of the structural error κ evaluated at $\hat{\theta}$. For the linear parameters, the derivative has an analytical form. For nonlinear parameters, the derivatives have to computed numerically.

3. Run a linearized first-stage-regression for each dependent variable, i.e., for each endogenous regressor, on the exogenous regressors X and the excluded instruments $A(Z)$.

$$\frac{\partial \kappa_{jt}(\hat{\theta})}{\partial \theta_k} = X_{jt}\pi_{1k} + A_j(Z_t)\pi_{2k} + \epsilon_{jtk} \tag{17}$$

   In our demand model, there are K endogenous variables corresponding to the K partial derivatives $\frac{\partial \xi}{\partial \theta_k}$ of the structural demand shock with respect to the non-linear preference parameters. In our supply model, the number of nonlinear parameters is equal to the number of estimated conduct parameters.

4. Test joint significance of π_{2k} using an appropriate F-test for each of the K first-stage regressions. This step is a generalization, of standard F-tests in linear IV regressions. Wright (2003) shows that at the true parameter value θ_0, one can use the same test logic

[22] We present the test for a general non-linear model and apply the same procedure for testing weak identification of our demand and supply model. The only difference between the two is in the definition of the structural error κ and potentially the choice of the instruments $A(Z)$. In our demand and supply models κ corresponds to ξ and ω respectively.

for the linearized first-stage regressions. Moreover, he shows that the same remains valid when evaluating the test at $\hat{\theta}$. For example, the null hypothesis $H_0 : \pi_{2k} = 0$ corresponds to complete non-identification of θ_k.

An important question is which F-test to use in Step 4. Standard F-tests, as reported by most linear IV regression software packages, can provide a first starting point. However, in models with multiple endogenous regressors, conventional F-tests can easily result in falsely rejecting non-identification. Angrist and Pischke (2008) (heceforth AP) propose a modified F-statistic that corrects for the presence of multiple endogenous regressors by profiling out the effects of the other $K-1$ endogenous regressors, and using only the variation in the projection residual when running the first-stage regression.

While single equation F-tests provide insights on whether a particular endogenous regressor is correlated with our instruments, these F-statistics need not be informative about identification of the model as a whole. In order to test whether all first-stage regressions are jointly significant, we combine the first-stage coefficients of all K regressions into a $dim\big(A(Z)\big) \times K$ matrix Ψ. Underidentification of the model is equivalent to Ψ being rank-deficient. Therefore, a natural choice for the null hypothesis of underidentification is $H_0 : rk(\Psi) = K-1$. A convenient and robust way to test for rank deficiency is to analyze the smallest singular value of Ψ. If the smallest singular value is statistically different from zero, we can reject underidentification. This logic has been formalized by Cragg and Donald (1993) and Kleibergen and Paap (2006) (henceforth KP). Intuitively, testing the rank of Ψ is equivalent to testing the local GMM-identification condition which requires that the $K \times K$-matrix $\mathbb{E}\left[G_0' W G_0\right]$ with $G_0 = \frac{\partial g(\theta_0)}{\partial \theta}$ has full rank. Noting that in our models $g(\theta) = \kappa(\theta) \cdot Z$ yields $G_0 = Z' \frac{\partial \kappa(\theta)}{\partial \theta}$. The matrix of first-stage coefficients $\Psi = (Z'Z)^{-1} Z' \frac{\partial \kappa(\theta)}{\partial \theta}$ contains the same information as $[G_0' W G_0]$ (up to a scaling factor that does not affect the rank). Therefore, testing the rank of Ψ is equivalent to testing the local identification condition of our GMM model.

Even when we can reject underidentification of our model, i.e., Ψ has full rank, the model may still be weakly identified. Endogenous regressors and excluded instruments might be correlated but only weakly resulting in Ψ having full rank but being close to singular. In such a case, estimation is likely to perform poorly. For example, estimates will be very sensitive to the selection of moments and the objective function can have several local minima. A suitable statistic to examine this type of weak identification of the model is the Cragg-Donald Wald statistic. Stock *et al.* (2002) discuss several definitions of performing poorly in various settings. For our models, we focus on the maximum relative bias as a measure for the performance of our instruments. If the Cragg-Donald Wald statistic exceeds the critical value we can reject the null hypothesis that our IV estimator has a bias of more than 5% (or

10%, or 20%) compared to the OLS estimator.[23]

Weak identification on the demand side Table 10 summarizes the results for a battery of weak identification tests for our demand model.

Table 10: Weak IV/Identification Tests: Demand Model

	$\frac{\partial \xi}{\partial \alpha} = p$	$\frac{\partial \xi}{\partial \Pi_1}$	$\frac{\partial \xi}{\partial \Pi_2}$	$\frac{\partial \xi}{\partial \Pi_3}$
Robust F-statistic	672.52	103.84	480.01	234.52
Robust AP-F-statistic	48.65	89.84	49.62	70.43
KP χ^2-statistic	679.61			
KP χ^2-p-value	0.00			
KP F-statistic	38.90			

Notes: The Kleibergen-Paap (KP) χ^2-statistic tests the null hypothesis of underidentification. The KP F-statistic is a heteroskedasticity-robust version of the Cragg-Donald F-statistic testing the null hypothesis of weak identification.

All standard first-stage F-statistics are substantially larger than 10. When examining the robust AP F-test we see a substantial drop in the statistic; therefore, controlling for multiple endogenous regressors is important. All of the test statistics are larger than the critical values by at least a factor of 4. The KP-χ^2-statistic for underidentification is very large with a p-value of less than 0.00001. Therefore, we can strongly reject underidentification of the model. The Kleibergen-Paap F-statistic generalizes the F-statistic by Cragg and Donald (1993) to models with heteroskedastic error terms. The KP F-statistic for weak identification exceeds 38. The critical value for a model with one endogenous regressor and 25 excluded instruments is less than 20 if we take as a reference point that the maximum relative IV-OLS bias should not exceed 5% at the 5% significance level.[24] Consequently, we can not only reject underidentification but also weak identification of our demand model very strongly.

Weak identification on the supply side Table 11 and 12 summarize the results from testing for weak identification in our supply model. Table 11 focuses on the small model

[23] A minor practical problem is that the critical values tabulated by Stock *et al.* (2002) are only available for rather special cases such as having only up to 3 endogenous regressors. Both our demand and supply model contain more nonlinear parameters. Therefore, we cannot formally compare our Cragg-Donald Wald statistic to the appropriate critical values. In our experience, models that seem robust and reasonable, i.e., results in estimates that are not sensitive to minor changes in the moments and that have low standard errors, should result in substantially larger test statistics than the critical values tabulated by Stock *et al.* (2002) for one or two endogenous regressors. Therefore, we judge the practical problem of not having the critical values readily available as not crucial.

