Estimating Industry Conduct Using Promotion Data*

The Evolution of Pricing Behavior in the U.S. RTE Cereal Industry

Christian Michel†
Jose Manuel Paz y Miño‡
Stefan Weiergraeber§

August 28, 2022

Abstract

We estimate the evolution of competition in the ready-to-eat 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' promotions with measures of products' relative isolation in the characteristics space. We find strong evidence for partial price coordination among cereal manufacturers in the beginning of our sample. After a merger in 1993 conduct becomes more competitive and on average consistent with multiproduct Nash pricing. The last part of our sample is characterized by even more aggressive pricing, implying median wholesale margins of less than 5%.
Keywords: Markups, Market Power, Conduct Estimation, Differentiated Products Markets
JEL Classification: L11, L41, C51

*This paper was previously circulated as *Estimating Industry Conduct in Differentiated Products Markets - The Evolution of Pricing Behavior in the RTE Cereal Industry*. We thank three anonymous referees, Daniel Ackerberg, John Asker, Allan Collard-Wexler, Pierre Dubois, Liran Einav, Alon Eizenberg, Juan Carlos Escanciano, Federico Ciliberto, Ying Fan, Rosa Ferrer, Jeremy Fox, Amit Gandhi, David Genesove, Gautam Gowrisankaran, Kenneth Hendricks, Günter Hitsch, Kate Ho, Jean-François Houde, Mitsuru Igami, Gaston Illanes, Kyoo-il Kim, Nathan Miller, Julie Mortimer, Massimo Motta, Emi Nakamura, Aviv Nevo, Volker Nocke, Aureo de Paula, Helena Perrone, Robert Porter, William Rogerson, Philipp Schmidt-Dengler, Johannes Schneider, Oleksandr Shcherbakov, Sandro Shelegia, André Stenzel, Andrew Sweeting, Yuya Takahashi, Otto Toivanen, Frank Verboven, Matthew 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 and the Barcelona GSE Seed Grant SG-2017-14. Weiergraeber gratefully acknowledges financial support from the IU OVPR Seed Grant 2018.
†Barcelona School of Economics. Email: christian.felix.michel@gmail.com.
‡Department of Social Sciences, Universidad Católica del Uruguay. Email: jose.pazymino@ucu.edu.uy.
§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. Empirically disentangling legitimate sources of market power, such as product differentiation, from anti-competitive behavior, such as coordinated pricing is thus an important task. This task is very difficult, however, because neither the intensity of competition nor marginal cost, which is another price determinant are commonly observed.

In a seminal paper, Bresnahan (1982) showed for a homogeneous goods model how demand rotators can be used to distinguish different oligopoly models. Berry and Haile (2014) formalized this intuition for a general class of models with differentiated products and show that it is, in principle, possible to empirically discriminate between different oligopoly models by exploiting variation in market conditions.[1] A practical problem, however, is that many of the instruments based on this logic tend to contain only little variation; therefore, estimation-based inference about the competitive intensity of an industry remains challenging in many applications.

Most studies that estimate *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. These restrictions can lead to inconsistent estimates of markups and marginal costs.

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

For our estimation, we use scanner data from the Dominick's Finer Food (DFF) database. The database includes detailed information on DFF's supermarkets located in the Chicago metropolitan area. In addition to detailed store-specific data on quantities, retail prices, and temporary promotions, our data contain information on wholesale prices. We analyze a five-and-a-half year span of data from 1991 to 1996. Our sample period includes two important events; first, the Post-Nabisco merger in January 1993 and second, a period starting in April

[1]Examples of 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 U.S. beer industry deviates from Nash pricing once one of them participates in the joint-venture.

1996 during which manufacturers massively decrease wholesale prices, which the business press referred to as a *price war*. A key focus of this paper is to quantify how pricing behavior changes following these events.

To motivate our structural model and our identification strategy, we conduct a series of reduced form regressions. We find that after the merger most prices increased slightly and that during the price war period wholesale prices are 7% lower than in the pre-merger period. Moreover, we show that how firms react to rival firms' promotions changes significantly over time. We illustrate using a stylized model that these patterns can be interpreted as evidence that industry conduct changed over time.

Naturally, the observed pricing trends can be due to several reasons, in particular, changes in demand, marginal cost, and industry conduct. In order to disentangle these channels, we estimate a structural empirical model. On the demand side, we use a random coefficients nested logit (RCNL) model in the style of Berry *et al.* (1995) (henceforth, BLP) and Nevo (2001). On the supply side, we use a flexible conduct framework similar to Miller and Weinberg (2017). We measure the degree of price coordination by a matrix of parameters that indicate the degree to which firms internalize their rivals' profits. Compared to a model that simply assumes a particular form of industry conduct, for example, multiproduct Nash pricing, as is often done in the literature, our model allows for substantially more heterogeneity of markups over time and across firms.[3] Therefore, it is likely to lead to more accurate predictions and more effective policy recommendations.

The novel instruments that we use to identify industry conduct are based on information on rival brands' temporary promotions interacted with measures of proximity in the characteristics space. These instruments follow the logic of Berry and Haile (2014) in the sense that, similarly to classical BLP-instruments, a rival's promotion affects a firm's marginal revenue curve by rotating its demand curve. Firms' responses to this variation in market conditions identify industry conduct.

The effects of a rival brand's promotions should be stronger the more consumers consider these products as substitutes. Therefore, we interact the market-specific number of rivals' promotions with the products' relative proximity in the characteristics space. The relative proximity feature mirrors the logic of the differentiation instruments recently proposed by Gandhi and Houde (2020) to identify consumers' substitution patterns.

Even though promotions are clearly endogenously set by firms, promotional activities can be considered as sequentially exogenous. In almost all consumer packaged goods (CPG) industries, manufacturers and retailers agree on promotions several months in advance. This is done for various reasons, for example, a sufficient supply of the product must be ensured and advertising brochures must be printed. Therefore, rivals' promotional activities in a

[3]For conciseness, we use the term *Nash pricing* to describe static multiproduct Bertrand-Nash pricing throughout the paper.

given time period are plausibly exogenous to innovations in product-specific demand and supply shocks. We present extensive evidence from the marketing literature and reduced form evidence from our data that support these timing assumptions in Section 2.3.

In contrast to typical BLP instruments, we do not require variation in the physical characteristics space or product entry and exit to construct a rich set of instruments. In addition, our empirical strategy does not rely on exogenous industry shocks. The data required to construct our instruments are readily available for many industries, and thus our empirical strategy has broad applicability. Finally, we conduct a series of weak identification tests and find that our proposed instruments indeed prove to be powerful for identifying industry conduct.

Our estimation results indicate that there are substantial changes in industry conduct over time. We find partially cooperative levels of conduct between firms in the beginning of our sample, with wholesale margins that are twice as large as those implied by Nash pricing.[4] Nine months after the Post-Nabisco merger conduct becomes more competitive and on average consistent with Nash pricing. When we allow our profit internalization parameters to differ across firms, we find that only the post-merger conduct of the smaller firms is consistent with Nash pricing, while the two market leaders, Kellogg's and General Mills still price more cooperatively. For the price war period, our estimates indicate that firms price even more aggressively, with median wholesale margins of less than 5%.

Finally, we use our model to conduct a series of counterfactual simulations. First, we decompose the observed price changes post-merger into the unilateral and coordinated effects of the merger. We find that the unilateral effects of the merger are minimal and that almost all of the observed price changes can be attributed to a change in overall industry conduct. Second, we simulate how the industry would have evolved if the price war had not taken place. In this scenario, prices would have been 9% higher, and consumer surplus for the Chicago area during the nine months of the price war period that our sample covers would have been US-$ 1 million lower. The corresponding firm profits would have been US-$ 1.1 million higher.

Related Literature. Our paper relates to several strands of the literature. First, it relies on the theoretical literature on the 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 that rotate the marginal revenue curve. We show one way in which their arguments can be applied to real-world industry data

[4] 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 do not suggest that our estimates provide evidence for anti-competitive behavior in the sense of violating antitrust laws.

and propose specific instruments that we find to be powerful for identifying detailed industry conduct patterns that are difficult to identify using established instruments.

Early work on industry conduct has mostly relied on estimating conjectural variations for the homogeneous good case, see, for example Bresnahan (1982) and Lau (1982). Corts (1999) critically discusses such approaches. He argues that the estimated parameters usually differ from the "as-if conduct parameters" and, therefore, that they do not necessarily reflect the economic parameters of interest. Nevo (1998) discusses the advantages and disadvantages of a direct conduct estimation compared to a non-nested menu approach for differentiated products industries. He argues that in practice, estimating detailed industry conduct directly using only a single demand rotator is impossible and proposes the use of selection tests for a "menu" of pre-specified models; see, for example, Rivers and Vuong (2002) and the recent working paper by Duarte *et al.* (2022). Backus *et al.* (2021) also propose a testing-based approach to empirically analyze industry conduct. They focus on distinguishing multiproduct Nash pricing from pricing patterns implied by the common ownership hypothesis using Nielsen data on the U.S. cereal industry from 2007 to 2016.

One advantage of these approaches compared to an estimation-based approach is that testing generally imposes lower requirements on the strength of the moment conditions. A disadvantage of a testing-based approach is that it requires the researcher to prespecify a fixed set of conduct models to test against each other. Moreover, in some cases the statistical power of these tests can be relatively weak and difficult to assess; see, for example, Shi (2015). This can be especially problematic when several detailed conduct patterns are tested against each other.

Our instruments can straightforwardly be incorporated into the testing-based approaches of Duarte *et al.* (2022) and Backus *et al.* (2021). A key advantage of our instruments is that we find them powerful enough for an estimation-based approach, which allows for quantifying more fine-grained conduct patterns. This is particularly advantageous for settings, in which the profit internalization parameters can be given a micro-founded structural interpretation.

Our analysis also contributes to a broader discussion about 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. 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: single-product Nash pricing, multiproduct Nash pricing, and joint profit maximization. He subsequently compares the different cost estimates with accounting data to select the most plausible specification, which he finds to be multiproduct Nash pricing. His sample period partly overlaps with the pre-merger period in our sample. However, his data has a different geographical coverage (65 U.S. cities) than ours (only the Chicago area). For the pre-merger period, our estimates are overall consistent with his results, but we provide additional insights. While our markup estimates are closer to those implied by Nash pricing than those implied by full collusion, even small differences in conduct can have a considerable impact on the estimated price-cost margins. We find that under Nash pricing, combined retailer and manufacturer gross margins would be around 40%, but that after allowing for a flexible conduct specification, median gross margins in the pre-merger period are 52%. Most importantly, Nevo (2001) assumes a time-invariant conduct, while we find significant changes in conduct over time.

There is also a growing interest in quantifying the heterogeneity of markups in both the macroeconomics and the trade literature, which usually relies on the production function approach, see for example, De Loecker *et al.* (2020). In a similar spirit, Döpper *et al.* (2021) analyze over 100 consumer products categories to quantify the evolution of markups over time using the demand inversion approach under the assumption of Nash pricing. We consider our analysis complementary to this literature by allowing heterogeneity in industry conduct to drive markup heterogeneity across firms and time.

The two papers most closely related to ours are Miller and Weinberg (2017), and Ciliberto and Williams (2014), who also employ an estimation-based approach to quantifying industry conduct. Miller and Weinberg (2017) assess the effects of a joint-venture on pricing behavior in the beer industry. They focus on estimating one parameter that measures the magnitude of mutual profit internalization between Anheuser-Busch InBev (ABI) and MillerCoors after the Miller-Coors joint-venture. Their model assumes industry-wide 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 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 multimarket contact between different airlines. They find that firms with a lower degree of multimarket 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 multimarket contact.[5]

2 Data and Industry Overview

In this section, we provide background information on the U.S. RTE cereal industry and describe our data. In addition, we conduct a series of reduced form regressions to provide evidence for changes in industry conduct over time and to guide our structural model.

2.1 Industry Overview

In the beginning of our sample period, six nationwide manufacturers dominate the U.S. RTE cereal industry. Table 5 in Appendix A.1 shows that the two largest firms, General Mills and Kellogg's, cover more than 75% of the market. The remainder of the market is split among the substantially smaller firms (Post, Nabisco, Quaker, and Ralston). More than 200 brands are available to consumers during the time span we analyze; however, the majority of sales can be attributed to the 25 most popular brands.

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. 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 the one before the merger. Table 5 in Appendix A.1 shows that the market is already highly concentrated before the merger, with an HHI of more than 2,500. However, the merger did not lead to a significant increase in the HHI. Figure 1 in Appendix A.1 illustrates that the prices of individual brands reacted heterogeneously to the merger. Kellogg's and Ralston increased prices the most, especially right after the merger. Wholesale prices of Post, Nabisco and Quaker increased only marginally and General Mills decreased them for many of its brands, especially starting in 1994.

In the spring of 1996, all cereal manufacturers massively decreased their wholesale prices nationwide. 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. Although we remain agnostic about the causes for the price war, there is anecdotal evidence that negative publicity and political pressure were important motivations for the price cuts. For example, in March 1995 two U.S. congressmen started a public campaign to reduce cereal prices, which received high media attention. This campaign was revived one year later right before the start of the large wholesale price cuts (Cotterill and Franklin, 1999). Figure 1 shows that the nationwide patterns are also present in our

[5]Although our paper focuses on estimating industry conduct for general industry settings, it is further related to the ex-post analysis of mergers, for example, Crawford *et al.* (2018), Ashenfelter *et al.* (2013), and Björnerstedt and Verboven (2016).

sample. 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 coordinated pricing rather than potential shifts in demand and marginal costs.

RTE cereals are distributed to consumers via supermarkets. An important feature of CPG industries is the prevalence of temporary promotions. A large literature in marketing and economics discusses manufacturers' and retailer's motivations for running promotions and how they are implemented in practice, see, for example, Anderson and Fox (2019).

In the following, we discuss several institutional features that motivate our model and our empirical strategy to quantify industry conduct. First, temporary promotions are strong rotators of (short-run) demand, for example, because promotions induce brand-switching or incentivize consumers to buy additional units (Anderson and Fox, 2019, p.501). Promotions generate these demand rotations through both temporary reductions in a product's retail price and demand-enhancing non-monetary effects, for example, because of an inclusion in a retailer's advertising brochure, better shelf space allocation or additional in-store promotion signs, see, for example, Anderson and Simester (1998) for the importance of sale-sign effects.

Second, promotions follow *sticky plans*, i.e., they are set at least several months in advance and almost never reversed (Anderson and Fox, 2019, p.541). This is to coordinate operations of the manufacturer and the retailer so that, for example, sufficient inventory is available during a promotion. Therefore, promotions usually cannot react to contemporaneous innovations in demand and cost shocks immediately, and can be considered sequentially exogenous.

Lastly, promotions are often funded via *trade spend* payments. In order to incentivize the retailer to pass through a promotion for a specific product, its manufacturer will usually grant a discount on the wholesale price for all units sold during the promotion period, which is financed from a *trade spend budget*, that is typically fixed for a longer time horizon and considered to come from a separate manufacturer account than payments associated with base wholesale prices (Anderson and Fox, 2019, p.514). To ensure that the retailer does not free-ride on lower promotion wholesale prices without passing-through these discounts to consumers, manufacturers use detailed *contingent contracts* (Anderson *et al.*, 2017).[6] Therefore, both retail and *net wholesale prices*, i.e., wholesale prices after trade spend discounts, often move in parallel during a promotion period and this variation in wholesale prices does not reflect variation in a manufacturer's marginal cost of production.

A key assumption that we exploit for our empirical strategy is that regular *base wholesale prices* can react to contemporaneous demand and cost shocks, while promotions and the promotion-induced price discounts financed by trade spend do not react to economic shocks

[6]For example, a manufacturer might require proof from the retailer that the agreed upon promotional activities were indeed implemented and that promotion-adequate quantities were sold during the promotion, for example, by providing scanner data.

but rather follow their predetermined sticky plans. Anderson *et al.* (2017) discuss extensive evidence for this pattern, and we provide additional supporting reduced form evidence for these assumptions from our data in Appendix A.7.

2.2 Data

Our main data come from the DFF scanner database covering the period from February 1991 to October 1996. It 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 27 brands from the six largest nationwide manufacturers. All of the products are offered throughout 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 fringe products, which are often only offered during an experimental phase.

We do not include private label products in our analysis, because our focus is on estimating the competitive interactions between national cereal manufacturers. In Appendix A.2 we provide additional details about the brands that we exclude from our analysis, in particular, about the private label products in our data. We show that for DFF private label sales are relatively small and stable; therefore, we argue that they are unlikely to affect our analysis. The 27 brands included in our sample cover 56% of all the cereals sold at DFF during our sample period. This coverage is comparable to other studies on the same industry and a similar time period, for example, Nevo (2000b, 2001).

Typically, each brand is offered in the form of two main UPCs (sizes) and temporarily features special UPCs that exhibit only small sales. For our estimation, we follow most of the literature and aggregate the data from the UPC-week to the brand-month level, mostly to alleviate concerns about measurement error and consumer stockpiling on a weekly level.

Information on temporary promotions is recorded on the UPC-store-week level, separately for two types of promotions: general price reduction sales and bonus buy/coupon promotions, i.e., we observe for how many weeks a given brand is on promotion within each month and store. We discuss detailed descriptives of the different promotion variables on several levels, in particular, comparing the UPC-week and the brand-month level, in Appendix A.3. Within a month, promotions are usually highly correlated across DFF stores, even though there is some variation within the chain in a given month, see Figure 10 in Appendix A.3.

Most importantly, Figures 7 and 8 reveal that, even after aggregating the data to the brand-month level, there is substantial variation across brands and months in the aggregate number of promotional activities, which provides us with significant shifters of consumers' demand. As we discuss in Section 4.1, our instruments for identifying industry conduct exploit this rich variation in promotion "breadth" across brands and over time. While we observe the price promotion "depth", i.e., by how much retail and wholesale prices decrease during a temporary promotion, we do not observe any depth measure of the non-monetary promotion components, such as the number of advertising brochures or shelf space allocation. We discuss the issue of promotion breadth versus promotion depth further in Appendix A.3.

A key advantage of the DFF data is that we observe the retailer's average acquisition costs for each product in each week. From this variable we compute a measure of wholesale prices for each brand and month. This variable is a weighted average of the wholesale prices for the products in the inventory; see Chevalier *et al.* (2003) for a discussion of this variable. Under the assumption that the retailer cannot carry forward significant inventory from month to month, we judge this wholesale price measure to be a reasonable approximation to the actual wholesale price paid by the retailer.

An important point is that the observed wholesale prices contain *trade spend* discounts, which are granted by manufacturers during a promotion period to incentivize the retailer to pass-through a promotion to the final consumer. As discussed extensively in Anderson and Fox (2019), trade spend discounts are unlikely to reflect variation in marginal costs of producing cereal; rather they are determined by sticky plans and complicated contingent contracts between manufacturers and retailers. Since our structural model will focus on base wholesale prices as manufacturers' strategic variable, we decompose the observed net wholesale prices into base wholesale prices and trade spend payments.

For our main specification, we compute base wholesale prices based on our UPC-week level data as follows. We assume that manufacturers set a chain-wide base wholesale price for each UPC in every month. Moreover, we assume that variation in observed wholesale prices during a promotion-week is fully due to pre-determined trade spend payments so that the base wholesale price does not change during a promotion-week. We then compute the base wholesale price during the promotion-weeks as the average of the wholesale price observed in the weeks immediately before and immediately after the promotion period. Afterwards, we aggregate this UPC-week level variable to the brand-month level, and compute the (average) trade spend on the brand-month level as the difference between the imputed base wholesale price and the observed total wholesale price.

The magnitude of the resulting trade spend variable is consistent with industry evidence. For example, we find that the total trade spend payments in our data are roughly 17% of manufacturers' revenue. Anderson and Fox (2019) report that this statistic amounts to 10% to 25% in a typical CPG industry. We provide additional details of this computation, as well as several alternatives and robustness checks, in Appendix A.4.

We complement the DFF data with input price data for commodities, such as sugar and various grains, as well as gasoline and electricity prices from Thomson Reuters Datastream and from www.indexmundi.com. Finally, we collect nutrition facts from the website www.calorieking.com and information on the different production and processing techniques for all brands from manufacturer websites. Throughout our analysis, we use prices deflated to 1991-US-$.

2.3 Reduced Form Analysis

In the following, we use reduced form regressions to investigate the patterns in our data in more detail and to provide evidence for our hypothesis that industry conduct changed following the industry events that we study. Our analysis is based on the idea that the promotion of a rival product should have different effects on a brand's own price under competition and collusion. In general, promotions of a rival's product increase the competitive pressure on a product because of two effects; first, a demand-enhancing sale-sign effect, that makes the promoted product more popular, for example, via shelf-space allocation or advertising brochures, and second, a retail price reduction that is triggered by a trade spend payment. From the perspective of the retailer, the trade spend acts like a short-term marginal cost reduction. In our application, the latter effect always dominates the first, so that promotions are not only associated with lower retail but also with lower net wholesale prices.

Under static Bertrand-Nash competition, a rival's promotion is therefore likely to lead to own-price cuts, since prices are strategic complements. Under joint profit maximization, however, each brand will take into account an additional term, which captures how a brand's price reaction affects the rival's profit. If a rival promotion increases this cross-derivative enough, it is possible that the rival promotion generates an own-price increase, which is much less likely to occur under competition.

Therefore, how a firm reacts to a rival brand's promotion can be informative about the intensity of price competition in the market. We provide a more formal discussion of these effects using a stylized model as well as a numerical illustration using a logit model in Appendix A.5.

Table 7 in Appendix A.6 summarizes the results from linear regressions that apply the above logic. We regress wholesale prices on the number of own-brand and rival promotions as well as several fixed effects.[7] The different columns differ in which promotion variables we include and in which wholesale price is analyzed. Columns 1 and 3 contain only general promotions, while columns 2 and 4 include both general and bonus buy promotions. Columns 1 and 2 use net wholesale prices (including trade spend) as the dependent variable, while columns 3 and 4 use base wholesale prices.

First, we find that after the Post-Nabisco merger wholesale prices tend to increase for

[7] For reasons that we explain in Section 4.1, we only count promotions of rivals that are close substitutes to the considered brand. We define two brands as close if the difference in their sugar contents is in the first decile of all sugar content differences. Using closeness definitions other than sugar or counting all rival promotions results in qualitatively similar results.

most firms, even though for many firms the total wholesale price increase is insignificant. Kellogg's and Ralston increase wholesale prices most by around 3% to 4%. General Mills is a notable exception and decreases prices in the post-merger period by 4% on average. Moreover, several brands exhibit a wholesale price decrease one year after the merger, see Figure 1 in Appendix A.1. During the price war period wholesale prices are roughly 10% lower than in the pre-merger period.

The effects of rival firms' promotions on wholesale prices change substantially across the three periods of our sample, which is in line with our hypothesis of changes in industry conduct, and the results are similar across all specifications. In the pre-merger period rival firms' general promotions and wholesale prices have a large and positive correlation. After the merger this effect becomes smaller. During the price war period general promotions have a large and significant negative effect on rivals' wholesale prices. Qualitatively, similar patterns hold for bonus buy/coupon promotions, although the coefficients are generally smaller and sometimes insignificant.

The effect of own-brand promotions on net wholesale prices is negative and significant, see columns 1 and 2 of Table 7. However, base wholesale prices are not significantly lower during promotion periods, see columns 3 and 4 of Table 7. This pattern is not surprising, because the net wholesale prices include trade spend payments during promotion periods, while base wholesale prices do not include these discounts.

