# Common Ownership and Competition in the Ready-To-Eat Cereal Industry*

PRELIMINARY AND INCOMPLETE
PLEASE DO NOT CITE OR CIRCULATE

Matthew Backus† Christopher Conlon‡ Michael Sinkinson§

November 21, 2020

**Abstract**

If firms maximize the value of their shareholders' portfolios, and shareholders also own shares in competing firms, firms may have an incentive to soften competition in the product market. This is is formalized as the common ownership model of firm conduct. We build on identification results for conduct testing (Berry and Haile, 2014) to devise a structural test for differentiated product markets and apply it to data from the ready-to-eat (RTE) cereal market. We find no evidence of higher prices through common ownership, though we show that the potential magnitude of such effects would be very large.

*Thanks to conference and seminar attendees and especially Steven Berry, Philip Haile, Jean-Francois Houde, Robin Lee, Aviv Nevo, and Ariel Pakes for thoughtful comments. All remaining errors are our own. The authors own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†Columbia University and NBER, matthew.backus@columbia.edu

‡New York University, cconlon@stern.nyu.edu

§Yale University and NBER, michael.sinkinson@yale.edu

## 1. Introduction

Publicly traded corporations, starting with the Dutch East India Company, have been the backbone of the global economy for hundreds of years. The fundamental promise of the corporation is that it maximizes shareholder value. The board of directors, with a fiduciary duty to represent the interests of shareholders, designs incentives and delegates decision-making to management. The goal of maximizing shareholder value aligns with that of maximizing profits, as shares are valued at the discounted present value of all future earnings.

More recently, concerns have been raised that through the rise of large index funds, ownership of public corporations has become concentrated among a small number of investors, and that this may distort the decisions firms make. In particular, the two largest asset managers, Blackrock and Vanguard, have a combined assets under management total of $10.2 trillion.[1] To put that number in context, the total market capitalization of the S&P 500 on December 31, 2017 was $23.9 trillion. This is concerning because the classic model of the corporation implicitly assumes shares are held by undiversified investors. Suppose a firm could take an action that would cause its own share price to rise, but that of a rival to fall, e.g. by obtaining some competitive advantage. In the classic model, the firm would take such an action, narrowly pursuing profits for itself. However, if the firm's shareholders also held shares in the rival, the firms incentives are less clear. This has led to regulatory interest, such as an Federal Trade Commission (FTC) hearing on the topic in December 2018 and a joint FTC and Department of Justice (DOJ) request for comments on proposed amendments to regulations around disclosure of minority stakes in firms in September of 2020.[2]

This issue has attracted significant attention since recent papers have claimed to see an empirical effect of common ownership on firm pricing (Azar et al., 2016, 2018a). These papers have argued that airfare and bank interest rates and fees are responsive to the level of common ownership in a market. The approach in these papers is to run reduced-form regressions of prices on concentration measures such as the Herfindahl-Hirschman Index (HHI) and the Modified HHI (MHHI), a measure that extends the Cournot foundations of HHI to allow for common ownership incentives. The problem with this approach, which is familiar to practitioners of (the no longer "new") empirical industrial organization, is that such measures can give spurious results when taken out of the narrow confines of symmetric Cournot competition. Especially in the case of HHI and MHHI, which are functions of market shares, any misspecification will introduce the classical endogeneity problem of price and quantity. In Appendix D, we offer an simulation of a simple differentiated products model to show how spurious results are likely, even when instrumenting for changes in common ownership.

[1] These figures are dated September 30, 2017. Blackrock reports $5.7 trillion AUM and Vanguard reports $4.5 trillion AUM.

[2] "FTC and DOJ Seek Comments on Proposed Amendments to HSR Rules and Advanced Notice of Proposed HSR Rulemaking." https://www.ftc.gov/news-events/press-releases/2020/09/ftc-doj-seek-comments-proposed-amendments-hsr-rules-advanced

As a result, many have challenged the empirical findings of an effect of common ownership.

In contrast, this paper considers testing the common ownership hypothesis in a standard differentiated product markets framework. We adapt the standard firm profit maximization approach to nest the common ownership hypothesis by including profit weights in the "ownership matrix" used by the firm to set optimal prices. Then, in the spirit of Bresnahan (1987), Nevo (2001), and Berry and Haile (2014), we propose a method for testing which of several different ownership matrices is consistent with product market data. This test is based on exclusion restrictions: under the true conduct assumption, marginal costs derived from estimates of the demand system should not respond to so-called "markup shifters." Under the incorrect model, however, the marginal cost estimates will be misspecified, and therefore mechanically related to the proposed shifters. In this spirit we can use first order conditions to show that our test should have power to reject misspecified conduct assumptions. The shifters we propose are based on demand rotations, cost shifters (of competitors' products), and, most pertinent to the common ownership hypothesis, proxies for common ownership incentives based on upstream financial structures.