[24] This is a relatively conservative rejection criterion. Often the maximum IV-OLS relative bias tolerated is set to a reference level of 10%.

with 3 conduct parameters. Table 12 displays the results for the more detailed model with 6 conduct parameters.

Table 11: Weak IV/Identification Tests: Supply Model 1

	$\frac{\partial\omega}{\partial\lambda_1}$	$\frac{\partial\omega}{\partial\lambda_2}$	$\frac{\partial\omega}{\partial\lambda_3}$
Robust F-statistic	244.50	341.60	173.15
Robust AP-F-statistic	128.24	184.77	114.02
KP χ^2-statistic	839.99		
KP χ^2-p-value	0.00		
KP F-statistic	85.77		

Notes: The Kleibergen-Paap (KP) χ^2-statistic tests the null hypothesis of underidentification. The KP F-statistic is a heteroskedasticity-robust version of the Cragg-Donald F-statistic testing the null hypothesis of weak identification.

Table 12: Weak IV/Identification Tests: Supply Model 2

	$\frac{\partial\omega}{\partial\lambda_1}$	$\frac{\partial\omega}{\partial\lambda_2}$	$\frac{\partial\omega}{\partial\lambda_3}$	$\frac{\partial\omega}{\partial\lambda_4}$	$\frac{\partial\omega}{\partial\lambda_5}$	$\frac{\partial\omega}{\partial\lambda_6}$
Robust F-statistic	286.90	301.11	220.51	305.88	134.00	588.53
Robust AP-F-statistic	172.69	51.47	68.04	34.90	54.46	138.90
KP χ^2-statistic	157.28					
KP χ^2-p-value	0.00					
KP F-statistic	15.80					

Notes: The Kleibergen-Paap (KP) χ^2-statistic tests the null hypothesis of underidentification. The KP F-statistic is a heteroskedasticity-robust version of the Cragg-Donald F-statistic testing the null hypothesis of weak identification.

First, we investigate the F-statistic of classical first stage regressions. We regress the endogenous variables, i.e., the derivatives of the structural cost shock ω with respect to the nonlinear (conduct) parameters, on our excluded instruments which are based on rivals' promotion activities interacted with relative proximity of products in the characteristics space. In all cases, the F-statistics massively exceed the rule-of-thumb critical value of 10. Next, we report F-statistics that take into account the presence of multiple endogenous regressors as initially proposed by Angrist and Pischke (2008). While the F-statistics generally become smaller by a factor of 2 in the small model and by a factor of 2 to 6 in the large model, they still consistently exceed the critical values by several orders of magnitude. We take this as strong evidence that our instruments shift the endogenous regressors substantially and therefore constitute strong instruments.

Finally, we analyze rank deficiency of the full matrix of first stage coefficients. For both supply models, we can strongly reject the null hypothesis of underidentification with KP-statistics of 840 and 157, respectively, resulting in p-values of less than 0.0001 for both models.

We also look at the KP-F-statistic, which is a heteroskedasticity-robust version of the Cragg-Donald Wald-statistic for weak identification. For the small model with 3 conduct parameters the test statistic is 86. This is substantially larger than the critical values computed by Stock *et al.* (2002) even in conservative cases such as when we allow for a 5% maximal IV bias relative to NLS at the 5%-significance level. For our large supply model with 6 conduct parameters, the KP F- statistic is substantially smaller pointing to more detailed conduct models being more difficult to identify with a fixed set of instruments. Nonetheless, the test statistic significantly exceeds the rule-of-thumb critical value of 10. Therefore, we conclude that even our larger supply model does not suffer from weak identification problems.

C Additional Estimation Results

Demand elasticities In random coefficient logit models, consumers' own- and cross-price elasticities can be computed according to the following formulas.

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

Table 13: Median Elasticities RC 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	-3.672	0.023	0.026	0.006	0.015	0.011	0.023	0.068	0.040	0.017	0.012	0.009	0.014
PO Raisin Bran	0.035	-3.182	0.030	0.011	0.023	0.024	0.025	0.065	0.068	0.033	0.012	0.018	0.030
PO Grape Nuts	0.032	0.024	-2.640	0.009	0.017	0.016	0.022	0.065	0.048	0.024	0.014	0.013	0.021
PO Honey Comb	0.014	0.017	0.016	-2.841	0.016	0.031	0.015	0.043	0.063	0.048	0.022	0.034	0.051
GM RaisinNutBran	0.028	0.028	0.025	0.013	-3.408	0.024	0.022	0.061	0.064	0.036	0.016	0.021	0.033
GM ApplCin Cheer	0.017	0.024	0.020	0.021	0.020	-2.889	0.019	0.048	0.077	0.055	0.020	0.037	0.056
GM Wheaties	0.030	0.022	0.024	0.009	0.016	0.016	-3.285	0.064	0.048	0.025	0.015	0.014	0.022
GM Cheerios	0.028	0.018	0.022	0.008	0.014	0.013	0.020	-3.524	0.041	0.021	0.017	0.012	0.018
GM HonNut Cheer	0.023	0.025	0.022	0.016	0.019	0.028	0.021	0.056	-3.136	0.042	0.018	0.026	0.041
GM Luck Charms	0.016	0.021	0.018	0.020	0.018	0.034	0.018	0.047	0.070	-3.114	0.021	0.034	0.053
GM Tot CoFlakes	0.014	0.009	0.013	0.011	0.009	0.015	0.013	0.044	0.036	0.025	-2.819	0.017	0.024
GM Trix	0.013	0.017	0.015	0.022	0.016	0.034	0.015	0.041	0.067	0.053	0.022	-3.056	0.057
KE Froot Loops	0.015	0.021	0.017	0.022	0.018	0.037	0.017	0.044	0.073	0.056	0.021	0.039	-2.897
KE Special K	0.015	0.010	0.014	0.012	0.010	0.016	0.013	0.046	0.038	0.027	0.024	0.018	0.026
KE Frost Flakes	0.017	0.022	0.019	0.020	0.018	0.033	0.018	0.048	0.070	0.049	0.020	0.033	0.051
KE Corn Pops	0.012	0.018	0.015	0.024	0.017	0.038	0.015	0.039	0.071	0.058	0.022	0.042	0.063
KE Raisin Bran	0.033	0.037	0.029	0.012	0.024	0.025	0.025	0.064	0.070	0.036	0.013	0.019	0.033
KE Corn Flakes	0.024	0.016	0.020	0.010	0.013	0.016	0.018	0.058	0.044	0.025	0.019	0.015	0.023
KE Honey Smacks	0.016	0.029	0.021	0.026	0.024	0.046	0.020	0.045	0.093	0.068	0.020	0.046	0.072
KE Crispix	0.014	0.009	0.013	0.012	0.010	0.015	0.013	0.044	0.036	0.026	0.024	0.017	0.025
KE Rice Krispies	0.016	0.010	0.015	0.011	0.010	0.015	0.014	0.048	0.038	0.025	0.023	0.016	0.024
RAL Chex	0.017	0.011	0.015	0.011	0.011	0.015	0.014	0.049	0.038	0.025	0.023	0.016	0.024
RAL Wheat Chex	0.032	0.023	0.025	0.008	0.016	0.015	0.022	0.065	0.046	0.022	0.014	0.012	0.019
RA Rice Chex	0.014	0.008	0.012	0.011	0.009	0.013	0.012	0.043	0.033	0.024	0.025	0.016	0.023
QU Quaker Oats	0.028	0.028	0.025	0.012	0.020	0.023	0.022	0.061	0.062	0.034	0.015	0.020	0.032
QU Capn Crunch	0.017	0.024	0.019	0.021	0.020	0.037	0.019	0.047	0.077	0.055	0.020	0.037	0.057
Outside good	0.031	0.019	0.023	0.008	0.014	0.012	0.021	0.064	0.040	0.019	0.015	0.011	0.017