In order to test formally whether the estimated changes over time are significant, we conduct several structural break tests for the promotion pass-through coefficients of both own and rival promotions for each of the four regression models in Table 7. Table 1 summarizes the results for total wholesale prices, which include trade spend payments.[8]

Each row tests the hypothesis that the promotion pass-through coefficient is constant over the three periods using F-tests. The first three rows correspond to testing the equality of the pass-through of rival promotions, which we can clearly reject in both models. In contrast, we cannot reject the equality of the pass-through coefficients of own-brand promotions over time.

A disadvantage of our F-tests is that in our regressions we implicitly need to provide the dates of the potential break points. To address this issue, we conduct additional tests that do not require prior knowledge of when the breaks in conduct occur. Towards this, we run a series of tests proposed by Bai and Perron (1998, 2003) to test when the effect of rival firms' promotions on a brand's wholesale price changes. We provide details about our testing procedure and the results in Appendix A.6.

For the price war period these tests align closely with the dates mentioned in industry sources. The business press typically declares the start of the nation-wide price war to be April 1996. Our structural break tests suggest February 1996 as the start date.

[8]The results for base wholesale prices are similar and summarized in Table 8 in Appendix A.6.

Table 1: Structural break tests - total wholesale prices

	$\beta(SE)$			$H_0 : pre = post = pw$	
	Pre-merger	Post-merger	Price war	F-statistic (df1,df2)	p-value
Promo (SG, rivals), model 1	8.7240	2.4414	-2.7381	11.0933 (2,1683)	0.0000
	(2.5680)	(0.8754)	(1.0324)		
Promo (SG, rivals), model 2	9.1778	2.2676	-2.3925	8.4951 (2,1677)	0.0002
	(2.8057)	(0.8784)	(1.1368)		
Promo (BC, rivals), model 2	-0.4443	2.5267	-0.7469	3.1954 (2,1677)	0.0412
	(0.8919)	(0.8857)	(1.9204)		
Promo (SG, own), model 1	-0.0022	-0.0019	-0.0012	1.8082 (2,1683)	0.1643
	(0.0007)	(0.0002)	(0.0003)		
Promo (SG, own), model 2	-0.0022	-0.0019	-0.0012	1.6309 (2,1677)	0.1961
	(0.0007)	(0.0002)	(0.0003)		
Promo (BC, own), model 2	-0.0002	-0.0003	-0.0004	0.4808 (2,1677)	0.6184
	(0.0001)	(0.0002)	(0.0003)		

Notes: The table summarizes the results from testing the equality of the own and rival promotion pass-through coefficients on total wholesale prices over time using F-tests. The coefficients in columns 1 to 3 come from the regressions in Table 7, columns 1 and 2 (standard errors in parentheses). SG *denotes general promotions.* BC *denotes bonus buy and coupon promotions.*

For the break following the merger the test results are less aligned with the actual merger event. In our tests, we find only weak evidence for a break around the time of the merger (January 1993). However, our tests find stronger evidence for structural breaks nine months after the merger (October 1993).

For our structural estimation we prefer to "let the data speak" about when conduct changes instead of relying only on industry information. Therefore, for our structural estimation, we define the start of the post-merger and the price war period as October 1993 and February 1996, respectively.

3 Empirical Model

There are several potential reasons for observing the price increases following the Post-Nabisco merger and the large reduction in wholesale prices three-and-a-half years later. For example, consumers' preferences 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. To disentangle 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 nested logit (RCNL) model with a specification that is similar to those in Nevo (2001) and Miller and Weinberg (2017). One key advantage of this model is that it allows for flexible substitution patterns. An accurate estimation of own- and cross-price elasticities is crucial in our application, since they are the most important determinants of a firm's pricing first-order conditions.

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

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

where x_{jt} denotes a K-dimensional vector of brand j's observable characteristics (including brand fixed effects, year fixed effects, and month-of-the-year fixed effects), p^r_{jt} denotes the retail price of product j in market t. All physical product characteristics in x_{jt}, such as sugar and fiber content, are time-invariant and therefore collinear with brand fixed effects.

As discussed extensively in the marketing literature, temporary promotions are important determinants of consumers' cereal choices through both direct price effects and non-monetary effects that increase the attractiveness of products on promotion. Our model captures direct price reductions in the observed retail price p^r_{jt}. To capture the non-monetary effects in our demand model in a parsimonious way, we include the number of promotions for product j in market t in the vector of observable product characteristics x_{jt}, which constitutes the only time-varying product characteristic.

We capture brand-market specific quality shocks that are unobservable to the researcher but observable to and equally valued by all consumers by ξ_{jt}. In addition, we assume that ξ follows an $AR(1)$-process so that $\xi_{jt+1} = \iota_D \xi_{jt} + \nu_{Djt+1}$. This specification allows for persistence in the structural demand error and, most importantly, enables us to form moment conditions based on the innovations of the process instead of its levels.

The coefficients α_i and β_i are individual-specific. They depend on the mean valuations (α, β), a vector of i's demographic variables, D_i, and Φ, a vector of parameters that measure how preferences vary with demographics, so that

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

In our main specification we include only income as a demographic characteristic and interact it with the constant, price, sugar, and fiber.

Finally, ϵ_{ijt} is an iid error term, which we model with a nested logit structure, such that

$\epsilon_{ijt} = \zeta_{igt} + (1-\rho)\tilde{\epsilon}_{ijt}$. The nesting parameter ρ captures the amount of correlation between the product-specific shocks within the same product group g. Our motivation for allowing for a nested logit structure is to obtain reasonable substitution patterns between the inside goods and the outside good. Therefore, we group all inside goods in one nest and the outside good in a separate nest.[9]

Consumers who do not purchase any cereal choose the outside good, the utility of which we normalize to ϵ_{i0t}.[10] Market share predictions are then given by

$$s_{jt} = \int_i \frac{\exp\big((\delta_{ijt} + \mu_{ijt})/(1-\rho)\big)}{\exp\big(I_{igt}/(1-\rho)\big)} \frac{\exp(I_{igt})}{\exp(I_{it})} dP_{it}, \tag{3}$$

where $\delta_{jt} = x_{jt}\beta + \alpha p^r_{jt} + \xi_{jt}$, $\mu_{ijt} = [p^r_{jt}, x_{jt}]' * \Phi D_i$, and I_{igt} and I_{it} are the inclusive values for consumer i from product group g and all products, respectively. The integral is taken over the distribution of consumer types in market t, P_{it}.

3.2 Supply Model

The J brands in the industry are produced by R multiproduct firms. We impose that wholesale marginal costs do not vary across stores and that manufacturers set a uniform wholesale price for all DFF stores for a specific brand each month; therefore, in our supply model a market is defined as a month instead of a month-store combination.[11]

We model wholesale marginal costs of producing cereals as a linear function of observable cost shifters w_{jt} and a brand-market specific cost shock ω_{jt}, that is unobserved by the researcher but known to the firms, so that $mc_{jt} = w_{jt}\gamma + \omega_{jt}$, where γ is a vector of marginal cost parameters to be estimated. In order to allow for flexible marginal costs, we include brand-half-year fixed effects and several input prices (for sugar, wheat, corn, oat, and rice) weighted with a cereal's content of each input in w_{jt}.[12] Analogous to our demand model, we allow for persistence in the unobserved cost shock and model ω as an $AR(1)$-process $\omega_{jt+1} = \iota_S \omega_{jt} + \nu_{Sjt+1}$.

Next, we describe the interplay between manufacturers and retailers, and the timing of promotion and price setting. It is important to note that we do not estimate the promotion part of the model. Rather we provide it to clarify our assumptions and to justify the validity of our identification strategy. The focus of our empirical supply model is on manufacturers' choice of base wholesale prices.

[9] Conceptually, grouping all inside goods into a nest is similar to including a random coefficient on the constant.

[10] Potential changes to the quality of the outside good are captured by our time fixed effects.

[11] Note that in contrast to many other retailers, DFF does not engage in uniform pricing, so that retail prices and promotions need not be perfectly synchronized across different DFF stores.

[12] The brand-half-year fixed effects also pick up most of the potential synergies arising from the Post-Nabisco merger. In a robustness check we also include a separate post-merger-merging firms dummy. The results are very similar and available upon request. For a more detailed discussion, see Appendix B.3.

In period $t-k$ (with $k \geq 1$), each manufacturer determines for each of its brands how often to put the brand on promotion. Furthermore, if product j is on promotion in period t, the manufacturer sets the trade spend payment td_{jt} also in period $t-k$. If brand j is not on promotion in period t, its trade spend is zero.

We assume that manufacturers have an exogenously given target for the promotion intensity of each brand and market. This target could come from a separate, potentially dynamic optimization problem, that we do not model. Furthermore, we assume that manufacturers make a take-it-or-leave-it offer with trade spend payments such that, in equilibrium, the retailer always accepts the offer by the manufacturers.

Promotions have two effects in our model. First, they have a perceived quality-enhancing effect that is captured by the inclusion of the brand-specific number of promotions in a market as an observable demand shifter in the consumer's utility function. Second, a promotion and the associated trade spend payment act like a temporary marginal cost reduction from the perspective of the retailer. This reduction has to be passed through to consumers in the form of a lower retail price in period t, because it is funded from the trade spend budget. The net effect of a brand's promotion on its total wholesale and retail price is determined by the sum of these two effects.[13]

In period t, a firm sets the base wholesale prices p^w_{jt} for all of its products j. We assume that retail prices are determined by the wholesale price plus a fixed retail markup, that is determined by the manufacturer and can differ across products and markets. The retailer is not allowed to respond independently to market conditions, however. Therefore, retail prices are given by $p^r_{jt} = p^{wb}_{jt} - td_{jt} + MU^r_{jt}$. The assumption of a fixed retailer margin is often used in the literature, see, for example, Miller and Weinberg (2017) and Backus *et al.* (2021), and several studies have rejected independent retail pricing models, see, for example, Bonnet and Dubois (2010).

In order to model deviations from Nash pricing, we follow the *profit internalization* approach similarly to Miller and Weinberg (2017) and Ciliberto and Williams (2014). We denote the degree to which brand i takes into account brand j's profits when setting its wholesale prices in market t by λ_{ijt}, which we treat as structural parameters. These parameters are arranged in an *internalization matrix* Λ_t, which generalizes the ownership matrix of zeros and ones in standard BLP-models. We do not restrict λ_{ijt} to the unit interval.

Therefore, our model can accommodate negative internalization parameters, which typically imply that pricing is more aggressive than Nash pricing. If interpreted structurally, a negative value indicates that firms derive a positive payoff from decreasing their rivals' profits. This is clearly counterintuitive, if firms consider only their static profits. As we discuss

[13] In our data we find that the retail price decreasing effect almost always dominates the quality-enhancing effect of promotions, so that during a promotion net wholesale prices are lower than in no-promotion periods, see Section 2.3 and Appendix A.3.

in Section 5, a plausible interpretation of a negative value is that it captures, in a reduced form, firm objectives other than static profits. A parameter $\lambda > 1$ would imply that a firm values its rivals' profits more than its own.

To keep the estimation tractable, we restrict the structure of Λ in an economically reasonable way. One of our primary goals is to quantify the evolution of conduct over time. Our structural break tests discussed in Section 2.3 suggest that pricing behavior indeed changed systematically across the pre-merger, the post-merger, and the price war period. Therefore, we estimate profit internalization parameters that change across but are constant within the three periods. We impose the standard assumption that after the merger, the merging firms fully internalize the profits of each other's product lines.[14] 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. 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.

A key challenge for our model is that we need to capture in a tractable way that a manufacturer commits in advance to lower its wholesale price during a promotion by paying an additional trade spend payment to the retailer.[15] We incorporate this commitment by assuming that brand j maximizes the following profit function in each period t:

$$\Pi_{jt} = (p_{jt}^{wb} - mc_{jt}) \underbrace{\sum_{g=1}^{G} w_t^g s_{jt}^g M_t^g}_{\tilde{s}_{jt}} + \sum_{k \neq j} \lambda_{jkt} (p_{kt}^{wb} - td_{kt} - mc_{kt}) \sum_{g=1}^{G} w_t^g s_{kt}^g M_t^g, \tag{4}$$

where s_{jt}^g is the market share of brand j in store g in month t, and $w_t^g = \frac{M_t^g}{\sum_l M_t^l}$ is the weight of store g in month t, which is computed as the share of the store-specific market size relative to the market size across all stores. $\tilde{s}_{jt}$ denotes the market share of brand j aggregated over all stores and p_{jt}^{wb} denotes the base wholesale price per unit of brand j in month t.

This profit function implies that the manufacturer ignores the direct costs of any applicable trade spend payment td_{jt} when setting its base wholesale price p_{jt}^{wb}. If instead brand j took into account its own trade spend commitment as part of its marginal cost, most static models would predict that the manufacturer reverts the promised trade spend with a higher base wholesale price. This would lead to total wholesale prices being higher during a promotion than in a no-promotion period with exactly the same characteristics.

Such a pattern would violate what we observe in the data, namely, that the total wholesale

[14] Our empirical strategy is flexible enough to easily incorporate models of partial within-firm internalization, as, for example, analyzed by Michel (2017).

[15] We judge a static model to be a reasonable simplification, because our data is aggregated to the monthly level, for which we believe that dynamic consumer behavior is much less relevant than for weekly data. In addition, the contingent contracts between manufacturers and the retailer are designed to prevent significant forward-buying by the retailer, see also our discussion in Appendix B.2.

price during a promotion decreases. From a practical perspective, it seems reasonable that manufacturers are willing to commit to this behavior in order to maintain a good relationship with the retailer, that is essential for the manufacturer to market its products to consumers.[16]

Note that base wholesale prices can react to contemporaneous shocks to marginal cost and demand, which, among others, is a function of the contemporaneous promotion pattern of all products. Furthermore, we assume that the manufacturer also takes into account that rival brands on promotion make trade spend payments to the retailer, which reduces the total profits that the manufacturer considers, if its profit internalization parameter is positive. If we assumed that a firm does not take into account its rivals' trade spend payment, it would internalize only a part of the rivals' profits, which would lead to a peculiar definition of conduct. Since the profit internalization parameters are the central object of our model, we prefer to keep their interpretation as standard as possible, i.e., have them work on the total profits of rivals.

Throughout, we treat the trade spend td_{jt} as exogenous and we do not model that firms could also adjust their promotions and trade spend as part of their conduct. Quantifying these aspects of firm behavior requires a substantially more complicated supply model, that considers endogenous promotions, trade spend, and base wholesale prices jointly. Such a model goes beyond the scope of this paper but is an interesting area for future research. We discuss the implications of this for the interpretation of our estimation results in Section 5.

Define $\Omega_{jkt} \equiv -\lambda_{jkt} * \frac{\partial \tilde{s}_{kt}}{\partial p_{jt}^{wb}}$, 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.[17] The first-order condition for brand j is given by

$$\tilde{s}_{jt} + td_{jt}\frac{\partial \tilde{s}_{jt}}{\partial p_{jt}^{wb}} + \sum_{k=1}^{J} \lambda_{jkt}(p_{kt}^{wb} - td_{kt} - mc_{kt})\frac{\partial \tilde{s}_{kt}}{\partial p_{jt}^{wb}} = 0. \tag{5}$$

The key difference to the first-order condition in a standard BLP model is the term $td_{jt}\frac{\partial \tilde{s}_{jt}}{\partial p_{jt}^{wb}}$ which comes from the fact that the firm does not consider the direct costs of the trade spend during a promotion period. This term will be zero by construction, if brand j is not on promotion. This first-order condition can be inverted as a function of the conduct matrix and the demand parameters to compute the vector of manufacturers' marginal costs of production for all products in market t.

Plugging in the marginal cost function 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

[16]For example, Anderson and Fox (2019) cite industry evidence that manufacturers consider the trade spend budget almost as sunk, because there is a strong sense in the industry that retailers are entitled to the trade spend.

[17]Note that $\frac{\partial \tilde{s}_{kt}}{\partial p_{jt}^{wb}} = \frac{\partial \tilde{s}_{kt}}{\partial p_{jt}^{r}}$, because we assume a model of fixed retail margins. In a model with independent retailer pricing the two derivatives would be different and depend on the exact assumption on the retailer behavior.

data, so that

$$\omega_{.t}(\theta_D, \gamma, \Lambda_t) = p^{wb}_{.t} - td_{.t} - w_{.t}\gamma - \Omega_t^{-1}\big(\theta_D, \Lambda_t, p^r_{.t}(p^w_{.t}), x_{.t}\big)\left(s_{.t}\big(\theta_D, p^r_{.t}(p^{wb}_{.t}), x_{.t}\big) + td_{.t} \cdot \frac{\partial \tilde{s}_{.t}}{\partial p^{wb}_{.t}}\right). \tag{6}$$

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

Two aspects, that are essential for our identification strategy, are noteworthy. First, we assume that firms cannot anticipate the innovations to marginal costs ν_{Sjt} before period t. We judge this to be a reasonable assumption given that we include a detailed set of fixed effects in the marginal cost function; therefore, the remaining cost shocks are plausibly hard to predict in advance. Second, contemporaneous promotions and trade spend are considered fixed by manufacturers and retailers, and they cannot be reversed. This is in line with the marketing literature on the institutional features of promotion setting and supported by our reduced form regressions, see Appendix A.7.

4 Identification & Estimation

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

4.1 Identification

Our identification strategy for industry conduct builds on the intuition first proposed by Bresnahan (1982) for homogeneous product industries and generalized by Berry and Haile (2014) for differentiated products markets. Their main insight is that a sufficiently rich set of rotators of the marginal revenue curve allows the researcher to discriminate among different oligopoly models. In a differentiated products model such rotators can be based on variables that are often used as standard BLP-instruments, for example, summary statistics of the characteristics of rival products, the number of available products, or other market characteristics.

In practice, however, these variables often do not exhibit a lot of variation so that, from an econometric perspective, standard BLP-instruments can be weak for identifying industry conduct. In addition, in many applications, including ours, the researcher does not even observe variation in the set of products offered, which makes these instruments collinear with brand fixed effects.

Our novel instruments for both our demand and our supply estimation are based on rival brands' promotions interacted with measures of proximity in the characteristics space.

These instruments follow the same logic as classical BLP instruments in the sense that rivals' promotions affect a firm's optimal markup by rotating its demand curve. Firms' responses to this variation in market conditions identify industry conduct.

Compared to most instruments employed in the existing literature, our instruments have several advantages. First, we do not require the availability of exogenous industry shocks, such as ownership changes, to identify industry conduct. 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 market-level data sets used in industrial organization or marketing; therefore, our empirical strategy can be easily applied to many markets and industries.

Before we discuss our specific instruments for both the demand and the supply model, we first discuss the general idea of our *promotion-differentiation* instruments and why they can serve as instruments on both the demand and the supply side. Formally, our promotion-differentiation instruments are defined as

$$z_{jt}^{x,k,w} = \sum_{i \in \mathcal{G}(j)} \mathbb{1}\left(|d_{ij}^x| < c_k^x\right) \cdot PROMO_{it}^w, \tag{7}$$

where x denotes a physical product characteristic, such as sugar or fiber content, $d_{ij}^x = x_i - x_j$ indicates how close products i and j are in dimension x, w denotes a type of promotional activity, for example, either *general* or *bonus buy*, and $\mathcal{G}(j)$ is the product portfolio that we consider for brand j. For example, we can consider either only the rival products owned by the same firm or all rival products owned by rival firms.

Intuitively, our instruments count the number of rival promotions in a given market but only consider those rivals that are sufficiently close according to some proximity measure. Since one typically has several options for x, k, w, and $\mathcal{G}$, this logic allows us to construct a large number of instruments, which can be used to identify rich patterns of industry conduct and consumers' substitution patterns.

Interacting the promotion intensity of rival brands with measures of the brands closeness in the characteristics space has a similar flavor as the differentiation instruments proposed by Gandhi and Houde (2020) for identifying consumers' substitution patterns. For our application, we find that an analogous logic yields powerful instruments for identifying industry conduct: 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 more competitive pressure than a distant product on sale; therefore, basing the instruments on the activity of close rivals should result in stronger instruments than relying on average statistics of all available products.[18]

[18]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.

In addition, our instruments need to be exogenous to the structural errors used to construct the moment conditions. Clearly, promotions are chosen by firms and are therefore endogenous. However, in almost all industries, including ours, decisions between retailers and manufacturers regarding whether a promotion for a particular market will occur in period t are made in advance, i.e., at the latest in $t-1$, while wholesale prices can still be flexible.

Under this timing assumption and the standard assumption that physical product characteristics are exogenous, our promotion-differentiation instruments are plausibly exogenous to innovations in both the demand and marginal cost shocks, as long as firms cannot anticipate the shock innovations. We extensively discuss the supporting evidence for our timing assumptions in Section 2 and provide additional supporting reduced form evidence from our data in Appendix A.7.

Given that we include several layers of fixed effects in both our demand and the marginal cost function, we judge it plausible that the innovations in the structural demand and supply shocks in period t cannot be anticipated by any firm before period t. As an additional "safety measure" we replace the observed promotions in Equation (7) with the number of predicted promotions from an auxiliary regression, which we explain below.

Next, we describe the specific instruments that we use for our demand and supply estimation, respectively. Conceptually, our demand model does not differ significantly from most of those used in the literature. Our main concern is to use strong instruments for prices and market shares to precisely identify both consumers' price sensitivity and the substitution patterns determined by the demographic interaction parameters and the nesting parameter. To do so we construct several sets of instruments.

First, we include brand dummies, year dummies, and month-of-the-year dummies as included instruments for themselves. Second, as instrument for retail prices we compute predicted wholesale prices. We obtain the predictions from a linear regression of observed net wholesale prices on the following regressors: brand dummies, month-of-the-year dummies, store dummies, a linear-quadratic time trend, gasoline and electricity prices, and input prices (for sugar, corn, wheat, rice, and oats) weighted with a cereal's respective grain content. In essence, our predicted wholesale price variable relies on input prices (cost shifters) as excluded instruments. An advantage of our price instrument compared to using all the excluded instruments in its raw form is that it combines the many variables more efficiently; therefore, our predicted wholesale prices have a flavor of Chamberlain's optimal instruments.[19]

Third, we instrument a brand's own promotion intensity in the current period, which we include as a demand shifter in the observed product characteristics x_{jt}, with a predicted promotion intensity in order to mitigate concerns that firms might be able to anticipate future

[19]Using predicted retail prices instead of predicted wholesale prices led to very similar results.

demand shocks.[20] For this regression, we use one- and two-period lags of the brand-store specific promotion counts and lagged input prices for various grains (weighted by the cereal's grain contents), electricity and gas prices, as well as brand dummies as regressors.[21]

Lastly, we employ a set of promotion-differentiation instruments as discussed above to identify the substitution patterns, see Equation (7). As product characteristics x we use sugar and fiber content, for the distance d_{ij}^x between products i and j along dimension x we use the 33%-percentile and the 66% percentile. Furthermore, we construct the instruments separately for (predicted) *general* and *bonus buy and coupon* promotions. Finally, we only consider brands that are owned by rival firms. This results in eight promotion-differentiation instruments in total.

On the supply side, the identification of the parameters in the marginal cost function γ is standard. We use brand-half year dummies and (weighted) input prices for sugar, corn, wheat, oat, and rice as included instruments.