Notes: Cell entries i (indexing row), j (indexing column), give the percent change in market share of brand i with a one percent change in the price of j. Each entry represents the median of the elasticities from the 3712 markets.

Table 14: Median Elasticities RC 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	0.020	0.040	0.010	0.051	0.036	0.004	0.010	0.035	0.005	0.007	0.006	0.022	0.014
PO Raisin Bran	0.020	0.076	0.022	0.087	0.038	0.011	0.010	0.034	0.005	0.008	0.005	0.033	0.029
PO Grape Nuts	0.023	0.053	0.015	0.056	0.037	0.006	0.011	0.038	0.005	0.007	0.006	0.024	0.019
PO Honey Comb	0.035	0.099	0.046	0.042	0.034	0.014	0.018	0.052	0.007	0.004	0.010	0.022	0.040
GM RaisinNutBran	0.026	0.078	0.026	0.067	0.038	0.011	0.013	0.041	0.006	0.007	0.007	0.028	0.030
GM ApplCin Cheer	0.033	0.115	0.049	0.060	0.036	0.017	0.016	0.048	0.007	0.005	0.008	0.028	0.047
GM Wheaties	0.024	0.055	0.017	0.052	0.037	0.006	0.012	0.041	0.006	0.007	0.007	0.023	0.020
GM Cheerios	0.027	0.046	0.014	0.041	0.037	0.005	0.013	0.043	0.006	0.006	0.007	0.020	0.016
GM HonNut Cheer	0.029	0.090	0.033	0.061	0.038	0.013	0.014	0.045	0.006	0.006	0.008	0.027	0.035
GM Luck Charms	0.034	0.107	0.046	0.052	0.036	0.016	0.017	0.050	0.007	0.005	0.009	0.025	0.043
GM Tot CoFlakes	0.037	0.052	0.021	0.022	0.032	0.005	0.019	0.057	0.008	0.004	0.011	0.013	0.019
GM Trix	0.035	0.110	0.052	0.044	0.033	0.016	0.018	0.051	0.007	0.004	0.009	0.022	0.045
KE Froot Loops	0.034	0.116	0.053	0.051	0.035	0.018	0.017	0.049	0.007	0.004	0.009	0.025	0.048
KE Special K	-2.735	0.055	0.022	0.025	0.034	0.006	0.019	0.056	0.008	0.004	0.011	0.014	0.020
KE Frost Flakes	0.033	-2.295	0.044	0.054	0.036	0.015	0.016	0.049	0.007	0.005	0.009	0.026	0.042
KE Corn Pops	0.035	0.118	-2.617	0.045	0.033	0.018	0.017	0.050	0.007	0.004	0.009	0.023	0.049
KE Raisin Bran	0.022	0.081	0.025	-3.045	0.038	0.011	0.010	0.035	0.005	0.008	0.005	0.033	0.031
KE Corn Flakes	0.030	0.054	0.018	0.038	-1.886	0.006	0.015	0.048	0.007	0.006	0.009	0.019	0.019
KE Honey Smacks	0.033	0.142	0.063	0.072	0.037	-2.898	0.016	0.046	0.007	0.005	0.008	0.032	0.059
KE Crispix	0.036	0.053	0.021	0.022	0.033	0.005	-2.572	0.057	0.008	0.004	0.011	0.014	0.019
KE Rice Krispies	0.035	0.053	0.020	0.025	0.034	0.005	0.018	-2.297	0.008	0.004	0.011	0.015	0.019
RAL Chex	0.035	0.053	0.020	0.026	0.034	0.005	0.018	0.055	-2.966	0.004	0.011	0.015	0.019
RAL Wheat Chex	0.023	0.050	0.015	0.052	0.037	0.006	0.011	0.039	0.006	-3.331	0.007	0.023	0.018
RA Rice Chex	0.037	0.048	0.019	0.020	0.032	0.005	0.019	0.057	0.008	0.004	-2.546	0.012	0.017
QU Quaker Oats	0.025	0.076	0.025	0.065	0.038	0.010	0.012	0.041	0.006	0.007	0.007	-2.556	0.029
QU Capn Crunch	0.033	0.116	0.049	0.059	0.036	0.018	0.016	0.048	0.007	0.005	0.008	0.028	-2.550
Outside good	0.024	0.044	0.013	0.042	0.036	0.004	0.012	0.040	0.006	0.006	0.007	0.020	0.015

Notes: Cell entries i (indexing row), j (indexing column), give the percent change in market share of brand i with a one percent change in the price of j. Each entry represents the median of the elasticities from the 3712 markets.

Table 15: Time-Brand-specific PCM

	Small model Pre-merger	Small model Post-merger	Small model Price War	Large model Pre-merger	Large model Post-merger	Large model Price War	MP-Nash MP-Nash
NAB Shred Wheat	0.32	0.33	0.30	0.34	0.35	0.30	0.28
PO Raisin Bran	0.39	0.40	0.40	0.42	0.42	0.40	0.32
PO Grape Nuts	0.45	0.47	0.42	0.49	0.49	0.43	0.39
PO Honey Comb	0.43	0.47	0.38	0.48	0.50	0.39	0.36
GM RaisinNutBran	0.37	0.39	0.32	0.36	0.39	0.32	0.32
GM ApplCin Cheer	0.44	0.50	0.38	0.43	0.49	0.38	0.38
GM Wheaties	0.37	0.39	0.31	0.36	0.39	0.31	0.33
GM Cheerios	0.33	0.36	0.29	0.32	0.36	0.29	0.30
GM HonNut Cheer	0.39	0.44	0.34	0.38	0.44	0.34	0.35
GM Luck Charms	0.41	0.45	0.34	0.40	0.44	0.35	0.35
GM Tot CoFlakes	0.41	0.44	0.38	0.41	0.44	0.38	0.38
GM Trix	0.41	0.46	0.36	0.40	0.45	0.36	0.36
KE Froot Loops	0.45	0.49	0.45	0.45	0.49	0.45	0.41
KE Special K	0.45	0.46	0.41	0.44	0.46	0.41	0.41
KE Frost Flakes	0.56	0.59	0.53	0.55	0.58	0.54	0.50
KE Corn Pops	0.51	0.54	0.49	0.50	0.54	0.49	0.46
KE Raisin Bran	0.42	0.44	0.39	0.41	0.43	0.39	0.38
KE Corn Flakes	0.67	0.62	0.56	0.66	0.62	0.56	0.59
KE Honey Smacks	0.50	0.52	0.47	0.49	0.51	0.47	0.43
KE Crispix	0.48	0.49	0.44	0.47	0.48	0.44	0.44
KE Rice Krispies	0.53	0.53	0.48	0.53	0.53	0.48	0.48
RAL Chex	0.40	0.42	0.33	0.43	0.43	0.33	0.34
RAL Wheat Chex	0.36	0.37	0.28	0.39	0.39	0.28	0.30
RA Rice Chex	0.45	0.47	0.39	0.49	0.49	0.39	0.39
QU Quaker Oats	0.47	0.53	0.41	0.51	0.56	0.41	0.40
QU Capn Crunch	0.48	0.56	0.39	0.53	0.60	0.39	0.40

Notes: The table entries reflect the brand-specific median price-cost margins for both the small and large model specification for the pre-merger, post-merger, and price-war periods, respectively, and for multi-product Bertrand-Nash pricing over the whole sample.

Table 16: Conduct Estimation with Synergies: Model Comparison

	Small Model			Large Model		
	Pre-merger	Post-merger	Price War	Pre-merger	Post-merger	Price War
All Firms	0.4900***	0.8419***	0.0000			
	(0.0080)	(0.0359)	(0.0110)			
Large Firms				0.3870***	0.7310***	0.0000
				(0.0661)	(0.0467)	(0.0014)
Small Firms				0.6626***	0.9902***	0.0011
				(0.0988)	(0.0236)	(0.0252)

Notes: The table entries reflect the conduct estimates for both the small and the large conduct specification. Standard errors are in parentheses, and account for two-step estimation. Number of observations: 96512.

Table 17: Marginal Cost Estimates with Synergies: Model Comparison

	Small model	Large model	MP-Nash
Time Trend	0.0042***	0.0043	
	(0.0009)	(0.0265)	
Elec. Price	0.0003	0.0000	
	(0.0128)	(0.0121)	
Synergy Dummy	−0.0081	−0.0069	
	(0.0344)	(0.0718)	
Median MC	0.085	0.085	0.099
Mean MC	0.082	0.082	0.095
St. Dev. MC	0.027	0.028	0.028
Median PCM	0.450	0.450	0.380
Mean PCM	0.450	0.450	0.392
St. Dev. PCM	0.095	0.094	0.080
Med. PCM pre	0.436	0.437	0.383
Med. PCM post	0.470	0.468	0.376
Med. PCM pw	0.389	0.389	0.389

Notes: The table entries reflect the marginal cost estimates from our small conduct model, the large conduct model, and under Bertrand-Nash pricing. For time-specific median price-cost margins, pre refers to pre-merger period, post to post-merger period, and pw to price war period. All estimations include brand fixed effects, store fixed effects and month-year fixed effects. Standard errors are in parentheses, and account for two-step estimation. Number of observations: 96512.

Table 18: Brand-specific PCM

	Small model	Large model	MP-Nash
NAB Shred Wheat	0.32	0.33	0.28
PO Raisin Bran	0.40	0.41	0.32
PO Grape Nuts	0.46	0.47	0.39
PO Honey Comb	0.45	0.47	0.36
GM RaisinNutBran	0.38	0.37	0.32
GM ApplCin Cheer	0.47	0.46	0.38
GM Wheaties	0.38	0.37	0.33
GM Cheerios	0.34	0.34	0.30
GM HonNut Cheer	0.42	0.41	0.35
GM Luck Charms	0.43	0.41	0.35
GM Tot CoFlakes	0.43	0.42	0.38
GM Trix	0.43	0.42	0.36
KE Froot Loops	0.47	0.46	0.41
KE Special K	0.45	0.45	0.41
KE Frost Flakes	0.57	0.56	0.50
KE Corn Pops	0.52	0.51	0.46
KE Raisin Bran	0.43	0.42	0.38
KE Corn Flakes	0.64	0.63	0.59
KE Honey Smacks	0.51	0.50	0.43
KE Crispix	0.48	0.47	0.44
KE Rice Krispies	0.53	0.52	0.48
RAL Chex	0.40	0.42	0.34
RAL Wheat Chex	0.37	0.38	0.30
RA Rice Chex	0.46	0.48	0.39
QU Quaker Oats	0.50	0.52	0.40
QU Capn Crunch	0.51	0.53	0.40

Notes: The table entries reflect the brand-specific median (across markets) price-cost margins for both the small and large model specification, and for multi-product Bertrand-Nash pricing.

Table 19: Time-Brand-specific PCM accounting for potential synergies

	Small model Pre-merger	Small model Post-merger	Small model Price War	Large model Pre-merger	Large model Post-merger	Large model Price War	MP-Nash MP-Nash
NAB Shred Wheat	0.32	0.34	0.30	0.33	0.35	0.30	0.28
PO Raisin Bran	0.39	0.41	0.40	0.41	0.42	0.40	0.32
PO Grape Nuts	0.45	0.47	0.42	0.47	0.49	0.42	0.39
PO Honey Comb	0.44	0.47	0.38	0.46	0.49	0.38	0.36
GM RaisinNutBran	0.37	0.40	0.32	0.36	0.39	0.32	0.32
GM ApplCin Cheer	0.44	0.50	0.38	0.43	0.49	0.38	0.38
GM Wheaties	0.37	0.40	0.31	0.36	0.39	0.31	0.33
GM Cheerios	0.33	0.36	0.29	0.33	0.35	0.29	0.30
GM HonNut Cheer	0.40	0.45	0.34	0.39	0.43	0.34	0.35
GM Luck Charms	0.41	0.45	0.34	0.40	0.44	0.34	0.35
GM Tot CoFlakes	0.42	0.45	0.38	0.41	0.44	0.38	0.38
GM Trix	0.42	0.46	0.36	0.40	0.45	0.36	0.36
KE Froot Loops	0.46	0.49	0.45	0.45	0.48	0.45	0.41
KE Special K	0.45	0.46	0.41	0.44	0.45	0.41	0.41
KE Frost Flakes	0.56	0.59	0.53	0.55	0.58	0.53	0.50
KE Corn Pops	0.51	0.55	0.49	0.50	0.53	0.49	0.46
KE Raisin Bran	0.42	0.44	0.39	0.42	0.43	0.39	0.38
KE Corn Flakes	0.67	0.62	0.56	0.66	0.61	0.56	0.59
KE Honey Smacks	0.50	0.52	0.47	0.49	0.51	0.47	0.43
KE Crispix	0.48	0.49	0.44	0.47	0.48	0.44	0.44
KE Rice Krispies	0.54	0.53	0.48	0.53	0.52	0.48	0.48
RAL Chex	0.40	0.42	0.33	0.42	0.43	0.33	0.34
RAL Wheat Chex	0.37	0.37	0.28	0.38	0.39	0.28	0.30
RA Rice Chex	0.45	0.48	0.39	0.47	0.49	0.39	0.39
QU Quaker Oats	0.47	0.53	0.41	0.50	0.55	0.41	0.40
QU Capn Crunch	0.48	0.57	0.39	0.51	0.59	0.39	0.40