In order to identify the conduct patterns, we consider the same promotion-differentiation instruments as for our demand estimation. However we make two modifications. First, since we estimate the supply model at the chain level, we use the predicted promotion counts aggregated to the chain level instead of the individual store level. Second, we also include instruments based only on other brands owned by the same firm. These instruments are relatively highly correlated with each other; therefore, among the instruments based on other brands owned by the same firm, we only keep the two instruments based on the 66th-percentile in the sugar dimension. This results in ten promotion-differentiation instruments for our supply estimation.

4.2 Estimation Algorithm

We estimate our model using the generalized method of moments (GMM) similarly to the seminal work by BLP. Following most of the literature, we estimate demand and supply parameters in two steps.

Demand estimation. Since our demand estimation is relatively standard, we relegate the details of the estimation algorithm to Appendix D. The key difference to many other studies is that we base our moment conditions on the innovations of the unobserved marginal cost shocks and not their levels.[22]

Our GMM estimator for the demand parameters θ^D minimizes the following objective

[20] The results using observed promotions instead of predicted ones are qualitatively similar and available upon request.

[21] Using predicted instead of observed promotion counts is essentially equivalent to instrumenting observed promotions with the variables that are included only in the auxiliary regression but not the main equation. In our case the excluded instruments are lagged promotions and lagged input prices.

[22] Similar approaches are used by Lee (2013) and Schiraldi (2011).

function

$$\hat{\theta}_D = \arg\min_{\theta} \nu_D(\theta)' Z_D \hat{W}_D^{-1} Z_D' \nu_D(\theta), \tag{8}$$

where $\hat{W}_D^{-1}$ is an estimate of the efficient weighting matrix based on parameter estimates obtained from a first-stage estimation with a 2SLS weighting matrix $E[Z_D' Z_D]^{-1}$ and Z_D denotes the demand instruments discussed in Section 4.1 and Appendix D.

Supply estimation. For the supply estimation we generalize the BLP-approach from a binary ownership matrix to a flexible conduct matrix. For a given guess for the supply side parameters $\theta_S = (\gamma, \lambda, \iota_S)$, we solve the stacked first-order conditions, given by Equation (5), for the unobserved cost shock ω for each brand and month. Afterwards, we compute the innovations in the marginal cost shocks ν_S as a function of the backed out ω-vector and the $AR(1)$-parameter ι_S.

Analogously to our demand estimation, we exploit orthogonality conditions between ν_S and a set of instruments Z_S described in Section 4.1 and Appendix D. The GMM estimator of our supply parameters is given by

$$\hat{\theta}_S = \arg\min_{\theta_S} \nu_S(\theta_S, \hat{\theta}_D) Z_S' \hat{W}_S^{-1} Z_S' \nu_S(\theta_S, \hat{\theta}_D), \tag{9}$$

where $\hat{W}_S$ is an estimate of the asymptotically efficient weighting matrix based on the parameter values obtained from the first-stage estimation using the 2SLS weighting matrix. We relegate additional details of the construction of our dynamic panel moments and potential alternative moments for both the demand and supply estimation to Appendix D.

5 Results

5.1 Demand Estimates

Table 2 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 in the consumer's utility function. Furthermore, we interact a consumer's income with preferences for the constant, price, sugar, and fiber content. In addition, we group all inside goods into one nest and the outside good into a separate nest.[23]

[23] We experimented extensively with alternative demand specifications that include additional demographic interactions and normally distributed random coefficients. The results are qualitatively similar, in particular, in terms of the implied price elasticities, which are the most important output of our demand model. Our larger demand models resulted in higher standard errors for some of the additional parameters, especially for the normally distributed random coefficients. We attribute this to the fact that our data contain information on only one local market and one retailer. We also estimated specifications that include interactions between a dummy for households with small children (less

Table 2: RCNL demand estimates: Main specification

	Mean	Income
Constant	0.178***	−2.801***
	(0.055)	(0.222)
Price	−15.475***	14.952***
	(0.241)	(0.919)
Sogginess	0.164***	
	(0.019)	
Sugar	−0.159*	1.349***
	(0.082)	(0.352)
Fiber	−0.606***	0.536***
	(0.023)	(0.156)
Promotions	0.607***	
	(0.045)	
Nesting parameter	0.381***	
	(0.069)	
AR(1) Coeff	0.849***	
	(0.032)	

*Notes: The estimation includes product-, month-of-the-year, and year fixed effects. Standard errors are in parentheses. *, **, and *** denote significance at the 10-, 5-, and 1-percent level, respectively. Number of observations: 97,092.*

All of our demand coefficients are precisely estimated and significant. The signs of the estimates for mean preferences are reasonable. The price coefficient is negative, and ceteris paribus, consumers prefer cereals with lower sugar and lower fiber content. Our estimated price-income coefficient is positive and significant indicating that high-income consumers are less price-sensitive. The constant-income interaction is negative so that high-income consumers have a lower preference for cereals overall. The demand for both sugar and fiber in cereals is positively correlated with income.

Promotions have a positive effect on consumers' purchase probabilities even after controlling for the lower retail price during a promotion. Finally, the error terms for the inside goods are substantially correlated as the positive nesting parameter (0.38) indicates.

Table 14 and Table 15 in Appendix E display the implied median price elasticities over all markets. The own-price elasticities are negative for all products with a median across markets and brands of roughly −4.5. We judge our elasticities to be in line with those in the literature.

For example, the magnitude of our elasticities is in between those of Nevo (2000b, 2001) and Backus *et al.* (2021) on the one hand and those of Meza and Sudhir (2010) on the other hand. The elasticities of Nevo (2000b, 2001) and Backus *et al.* (2021) imply somewhat less

than 10 years old) with the preference for sugar and experienced similar precision issues, which we attribute to the fact that within our data we observe relatively little variation in the share of family's with children across DFF stores.

elastic demand than our estimates. These papers use different samples that covers data from several U.S. cities and several retailers, while our data cover only one retailer in the Chicago area. In addition Backus *et al.* (2021) study a different time period (2008-2018) than we do. Meza and Sudhir (2010) estimate a demand model that focuses on the effects of private label brands using DFF data from an earlier time period (1990-1991). They find own-price elasticities between -5 and -10 for most national brands, which implies more elastic demand than our estimates.

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 tend to be high among the signature products of Kellogg's, for example, between *Corn Flakes* and *Frosted Flakes*. In addition, we generally observe strong substitution among products with similar characteristics, for example, among sugary cereals, such as *Kellogg's Frosted Flakes*, *Kellogg's Rice Krispies*, and *Quaker Cap'n Crunch*, or among low-sugar cereals, such as *GM Cheerios*, *Nabisco Shredded Wheat* and *Quaker Oats*. Finally, the diversion to the outside good is on average 45%.

Overall, 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 Nash pricing our model predicts negative marginal costs only for less than 0.04 percent of our observations.

5.2 Supply Estimates

On the supply side we focus on two different specifications. In our small model, we estimate a profit internalization matrix with three 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 the pre-merger and the post-merger period 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. Since we only observe nine months of the price war, we estimate only one parameter for this period; therefore, our large model features five profit internalization parameters. This specification allows us to capture that industry leaders might have different pricing incentives than smaller competitors.

For our small model, we find significant internalization between firms pre-merger, with an estimate of 0.64, see Table 3. If interpreted structurally, this parameter indicates that a firm values US-$ 1 profit of a rival firm as much as US-$ 0.64 of its own profits. For the

Table 3: Conduct estimates: Model comparison

	Small Model Pre-merger	Small Model Post-merger	Small Model Price War	Large Model Pre-merger	Large Model Post-merger	Large Model Price War
All Firms	0.642***	0.178	−1.424***			−1.378***
	(0.118)	(0.132)	(0.431)			(0.468)
KE and GM				0.632***	0.464**	
				(0.169)	(0.190)	
Other Firms				0.719***	−0.063	
				(0.205)	(0.194)	
AR(1) Coeff		0.048			0.082**	
MC Prod	0.096	0.125	0.139	0.095	0.123	0.139
Manuf MU	0.414	0.188	0.045	0.415	0.187	0.048

Notes: The table displays the conduct estimates for both the small and the large conduct specification. Standard errors are in parentheses and account for two-step estimation. The last two rows display estimated medians (across brands and months) of manufacturer marginal costs (in US-$ per ounce, including trade spend payments) and total wholesale margins, defined as $(p^w - mc^w)/p^w$, where p^w denotes the net wholesale price and mc^w is the marginal cost of production. Number of observations: 1,674.

post-merger period, the internalization parameter decreases to 0.18, which is not statistically different from zero.[24] Therefore, the estimated pricing behavior in the post-merger period is consistent with Nash pricing.

In the price war period, the profit internalization parameter drastically decreases further, with an estimate that is negative (−1.42) and significantly lower than zero. If interpreted structurally, this parameter indicates that firms derive a positive payoff from decreasing their rivals' profits. This is clearly counterintuitive, if firms focus purely on their static profits. A plausible interpretation is that our parameter captures, in a reduced form, a temporary punishment phase in the style of Green and Porter (1984).[25]

We perform several additional t-tests to test whether the internalization parameters are statistically different from joint profit maximization, i.e., a parameter of one, and whether they are statistically different over time. For the small model, we reject both joint profit maximization in any period and time-invariant conduct. Table 18 in Appendix E presents the detailed results of all hypothesis tests.

The results for our large model are overall in line with the ones from the small model. There are several notable differences, however. First, for the small firms pre-merger pricing

[24] Recall that, motivated by the results of our structural break tests, we define the post-merger period to start nine months after the merger is consummated, see Appendix A.6 for the details.

[25] We also estimated our supply model restricting all internalization parameters to the unit interval. These estimates are qualitatively different. The price war parameter is always estimated at the lower bound of zero, and for the post-merger period we often hit the upper bond of one. We prefer the unrestricted model specification, because it allows for a broader range of industry conduct.

is not statistically different from joint profit maximization, see row 2 in Table 18.

Second, during the post-merger period we reject Nash pricing for the large firms (Kellogg's and General Mills). However, we cannot reject Nash pricing for the smaller firms, see row 1 of Table 18. Third, we cannot reject that the internalization parameter for the large firms is constant across the pre-merger and the post-merger period, see row 4 of Table 18. Fourth, the internalization parameters for small and large firms are statistically different (at the 10%-level) across firm groups only in the post-merger period but not in the pre-merger period, see row 6 of Table 18.

For the small and the large model the Sargan-Hansen test statistics have p-values of 0.27 and 0.38, respectively; therefore, we do not reject the null hypothesis of the joint validity of the moment conditions for either model.

When interpreting our results with respect to policy recommendations, several caveats should be noted. First, we do not suggest that positive profit internalization parameters necessarily provide evidence that firms violated antitrust laws. Second, we do not claim that the merger caused the shifts in industry conduct. Instead, the focus of our model is to detect and measure systematic changes in pricing patterns associated with these events.

Third, our model focuses on conduct in base wholesale prices and takes promotions as given. Endogenizing promotions would require a much more complicated dynamic supply model. In our reduced form regressions we find that the trade spend payment associated with a single promotion activity does not statistically change over time, see Table 1. However, there is a slight increase in the total number of promotional activities over time, especially during our price war period. As a consequence, base prices and promotions move in parallel in our application, i.e., as base wholesale prices become more competitive, we also observe more promotions. Therefore, we argue that the qualitative trend in conduct that we find should be robust to endogenizing the promotion part of our model.

At first sight, our estimated conduct pattern might seem surprising. Many standard models, for example, a Cournot model with homogeneous products and symmetric firms, would predict that collusion becomes more sustainable after a merger. These theoretical predictions do not carry over to a more general setting with product differentiation and asymmetric firms, see, for example, Davis and Huse (2010) and Ivaldi and Lagos (2017).[26]

In addition, it could be that price coordination became harder over time for exogenous reasons that we do not model. The merger could simply have been an attempt by Post and Nabisco to sustain the existing price coordination in a different form.

The negative profit internalization parameter in the price war period might be surprising as well, because it implies that firms price more aggressively than static Bertrand-Nash at the end of our sample period.

[26] Both papers use repeated-game models to illustrate that after a merger non-merging firms always find it harder to collude, while the effect on the collusion incentives of the merging firms is not clear.

To put our estimates better into context, we compute the implied markups and marginal costs.[27] We find substantial heterogeneity in markups over time and across brands, see the last row of Table 3 for industry-level statistics and Table 17 in Appendix E for product-specific wholesale price-cost margins over time. In the pre-merger period, wholesale margins are almost twice as large as the ones implied by Nash pricing. After the merger, our estimated margins are relatively close to Bertrand-Nash margins. In the price war period median wholesale margins are less than 5%. This indicates that pricing strategies during the price war are closer to marginal cost pricing than to Bertrand-Nash, even though there is heterogeneity across brands. General Mills and and Post charge the highest margins during the price war period, around 10%. Kellogg's prices some of its brands even slightly below marginal costs, see Table 17 in Appendix E. Although our model does not allow us to investigate this aspect formally, a plausible interpretation of these margins is that manufacturers indeed engaged in a temporary price war, which could have been part of a dynamic price coordination strategy in the spirit of Green and Porter (1984). This interpretation is also consistent with industry evidence that cereal prices increased again a few months after the end of our sample period.[28]

Under the assumption of multiproduct Nash pricing, the median marginal costs implied by our model are US-$ 0.125 per serving. For our small conduct specification, marginal cost estimates in the pre-merger period are substantially lower (US-$ 0.096) and increase over time to US-$ 0.125 and US-$ 0.139 in the post-merger and price war period, respectively.[29]

6 Counterfactual Simulations

We use our structural model to simulate how changes in industry conduct would affect consumer surplus and manufacturers' pricing. All of our counterfactual results are based on the small conduct specification with three profit internalization parameters.[30]

We start by decomposing the price changes in the post-merger period into the *unilateral* and *coordinated* effects of the merger. To do so, we start by simulating the industry assuming that the Post-Nabisco merger is not consummated and conduct remains at the estimated pre-merger level, i.e., $\lambda_{Post} = \hat{\lambda}_{Pre} = 0.64$ (CF1). Afterwards, we simulate the industry in the post-merger period assuming that the post-merger conduct remains at the pre-merger level

[27]Throughout the paper, we report costs and firm profits including trade spend payments. That is, the marginal costs in Table 3 should be interpreted as the sum of the marginal costs of production and the trade spend payments that the manufacturer makes to the retailer to implement the promotion plan. Analogously, the markups are based on total and not base wholesale prices.

[28]See, for example, https://www.nytimes.com/1997/07/04/business/stocks-of-cereal-companies-up-on-possible-price-rises.html.

[29]This marginal cost evolution is consistent with public remarks by Kellogg's that production costs increased during this time period, see, for example, https://www.nytimes.com/1993/08/10/business/consumers-wake-up-to-increases-in-cereal-prices.html. Our large conduct specification yields marginal costs estimates that are very similar on average but more heterogeneous across brands.

[30]The results using the large model are similar and available upon request.

and the Post-Nabisco merger takes place, i.e., $\lambda_{Post} = \hat{\lambda}_{Pre} = 0.64$ (CF2). Next, we simulate the post-merger period for the case in which conduct changes as estimated and the Post-Nabisco merger does not take place, i.e., $\lambda_{Post} = \hat{\lambda}_{Post} = 0.18$ (CF3). Finally, in CF4 we simulate the industry in the price war period assuming that the price war does not occur, i.e., firms continue to price as in the post-merger period with $\lambda_{Pw} = \hat{\lambda}_{Post} = 0.18$.

Table 4 summarizes our results for changes in net wholesale prices, i.e., including exogenous trade spend as observed in the data, retail prices, consumer surplus, sold quantities, and firm profits.[31] All reported price changes are averages over all markets in the respective period analyzed. For consumer surplus, quantities, and firm profits, we report the total changes aggregated over all months of the period analyzed and all 58 DFF stores in our sample.

Table 4: Summary of counterfactual results

	CF1 Post merger	CF2 Post merger	CF3 Post merger	CF4 Price War
$\Delta\ p_w$ (in %)	6.79	6.81	-0.54	8.85
$\Delta\ p_r$ (in %)	6.81	6.84	-0.50	8.87
Δ CS (total, in mio-USD)	-1.94	-1.95	0.21	-1.02
Δ quantities (total, in mio-oz)	-18.95	-19.03	2.15	-10.00
Δ firm profit (total, in mio-USD)	1.06	1.07	-0.08	1.10

Notes: CF1 = No merger and no conduct change in post-merger period ($\lambda_{Post} = \hat{\lambda}_{Pre} = 0.64$). CF2 = Merger and no conduct change in post-merger period ($\lambda_{Post} = \hat{\lambda}_{pre} = 0.64$). CF3 = No merger and conduct change in post-merger period ($\lambda_{Post} = \hat{\lambda}_{Post} = 0.18$). CF4 = Same conduct as in post-merger in price war period ($\lambda_{Pw} = \hat{\lambda}_{Post} = 0.18$).

The first three columns summarize the results from our post-merger counterfactuals. We use column 1 as a benchmark in which both the ownership structure and the estimated profit internalization parameter in the post-merger period are the same as in the pre-merger period. In this scenario prices increase significantly, by 7%, consumer surplus decreases by US-$ 2 million, and total firm profits increase by US-$ 1 million over the two-and-a-half years of our post-merger period.

Column 2 and 3 reveal that almost all of the change in the post-merger period is due to the change in industry conduct and not the unilateral effects of the Post-Nabisco merger. If the Post-Nabisco merger is consummated but post-merger conduct remains at the pre-merger level of 0.64 instead of the estimated 0.18, prices increase only marginally more compared to the case in column 1, see column 2. Analogously, consumer surplus and profit changes are almost identical to the ones in column 1. If conduct changes as estimated, but the merger does not occur, prices are only marginally lower than in the observed data, by 0.5%, see column 3. Consumer surplus and firm profits remain almost the same as in the observed

[31] As a measure of consumer surplus, we estimate the compensating variation, i.e., the dollar amount such that consumers would be equally well off in both the observed industry state and the counterfactual simulation.

data as well.

In summary, our post-merger counterfactuals provide evidence that coordinated effects can be much more important than unilateral effects for quantifying the effects of a merger, and that conduct post-merger can be substantially more aggressive than before the merger.

Our last counterfactual confirms that firms price very aggressively during the price war period. If conduct in the price war period remains at the post-merger level of 0.18, both retail and wholesale prices are 9% higher. Consumer surplus is lower by roughly US-$ 1 million and firm profits are US-$ 1.1 million higher when aggregated over the price war period, which we define as the last nine months of our sample.

7 Conclusion

In this paper, we estimate the evolution of competition in the U.S. 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 it 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' isolation in the characteristics space with data on rival firms' temporary promotional activities. Intuitively, our identification of industry conduct is based on two ideas: First, rivals' promotions act as sequentially exogenous demand rotators. Second, a firm's markups should 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 in 1993 and second, a period of large wholesale price cuts in 1996. Our estimation results suggest that in the beginning of our sample, the industry was characterized by substantial price coordination, with wholesale margins that are almost twice as large as those implied by Nash pricing. Nine months after the Post-Nabisco merger, price coordination decreased to a level that is on average not statistically different from Nash pricing. When allowing conduct to differ across firms, we find that only the small firms revert to Nash pricing. The industry leaders (Kellogg's and General Mills) continue to exhibit a profit internalization parameter larger than zero. Our conduct estimates for the last months of our sample indicate even more aggressive pricing, with median wholesale margins of less than 5%.

These results indicate that a significant percentage of the markups of national cereal manufacturers during the the early 1990s can be attributed to cooperative industry behavior, and most importantly, that there are substantial changes in conduct over time.

A well-known critique of the *profit internalization approach* in general is that the estimated parameters ultimately constitute only a reduced form approximation to a more structural model of firm behavior, for example, in the form of a repeated game. While the development of such a framework goes beyond the scope of this paper, it is a promising area for future research.[32] Our empirical strategy and the rich set of instruments that we propose are likely to be easy to adapt to these more complicated settings. In particular, a structural repeated game model is likely to contain more parameters than ours. The empirical results from our application provide first evidence that our instruments may work well for estimating markups in such high-dimensional models. In addition, our instruments can straightforwardly be incorporated into testing-based approaches to quantify industry conduct, as, for example, proposed by Backus *et al.* (2021) and Duarte *et al.* (2022).

Recently, there has also been an increased interest in the evolution of markups over time from a macroeconomic perspective. De Loecker *et al.* (2020) document a substantial increase in markups from 1980 onwards for the U.S. economy by using a production function approach. Our approach can be seen as complementary to this literature. By focusing on estimating the competitive interactions between firms within an industry, one can gain detailed insights into the extent to which both 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.

[32]See, for example, Miller *et al.* (2021) for a micro-founded approach to model oligopolistic price leadership in the beer industry.

Anderson, E., Malin, B. A., Nakamura, E., Simester, D. and Steinsson, J. (2017). Informational rigidities and the stickiness of temporary sales. *Journal of Monetary Economics*, **90**, 64–83.

Anderson, E. T. and Fox, E. J. (2019). How price promotions work: A review of practice and theory. In *Handbook of the Economics of Marketing*, vol. 1, Elsevier, pp. 497–552.

— and Simester, D. I. (1998). The role of sale signs. *Marketing Science*, **17** (2), 139–155.

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

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, **68** (1), 29–51.

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.

Backus, M., Conlon, C. and Sinkinson, M. (2021). *Common ownership and competition in the ready-to-eat cereal industry*. Tech. rep., National Bureau of Economic Research.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pp. 47–78.

— and — (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, **18** (1), 1–22.

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.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, **87** (1), 115–143.

Bonnet, C. and Dubois, P. (2010). Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance. *The RAND Journal of Economics*, **41** (1), 139–164.

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

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.

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.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics*, **51** (4), 1108–1161.

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. (2018). The welfare effects of vertical integration in multichannel televistion. *Econometrica*, **86** (3), 891–954.

Davis, P. and Huse, C. (2010). Estimating the "coordinated effects" of mergers. *Competition Commission, January*.

De Loecker, J., Eeckhout, J. and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, **135** (2), 561–644.

Döpper, H., MacKay, A., Miller, N. and Stiebale, J. (2021). Rising markups and the role of consumer preferences. *Available at SSRN 3939126*.

Duarte, M., Magnolfi, L., Sølvsten, M. and Sullivan, C. (2022). *Testing firm conduct*. Tech. rep., working paper, University of Wisconsin-Madison.

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.-F. (2020). Measuring substitution patterns in differentiated-products industries. *NBER Working Paper*, (w26375).

—, Lu, Z. and Shi, X. (2020). Estimating demand for differentiated products with zeroes in market share data. *Available at SSRN 3503565*.

Green, E. J. and Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica: Journal of the Econometric Society*, pp. 87–100.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In *The economics of new goods*, University of Chicago Press, pp. 207–248.

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

Ivaldi, M. and Lagos, V. (2017). Assessment of post-merger coordinated effects: Characterization by simulations. *International Journal of Industrial Organization*, **53**, 267–305.

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

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

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, **103** (7), 2960–3000.

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., Sheu, G. and Weinberg, M. C. (2021). Oligopolistic price leadership and mergers: The united states beer industry. *American Economic Review*, **111** (10), 3123–59.

— 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. (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal*, **5** (1), 1–39.

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

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

Schiraldi, P. (2011). Automobile replacement: a dynamic structural approach. *The RAND Journal of Economics*, **42** (2), 266–291.

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

Shi, X. (2015). A nondegenerate Vuong test. *Quantitative Economics*, **6** (1), 85–121.