Notes: The table entries reflect the brand-specific median price-cost margins for both the small and large model specification for the pre-merger, post-merger, and price-war periods, respectively, and for multi-product Bertrand-Nash pricing over the whole sample.

D Goodness-of-Fit Graphs

In this section, we provide several representative graphs to illustrate that both our demand and supply model fit the data reasonably well. We compare the observed data for market shares and wholesale prices to our model predictions at various levels of aggregation. Our model predictions are computed using the estimated demand and supply parameters and setting the structural error terms $\Delta\xi$ and $\Delta\omega$ to zero.

Figure 2 displays observed and predicted brand-level market shares for two representative brands averaged across all stores over time. Figure 3 displays predicted and observed wholesale prices for the same brands. Figures 4 and 5 display a comparison of predicted and observed total, i.e., aggregated over all brands, cereal sales and wholesale prices for two representative stores over time. Figures 6 and 7 display the distribution of our structural error terms implied by our parameter estimates.

Figure 2: Goodness-of-fit: Brand-level market shares

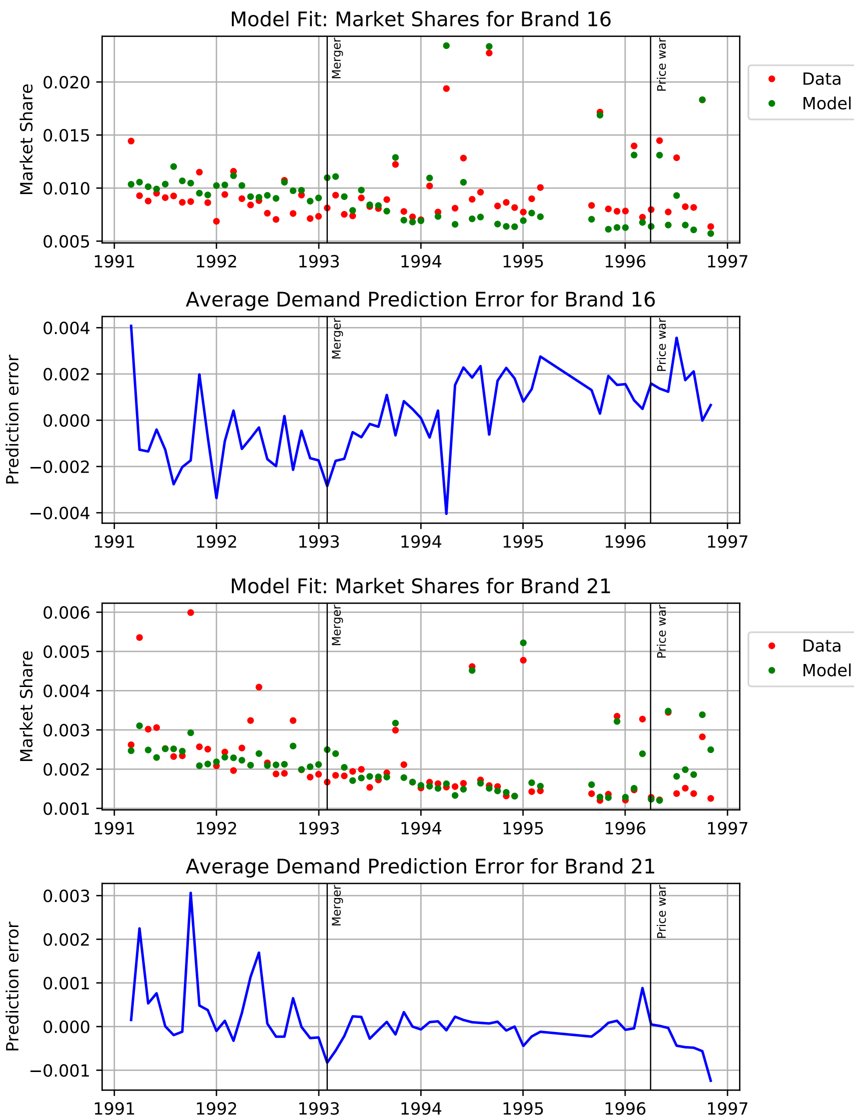


Figure 3: Goodness-of-fit: Brand-level wholesale prices

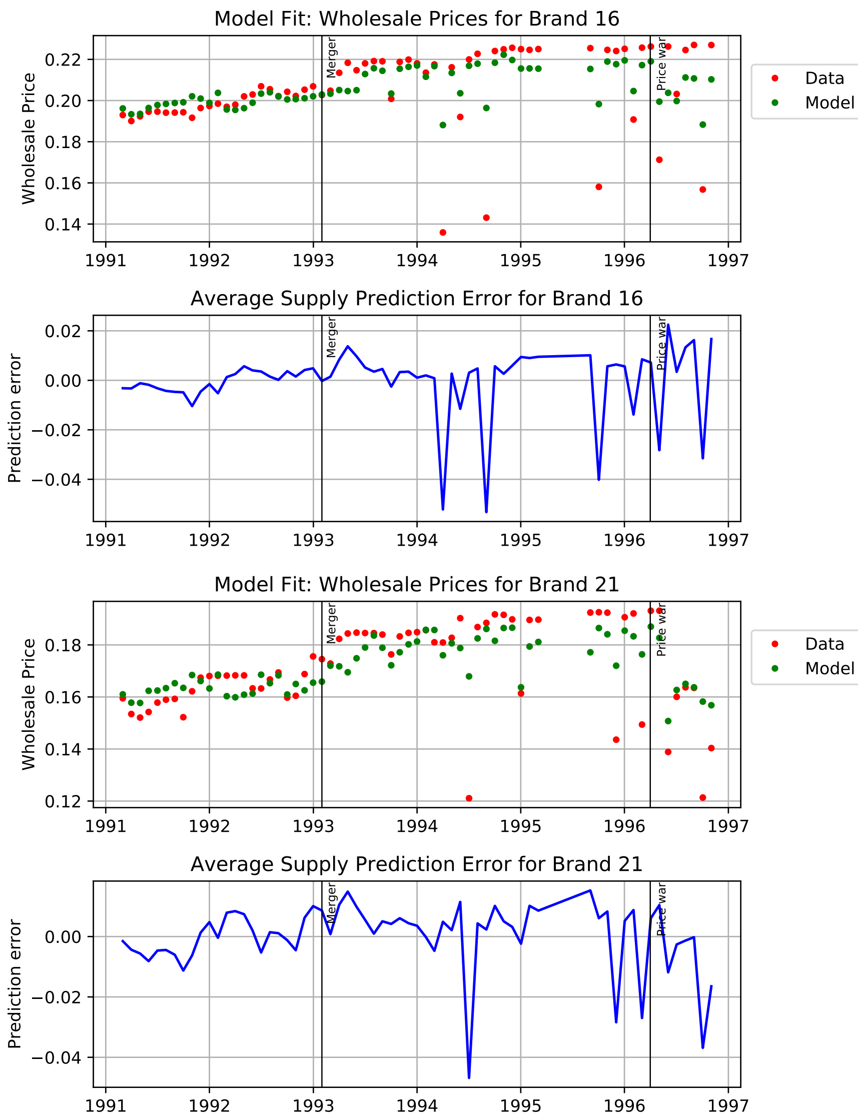


Figure 4: Goodness-of-fit: Store-level sales

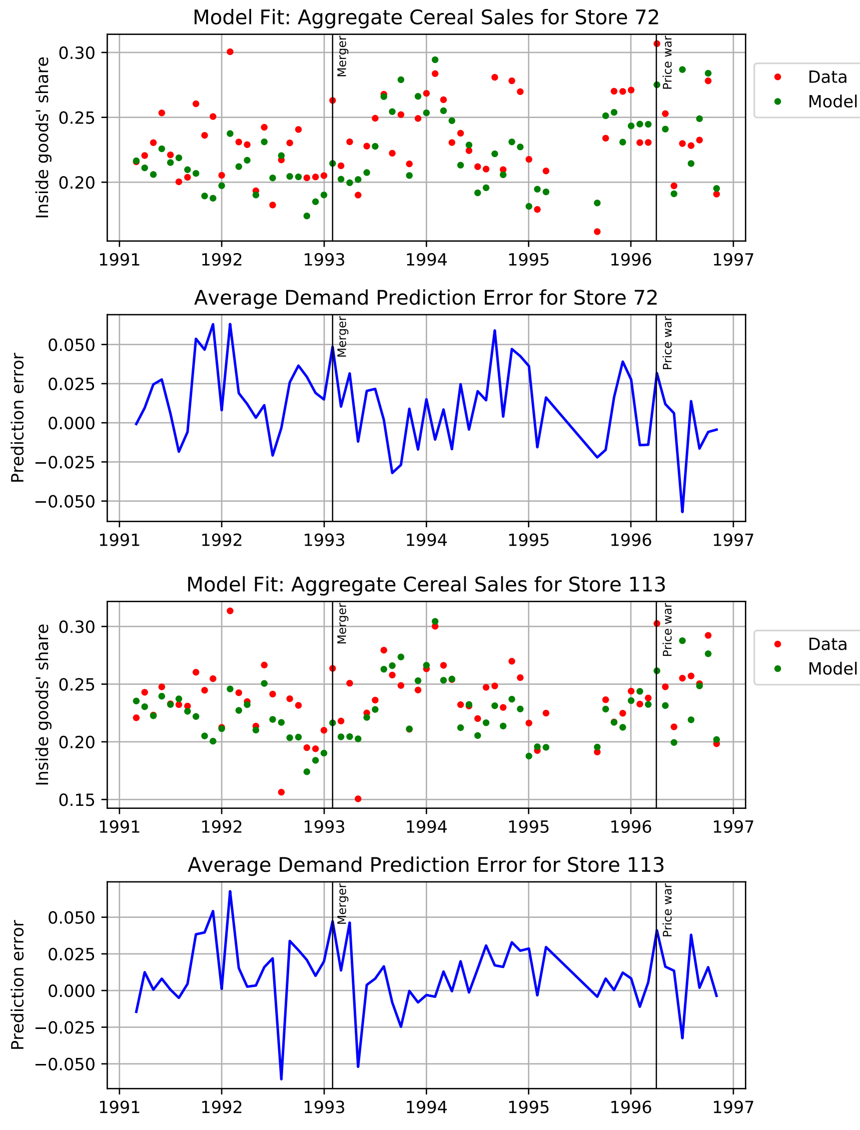


Figure 5: Goodness-of-fit: Store-level wholesale prices

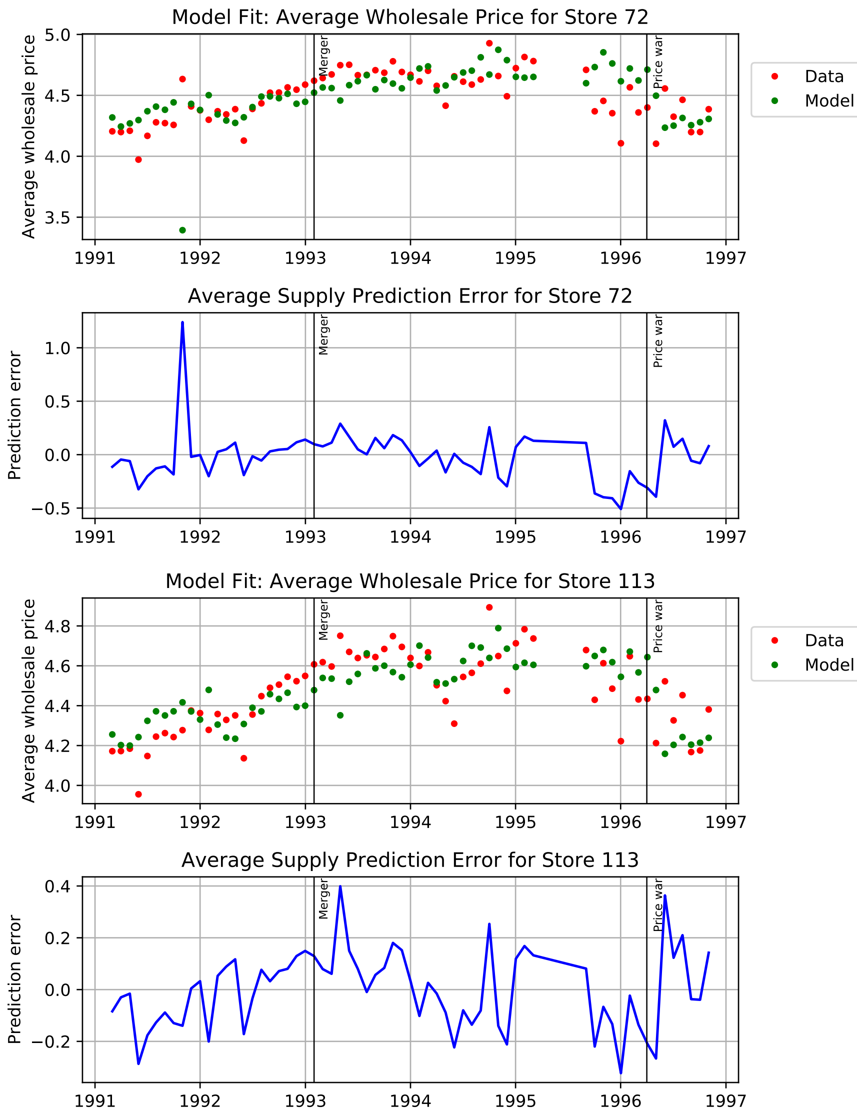


Figure 6: Goodness-of-fit: Distribution of structural demand errors ξ

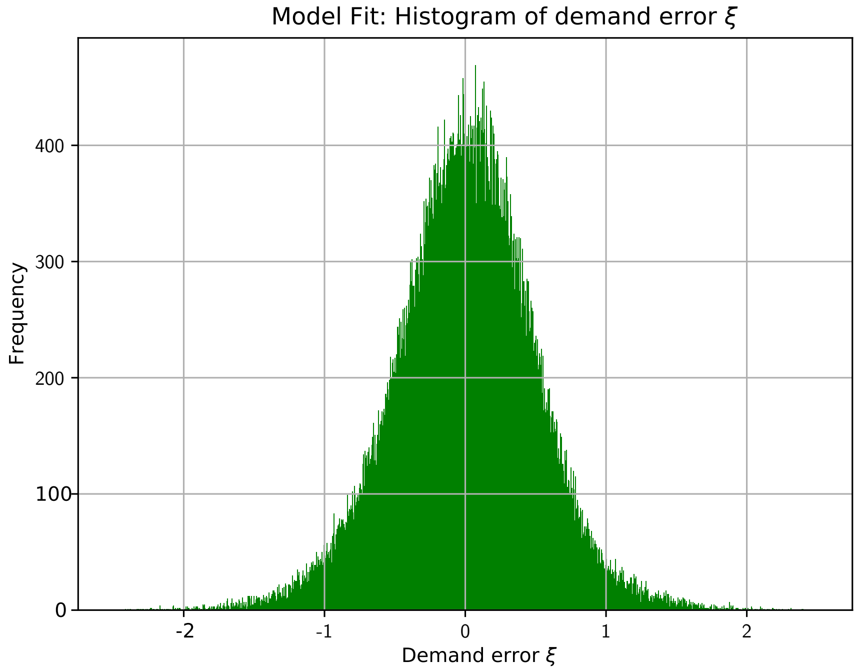
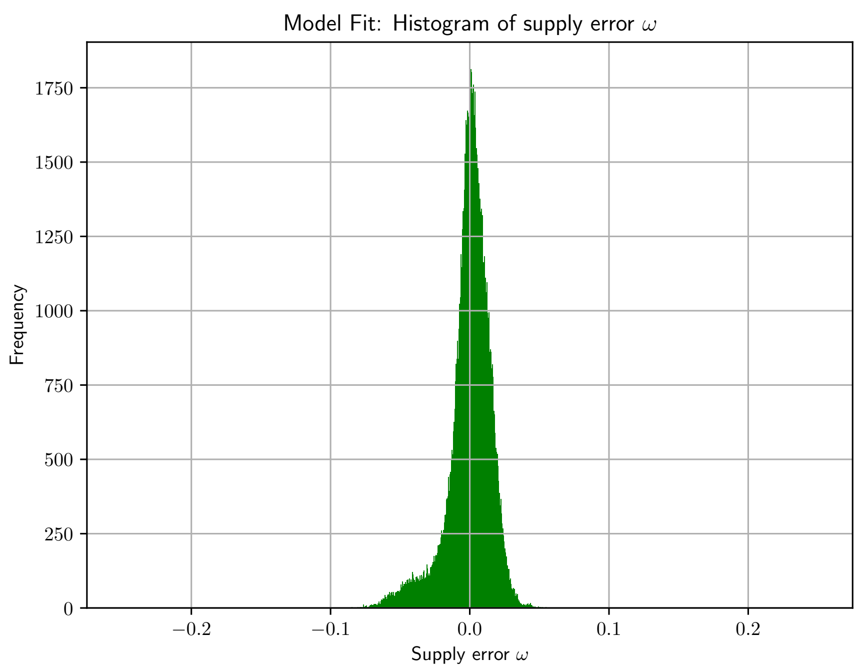


Figure 7: Goodness-of-fit: Distribution of structural supply errors ω



E Numerical Details of the Estimation Algorithm

In this section, we provide numerical details about our estimation algorithm and the software routines used. We estimate the demand side and the supply side in two steps. In principle, it is possible to estimate demand and supply jointly, which generally leads to efficiency gains because it exploits cross-model restrictions and correlations. We choose to estimate the models separately for several reasons. Joint estimation is computationally more intensive and with our very large sample we do not suffer from imprecise parameter estimates. The main reason for not estimating the model jointly is robustness. As demand and supply models differ in several important aspects, we achieved much more robust results by employing different optimization algorithms for each model.

Demand side For estimating our demand model, we use the nested fixed point routine proposed by BLP. Random coefficient demand models can be numerically difficult and computationally intensive to estimate as extensively discussed by Knittel and Metaxoglou (2014). Appropriate choices of optimizers and tolerance levels for such highly nonlinear models are crucial, see also the online appendix of Goldberg and Hellerstein (2013). In our baseline model, we estimate 3 non-linear coefficients capturing heterogeneity across different demographic consumer types. We also estimate much larger demand models with up to 12 non-linear parameters to perform extensive robustness checks.[25]

When computing the model's market share predictions, we simulate 500 consumers per market using Halton draws. Train (2000) demonstrates that Halton draws are much more efficient in simulating the integral over the consumer population than naive random sampling. In line with the recommendations of Knittel and Metaxoglou (2014) and Dubé *et al.* (2012), we set the convergence criterion for the contraction mapping very tight. We stop the mapping, when the sup-norm of the change in the mean utilities δ between two iterations is less than 10^{-9}.