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.

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

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 Data Management and Reduced Form Evidence

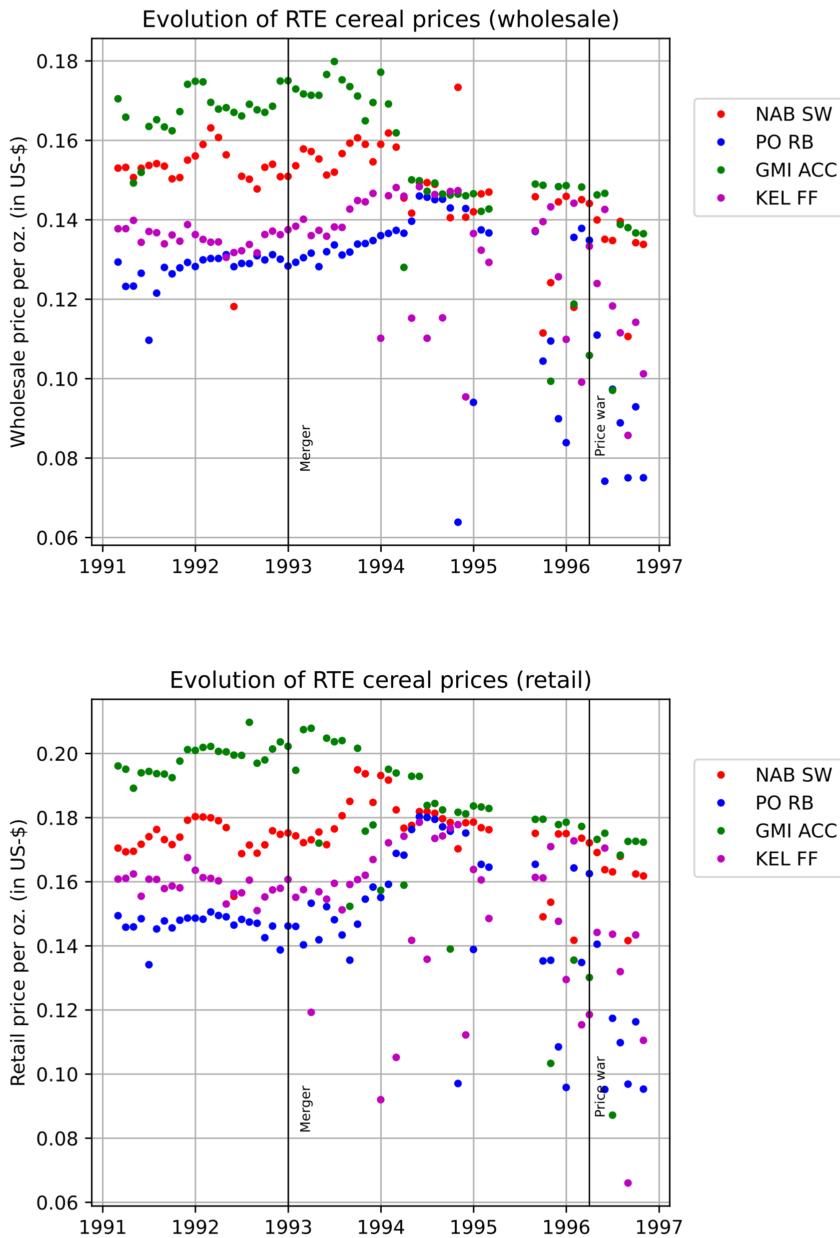
A.1 Market Shares and Price Evolution

Table 5: Market share evolution

	GMI	KEL	POS	NAB	QUA	RAL	HHI
1991	32.2	47.1	7.8	0.9	8.8	3.3	3400
1992	29.4	47.7	9.7	1.5	8.7	3.0	3317
1993	27.9	48.8	8.9	0.0	11.0	3.4	3367
1994	25.1	49.1	9.8	0.0	12.8	3.1	3311
1995	33.3	43.9	10.8	0.0	9.0	3.1	3243
1996	27.3	49.9	10.5	0.0	9.7	2.6	3447

Notes: The table summarizes the firm-specific volume-based market shares across all stores in our data set for each year, as well as the Herfindahl-Index (HHI). 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.

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 General Mills Apple Cinnamon Cheerios.

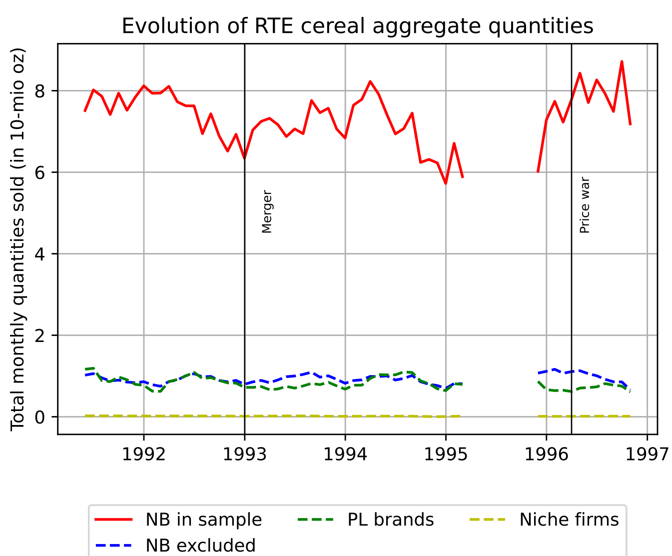
A.2 Details on Sample Selection and Role of Private Label Products

In this appendix, we provide background information on the evolution of various products that we pool into the outside good. We start by discussing what fraction of the overall cereal market is captured by our data. Afterwards, we present statistics on the evolution of prices and quantities of various product groups in the cereal segment that we exclude from our sample. Based on these descriptive statistics we justify the exclusion of these products and argue that our sample selection does not affect the essence of our estimation results.[33]

The 27 brands in our sample cover 56% of the overall sales in the cereal category at DFF stores. The remaining 44% of cereal sales can be attributed to the following product groups: (1) national niche manufacturers, (2) niche brands by the large national manufacturers, (3) private label cereals, and (4) products that we judge to be in a separate market from RTE cereals. With our product selection, we cover a similar share as other studies on the same industry and a similar time period, in particular, Nevo (2001, 2000b). We include all brands from his sample and in addition add the brands from the national manufacturer *Ralston*. In the following, we discuss the role of each of the excluded product groups in more detail.

First, our data cover two additional national manufacturers: *Sunbelt* and *Kashi*. Both firms maintain a focus on homemade-style, environmentally-friendly-produced products that are marketed to health-conscious consumers.[34] In addition to having only a minuscule share of the market (see the yellow line in Figure 2) and low availability, we argue that many consumers consider these products as separate from the mainstream RTE cereal market that we analyze.

Figure 2: Evolution of aggregate quantities for different product groups



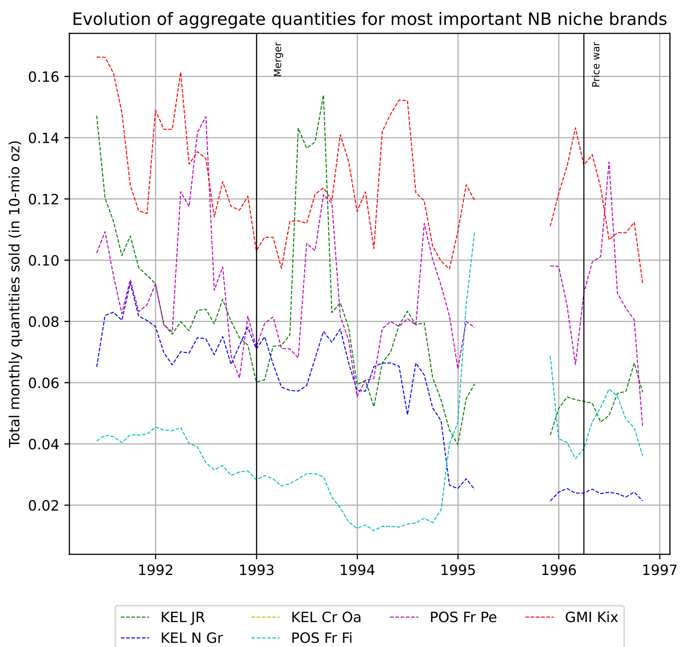
Notes: The figure displays the evolution of the aggregate quantities sold for the national brands (NBs) included in our sample, the most important NBs excluded, the most important private label brands (PLs) and the most important brands from niche manufacturers, respectively.

[33]Note that we exclude five months in 1995 from our sample because of a substantial amount of missing data in the DFF database during these months.

[34]Most of their products fall into the category of whole wheat or oat cereals and granolas. Most of Kashi's products are also certified organic.

Second, our raw data contain information on over 200 brands by the national cereal manufacturers that were offered at some point during our sample period in some stores. However, none of these products achieved a significant presence in the market. A typical product in this subsample has only a very small market share (typically less than 1% of the total inside good share) and is available in only about 15% of the week-store combinations, most likely due to the product being only offered in an experimental phase. When considering all these brands together, their sales represent less than 15% of the national brands in our sample.

Figure 3: Evolution of aggregate quantities: Selected NB niche brands

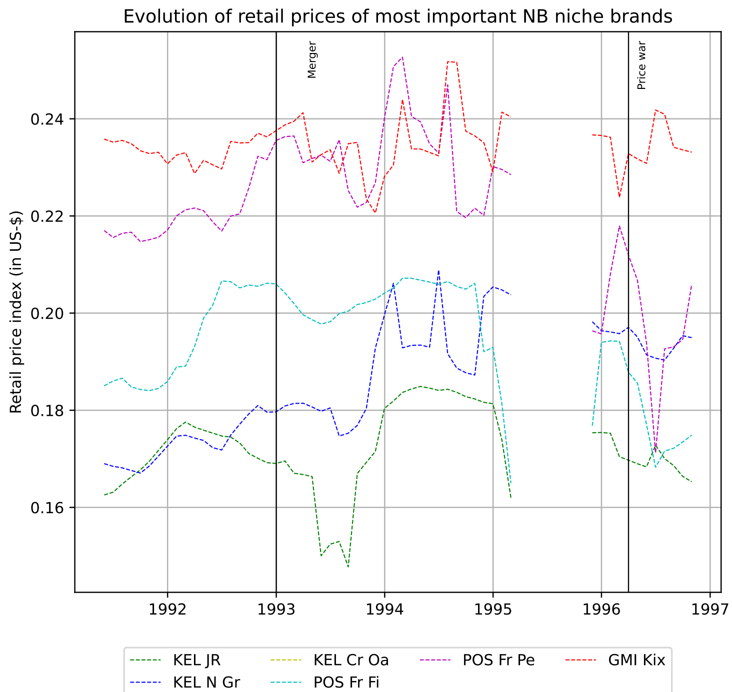


Notes: The figure displays the evolution of the aggregate quantities sold for selected niche brands (by major national manufacturers) that are excluded from our sample.

Therefore, while there is in principle a huge set of products that we could incorporate in the model, this right tail of products is extremely scattered and the associated price and quantity data is likely to be measured with significant measurement error, which in our experience makes it hard to extract meaningful information from these products.

Figure 3 illustrates the quantity evolution of the most popular niche brands by the national manufacturers. Overall, the shares exhibit significant fluctuations around a relatively stable mean. In addition, Figure 4 illustrates the price evolution of the niche brands of the national manufacturers. While these brands roughly follow the same trends as our in-sample brands (a

Figure 4: Evolution of average prices: Selected NB niche brands



Notes: The figure displays the evolution of the average (retail) price of selected niche brands (by major national manufacturers) that are excluded from our sample.

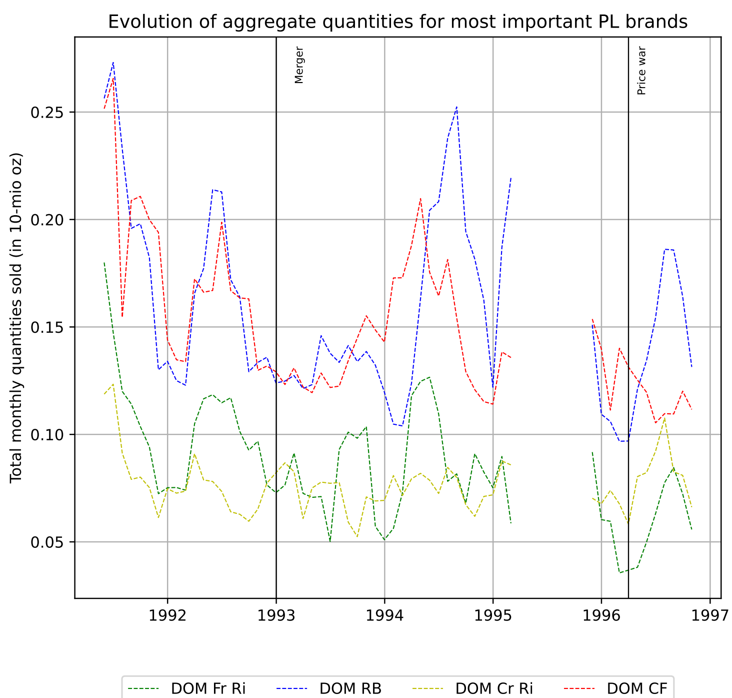
slight increase post-merger and a drop in the price war period), the trends are less pronounced and prices are more volatile. Therefore, we believe that we do not miss any significant trends in the industry by excluding these brands. Instead, restricting the sample to well-established brands helps us to keep measurement error in our estimation reasonably low.

Third, our data contain information on more than 15 private label DFF cereals. Figure 2 illustrates that private label cereals at DFF stores did arguably not play an important role during our sample period, with combined private label sales amounting to less than 15% of the sales of the national brands in our sample. Most importantly, the evolution of private label quantities remains reasonably stable over time.[35] Note that, in order to accommodate any remaining concerns about changes in the outside good, we include year-dummies for the inside goods in our demand model, see Section 3.

Figure 5 provides more information on the evolution of the quantities of the four most popular private label brands at DFF. There is some evidence that some private label brands

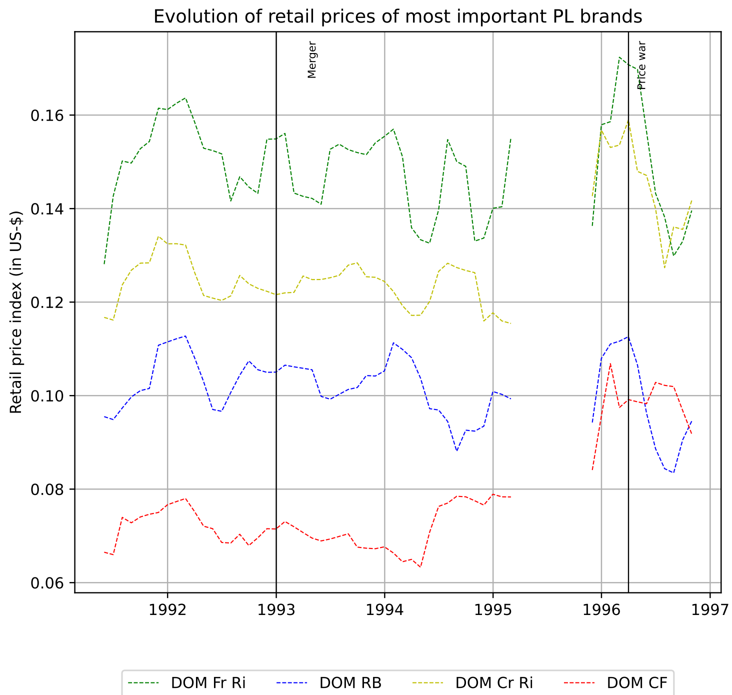
[35]On the national level private label products gained somewhat more importance during our sample period. As illustrated in Figure 2 this trend is not observed at DFF stores to the same extent.

Figure 5: Evolution of aggregate quantities: Selected PL brands



Notes: The figure displays the evolution of the aggregate quantities sold for selected PL brands that are excluded from our sample.

Figure 6: Evolution of average prices: Selected PL brands



Notes: The figure displays the evolution of the average (retail) price of selected private label brands (PLs) that are excluded from our sample.

gained market share in the post merger period, for example, DOM Raisin Bran and DOM Corn Flakes between 1994 and 1995. However, the quantity evolution is fairly volatile, and relatively flat for other private label brands.

Figure 6 displays the evolution of prices for the most important private label brands. In line with the aggregate quantity evolution discussed above, there is not an obvious trend in the price evolution, even though arguably some private label brands did not decrease their prices in the price war period much.

Fourth, a small share of the products officially labeled as cereals comes from products that most consumers would not consider a suitable substitute to RTE cereals, for example, Pop Tarts.

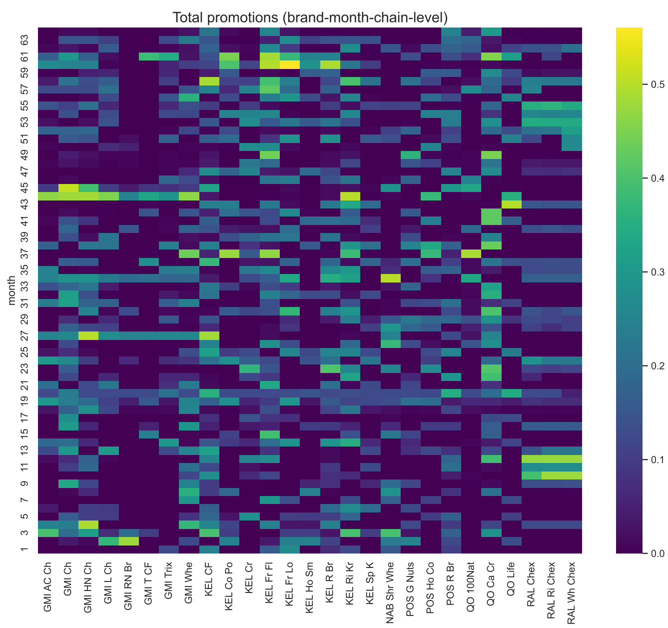
A.3 Distribution of Promotional Activities and Prices

In this appendix, we describe several characteristics of our promotion data and how these patterns provide us with suitable variation to construct instruments to identify industry conduct.

Figures 7 and 8 illustrate the distribution of the number of promotions for each product at different levels of aggregation using heatmaps. As our main measure of promotion "breadth" we count the number of promotions as follows. We start with the data on the UPC-week level and declare a brand to be on promotion if at least one of its UPCs is on promotion. We then aggregate this dummy variable from the brand-store-week level over all stores and all weeks of a given month to obtain the promotion count on the brand-month level.[36]

For the illustration, we scale the promotion intensity to be between zero and one with zero indicating that a product is never on promotion, and one denoting that the product was on promotion at every opportunity (week, store, and UPC). In all heatmaps, brighter colors indicate a higher promotional intensity than darker ones.

Figure 7: Distribution of promotional activities at the month-brand level

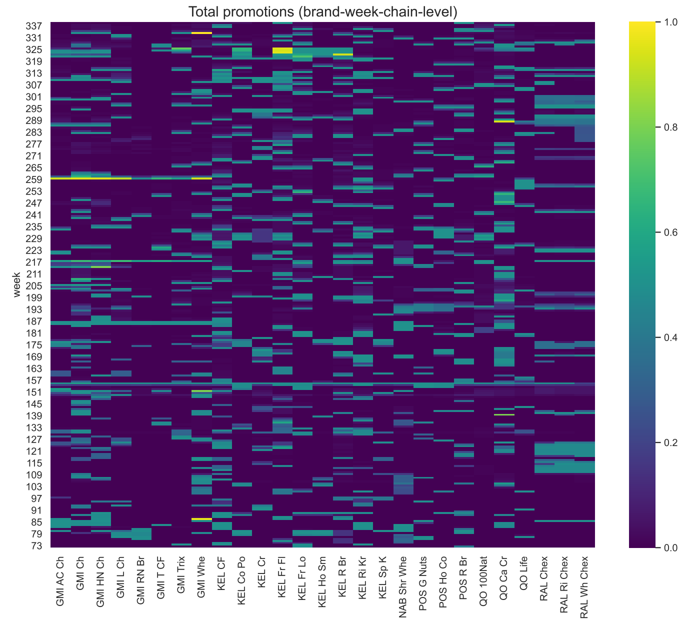


Notes: The figure displays a heatmap of the distribution of the total promotion intensity for each brand (on the x-axis) over time (on the y-axis) aggregated over all stores in our sample on a monthly level.

In Figure 7 we plot the total number of promotional activities aggregated over all UPCs for a given brand and all 58 stores in our sample for each brand-month combination. This figure illustrates that, even though promotions are often a weekly activity, there is substantial variation in promotional activity both across time and brands at the brand-month aggregation level, which is the aggregation level that we use for our structural supply estimation. While some firms (General Mills and Ralston) tend to coordinate promotions across their brands, others (Kellogg's, PostNabisco, and Quaker) put brands on promotion more asynchronously.

[36]Counting the number of promotions slightly differently, such as counting promotions for all UPCs of the same brand or counting only the promotions of a brand's main UPC led to qualitatively similar statistics and did not affect the results of our estimation significantly.

Figure 8: Distribution of promotional activities at the week-brand level



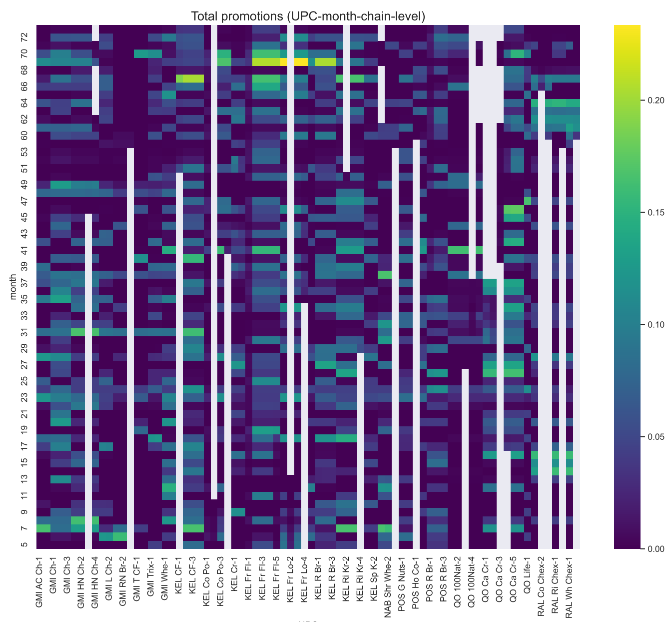
Notes: The figure displays a heatmap of the distribution of the total promotion intensity for each brand (on the x-axis) over time (on the y-axis) aggregated over all stores on a weekly level.

In Figure 8 we plot the total number of promotions for each brand-store combination for each week in our sample. Qualitatively, we observe the same patterns as with the monthly aggregation; therefore, we argue that we do not lose significant variation in our promotion variables, when aggregating to the monthly level.

In Figure 9 we disaggregate our measure of promotion intensity to the UPC level for each month, aggregated over all stores. Most importantly, we observe that different UPCs of the same brand have a strong tendency to be on promotion at the same time in DFF stores; therefore, we argue that UPC composition effects do not substantially affect the use of promotion intensity to construct our instruments.