For minimizing the GMM objective function, we use the gradient-based optimizer *SOLVOPT*. To exploit the full power of gradient-based optimization methods, we compute gradients analytically based on the code provided by Nevo (2000a). Similar to the simulation studies in Knittel and Metaxoglou (2014), we find *SOLVOPT* to be very powerful and robust in minimizing the GMM objective function in comparison to other gradient- and non-gradient based optimizers. As tolerance for a minimum of the GMM objective function value we set

[25] Generally, we find that our baseline demand model provides a good fit to the data. The larger models result in very similar price elasticities which are the most important output of the demand estimation. Detailed results are available on request.

the norm of the objective functions' gradient to 10^{-6}. By using multiple starting values, we verify that the obtained minimum of the GMM function is indeed a global minimum.

Supply side For the estimation of the supply model, we generalize the algorithm proposed by BLP to allow for a flexible ownership/internalization matrix. The algorithm can be decomposed into four steps (2.-5.) as follows.

1. **Estimate the demand parameters θ and compute $\frac{\partial s()}{\partial p}$ to compute aggregate own- and cross-price elasticities as described in the previous paragraph.**

2. **Pick a guess for the non-linear supply parameters $vec(\Lambda)$.**

3. **Back out marginal costs given a guess for $vec(\Lambda)$, and $\frac{\partial s()}{\partial p}$ from the demand estimation.** Combining the price elasticities from Step 1 and the pick of $vec(\Lambda)$ from Step 2, we can compute marginal costs for each product and market. Since our marginal cost equations are linear, we can estimate the marginal cost parameters γ by simple linear regressions, which allows us to compute the unobservable marginal cost shock ω for each product and market.

4. **Compute supply-side GMM objective function.** Based on the values for ω backed out in Step 3, we compute the supply side moments which are based on orthogonality conditions between ω and appropriate instruments, and aggregate the GMM criterion function for the parameter guess $vec(\Lambda)$.

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

Compared to the demand model, the supply side is computationally lighter because it does not require solving a contraction mapping for every parameter guess. However, because we have to invert the system of firms' pricing FOCs, the analytical computation of the gradient of the GMM function is much more difficult. Since gradient-based optimization easily loses its power when gradients have to be computed numerically, we revert to non-gradient methods for estimating our supply model.

We find that a finite-descent accelerated random search (ARS) algorithm, as proposed by Appel *et al.* (2004), is very efficient and results in robust estimates. Especially for our larger supply models with a high number of (non-linear) conduct parameters, ARS typically outperforms other non-gradient based approaches such as the simplex-based Nelder-Mead algorithm. For the ARS routine, we use a contraction factor of 2 and a convergence criterion of 10^{-8}. In order to make sure that we find the global minimum, we search for up to 12 local minima of the objective function.

F Standard Error Adjustments

Two-step standard errors adjustment Because we estimate demand and supply in separate steps, we have to account for the two-step nature of our estimation when computing the standard errors of the supply parameters. The correction takes into account the sensitivity of the supply moments with respect to the demand estimates and their variance. The general procedure for obtaining standard errors in this setting is outlined for example by Wooldridge (2010, Chapters 12.5.2). The asymptotic variance-covariance matrix of the GMM estimator for the supply side parameters $\hat{\theta}_S$ can be written as:

$$var(\hat{\theta}_S) = \left[J_S(\hat{\theta}_S, \hat{\theta}_D)' W_S J_S(\hat{\theta}_S, \hat{\theta}_D)\right]^{-1} J_S(\hat{\theta}_S, \hat{\theta}_D)' W_S S_S W_S J_S(\hat{\theta}_S, \hat{\theta}_D) \left[J_S(\hat{\theta}_S, \hat{\theta}_D)' W_S J_S(\hat{\theta}_S, \hat{\theta}_D)\right]^{-1}$$

where $J_S(\cdot)$ denotes the Jacobian of the l_2 supply side moments with respect to the k_2 supply parameters, W_S is the supply side weighting matrix and S_S denotes the $l_2 \times l_2$ matrix containing the outer product of the l_2 supply side moments $g_\omega(\cdot) = \omega(\hat{\theta}_S, \hat{\theta}_D) Z_S$.

When demand and supply parameters are estimated in two separate steps, the standard formula underestimates the variance of the supply side parameters. In order to obtain correct standard errors, S_S has to be modified to take into account the sensitivity of the supply moments with respect to the demand parameters. In our model, S_S has to be replaced with

$$\tilde{S}_S = \left[g_\omega(\hat{\theta}_S, \hat{\theta}_D) + F g_\xi(\hat{\theta}_S, \hat{\theta}_D)\right] \left[g_\omega(\hat{\theta}_S, \hat{\theta}_D) + F g_\xi(\hat{\theta}_S, \hat{\theta}_D)\right]'$$

where g_ω contains the $l_2 \times n$ supply moments and g_ξ contains the $l_1 \times n$ demand moments both evaluated at the estimated parameter values $(\hat{\theta}_D, \hat{\theta}_S)$. The sensitivity of the supply moments with respect to the demand parameters is captured by the $l_2 \times l_1$ matrix F

$$F = J_{SD}(\hat{\theta}_S, \hat{\theta}_D) \left[J_{DD}(\hat{\theta}_S, \hat{\theta}_D)' W_D J_{DD}(\hat{\theta}_S, \hat{\theta}_D)\right]^{-1} J_{DD}(\hat{\theta}_S, \hat{\theta}_D)' W_D$$

where $J_{SD}(\cdot)$ contains the derivatives of the l_2 supply moment conditions with respect to the k_1 demand parameters evaluated at the estimated demand and supply parameters. $J_{DD}(\cdot)$ denotes the derivatives of the l_1 demand moments with respect to the k_1 demand parameters and W_D is the $l_1 \times l_1$ is the weighting matrix used in the demand estimation.

Delta method standard error adjustment We use an exponential transformation to restrict our conduct parameters λ to lie between 0 and 1. We estimate a parameter λ_e such that the actual model parameter is

$$\lambda = g(\lambda_e) = \frac{exp(\lambda_e)}{1 + exp(\lambda_e)}.$$

The standard GMM variance formula provides us with an estimate of the standard errors of λ_e. In order to compute standard errors for our actual conduct parameters λ, we apply the delta method which states that the variance of a continuous function $g(\cdot)$ of a random variable X is given by

$$var\big[g(X)\big] = \big[g'(X)\big]^2 var\big(X\big).$$

The derivatives of our functional transformation with respect to the estimated parameter λ_e are

$$g'(\theta) = \frac{exp(\lambda_e)}{1 + exp(\lambda_e)} - \frac{exp(\lambda_e)^2}{[1 + exp(\lambda_e)]^2} = \frac{exp(\lambda_e) + exp(\lambda_e)^2 - exp(\lambda_e)^2}{[1 + exp(\lambda_e)]^2} = \frac{exp(\lambda_e)}{[1 + exp(\lambda_e)]^2}.$$