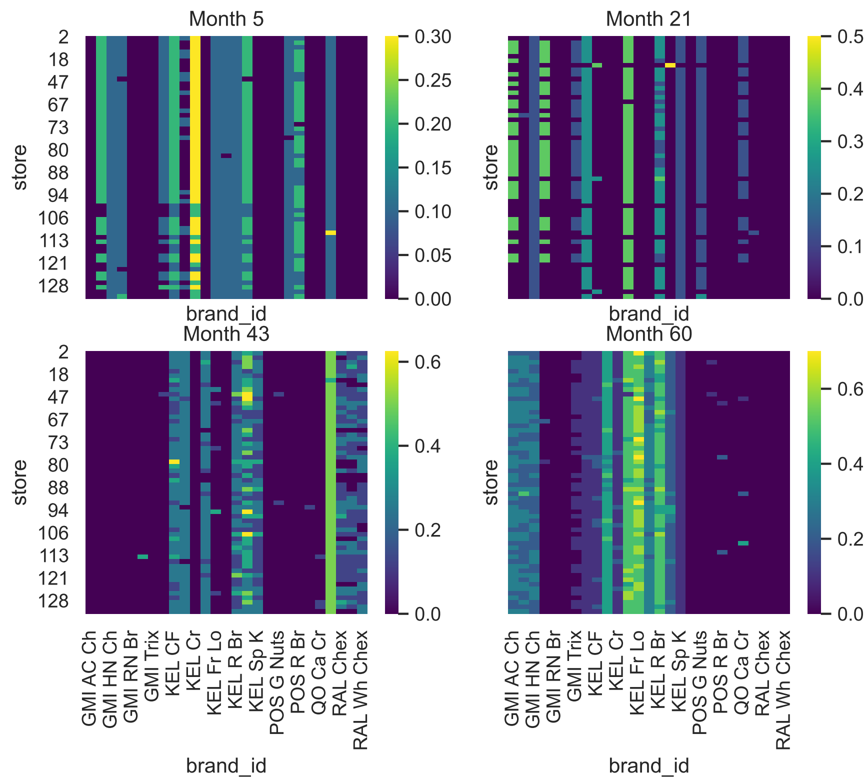
Finally, we illustrate the coordination of promotions across different stores. Figure 10 highlights that promotions are often very similar across different stores, i.e., products tend to go on promotion at the same time. However, this correlation is not perfect, in particular, for Kellogg's brands. We take this as evidence that, even though our main variation in promotion intensity occurs over time and across brands, there is some variation across stores. This provides us with an additional layer of variation to use our instruments for the demand estimation, which we conduct at the month-store-brand level.

Figure 9: Distribution of promotional activities at the UPC-month level



Notes: The figure displays a heatmap of the distribution of the total promotion intensity for each combination of UPC (on the x-axis) and month (on the y-axis) aggregated over all stores.

Figure 10: Distribution of promotional activities across stores for representative months



Notes: The figure displays a heatmap of the distribution of the total promotion intensity across brands (on the x-axis) and stores (on the y-axis) for several representative months of our sample.

In the following, we provide additional descriptive statistics on the promotion frequency (promotion breadth) and promotional price changes (promotion depth) at the chain-month-brand level. Table 6 compares the statistics across the different brands in our sample. The first four columns show the share of no-promotion months, the average retail price, the average (total) wholesale price, as well as the average retail margin during months, in which the specific brand is not on promotion in any DFF store. Columns 5 to 13 present the analogous variables for the subset of months in which a given brand is on promotion in at least one store and week during a given month. In addition, we compare the following variables across brands: the share of months in which the brand is on promotion at least in one store and week (Share) as a measure of the extensive margin of promotion breadth, the average share of store-week units that participate in the promotion conditional on some promotion taking place during this month (Pr Int) as a measure of the intensive margin of promotion breadth, the average retail discount, i.e., by how much retail prices are lower on average during a promotion period than during a no-promotion period, and the analogous average (total) wholesale price discount. We interpret the latter two variables as a measure of the monetary depth of a promotion.[37] Finally, we compare the average time between two promotional activities (Avg spell), and the average duration of a promotion (Avg dur).

Overall, we find significant heterogeneity across brands in how often a product is promoted. While the average brand is promoted roughly 60% of the time at least somewhere within the DFF chain, a few General Mills brands (Raisin Nut Bran and Total Cornflakes) are relatively rarely promoted. Several Kellogg's products, for example, Rice Krispies and Frosted Flakes, are on promotion somewhere at DFF nearly 90% of all months. Most importantly, we do not observe significant trends in these numbers over time, see also Figure 7. Conditional on being on promotion, there is some variation across brands in the overall promotion breadth (see column Pr Int) with the average brand being on promotion in roughly 25% of all store-week bins. However, there is significant heterogeneity. For example, several sugary cereals (Quaker Cap'n Crunch and Kellogg's Frosted Flakes) are on average on sale in more than 33% of all store-week combinations conditional on being on promotion at DFF. In contrast, several brands that are marketed as healthy cereals (Kellogg's Special K and Quaker Oats) tend to be promoted less heavily with an average intensity of 0.18 and 0.16, respectively.

We interpret our measures of retail and wholesale price discounts as a measure of promotion depth. On average, retail prices are roughly 10% lower during promotion periods. However, there is some heterogeneity across brands. While some of the rare-promoters (General Mills Raisin Nut Bran and Total Cornflakes) hardly reduce prices, some popular kids cereals (Quaker Cap'n Crunch and Kellogg's Cornpops) decrease retail prices by almost 15%

[37]Unfortunately, we do not observe any information on the non-monetary depth of promotions, such as the prominence on the retailer's shelf space, or advertising intensity.

on average during a promotion. Similar patterns are observed for wholesale price discounts although they are slightly less pronounced than the ones for retail prices. The smaller share of promotion periods for some brands translates into longer spells between two promotional activities. While rare promoters on average wait three to five months after the end of one promotion until the next one occurs, frequent promoters only exhibit no-promotion gaps of one to two months on average. Consistent with this pattern, frequent promoters on average conduct uninterrupted promotions for more than six months, while rare promoters incur promotion stretches of two months or less. This duration variable should be interpreted with caution, however, because by construction it suffers from a time-aggregation issue. For example, in our monthly data it is possible that a brand goes on promotion only in the first week of every month, but not during the remaining three weeks of each month, which our comparison would pick up as an interrupted promotion stretch. In contrast, a brand that goes on promotion for four uninterrupted weeks in month 1, and is not on promotion in any week of month 2 will be recorded as a promotion duration of only one month.[38] Given that our instruments mostly exploit variation in total promotion breadth, as measured by the number of total promotional activities on the chain-month level, we judge this problem to be of little relevance for our identification strategy. In addition, it is important to keep in mind that we use the promotion data mostly to construct our instruments, which need to satisfy the familiar relevance and exogeneity conditions. However, our instruments need not be perfect, i.e., they can contain measurement error without affecting the consistency of our parameter estimates.

Retail price dispersion across stores. One difference between DFF and most other retailers is that DFF uses zone-pricing instead of uniform retail pricing. We provide an illustration of the across-store retail price variation separately for promotion and no-promotion observations in Figure 11. We plot the distribution of the spread, i.e., the maximum price for a given brand-month combination minus the minimum price across all stores.

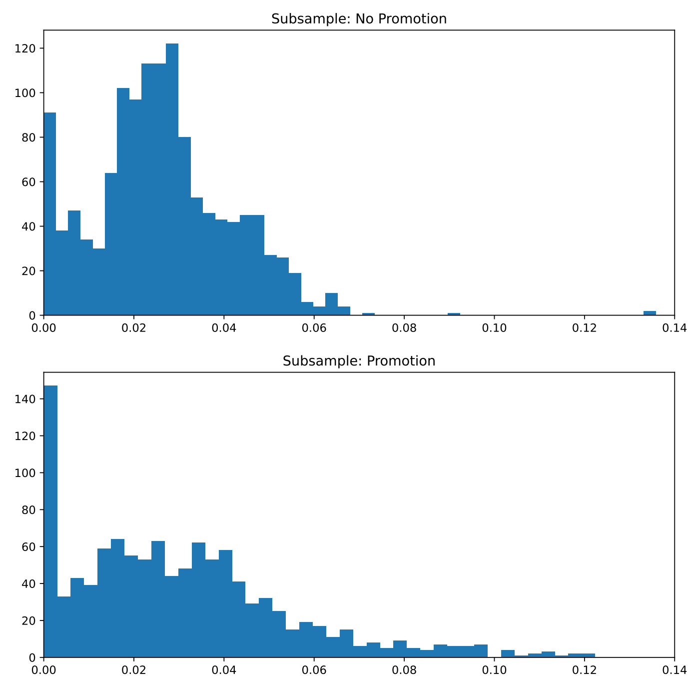
We observe that there is some variation across stores for a given brand-month combination both during promotion and no-promotion periods.[39] The mean retail price spread is roughly US-$ 0.03, which corresponds to approximately 20% of the average retail price.

While the across-store price variation helps somewhat for the demand estimation, it is not crucial, and the variation across brands and months is generally sufficient for our empirical strategy. In particular, our supply estimation is conducted at the chain level, which relies on variation across brands and over time exclusively.

[38] The analogous descriptive statistics on a weekly level look similar and do not provide much additional insights. They are available upon request.

[39] Naturally, part of this across-store variation is due to aggregation biases, because we aggregate from the UPC-week to the brand-month level. However, a non-negligible share of the dispersion persists on UPC-week level.

Figure 11: Distribution of retail price spreads across stores



Notes: The figure illustrates the distribution of the price dispersion across stores. *Spread* is measured as the difference between the maximum and the minimum retail prices observed across stores for a given month and brand. Retail price is measured in US-$ per ounce. The upper panel visualizes the distribution for the subsample of brand-store-month combinations that are not on promotion. The lower panel visualizes the distribution for the subsample of observations that are on promotion. The mean spreads for the no-promotion and the promotion sample are 0.0262 amd 0.0303, respectively.

Table 6: Comparison of no-promotion and any-promotion periods (monthly level)

	No-promotion periods				Any-promotion periods								
Brand	Share	Avg p^r	Avg p^w	Avg rm	Share	Pr Int	Avg p^r	p^r disc.	Avg p^w	p^w disc.	Avg rm	Avg spell	Avg dur
GMI AC Ch	0.45	0.19	0.16	0.19	0.55	0.28	0.18	0.90	0.15	0.94	0.13	2.00	2.50
		(0.01)	(0.01)	(0.04)		(0.25)	(0.03)		(0.02)		(0.17)	(1.11)	(1.87)
GMI Ch	0.19	0.20	0.17	0.17	0.81	0.33	0.18	0.91	0.16	0.93	0.08	1.22	5.22
		(0.01)	(0.01)	(0.03)		(0.24)	(0.03)		(0.02)		(0.16)	(0.44)	(3.87)
GMI HN Ch	0.28	0.19	0.16	0.17	0.72	0.31	0.17	0.91	0.15	0.95	0.10	1.21	2.57
		(0.01)	(0.01)	(0.04)		(0.26)	(0.02)		(0.02)		(0.13)	(0.43)	(1.70)
GMI L Ch	0.38	0.22	0.18	0.18	0.62	0.24	0.20	0.91	0.17	0.93	0.11	2.30	4.00
		(0.01)	(0.02)	(0.06)		(0.21)	(0.03)		(0.03)		(0.14)	(1.42)	(3.16)
GMI RN Br	0.67	0.18	0.15	0.17	0.33	0.19	0.17	0.98	0.15	1.00	0.12	3.10	2.10
		(0.01)	(0.01)	(0.03)		(0.24)	(0.01)		(0.01)		(0.12)	(2.60)	(1.20)
GMI T CF	0.67	0.24	0.20	0.17	0.33	0.23	0.23	0.96	0.20	0.98	0.13	4.56	2.33
		(0.01)	(0.02)	(0.03)		(0.25)	(0.02)		(0.02)		(0.10)	(4.19)	(1.22)
GMI Trix	0.31	0.25	0.21	0.17	0.69	0.19	0.22	0.91	0.20	0.94	0.11	1.90	3.70
		(0.01)	(0.01)	(0.02)		(0.19)	(0.03)		(0.03)		(0.16)	(2.56)	(2.00)
GMI Whe	0.16	0.18	0.14	0.19	0.84	0.27	0.16	0.93	0.14	0.95	0.13	1.00	5.44
		(0.01)	(0.00)	(0.02)		(0.25)	(0.02)		(0.02)		(0.13)	(0.00)	(7.16)
KEL CF	0.09	0.12	0.10	0.17	0.91	0.32	0.11	0.96	0.10	0.96	0.07	0.83	8.00
		(0.01)	(0.01)	(0.02)		(0.25)	(0.02)		(0.02)		(0.23)	(0.41)	(7.27)
KEL Co Po	0.45	0.22	0.19	0.17	0.55	0.29	0.19	0.88	0.17	0.90	0.08	2.33	2.83
		(0.01)	(0.01)	(0.04)		(0.25)	(0.03)		(0.02)		(0.17)	(1.30)	(1.85)
KEL Cr	0.38	0.22	0.18	0.16	0.62	0.24	0.21	0.96	0.17	0.94	0.13	1.92	3.33
		(0.01)	(0.01)	(0.04)		(0.20)	(0.02)		(0.02)		(0.10)	(1.31)	(2.35)
KEL Fr Fl	0.12	0.16	0.14	0.16	0.88	0.35	0.15	0.93	0.13	0.95	0.07	1.17	6.17
		(0.01)	(0.00)	(0.03)		(0.27)	(0.02)		(0.01)		(0.23)	(0.75)	(6.88)
KEL Fr Lo	0.30	0.23	0.19	0.16	0.70	0.31	0.19	0.87	0.17	0.90	0.01	1.80	4.00
		(0.01)	(0.01)	(0.10)		(0.26)	(0.04)		(0.03)		(0.21)	(1.55)	(2.54)
KEL Ho Sm	0.59	0.19	0.16	0.18	0.41	0.24	0.17	0.91	0.15	0.91	0.10	2.64	1.86
		(0.01)	(0.01)	(0.04)		(0.17)	(0.02)		(0.02)		(0.13)	(1.82)	(1.03)
KEL R Br	0.11	0.15	0.13	0.17	0.89	0.25	0.14	0.93	0.12	0.97	0.09	1.00	7.17
		(0.01)	(0.00)	(0.05)		(0.23)	(0.02)		(0.02)		(0.17)	(0.63)	(6.11)
KEL Ri Kr	0.12	0.18	0.15	0.18	0.88	0.30	0.17	0.94	0.15	0.96	0.09	1.00	8.00
		(0.01)	(0.01)	(0.01)		(0.27)	(0.02)		(0.02)		(0.16)	(0.58)	(11.97)
KEL Sp K	0.44	0.22	0.19	0.17	0.56	0.18	0.21	0.94	0.18	0.94	0.13	1.80	2.40
		(0.01)	(0.00)	(0.03)		(0.14)	(0.03)		(0.02)		(0.11)	(0.86)	(1.59)
NAB Shr Whe	0.48	0.18	0.15	0.18	0.52	0.29	0.17	0.98	0.15	1.00	0.15	2.64	3.00
		(0.01)	(0.01)	(0.05)		(0.26)	(0.01)		(0.01)		(0.07)	(1.69)	(2.00)
POS G Nuts	0.47	0.13	0.11	0.19	0.53	0.25	0.13	0.95	0.10	0.95	0.13	1.65	2.00
		(0.01)	(0.01)	(0.02)		(0.16)	(0.01)		(0.01)		(0.08)	(0.70)	(1.22)
POS Ho Co	0.45	0.22	0.18	0.18	0.55	0.26	0.19	0.90	0.17	0.91	0.07	2.08	2.69
		(0.02)	(0.01)	(0.03)		(0.21)	(0.04)		(0.03)		(0.21)	(1.26)	(2.56)
POS R Br	0.17	0.16	0.13	0.16	0.83	0.22	0.14	0.91	0.12	0.91	0.12	1.14	4.57
		(0.01)	(0.01)	(0.05)		(0.19)	(0.02)		(0.02)		(0.13)	(0.69)	(2.57)
QO 100Nat	0.41	0.14	0.12	0.17	0.59	0.16	0.12	0.91	0.10	0.90	0.12	3.57	4.86
		(0.01)	(0.01)	(0.03)		(0.23)	(0.02)		(0.02)		(0.18)	(2.44)	(5.34)
QO Ca Cr	0.17	0.19	0.15	0.23	0.83	0.38	0.16	0.84	0.13	0.89	-0.01	1.50	5.67
		(0.01)	(0.01)	(0.04)		(0.29)	(0.03)		(0.03)		(0.28)	(0.84)	(4.80)
QO Life	0.55	0.15	0.12	0.18	0.45	0.27	0.14	0.92	0.11	0.92	0.12	2.62	2.23
		(0.01)	(0.01)	(0.04)		(0.25)	(0.01)		(0.02)		(0.11)	(1.61)	(1.59)
RAL Chex	0.45	0.23	0.18	0.22	0.55	0.31	0.21	0.92	0.17	0.90	0.12	2.25	2.67
		(0.01)	(0.01)	(0.04)		(0.23)	(0.03)		(0.03)		(0.17)	(1.06)	(1.61)
RAL Ri Chex	0.47	0.23	0.19	0.20	0.53	0.31	0.21	0.93	0.17	0.90	0.13	2.23	2.62
		(0.01)	(0.01)	(0.03)		(0.25)	(0.03)		(0.03)		(0.17)	(1.01)	(1.39)
RAL Wh Chex	0.41	0.17	0.14	0.22	0.59	0.30	0.16	0.93	0.12	0.90	0.16	1.92	2.92
		(0.01)	(0.01)	(0.04)		(0.26)	(0.02)		(0.02)		(0.17)	(0.86)	(1.55)
Average	0.36	0.19	0.16	0.18	0.64	0.27	0.17	0.92	0.15	0.93	0.10	1.98	3.88
		(0.01)	(0.01)	(0.04)		(0.23)	(0.02)		(0.02)		(0.15)	(1.26)	(3.27)

Notes: Comparison of descriptive statistics of chain-level price and promotion variables across promotion and no-promotion periods months. Standard deviations in parentheses where meaningful. Columns 1 to 4 contain statistics for the no-promotion periods. Columns 5 - 13 contain statistics for any-promotion periods. Share *is the fraction of store-month combinations that are / are not on promotion.* Avg p^r *is the average retail price.* Avg p^w *is the average wholesale price.* Avg rm *is the average retail margin.* Pr Int *indicates how many store-week combinations of a month-chain combination participate in the promotion.* p^r disc. *is the average retail price discount observed during a promotion.* p^w disc. *is the average wholesale price discount observed during a promotion.* Avg spell *denotes the average time span (in months) between two promotional activities.* Avg dur *indicates how long (in months) a promotional activity lasts on average.*

A.4 Computation of Base Wholesale Prices and Trade Spend Payments

In this appendix, we provide the details on how we compute our measure of base wholesale prices from the observed wholesale prices, which include trade spend payments during a promotion period. Formally, let p_{jt}^{wb} be the base wholesale price of brand j in market t, and let p_{jt}^{w} denote the net (or total) wholesale price, which includes a trade spend discount td_{jt}, of brand j in market t.

Note that for brand-market combinations in which there is no promotion, the trade spend payment is by construction zero and we assume that we observe base wholesale prices in the data, so that $p_{jt}^{wb} = p_{jt}^{w}$. During a promotion period, we assume that the manufacturer pays a constant trade spend payment td_{jt} per unit sold, so that $p_{jt}^{w} = p_{jt}^{wb} - td_{jt}$.

For our main specification, we compute the base wholesale prices from the wholesale price data on the UPC-week level. We assume that each brand sets a chain-wide base wholesale price for each month, and that variation in the wholesale price during a promotion-week is due to a trade spend payment. Moreover, we assume that the base wholesale price during a promotion week does not differ from the base wholesale prices in the surrounding weeks.

Therefore, we impute the base wholesale price for a UPC in a given store during a promotion week as the average between the observed wholesale price during the weeks immediately before and immediately after the promotion. After having computed the base wholesale prices on the UPC-store-week level, we aggregate them to the brand-month level. In order to obtain the trade spend payments we subtract the observed net wholesale prices from the imputed base wholesale prices on the brand-month level.

Note that we do not explicitly model how manufacturers set trade spend payments and budgets. In line with industry evidence (Anderson and Fox, 2019), we assume that the trade spend payments come from a separate fixed trade spend budget, which the manufacturers consider as separate from the brand's base account. We discuss the implications for the interpretation of our model and our estimation results in Sections 3 and 5, respectively.

We experimented extensively with alternative ways of computing the trade spend variable and base wholesale prices; for example, by imputing base wholesale prices on the brand-month level instead of the week-UPC level or by computing base wholesale prices using the predictions from flexible regression models. Overall, these approaches led to very similar trade spend values and did not affect the results of our structural estimation significantly.

A.5 Model-based Intuition for Reduced Form Regressions

In this appendix, we consider a stylized version of our supply model, which we develop in Section **??**, with two single-product firms. Firms set base wholesale prices (p_j^{wb}) taking a value of the trade spend (td_j), retail markup (mu_j^r), and marginal costs of production (mc_j) as given. Potential price coordination among firms is captured by the profit internalization parameter λ. Equilibrium base wholesale prices are characterized by the following first order conditions.

$$p_1^{wb} = mc_1 - \left(\frac{\partial s_1}{\partial p_1^r}\right)^{-1} \cdot s_1 - \lambda \frac{\partial s_2}{\partial p_1^r} \left(\frac{\partial s_1}{\partial p_1^r}\right)^{-1} \left(p_2^{wb} - td_2 - mc_2\right)$$
$$p_2^{wb} = mc_2 - \left(\frac{\partial s_2}{\partial p_2^r}\right)^{-1} \cdot s_2 - \lambda \frac{\partial s_1}{\partial p_2^r} \left(\frac{\partial s_2}{\partial p_2^r}\right)^{-1} \left(p_1^{wb} - td_1 - mc_1\right),$$

where $p_j^r = mu_j^r + p_j^{wb} - td_j$. From the above equations, it is clear that the value of λ has an effect on equilibrium wholesale prices. Next, we analyze how a firm's base wholesale price changes when the rival firm goes on promotion and how this reaction depends on the level of price coordination λ. For illustrational purposes, we consider two cases assuming that firm 2 goes on promotion:

1. Competition ($\lambda = 0$): A rival promotion has several effects. First, it decreases s_1, both via sale-sign demand effects which make the product on promotion more attractive and through a reduction in p_2^r (as td_2 increases). A promotion can either increase or decrease $\frac{\partial s_1}{\partial p_1^r}$ so that the overall effect of the rival promotion is theoretically ambiguous. However, we judge it likely –and consistent with our data– that a promotion of product 2 will make demand for product 1 more elastic, so that the the wholesale price for product 1 should decrease under competition. More importantly, the own price derivative is not crucial for comparing the differential effects under competition and collusion, because this term enters both terms equally.

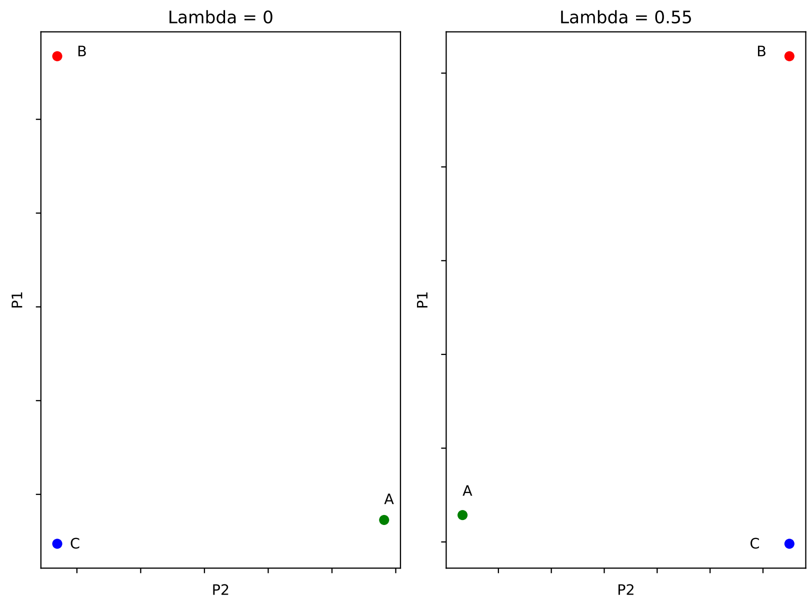
2. Collusion ($\lambda > 0$): The same effects as in the case of $\lambda = 0$ remain. However, there is an additional term present: $\left(p_2^{wb} - td_2 - mc_2\right) \frac{\partial s_2}{\partial p_1^r}$, which is also affected by the rival promotion. By construction, a promotion of firm 2 decreases $(p_2^{wb} - td_2 - mc_2)$, because td_2 increases. However, it is possible that $\frac{\partial s_2}{\partial p_1^r}$ increases relatively more during a promotion of firm 2. This can result in firm 1 increasing its price. Therefore, it is more likely that a promotion of firm 2 leads to an increase in the wholesale price of firm 1 if the firms engage in coordinated pricing.

A natural caveat with the above reasoning is that every potential change in prices can be rationalized by flexible changes in the own- and cross-derivatives of the demand function. This motivates us to estimate a structural model to disentangle the differential effects of changes in demand and strategic pricing incentives.

To conclude this section, we provide a numerical illustration of the above reasoning for a parametrized logit demand model, which can generate the patterns discussed above. We continue to focus on a setting with two single-product firms. Let the utility from good j be $u_j = \bar{V}_j - \alpha p_j^r + \epsilon_j$, $p_j^r = p_j^{wb} + mu_j^r - td_j$, $mc_1 = mc_2 = 1$, $mu_1^r = mu_2^r = 0$, $\bar{V}_j = V_j$ when there is no promotion, $td_j = 0$ and $\bar{V}_j = 1.08 \cdot V_j$. When j is on promotion, $td_j > 0$ and

$V_1 = V_2 = 2.6$, $\alpha = 0.50$ and $\lambda = 0.55$. Furthermore, we assume that during a promotion of firm 1 trade spend is equal to $td_1 = V_1 \cdot 0.10$. These parameters imply that the effect of product j's own promotional activity increases its perceived quality from V_j to $1.08 \cdot V_j$, while decreasing the retail prices by td_j. The effects of firm 1's promotion on base and net wholesale prices are illustrated in Figure 12. The left and right panel indicate the effects under $\lambda = 0$ and $\lambda = 0.55$, respectively. During a promotion of firm 1 the base wholesale price equilibrium moves from point A to point B. Under competition, we observe that firm 2's base wholesale price decreases, while it increases under coordinated pricing, i.e., for $\lambda > 0$. Finally, point C visualizes the net wholesale prices that the retailer will face after the trade spend payment is taken into account. Consistent with the patterns in our data, firm 1's net wholesale price is smaller than the initial base wholesale price in point A.

Figure 12: Illustration of effects of rival promotion on wholesale prices



Notes: The figure illustrates equilibrium prices under Nash pricing with no rival profit internalization and partially coordinated pricing with an internalization parameter of 0.55 in the left and right panel, respectively. Prices of firm 2 and firm 1 are on the x-axis and the y-axis, respectively. Points A (green) indicate the equilibrium base wholesale prices when no product is on promotion. Points B (red) indicate equilibrium base wholesale prices when firm 1 is on promotion. Points C (blue) indicate equilibrium net wholesale prices when firm 1 is on promotion.

A.6 Reduced Form Results

Testing for structural breaks with unknown dates. Our F-tests discussed in Section 2.3 have the disadvantage that one needs prior knowledge about the potential break points in conduct. To address this problem and to gain additional insights about when exactly the breaks in conduct occur, we conduct a series of tests proposed by Bai and Perron (1998,

Table 7: Reduced form analysis: Wholesale prices

	(1) Net - Only SG	(2) Net - SG BC	(3) Base - Only SG	(4) Base - SG BC
Post-merger KEL	0.0465***	0.0319*	0.0678***	0.0569***
	(0.0132)	(0.0142)	(0.0076)	(0.0079)
Post-merger RAL	0.0473	0.0371	0.0403***	0.0283*
	(0.0261)	(0.0258)	(0.0109)	(0.0114)
Post-merger QUA	0.0196	0.0197	0.0458***	0.0415***
	(0.0200)	(0.0201)	(0.0093)	(0.0095)
Post-merger GMI	-0.0476***	-0.0678***	-0.0234**	-0.0336***
	(0.0124)	(0.0152)	(0.0075)	(0.0080)
Post-merger POSTNAB	0.0149	-0.0002	0.0547***	0.0458***
	(0.0160)	(0.0186)	(0.0092)	(0.0105)
Price war period	-0.0928***	-0.1003***	-0.0696***	-0.0709***
	(0.0189)	(0.0196)	(0.0141)	(0.0142)
Promo (SG, own), pre-merger	-0.0022***	-0.0022**	-0.0004	-0.0004
	(0.0007)	(0.0007)	(0.0002)	(0.0002)
Promo (SG, own), post-merger	-0.0019***	-0.0019***	0.0001	0.0001*
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Promo (SG, own), price war	-0.0012***	-0.0012***	0.0000	0.0000
	(0.0003)	(0.0003)	(0.0001)	(0.0001)
Promo (same firm), pre-merger	0.0000	0.0000	0.0002**	0.0001**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Promo (same firm), post-merger	0.0001**	0.0001**	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Promo (same firm), price war	0.0000	0.0001	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Promo (SG, rivals), pre-merger	8.7240***	9.1778**	1.2889	1.1669
	(2.5680)	(2.8057)	(1.7064)	(1.7078)
Promo (SG, rivals), post-merger	2.4414**	2.2676**	0.1696	0.3296
	(0.8754)	(0.8784)	(0.5366)	(0.5372)
Promo (SG, rivals), price war	-2.7381**	-2.3925*	-3.7304***	-3.8676***
	(1.0324)	(1.1368)	(0.8482)	(0.9311)
Promo (BC, own), pre-merger		-0.0002***		-0.0001*
		(0.0001)		(0.0000)
Promo (BC, own), post-merger		-0.0003*		0.0001
		(0.0002)		(0.0001)
Promo (BC, own), price war		-0.0004		0.0003*
		(0.0003)		(0.0001)
Promo (BC, rivals), pre-merger		-0.4443		0.5734
		(0.8919)		(0.3664)
Promo (BC, rivals), post-merger		2.5267**		1.4505**
		(0.8857)		(0.4546)
Promo (BC, rivals), price war		-0.7469		0.9354
		(1.9204)		(1.1007)
Observations	1728	1728	1728	1728
R-square	0.75	0.76	0.90	0.91

Notes: SG indicates general promotions, BC indicates bonus buy/coupon promotions. Standard errors in parentheses. Data is aggregated at the brand-month level. All regressions include brand fixed effects.

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 8: Structural break tests - base wholesale prices

	$\beta(SE)$			$H_0 : pre = post = pw$	
	Pre-merger	Post-merger	Price war	F-statistic (df1,df2)	p-value
Promo (SG, rivals), model 3	1.2889	0.1696	-3.7304	8.6456 (2,1683)	0.0002
	(1.7064)	(0.5366)	(0.8482)		
Promo (SG, rivals), model 4	1.1669	0.3296	-3.8676	8.1656 (2,1677)	0.0003
	(1.7078)	(0.5372)	(0.9311)		
Promo (BC, rivals), model 4	0.5734	1.4505	0.9354	1.4511 (2,1677)	0.2346
	(0.3664)	(0.4546)	(1.1007)		

Notes: The table summarizes the results from testing the equality of the rival promotion pass-through coefficients on base wholesale prices over time using F-tests. The coefficients in columns 1 to 3 come from the regressions in Table 7, columns 3 and 4 (standard errors in parentheses). SG *denotes general promotions.* BC *denotes bonus buy and coupon promotions.*

2003). These tests do not require us to pre-specify the structural break dates. Instead, they will provide us with the dates when the coefficient of rivals' promotions on own wholesale prices changes.

We specify a regression model analogous to the wholesale price regressions in Table 7. However, since the test by Bai and Perron (1998, 2003) is designed for a single time series, we run the tests separately for each brand. Moreover, we need to specify how many structural breaks should be tested for. We present the results for testing for three structural break points in Figure 13.[40] The x-axis displays the months of our sample. The y-axis denotes the number of brands for which we detect a structural break in each month.

For the price war period the test results align closely with the price war start date mentioned in industry sources. The business press typically declares the start of the nation-wide price war to be April 1996. Our structural break tests suggest February 1996 as the start date.

For the break following the merger, the results are less aligned with the actual merger event. In our tests we find only very weak evidence for a structural break around the time of the merger (January 1993). Only three brands exhibit break points in the first quarter of 1993. These three isolated breaks are likely not indicative of an industry-wide change in conduct. The first significant cluster of brand-specific break points starts in October 1993, which is nine months after the merger is consummated. Specifically, between October 1993 and March 1994, we find a structural break for the majority of brands. We find it plausible that the effects of the merger on industry-wide conduct may need time to materialize.

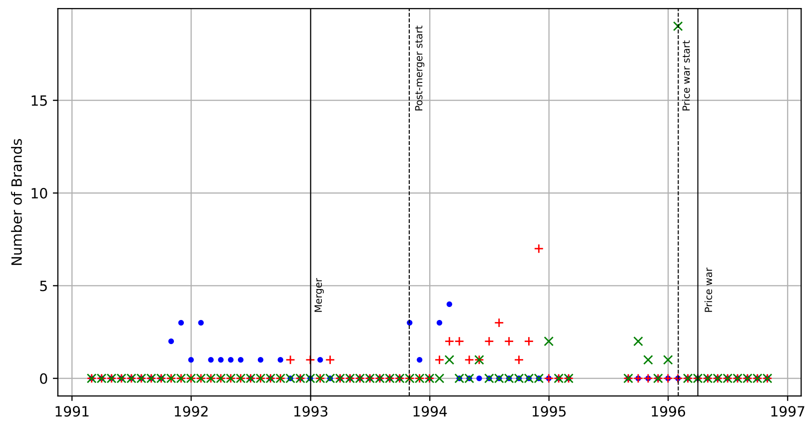
Therefore, and because industry evidence and our structural break tests do not fully align with each other, one could rationalize a range of different start dates for our post-

[40]The results when testing for a different number of breaks are similar. Figure 13 displays the results when using the total wholesale price as the dependent variable. The results are similar when we use the base wholesale price instead.

merger period. For our main analysis we prefer to "let the data speak" about when industry conduct changes instead of exclusively relying on industry evidence about the event dates.

For the structural estimation, we set the start date of the post-merger period and the price war period to October 1993 and February 1996, respectively. We experimented with moving the start date of the post-merger period between January 1993 and October 1993. Qualitatively, this did not affect our conduct estimates.

Figure 13: Unknown structural break dates: Summary of Bai and Perron (1998, 2003)-test results



Notes: The figure summarizes for how many brands the test by Bai and Perron (1998, 2003) detects a structural break in the effect of rivals' promotions on net wholesale prices in any given month. The tests are run separately for the time series of each brand at the chain level. Blue dots, red crosses, and green crosses denote the first, second, and third structural break for each brand, respectively. The solid vertical lines indicate the dates for the merger and the price war based on institutional evidence. The dashed vertical lines indicate how we define the start of the post-merger and the price war period based on the results of the structural break tests displayed in this figure.

A.7 Reduced Form Evidence for Validity of Modeling Assumptions

In this appendix, we present reduced form evidence for our modeling assumptions.

Timing assumptions. The essential timing assumption in our model is that promotions in period t are set at the latest in period $t-1$, while (base) wholesale prices for period t are flexible until period t.[41]

Formally, we require that base wholesale prices $p_{jt}^{wb} = f(\mathcal{I}_t)$, where $\mathcal{I}_t$ denotes all information available in period t, i.e., all contemporaneous demand and cost shifters. In contrast, the number of promotions (and trade spend payments) in period t, $Promo_{jt} = f(\mathcal{I}_{t-1})$, where $\mathcal{I}_{t-1}$ contains only information available up to $t-1$. Shocks that cause $\mathcal{I}_{t-1}$ to be different

[41] 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 to charge different wholesale prices. 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).

from $\mathcal{I}_t$ provide the variation in the data necessary to make our instruments for identifying industry conduct work.

Because there is little hard evidence on the contracts between manufacturers and retailers, we provide support from the data for these timing assumptions. A testable implication of this timing assumption is that wholesale prices in period t should be a function of all information available in period t, and that the number of promotions in period t, is only a function of information available up to $t-1$. This implies that, if our timing assumptions are satisfied, one would expect that wholesale base prices react to demand and cost shocks immediately. In addition, if promotions are predetermined, they should not react to contemporaneous shocks. Instead, one would expect that promotions adjust with a lag.

We investigate these hypotheses in a series of reduced form regressions. Specifically, we regress both base and net wholesale prices on several cost shifters that should affect the pricing decisions of manufacturers, in particular, input prices for sugar, rice, and corn weighted by the respective content in a given product, the gasoline price interacted with a production facility's distance to the Chicago area, and the electricity price in the Chicago area (3-months-moving average). Throughout, we control for brand and time fixed effects, and we run the regressions at the chain level. Afterwards, we conduct analogous regressions with the number of contemporaneous promotions on the brand-month level as the dependent variable. Finally, we repeat this regression, replacing the contemporaneous promotion intensity measure with the number of promotions in future periods (one to six months into the future). The associated results are summarized in Table 9.

Column (1) summarizes the results from regressing logged net wholesale prices on various cost shifters to illustrate that wholesale prices react immediately to contemporaneous cost shocks. Column (2) repeats the regression with base wholesale prices as the dependent variable and we find almost identical significance patterns as in Column (1).

Column (3) reveals that promotions in the current period are not affected by contemporaneous cost shocks, which provides evidence that promotional activities are not adjusted immediately. Columns (4) to (9) show that future promotions are affected by cost shocks today, however. Therefore, while promotions are clearly endogenous, Table 9 provides evidence that they are plausibly sequentially exogenous to future innovations in manufacturers' supply shocks.

These patterns are consistent with firms being able to react to different types of shocks in different ways. For example, grain prices today affect the promotion intensity one and two months in the future, the gasoline price interacted with factory distances affects promotion intensity two to six months from today, and electricity prices have an effect on the number of promotions four to six months into the future. Overall, we interpret these regressions as support for the validity of our timing assumptions. Note that the staggered effect of different cost shifters does not invalidate our instruments. Instead, all that matters is that it takes time for promotions to be adjusted so that they are plausibly uncorrelated with the innovations in the structural cost shock that we use in our moment conditions.

Non-monetary effects of promotions on consumer demand. To investigate the presence of non-monetary sale-sign effects of promotions on consumer demand, we regress the logged quantities sold on brand fixed effects, a linear-quadratic time trend, and statistics of own brand and rival firms' promotional activities. Table 10 summarizes the associated results. The main purpose of these quantity regressions is to illustrate that –even after controlling for the retail prices paid by consumers– the pattern of product-specific promotions in a market has a significant effect on consumer choices. In particular, both general promotions and bonus buy promotions for a brand increase consumer demand significantly. We interpret this as evidence for the presence of considerable non-price effects (advertising intensity, brochures, shelf space allocations, or signs for products on sale) of promotions, which shift/rotate consumer demand.

Table 9: Reduced form evidence supporting our timing assumptions

	(1) p^w	(2) p^{wb}	(3) Promo in t	(4) Promo in $t+1$	(5) Promo in $t+2$	(6) Promo in $t+3$	(7) Promo in $t+4$	(8) Promo in $t+5$	(9) Promo in $t+6$
Sugar price	0.00171***	0.000667***	-0.00162	-0.00486**	-0.00563***	-0.00462***	-0.00343*	-0.00181	-0.00446**
	(3.85)	(2.58)	(-0.90)	(-2.07)	(-3.17)	(-3.29)	(-1.99)	(-0.99)	(-2.21)
Corn price	0.00000649	0.00000915***	0.0000299	0.0000842**	0.0000667*	0.0000340	0.0000293	-0.00000273	0.0000622
	(1.46)	(3.63)	(0.85)	(2.35)	(1.79)	(1.16)	(0.79)	(-0.09)	(1.22)
Rice price	0.00000653**	0.00000445***	0.00000715	0.0000340***	0.0000247**	0.0000279	0.0000402***	0.0000144	0.00000958
	(2.54)	(3.59)	(0.61)	(9.00)	(2.49)	(1.18)	(6.32)	(1.03)	(0.80)
Gasoline price	0.000938**	0.00149***	-0.00224	-0.0000700	0.00335**	0.00392**	-0.000525	-0.00691***	-0.00727***
	(2.37)	(6.46)	(-1.56)	(-0.05)	(2.28)	(2.12)	(-0.29)	(-4.31)	(-5.47)
Electricity price			0.00137	-0.00119	0.000293	0.00259	0.00592*	0.00722**	0.00782***
			(0.91)	(-0.68)	(0.16)	(1.11)	(1.79)	(2.69)	(4.03)
Observations	1728	1728	1728	1701	1674	1647	1620	1593	1566
R-square	0.73	0.91	0.10	0.09	0.09	0.09	0.09	0.09	0.10

t statistics in parentheses
Notes: The table summarizes the results from regressing several measure of wholesale prices and promotions on various cost shifters and fixed effects.
All specifications are estimated at the chain level. p^w and p^{wb} denote total and base wholesale prices, respectively.
Promo in t denotes current promotion intensity. *Promo in* $t+j$ denotes promotion intensity j months into the future.
Sugar price, *Corn price*, and *Rice price* denote contemporaneous sugar, corn, and rice prices weighted with a cereals content of each grain, respectively.
Gasoline price denote the contemporaneous gasoline price interacted with the distance between the manufacturer's production plant and the Chicago area.
Electricity price denote the contemporaneous electricity price in the Chicago area.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Reduced form analysis: Quantities sold

	(1) Baseline	(2) Baseline w/ BC	(3) Lagged promos
Log Retail Price	-1.7703***	-1.7727***	-1.7959***
	(0.1032)	(0.1444)	(0.1054)
Promo (SG, own), pre-merger	0.0029**	0.0028*	0.0027**
	(0.0009)	(0.0012)	(0.0010)
Promo (SG, own), post-merger	0.0028***	0.0027***	0.0026***
	(0.0005)	(0.0006)	(0.0005)
Promo (SG, own), price war	0.0018*	0.0018*	0.0018*
	(0.0008)	(0.0007)	(0.0007)
Promo (BC, own), pre-merger	0.0014***	0.0014***	0.0015***
	(0.0003)	(0.0003)	(0.0003)
Promo (BC, own), post-merger	0.0021***	0.0022***	0.0021***
	(0.0004)	(0.0003)	(0.0004)
Promo (BC, own), price war	-0.0000	-0.0000	0.0000
	(0.0005)	(0.0006)	(0.0005)
Promo (same firm), pre-merger	0.0001	0.0002	0.0000
	(0.0004)	(0.0004)	(0.0005)
Promo (same firm), post-merger	-0.0003*	-0.0003*	-0.0002*
	(0.0001)	(0.0001)	(0.0001)
Promo (same firm), price war	-0.0001	-0.0001	-0.0001
	(0.0001)	(0.0001)	(0.0001)
Promo (SG, rivals), pre-merger	-31.5933**	-30.5942*	-29.1837**
	(10.0523)	(11.4010)	(10.3511)
Promo (SG, rivals), post-merger	-6.9483**	-6.5275**	-6.7376*
	(2.2062)	(1.8434)	(2.5658)
Promo (SG, rivals), price war	-4.1028*	-6.1626	-6.1113**
	(1.8599)	(3.5360)	(2.1132)
Promo (BC, rivals), pre-merger		-2.7141	-2.3016
		(1.6815)	(1.8548)
Promo (BC, rivals), post-merger		-4.3486*	-4.3669
		(2.0383)	(2.8780)
Promo (BC, rivals), price war		7.3440*	7.3302
		(3.1963)	(4.7883)
Promo (SG, own), lag 1			-0.1107
			(0.0640)
Promo (BC, own), lag 1			-0.0863
			(0.0574)
Promo (SG, own), lag 2			-0.0101
			(0.0512)
Promo (BC, own), lag 2			-0.0628
			(0.0523)
Promo (SG, own), lag 3			-0.0265
			(0.0709)
Promo (BC, own), lag 3			-0.0285
			(0.0503)
Promo (SG, own), lag 4			0.0199
			(0.0548)
Promo (BC, own), lag 4			0.0500
			(0.0491)
Observations	1728	1728	1620
R-square	0.89	0.90	0.90

Notes: All estimations include brand fixed effects and a linear-quadratic time trend.
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. SG and BC denote general promotions and bonus buy/coupon promotions, respectively. Column (2) adds rival firms' BC promotions as regressors. Column (3) adds lags of a brand's promotions as regressors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B Discussion of Model Setup and Assumptions

In this appendix, we provide additional examples and discuss several assumptions of our model setup.

B.1 Example of Conduct Matrix with Three Firms

In order to illustrate the role of the profit internalization matrix in our model, assume that there are three single-product firms. If each firm equally internalizes its pricing externalities on every rival, the pre-merger 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 firms 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 engage in Nash pricing, we expect λ^{PW} to be zero.

B.2 Static Demand

We abstract from dynamic consumer behavior for several reasons. In principle, our supply model and our 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 computationally intensive. A dynamic model would therefore have to heavily compromise in this dimension. In our application, we judge accounting for consumer heterogeneity to be more important for estimating consumers' substitution patterns than dynamic storage behavior. We use data at the month level for which dynamic behavior is arguably much less relevant than for weekly data. To further support our consumer myopia assumption, we present evidence that storage behavior does not play a significant role in our sample. Specifically, we regress the quantities sold of a given brand in a given month on a brand's aggregate promotion intensity in previous months. The associated results are displayed in column (3) of Table 10 in Appendix A.7. While current brand-specific promotions have a large effect on the quantities sold, lagged promotional activities for the same brand do not significantly affect demand in the current period.

B.3 Discussion of Potential Synergies

Note that we do not use the ownership change as an instrument, so that the occurrence of synergies does in principle not pose a problem for our empirical analysis. Our identification strategy would lead to biased estimates only if our instruments are correlated with the innovations in the structural cost shock ν_S. This would be the case if there are synergies that are absorbed into the innovations of the unobservable cost shock and these synergy effects 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 cost shock innovations, and rival firms anticipate these future shocks and therefore systematically change their promotional activities. Given that we include a battery of fixed effects in the marginal cost function, in particular, brand-half year fixed effects, and construct our moments based only on the innovations instead of the levels of the cost shocks, we argue that our error term ν_{Sjt} contains only shocks that are hard for j's rivals to anticipate when setting their promotions for period t. Furthermore, we are not aware of any industry evidence for these kinds of shifts in manufacturers' strategies after the merger, nor do we find any support for such behavior in our data.

Furthermore, we have not found any industry evidence suggesting that the Post-Nabisco merger caused significant marginal cost synergies. For example, cost synergy considerations have not been of significant importance during the merger case.[42] 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 of such savings is 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

[42] 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.

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. Moreover, the merging firms' products use different production technologies. Post's products primarily require flaking and baking processes, while Nabisco's products mainly rely on shredding.

To address potential remaining concerns about merger-related synergies, we run a robustness check in which we include a post merger-merging firm dummy in the marginal cost function. The results are unaffected and available upon request.

B.4 Consumers' Retailer Choice

We focus on data from a single retailer, i.e., DFF. This allows us to exploit available wholesale price data. The downside of this approach is that we cannot analyze substitution to different supermarket chains. 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.

B.5 Market Size Definition and Computation of Market Shares and Price Variables

We define a unit of cereal as a 1 ounce serving of a specific brand. The total overall market size is defined as one serving per capita every second day times the mean store-specific number of total customers that visit a DFF store per month.[43]

For our estimations, we include all package sizes (UPCs) between 10 and 32 ounces for the brands in our sample, and calculate aggregate quantities and the average price per ounce for each brand. If there are multiple UPCs for a brand in a market we aggregate them to the brand level by adding the total amount of ounces sold across all UPCs of that brand for a given market. We compute the brand-level market share as the total quantity sold divided by our market size measure. The remainder, i.e., one minus the sum of the inside good market shares for a given market, yields the market share of the outside good. For the price variable, we proceed analogously. That is, we compute the total expenditure for a given brand and divide this number by the total number of ounces sold. This results in a quantity-weighted average of the UPC-level prices. We prefer a model on the brand-level

[43] We find the empirical results to be robust to using a time-variant market size specification, and to changing the market size by factors $\frac{1}{3}$, $\frac{1}{2}$, 2, and 3, respectively. The demand results for the alternative market specifications are available upon request.

to one on the UPC-level, because the latter would result in a demand system with more products. Many of these products would often have a market share of zero, and it is not fully clear from the DFF data, whether such a product was not available in the market or whether it was available but not sold. Solving this *zero-problem* is not trivial and would substantially complicate how we approach the demand estimation, see, for example, Gandhi *et al.* (2020).

C Weak Identification Tests

In this appendix, we illustrate 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, the instruments can still be weak enough to result in very sensitive estimates and high standard errors.[44]

In order to overcome the first problem, we adapt a testing procedure recently proposed by Gandhi and Houde (2020) 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 (2020). To the best of our knowledge, this procedure has so far not been used to test for weak identification of conduct patterns.

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

$$\begin{aligned}\kappa_{jt}(\theta) &= \kappa_{jt}(\theta_0) + \sum_{k=1}^{K}(\theta_k - \theta_{0k})\frac{\partial \kappa_{jt}(\theta_0)}{\partial \theta_k} + \upsilon_{jt} \qquad (10)\\ &= \kappa_{jt}(\theta_0) + J_{jt}(\theta_0)b + \upsilon_{jt}, \qquad (11)\end{aligned}$$

where J denotes the Jacobian stacking all the partial derivatives with respect to each pa-

[44] A popular rule-of-thumb criterion is that the F-statistic is larger than 10.

rameter 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.[45] 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 be 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 $A(Z)$.

$$\frac{\partial \kappa_{jt}(\hat{\theta})}{\partial \theta_k} = X_{jt}\pi_{1k} + A_j(Z_t)\pi_{2k} + \epsilon_{jtk} \tag{12}$$

   In our demand model, there are K endogenous variables corresponding to the K partial derivatives $\frac{\partial \nu_D}{\partial \theta_k}$ of the innovations in the structural demand shocks with respect to the non-linear preference parameters plus one for the $AR(1)$-coefficient in the ξ-process. In our supply model, the number of nonlinear parameters is equal to the number of profit internalization parameters plus one for the $AR(1)$-coefficient in the ω-process.

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 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 starting point. However, in models with multiple endogenous regressors, conventional F-tests can easily result in falsely rejecting non-identification. Angrist and Pischke (2008) (henceforth, AP) propose a modified F-statistic that corrects for the presence of multiple endogenous regressors by profiling out the

[45] 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 ν_D and ν_S, respectively.

effects of the other $K-1$ endogenous regressors and using only the variation in the projection residual when running the first-stage regression. This test statistic has been further refined by Sanderson and Windmeijer (2016) (henceforth, SW) and we report their version of the F-statistic for testing for weak identification of a single regressor in row *Robust AP-SW-F-statistic* in Tables 11 to 13.

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, which can result 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.[46]

[46] 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., result 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. This is always the case in our application. Therefore, we judge the practical problem of not having the critical values readily available as not crucial.

Weak identification of the demand model. Table 11 summarizes the results of our weak identification tests for the demand model.

Table 11: Weak IV tests: Demand model

	$\frac{\partial \nu_D}{\partial \alpha}$	$\frac{\partial \nu_D}{\partial \Pi_1}$	$\frac{\partial \nu_D}{\partial \Pi_2}$	$\frac{\partial \nu_D}{\partial \Pi_3}$	$\frac{\partial \nu_D}{\partial \Pi_4}$	$\frac{\partial \nu_D}{\partial \rho}$	$\frac{\partial \nu_D}{\partial \iota_D}$
Robust F-statistic	59.02	18.21	37.53	35.05	25.58	71.77	193.75
Robust AP-SW-F-statistic	31.20	15.56	22.75	26.52	22.71	22.26	74.64
KP χ^2-statistic	292.97						
KP χ^2-p-value	0.00						
KP F-statistic	10.63						

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-SW F-test we see a substantial drop in the statistic for most parameters; therefore, controlling for multiple endogenous regressors is important. All of the test statistics remain larger than the critical values. 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 10. Consequently, we can reject both underidentification and weak identification of our demand model.

Weak identification of the supply model. Table 12 and 13 summarize the results from testing for weak identification in our supply models. Table 12 focuses on the small specification with three profit internalization parameters. Table 13 displays the results for the more detailed specification with five internalization parameters.

First, we investigate the F-statistic of classical first stage regressions. We regress the endogenous variables, i.e., the derivatives of the innovations in the structural cost shocks ν_S with respect to the profit internalization parameters, on our excluded instruments which are based on rivals brands' 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 by orders of magnitude. Next, we report F-statistics that take into account the presence of multiple endogenous regressors as initially proposed by Angrist and Pischke (2008) and refined by Sanderson and Windmeijer (2016). While the F-statistics generally become smaller, they still consistently exceed the critical values substantially. We take this as strong evidence that our instruments shift the endogenous regressors sufficiently and therefore constitute strong instruments.

Table 12: Weak IV tests: Supply model (Small)

	$\frac{\partial \nu_S}{\partial \lambda_1}$	$\frac{\partial \nu_S}{\partial \lambda_2}$	$\frac{\partial \nu_S}{\partial \lambda_3}$	$\frac{\partial \nu_S}{\partial \iota_S}$
Robust F-statistic	2612.06	6475.77	33659.89	13647.03
Robust AP-SW-F-statistic	2426.63	6110.19	33113.59	7410.76
KP χ^2-statistic	1373.15			
KP χ^2-p-value	0.00			
KP F-statistic	1987.61			

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 13: Weak IV tests: Supply model (Large)

	$\frac{\partial \nu_S}{\partial \lambda_1}$	$\frac{\partial \nu_S}{\partial \lambda_2}$	$\frac{\partial \nu_S}{\partial \lambda_3}$	$\frac{\partial \nu_S}{\partial \lambda_4}$	$\frac{\partial \nu_S}{\partial \lambda_5}$	$\frac{\partial \nu_S}{\partial \iota_S}$
Robust F-statistic	1988.10	5149.33	4406.99	4406.99	14760.91	4475.72
Robust AP-SW-F-statistic	1569.03	4912.15	2867.24	3303.71	4799.31	2853.11
KP χ^2-statistic	1309.38					
KP χ^2-p-value	0.00					
KP F-statistic	1164.05					

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.

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 1373 and 1309, respectively, resulting in p-values of less than 0.0001 for both models.

We also compute the KP-F-statistic, which is a heteroskedasticity-robust version of the Cragg-Donald Wald-statistic for weak identification. For the small model with three profit internalization parameters the test statistic is 1988. This is significantly 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 five profit internalization parameters, the KP F-statistic is smaller (1164). Nevertheless, the test statistic exceeds all conventional critical value by orders of magnitude. Therefore, we conclude that even our larger supply model does not suffer from weak identification problems.

D Details on Estimation Algorithm

In this appendix, we provide additional details on our estimation algorithm, which generalizes the typical BLP approach to using a flexible set of dynamic panel moments similarly to Lee (2013) and Schiraldi (2011). The key generalization on the supply side is that we use a flexible conduct matrix instead of a binary ownership matrix when backing out marginal costs. As most other papers, we estimate the demand side and the supply side in two steps.[47]

Demand estimation. For a given guess of the nonlinear demand parameters (four demographic interaction parameters, one nesting parameter, and one $AR(1)$-parameter) 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.

When computing the model's market share predictions, we simulate 500 consumers per market using Halton draws. Train (2000) demonstrates that Halton draws can be much more efficient in simulating the integral over the consumer population than Monte Carlo sampling.

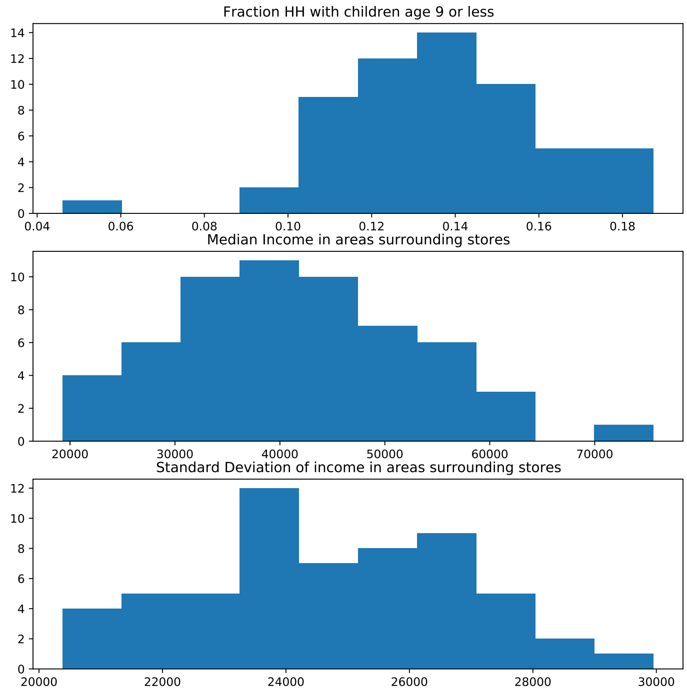
DFF provides data for the demographic distributions of the areas around each store. The demographic data are constant over time but differ across stores. Specifically, DFF reports the median income as well as the standard deviation of the income distribution of the population around each store.[48]

We assume that income follows a log-normal distribution with mean and variance parameters that are consistent with the reported median and standard deviation of each store in

[47] 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. Because of our reasonably large sample, we judge the efficiency gains to be less important than the robustness associated with a two-step estimation.

[48] Figure 14 summarizes the distributions of our demographic variables.

Figure 14: Histogram of demographics in the DFF data



Notes: The figure displays histograms of the most important demographic variables in the DFF data (share of households with children, median income, standard deviation of income in the top, middle, and bottom panel, respectively).

the DFF data. We draw the income of our simulated consumers from this distribution by transforming the uniformly distributed Halton draws. Next, for each consumer, we compute the difference between the log of her income draw and the log of the mean income across all stores. This individual income deviation (measured in logged US-\$) is interacted with the product characteristics and the income-interaction parameters.[49]

In line with the recommendations of Dubé *et al.* (2012) and Conlon and Gortmaker (2020), 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^{-12}.

As first proposed by Nevo (2001), we profile out all linear parameters contained in δ so that we have to optimize numerically only over the nonlinear coefficients. This procedure needs to be slightly adjusted, because our moment conditions are based on the innovations in the ξ-process and not its levels. In a standard BLP-model, i.e., one with moment conditions based on ξ instead of its innovations, the estimates of the linear parameters as a function of

[49]We experimented with different ways of simulating income. Overall, we found that the results are not affected much by how exactly we simulate and transform the income draws.

the nonlinear parameters θ are given by

$$\hat{\beta} = \left(X'ZWZ'X\right)^{-1}\left(X'ZWZ\delta(\theta)\right),$$

where X, Z, and W denote the matrices of product characteristics, instruments, and weighthing matrix, respectively and the mean utilities have the familiar structure $\delta_{jt} = X_{jt}\beta + \xi_{jt}$. For our model with dynamic panel moments, the formula changes to

$$\hat{\beta} = \left(\tilde{X}'ZWZ'\tilde{X}\right)^{-1}\left(\tilde{X}'ZWZ\tilde{\delta}(\theta)\right),$$

where $\tilde{X}$ and $\tilde{\delta}$ are the pseudo-differenced versions of X and δ, i.e., the stacked versions of $\tilde{\delta}_{jt} = \delta_{jt} - \iota_D\delta_{jt-1}$ and $\tilde{X}_{jt} = X_{jt} - \iota_D X_{jt-1}$. Therefore, the linear parameters are not only a function of the nesting parameter and the demographic interaction parameters but also the $AR(1)$-parameter ι_D.

In our application, product characteristics do not change across markets; therefore, we cannot include the time-invariant product characteristics, such as sugar and fiber content, in the profiling matrix directly. We follow Nevo (2001) and back out the mean preferences for each time-invariant product characteristic in a separate post-estimation step, which decomposes the estimated brand fixed into the effects of the various product characteristics.

After having computed the levels of the structural demand errors ξ, we compute the innovations ν_D of the ξ-process as a function of the $AR(1)$-parameter. Finally, we interact ν_D with the instruments Z_D discussed in Section 4.1.

In order to improve the efficiency of our estimation, we build on insights from the dynamic panel literature, see, for example, Arellano and Bover (1995) and Blundell and Bond (1998). In particular, we include our excluded instruments (predicted wholesale prices and rival-promotion-based differentiation instruments) not only in levels but also in first differences. Moreover, we interact the predicted own-promotion variable and the rival-promotion-based instruments in levels and first differences with dummies for our three time periods (pre-merger, post-merger, price war). Finally, to identify the $AR(1)$-parameter of the ξ-process, we include one- and two-period lags of the implied mean utilities δ in Z_D similarly to Lee (2013).

For minimizing the GMM objective function, we use a Nelder-Mead line search algorithm.[50] As stopping criterion for the Nelder-Mead routine we set the step size in the pa-

[50] We also estimated our demand model with a gradient-based optimizer and obtained identical results. Dubé *et al.* (2012) discuss that gradient-based optimization of BLP-style models can have substantial advantages. To exploit the full power of gradient-based optimization methods, one has to compute the gradient of the objective function, ideally analytically. For a random coefficients nested logit model and dynamic panel moments the gradient is more complicated than for a standard random coefficient logit model as, for example, in Nevo (2000a). Even though the gradient of our model is still tractable, we found the Nelder-Mead algorithm to be comparable to the gradient-based optimization in terms of speed when we start from the same starting values.

rameter change and the change in the function value to 10^{-8} and 10^{-4}, respectively. By using multiple starting values, we verify that the obtained minimum of the GMM function is indeed a global minimum. For computing the standard errors and the optimal weighting matrix in the second step, we assume that the error terms ν_{Djt} are *iid* across brands, stores, and months.

We also estimated larger demand models with up to 8 non-linear parameters as robustness checks. The larger models resulted in similar price elasticities. However, larger models exhibited significantly larger standard errors and we prefer to use a smaller model, that is more precisely estimated and robust, rather than a potentially noisy large model. We conjecture that with our data, that only relies on data from one retailer and one local market, it is hard to estimate much larger random coefficient models precisely.

Supply estimation. For the estimation of the supply model, we generalize the algorithm proposed by BLP to allow for a flexible internalization matrix. The algorithm can be decomposed into five steps (2.-6.) 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 above.**

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

3. **Back out marginal costs given a guess for the profit internalization parameters in Λ, and $\frac{\partial s()}{\partial p}$ from the demand estimation.** Combining the price elasticities from Step 1 and the parameter guess for Λ from Step 2, we compute the implied production marginal costs for each product and market based on Equation (5). Since our marginal cost functions are linear, we can profile out the marginal cost parameters γ using 2SLS regressions, similarly to Nevo (2000a). As for our demand estimation, the dynamic panel moments require a slight modification to the profiling out formula.

   Let X, Z, and W denote the matrices of marginal cost shifters, supply side instruments, and weighthing matrix, respectively and assume that marginal costs are linear, such that $mc_{jt} = X_{jt}\gamma + \omega_{jt}$ with $\omega_{jt} = \iota_S \omega_{jt-1}$. For our model with dynamic panel moments, the formula for the linear parameters is

   $$\hat{\gamma} = \left(\tilde{X}'ZWZ'\tilde{X}\right)^{-1}\left(\tilde{X}'ZWZ\tilde{mc}(\lambda)\right),$$

   where $\tilde{X}$ and $\tilde{mc}$ are the pseudo-differenced versions of X and mc, i.e., the stacked versions of $\tilde{mc}_{jt} = mc_{jt} - \iota_S mc_{jt-1}$ and $\tilde{X}_{jt} = X_{jt} - \iota_S X_{jt-1}$. Therefore, the linear parameters are not only a function of the profit internalization parameters λ but also of the $AR(1)$-parameter ι_S.

This step allows us to compute the unobservable marginal cost shock ω for each product and market.

4. **Compute innovations in ω-process.** Next, we compute the innovations of the shock process, ν_S as a function of the parameter guess for ι_S and the backed-out vector of unobserved marginal cost shocks ω.

5. **Compute GMM objective function.** Based on the values for ν_S backed out in Step 4, we compute the supply moments which are based on orthogonality conditions between ν_S and the instruments discussed in Section 4.1. Finally, we aggregate the moment conditions to obtain the GMM criterion function for the parameter guess $vec(\Lambda, \iota_S)$. Analogously to our demand estimation, we build on the insights from the dynamic panel literature to improve the efficiency of our estimation. Specifically, we include our promotion-differentiation instruments not only in levels but also in first differences. Finally, we also include one- and two-period lagged marginal costs as instruments to identify the $AR(1)$-parameter in the ω-process.

6. **Repeat steps 2-5 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. Similarly as for the demand side, we did not experience significant advantages of using a gradient-based optimization in our application. Therefore, we revert to derivative-free Nelder-Mead simplex method for estimating our supply model.[51] We use the same stopping criterion as for our demand estimation. For computing the standard errors and the optimal weighthing matrix in the second step, we assume that the error terms ν_{Sjt} are *iid* across brands and months.

Third-stage GMM as robustness check. For both the demand and the supply estimation, we use lagged values of the mean utilities δ and the implied marginal costs on the demand and supply side, respectively, in order to identify the $AR(1)$-coefficient of the structural error process, similarly to Lee (2013). This has some implications for the weighting matrices. For our first-stage weighting matrix, which is essentially the familiar 2SLS weighting matrix $(Z'Z)^{-1}$, we need to pick initial instrument vectors for the lagged mean utilities and the lagged marginal costs. On the demand side, our first-stage weighting matrix includes the lagged mean utilities from a standard logit model. For the second stage weighting matrix we use the mean utilities implied by the first stage estimates.

Analogously, we used the lagged marginal cost vector implied by Nash pricing for constructing the first-stage weighting matrix on the supply side. For the second stage weighting

[51] As for the demand model, we also estimate our supply model using gradient-based methods and using different starting values to ensure that we find the global minimum of the objective function.

matrix on the supply side, we used the lagged marginal cost vector implied by the first stage estimates.

In some specifications, we experienced that the first stage and the second stage differed a bit more each other than what one would expect in a GMM estimation without these additional dynamic panel moments. We suspect that this is due to the fact that the logit mean utilities and the Nash marginal costs could differ significantly from the mean utilities and marginal cost estimates implied by the full model. In our main specifications, we found this issue not to be very relevant.

As an additional robustness check, we run a third-stage GMM, which uses the estimates from the second stage to construct an updated efficient weighting matrix.[52]

In our application, the first and second stage are already relatively close, and the second and third stage estimates are very close in all of our specifications. Throughout the paper, we report the estimates from the second stage.[53]

D.1 Standard Error Adjustments

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, Chapter 12.5.2). The asymptotic variance-covariance matrix of the one-step 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 $g_{\nu_S}(\cdot) = \nu_S(\hat{\theta}_S, \hat{\theta}_D) Z_S$.

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

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

[52] An alternative approach to this issue is to continuously update the weighting matrix with each iteration. We experimented with this and obtained similar results, but we encountered similar issues as with a full-fledged continuously updating GMM, namely, that the estimation took significantly longer to converge and was somewhat more unstable numerically.

[53] The detailed comparison of our estimates across the first, second, and third stage estimation are available upon request.

where g_{ν_S} is $l_2 \times n_S$ and contains the observation-level information for the supply moments and g_{ν_D} is $l_1 \times n_D$ and contains observation-level information on the 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.

E Additional Estimation Results

Demand elasticities. In our random coefficients nested logit model, consumers' own- and cross-price elasticities can be computed according to the following formulas.

$$\eta_{jkt} = \begin{cases} \frac{p_{jt}}{s_{jt}} \int \alpha_i s_{ijt} \left(\frac{1}{1-\rho} - s_{ijt} - \frac{\rho}{1-\rho} s^c_{ijt} \right) dP_D(D) & \text{for } j = k \\ -\frac{p_{kt}}{s_{jt}} \int \alpha_i \left(s_{ijt} + \frac{\rho}{1-\rho} s^c_{ijt} \right) s_{ikt} dP_D(D) & \text{for } j \neq k, \end{cases}$$

where s^c_{ijt} denotes the market share of product j among consumers of type i conditional on i choosing one of the inside goods and integration is taken with respect to the distribution of consumer demographics D.

Table 14: Median Elasticities RCNL Model (Part 1)

	NAB ShW	PO RBr	PO GNu	PO HCb	GM RNB	GM ACC	GM Whe	GM Che	GM HNC	GM LCh	GM TCF	GM Tri	KE FrL
NAB Shred Wheat	-4.551	0.057	0.080	0.036	0.045	0.052	0.070	0.206	0.145	0.065	0.065	0.043	0.075
PO Raisin Bran	0.026	-3.524	0.074	0.036	0.043	0.050	0.065	0.197	0.140	0.064	0.063	0.043	0.073
PO Grape Nuts	0.030	0.060	-4.140	0.033	0.046	0.052	0.077	0.214	0.147	0.055	0.061	0.037	0.068
PO Honey Comb	0.023	0.049	0.060	-4.229	0.040	0.046	0.058	0.175	0.128	0.061	0.059	0.041	0.068
GM RaisinNutBran	0.025	0.054	0.069	0.035	-4.057	0.050	0.065	0.196	0.139	0.065	0.064	0.044	0.075
GM ApplCin Cheer	0.024	0.052	0.065	0.034	0.042	-4.133	0.063	0.192	0.137	0.064	0.062	0.043	0.072
GM Wheaties	0.027	0.056	0.082	0.035	0.045	0.053	-4.369	0.207	0.144	0.063	0.064	0.043	0.073
GM Cheerios	0.025	0.053	0.071	0.035	0.043	0.050	0.066	-4.325	0.142	0.064	0.063	0.043	0.074
GM HonNut Cheer	0.025	0.053	0.068	0.035	0.042	0.050	0.064	0.198	-4.040	0.064	0.062	0.043	0.073
GM Luck Charms	0.019	0.041	0.045	0.029	0.034	0.039	0.048	0.150	0.111	-3.412	0.052	0.035	0.062
GM Tot CoFlakes	0.022	0.047	0.055	0.033	0.038	0.045	0.056	0.172	0.126	0.061	-4.520	0.041	0.069
GM Trix	0.018	0.040	0.044	0.029	0.033	0.038	0.047	0.147	0.108	0.053	0.051	-3.776	0.060
KE Froot Loops	0.022	0.047	0.055	0.031	0.038	0.043	0.055	0.167	0.123	0.057	0.057	0.038	-3.922
KE Special K	0.025	0.052	0.068	0.035	0.042	0.049	0.062	0.190	0.137	0.064	0.062	0.043	0.073
KE Frost Flakes	0.028	0.059	0.095	0.034	0.045	0.052	0.074	0.213	0.146	0.059	0.062	0.040	0.071
KE Corn Pops	0.025	0.053	0.068	0.034	0.042	0.048	0.064	0.188	0.136	0.063	0.062	0.042	0.074
KE Raisin Bran	0.027	0.057	0.085	0.035	0.045	0.052	0.070	0.207	0.145	0.062	0.063	0.042	0.074
KE Corn Flakes	0.029	0.057	0.137	0.026	0.041	0.046	0.076	0.199	0.131	0.042	0.047	0.026	0.052
KE Honey Smacks	0.024	0.052	0.065	0.035	0.041	0.048	0.062	0.184	0.134	0.064	0.062	0.043	0.073
KE Crispix	0.026	0.056	0.075	0.036	0.044	0.051	0.067	0.200	0.142	0.064	0.064	0.043	0.074
KE Rice Krispies	0.028	0.060	0.098	0.033	0.045	0.051	0.075	0.209	0.144	0.058	0.060	0.039	0.069
RAL Chex	0.023	0.050	0.063	0.034	0.041	0.046	0.058	0.176	0.128	0.061	0.059	0.041	0.070
RAL Wheat Chex	0.028	0.057	0.086	0.036	0.046	0.052	0.070	0.206	0.144	0.063	0.063	0.042	0.073
RAL Rice Chex	0.023	0.050	0.063	0.034	0.041	0.046	0.058	0.176	0.128	0.061	0.059	0.041	0.069
QU Quaker Oats	0.029	0.060	0.104	0.033	0.046	0.052	0.076	0.218	0.147	0.057	0.061	0.038	0.068
QU Capn Crunch	0.026	0.055	0.082	0.034	0.043	0.050	0.067	0.197	0.140	0.060	0.061	0.040	0.071
QU Life	0.029	0.060	0.103	0.034	0.046	0.052	0.077	0.216	0.146	0.058	0.062	0.038	0.069
Outside good (div.)	0.439	0.460	0.454	0.440	0.438	0.437	0.438	0.449	0.446	0.433	0.432	0.430	0.444

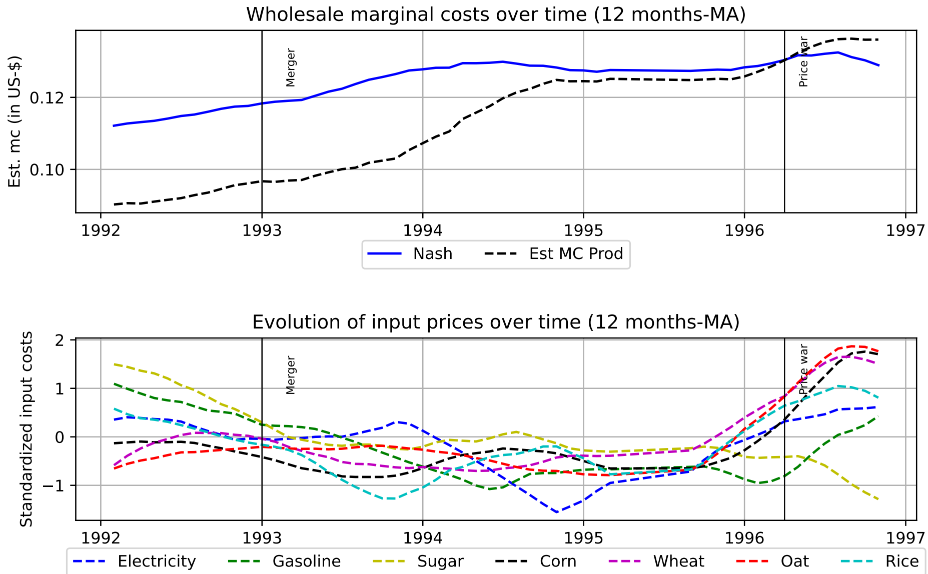
Notes: Cell entries i (indexing row), j (indexing column) give the percent change in market share of brand i associated with a one percent change in the price of j. Each entry represents the median of the elasticities across the 3596 markets. The last row displays the predicted diversion to the outside good, i.e., the percentage of consumers who substitute from a specific inside good to the outside good (as a percentage of all who substitute away) when a product increases its price.

Table 15: Median Elasticities RCNL 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	QU Li
NAB Shred Wheat	0.111	0.233	0.072	0.145	0.158	0.020	0.060	0.166	0.023	0.020	0.032	0.076	0.078	0.043
PO Raisin Bran	0.107	0.217	0.070	0.135	0.142	0.020	0.058	0.154	0.022	0.019	0.031	0.069	0.075	0.040
PO Grape Nuts	0.111	0.292	0.071	0.163	0.275	0.020	0.065	0.212	0.023	0.023	0.031	0.101	0.086	0.057
PO Honey Comb	0.098	0.183	0.065	0.116	0.101	0.018	0.051	0.128	0.020	0.017	0.029	0.058	0.065	0.033
GM RaisinNutBran	0.107	0.208	0.069	0.132	0.123	0.019	0.056	0.146	0.022	0.019	0.031	0.065	0.073	0.037
GM ApplCin Cheer	0.103	0.200	0.068	0.127	0.113	0.019	0.054	0.139	0.021	0.018	0.029	0.062	0.070	0.036
GM Wheaties	0.110	0.239	0.071	0.144	0.168	0.020	0.060	0.171	0.023	0.021	0.032	0.078	0.077	0.044
GM Cheerios	0.106	0.209	0.069	0.135	0.133	0.019	0.056	0.148	0.022	0.019	0.030	0.068	0.072	0.039
GM HonNut Cheer	0.105	0.204	0.068	0.132	0.123	0.019	0.055	0.144	0.021	0.018	0.030	0.065	0.071	0.037
GM Luck Charms	0.083	0.145	0.056	0.096	0.065	0.016	0.044	0.101	0.017	0.014	0.024	0.044	0.055	0.025
GM Tot CoFlakes	0.095	0.169	0.063	0.112	0.085	0.018	0.050	0.117	0.020	0.016	0.028	0.052	0.063	0.030
GM Trix	0.081	0.141	0.054	0.094	0.063	0.015	0.043	0.097	0.017	0.013	0.024	0.042	0.052	0.024
KE Froot Loops	0.093	0.170	0.061	0.110	0.088	0.017	0.048	0.117	0.019	0.016	0.027	0.054	0.062	0.030
KE Special K	-4.813	0.196	0.068	0.129	0.122	0.019	0.055	0.141	0.022	0.018	0.030	0.064	0.070	0.037
KE Frost Flakes	0.111	-4.369	0.071	0.156	0.225	0.020	0.063	0.199	0.023	0.022	0.032	0.091	0.083	0.051
KE Corn Pops	0.103	0.205	-4.666	0.128	0.120	0.019	0.056	0.143	0.021	0.018	0.030	0.064	0.070	0.037
KE Raisin Bran	0.111	0.250	0.070	-3.785	0.184	0.020	0.061	0.174	0.023	0.021	0.032	0.081	0.080	0.046
KE Corn Flakes	0.101	0.335	0.060	0.169	-4.052	0.017	0.064	0.262	0.020	0.024	0.028	0.125	0.081	0.069
KE Honey Smacks	0.103	0.195	0.067	0.125	0.111	-4.015	0.055	0.137	0.022	0.018	0.030	0.062	0.070	0.035
KE Crispix	0.109	0.220	0.071	0.138	0.141	0.020	-5.150	0.157	0.023	0.020	0.032	0.071	0.076	0.041
KE Rice Krispies	0.112	0.277	0.069	0.156	0.240	0.020	0.064	-5.119	0.023	0.023	0.032	0.095	0.082	0.052
RAL Chex	0.099	0.183	0.065	0.119	0.107	0.018	0.053	0.131	-4.557	0.018	0.030	0.060	0.067	0.034
RAL Wheat Chex	0.112	0.243	0.071	0.146	0.180	0.020	0.062	0.175	0.024	-4.401	0.033	0.080	0.079	0.046
RAL Rice Chex	0.100	0.183	0.065	0.118	0.108	0.018	0.053	0.132	0.022	0.018	-4.556	0.060	0.067	0.034
QU Quaker Oats	0.111	0.288	0.070	0.161	0.266	0.020	0.064	0.207	0.023	0.023	0.031	-4.105	0.084	0.054
QU Capn Crunch	0.106	0.232	0.068	0.141	0.160	0.019	0.059	0.160	0.022	0.020	0.030	0.077	-4.175	0.044
QU Life	0.112	0.286	0.071	0.163	0.258	0.020	0.065	0.210	0.023	0.023	0.032	0.098	0.085	-4.497
Outside good (div.)	0.432	0.512	0.444	0.453	0.463	0.440	0.437	0.445	0.426	0.434	0.426	0.451	0.508	0.447

Notes: Cell entries i (indexing row), j (indexing column) give the percent change in market share of brand i associated with a one percent change in the price of j. Each entry represents the median of the elasticities across the 3596 markets. The last row displays the predicted diversion to the outside good, i.e., the percentage of consumers who substitute from a specific inside good to the outside good (as a percentage of all who substitute away) when a product increases its price.

Figure 15: Evolution of predicted marginal costs for different model specifications



Notes: The top figure displays the evolution of the estimated manufacturer marginal costs over time predicted by two different models: static multiproduct Nash pricing (solid line) and our small conduct specification (dashed line). The figure is based on the median over all brands for a given month and a moving average over a rolling 12-months window. The bottom figure displays the evolution of the 12-months moving average of various input prices over time.

Evolution of marginal cost estimates. As an additional validation of our estimates, we compare marginal cost predictions from our two conduct models with the ones obtained under the assumption of Nash pricing. Figure 15 illustrates the evolution over time of the median marginal costs implied by two different models. Under the assumption of Nash pricing (solid line), we obtain relatively stable marginal costs over time. The implied marginal costs from our conduct models exhibit a slightly different pattern (dashed line). In particular, they increase in the beginning of the post-merger period, and increase further during the price war period, while Nash pricing predicts relatively constant marginal costs during the price war period and even a slight decrease at the very end of our sample.

Ideally, one would like to compare these predicted marginal costs to an observed counterpart. Unfortunately, it is extremely hard to obtain such measures from the data. Therefore, we plot the evolution of several important input prices (corn, wheat, rice, oat, sugar, electricity, and gasoline) in the bottom panel of Figure 15. It is important to note that input prices are not a perfect proxy for economic marginal costs. However, especially for the price war period, we find that the predictions from our estimated conduct pattern are more consistent with input price data than the ones from a Bertrand-Nash model: There is a sharp increase in input prices shortly before and during the price war.[54] Overall, this pattern is more difficult

[54]Sugar constitutes an exception as its price drops monotonically (except for a short period in 1994).

Table 16: Summary of marginal cost and margin estimates

	Small model Pre-merger	Small model Post-merger	Small model Price War	Large model Pre-merger	Large model Post-merger	Large model Price War	Nash
MC prod	0.0957	0.1252	0.1391	0.0947	0.1233	0.1390	0.1251
Total MU	0.5167	0.4021	0.3344	0.5240	0.4007	0.3340	0.3947
Manuf MU	0.4141	0.1879	0.0448	0.4148	0.1869	0.0484	0.2147

Notes: The table summarizes estimates for manufacturer marginal costs (in US-$ per ounce, including trade spend payments), total markups (assuming a retailer marginal cost of zero), and wholesale markups defined as $(p^w - mc^w)/p^w$, where p^w denotes the net wholesale price and mc^w is the marginal cost of production. The numbers are medians over all brands and months separately for the pre-merger, post-merger, and price-war period. The last column displays the statistics assuming multiproduct Nash pricing over the whole sample.

to reconcile with the marginal costs predicted by Nash pricing. We believe that these results provide further support for our conduct specifications.

Finally, our marginal cost estimates are overall in line with the results of Nevo (2000b), who finds an average marginal cost per serving of roughly 10.7 cents at the end of 1992 under the assumption of multiproduct Nash pricing. This number is in between our average marginal cost estimates for multiproduct Nash and the estimated conduct model, for which we find marginal costs of 11.9 and 9.8 cents per serving, respectively. When comparing our results to Nevo's one should keep in mind that he uses a different geographic sample. His marginal cost estimates are based on the median across 45 U.S. cities, while ours are only based on data from the Chicago area and one retailer.

Conduct estimates: Hypothesis testing. In the following, we present results from a series of statistical hypotheses tests (t-statistics) for the equality of the profit internalization parameters over time and across firms for both our small and our large conduct model. Table 18 summarizes the associated results.

Row 1 displays the result of testing $H_0 : \lambda_t = 0$, which is equivalent to Nash pricing. While we reject the null hypothesis for the pre-merger period and the price war period, we cannot reject Nash pricing in the post-merger period for our small model with homogeneous conduct and in the large model for the four smaller firms. For the large firms, the large model indicates that pricing is more cooperative than Bertrand-Nash in the post-merger period.

Row 2 shows the results of testing $H_0 : \lambda_t = 1$, which implies joint profit maximization. We reject this hypothesis in all periods for the small model, and for the large model during the post-merger and the price war period for all firms. During the pre-merger period, the large model indicates that the pricing behavior of the smaller firms is consistent with joint profit maximization.

Rows 3 to 6 test the equality of different pairs of profit internalization parameters. For the small model (row 3) we reject clearly that the parameters in any two periods are equal. The corresponding results for the large model (rows 4 and 5) are similar. However, we cannot reject the hypothesis that the internalization parameters for Kellogg's and General Mills are equal before and after the merger, see the first t-statistic (1.55) in row 4.

Finally, row 6 tests whether the profit interalization parameters of our large model are statistically different across firm groups within a given time period. We can reject the equality of the internalization parameters for the small and large firms during the post-merger period at the 10%-level. In the pre-merger period, however, the internalization parameters are not statistically different across firm groups.

Relationship to other industry studies. Because the RTE cereal industry has been studied extensively, it is useful to relate our results to those in the literature. The work of Nevo

Table 17: Time-brand specific wholesale price-cost margins

	Small model Pre-merger	Small model Post-merger	Small model Price War	Large model Pre-merger	Large model Post-merger	Large model Price War	Nash
NAB Shred Whe	0.3459	0.1805	0.0717	0.3715	0.1571	0.0722	0.1657
PO Raisin Bra	0.4621	0.2874	0.1330	0.4905	0.2565	0.1325	0.2575
PO Grape Nuts	0.4563	0.1804	0.1020	0.4853	0.1368	0.1021	0.2035
PO Honey Comb	0.3453	0.2171	0.0672	0.3649	0.1886	0.0677	0.1978
GM RaisinNutB	0.4334	0.2573	0.1052	0.4336	0.2773	0.1070	0.2611
GM ApplCin Ch	0.4270	0.2415	0.0953	0.4283	0.2581	0.0971	0.2541
GM Wheaties	0.4170	0.1997	0.0660	0.4173	0.2226	0.0679	0.2125
GM Cheerios	0.3836	0.1868	0.0417	0.3839	0.2059	0.0400	0.1962
GM HonNut Che	0.4172	0.2245	0.0521	0.4173	0.2510	0.0498	0.2350
GM Luck Charm	0.4709	0.3073	0.1745	0.4712	0.3270	0.1779	0.3314
GM Tot CoFlak	0.3500	0.1743	0.0731	0.3500	0.1974	0.0751	0.2191
GM Trix	0.4167	0.2471	0.1178	0.4168	0.2732	0.1212	0.2704
KE Froot Loop	0.4115	0.2737	−0.0886	0.4120	0.2855	−0.0917	0.2752
KE Special K	0.3455	0.1954	0.0293	0.3462	0.2117	0.0297	0.2093
KE Frost Flak	0.4223	0.1293	−0.0163	0.4242	0.1443	−0.0193	0.2198
KE Corn Pops	0.3625	0.1854	−0.0067	0.3640	0.2024	−0.0040	0.2109
KE Raisin Bra	0.4876	0.2605	−0.0413	0.4900	0.2787	−0.0415	0.2825
KE Corn Flake	0.5071	0.1299	−0.1082	0.5087	0.1175	−0.1108	0.2374
KE Honey Smac	0.4256	0.2504	−0.0200	0.4265	0.2609	−0.0181	0.2569
KE Crispix	0.3589	0.1742	0.0108	0.3598	0.1885	0.0143	0.1988
KE Rice Krisp	0.3799	0.1405	−0.0281	0.3818	0.1479	−0.0270	0.1812
RAL Chex	0.3316	0.1996	0.0764	0.3540	0.1770	0.0791	0.1860
RAL Wheat Che	0.4122	0.1960	0.0718	0.4379	0.1526	0.0734	0.1837
RAL Rice Chex	0.3323	0.2032	0.1134	0.3549	0.1823	0.1150	0.1897
QU Quaker Oat	0.4588	0.1704	−0.0207	0.4868	0.1424	−0.0175	0.2073
QU Life	0.3771	0.1662	−0.2004	0.4020	0.1176	−0.1978	0.1710
QU Capn Crunc	0.4053	0.1617	0.0089	0.4301	0.1399	0.0097	0.1843

Notes: The table entries reflect brand-specific median (across months) total wholesale price-cost margins for both the small and the large model specifications separately for the pre-merger, post-merger, and price-war periods, as well as under the assumption of multiproduct Bertrand-Nash pricing over the whole sample.

Table 18: Conduct estimates: Hypothesis tests

Row	Model	$\lambda_{Pre} = 0$	$\lambda_{Post} = 0$	$\lambda_{Pw} = 0$
1	Small	5.45***	1.35	-3.30***
	Large	3.73*** 3.51***	2.44** -0.33	-2.94***
		$\lambda_{Pre} = 1$	$\lambda_{Post} = 1$	$\lambda_{Pw} = 1$
2	Small	-3.04***	-6.21***	-5.62***
	Large	-2.17** -1.37	-2.82*** -5.49***	-5.08***
		$\lambda_{Pre} = \lambda_{Post}$	$\lambda_{Post} = \lambda_{PW}$	$\lambda_{Pre} = \lambda_{PW}$
3	Small	7.09***	3.97***	4.88***
		$\lambda^{KE,GM}_{Pre} = \lambda^{KE,GM}_{Post}$	$\lambda^{KE,GM}_{Post} = \lambda^{KE,GM}_{PW}$	$\lambda^{KE,GM}_{Pre} = \lambda^{KE,GM}_{PW}$
4	Large	1.55	4.22***	4.32***
		$\lambda^{Rest}_{Pre} = \lambda^{Rest}_{Post}$	$\lambda^{Rest}_{Post} = \lambda^{Rest}_{PW}$	$\lambda^{Rest}_{Pre} = \lambda^{Rest}_{PW}$
5	Large	5.99***	2.79***	4.29***
		$\lambda^{KE,GM}_{Pre} = \lambda^{Rest}_{Pre}$	$\lambda^{KE,GM}_{Post} = \lambda^{Rest}_{Post}$	
6	Large	0.31	1.88*	

*Notes: The table summarizes the t-statistics associated with the hypothesis tests described above. Standard errors account for two-step estimation. *, **, *** denote significance of the test statistic at the 10%, 5%, and 1%-level, respectively. Number of observations: 1,674.*

(2000b) and Nevo (2001) is of particular interest. Nevo (2000b) simulates the effects of different hypothetical horizontal mergers using only pre-merger data. Assuming multiproduct Nash pricing before and after the merger, he finds that, in the absence of considerable cost synergies for the merging firms, the merger between Post and Nabisco leads to an increase in prices and a decrease in consumer surplus. Our focus is on estimating the evolution of conduct over time using pre- and post-merger data. Most importantly, in our model specifications, changes in markups cannot only be explained by the *unilateral effects* of the merger but also by changes in industry conduct in the post-merger period (*coordinated effects*).

Nevo (2001) measures market power in the RTE cereal industry. His sample contains data from 65 U.S. cities covering a period from 1988 to 1992. Time-wise, this partially overlaps with our pre-merger period. Our demand estimates differ somewhat from his estimates for several, but related, reasons. Nevo (2001, 2000b) uses a much richer data set covering many U.S. regions and several retailers. In contrast, we have to rely on data from only one retailer and one geographic market. The richer data structure allows him to estimate substantially more random coefficients and demographic interactions than what we are able to do with our data.

The advantage of our data is that we can rely on wholesale price data, which allows us to directly model manufacturer pricing instead of relying only on retail price data. For example, during parts of our sample period, retail and wholesale prices for some brands move in opposite directions, in particular in 1994.

The limitations of our data force us to use a less general random coefficient specification. For example, we only incorporate four income-interactions and group all the inside goods into one nest instead of estimating normally-distributed random coefficients on various product characteristics. In addition, Nevo uses Hausman (1996) instruments based on prices from other regions, and a slightly different product set.[55]

Finally, Nevo estimates his demand model on quarterly data, while we use monthly data; therefore, one could argue that our estimates describe more of a short-run elasticity than the elasticities that Nevo reports for the quarterly level. The different time aggregation could explain, at least partially, why our demand is more elastic than Nevo's.

If one believes that our demand model overestimates the price elasticities, our conduct estimates would provide a lower bound on the profit internalization parameter, but the qualitative trend of industry conduct becoming more competitive over time should still be robust. We believe that the advantages of conducting our analysis using wholesale price data outweigh the drawbacks of using a more restrictive demand specification. Replicating our empirical approach using a more comprehensive data set, such as the ones from IRI used by Nevo or Nielsen as used by Backus *et al.* (2021) is a promising avenue for future research.

To select among different forms of industry conduct, Nevo (2001) compares the recov-

[55] We include all products used by Nevo, and we additionally include products from the manufacturer Ralston.

ered marginal cost for different pre-specified conduct models with accounting cost data under the assumption of vertical integration, i.e., joint profit maximization between retailers and manufacturers. Comparing three different conduct assumptions (single-product Nash, multiproduct Nash, and joint ownership of all products), he finds that multiproduct Nash pricing provides the best fit to the industry accounting data, resulting in a combined retailer-manufacturer price-cost margin of 42.2 percent compared to 35.8 percent under single-product Nash and 72.6 percent under joint ownership of all brands. For the part of our sample period that overlaps with his sample period, i.e., the pre-merger period, the estimated markups from both of our conduct models are approximately 50% and therefore, closer to Nevo's multiproduct Nash specification than to his other two considered options.

However, our results indicate a significantly positive but moderate level of cooperative conduct during this period that results in estimated total margins that are 30% larger than the ones implied by Nash pricing.[56]

[56] Assuming multiproduct Nash pricing, we find an implied gross margin for the pre-merger period of approximately 40 percent, which is only slightly lower than Nevo's (2001).