We apply our approach to the market for ready-to-eat cereals using scanner-level data over a ten year time horizon. The chosen industry is an ideal one for this question as it exhibits variation in ownership concentration across firms, as well as over time, in addition to a high level of product market concentration to begin with ($C_4$ is approximately 85%). For example, The Kellogg Foundation is a major (undiversified) shareholder of Kelloggs, Inc, although its share has fallen over time from over 30% in 2000 to just under 20% in 2017. We also observe that the products sold under the Post brand changed hands several times, from being a part of Kraft foods, to being a part of Ralcorp, and finally to being an independent firm, all of which affect the extent to which its shares are commonly owned.

In this context we illustrate the economic content of different conduct assumptions on firm price-setting in ready-to-eat cereal: perfect competition, firm-level profit maximization, multi-product monopoly, and variants of the common ownership hypothesis. We implement the test derived in Section 3 using the rich demand data and marginal cost instruments and find that the firm-level profit maximization is the most consistent with the data. We then construct counterfactual predictions of prices with and without common ownership in order to quantify the economic significance of the hypothesis, and find that the potential economic implications of the common ownership hypothesis are large: if firm conduct were as assumed in most common ownership work, prices would rise by more than they would if any pair of the four main firms were to merge.

This paper contributes to the evolving literature on common ownership in several ways. First, it documents the issues with using simpler empirical approaches to assess common ownership effects in more realistic product markets. Second, it provides a framework for including common ownership effects in structural empirical research on differentiated product markets and show how to test the common ownership hypothesis against other models of conduct. Third, it assesses the impact of

common ownership in a well-studied product market and quantifies the potential impact of different forms of common ownership.

### 1.1. Related Literature

Empirical testing of firm conduct is an endeavor with a storied history in industrial organization (IO). Modern methods build on advances in structural modeling of demand and firm pricing. From a methodological standpoint, Berry and Haile (2014) iterate on the insight of Bresnahan (1982) to show that models of conduct are testable if demand is identified subject to restrictions on the cost structure. Examples of this approach include Bresnahan (1987), Nevo (2001), and Miller and Weinberg (2017).[3] With respect to this literature, we pose the common ownership hypothesis as an alternative model of conduct, and employ some of the same identification results that have been used to test for collusion and portfolio pricing.

The theoretical foundations of the common ownership hypothesis are not new. Rotemberg (1984) offered the earliest model in which diversification of shareholders affects the character of imperfect competition in product markets. Bresnahan and Salop (1986) and O'Brien and Salop (2000) treats the same question, characterizing "partial mergers," where diversified ownership by shareholders and cross-ownership among firms effects degrees of control. In particular, Bresnahan and Salop (1986) introduced the modified HHI (MHHI) concentration measure to capture such partial control.

The recent revival of interest in the common ownership hypothesis follows from a handful of empirical studies that appear to show large effects on pricing in product markets. The main two instigating endeavors are Azar et al. (2016), which studied the effects of common ownership on bank fees, and Azar et al. (2018a), which uses the Blackrock acquisition of Barclays as an instrument to study the effect of common ownership on airfares. Neither of the above papers address the question of finding a mechanism – precisely how do large common investors affect prices? Antón et al. (2020) propose that it is through executive compensation schemes with respect to firm performance. They exploit index-entry events that change common ownership and show that they are correlated with attributes of corporate compensation in a way that is consistent with reduced competition in product markets.

This literature has not been without criticism, in particular for its descriptive approach. The explanatory variables of interest are concentration indices such as MHHI, an approach to studying markets that has fallen out of favor in industrial organization. Besides the obvious concerns for identification these concentration measures have no monotonic relationship to the main outcome of interest – product market prices – as O'Brien (2017) reminds us. Notwithstanding these con-

[3] Nevo (2001) also studies ready-to-eat cereal to test models of conduct and is the most proximate point of departure for our endeavor. We iterate on that seminal work by using an updated sample (1990-1996 vs 2000 - 2015), more stores (6 versus thousands), more detailed characteristics data, store-specific measurement of market size and, most importantly, an alternative conduct hypothesis.

cerns about the empirical results, this burgeoning literature has attracted the interest of the legal community (Elhauge, 2016) and antitrust authorities, and there is already interest in regulatory remedies (Posner et al., 2017). The literature is further reviewed in a related paper by the authors (Backus et al. (2019)). It is a pressing concern then, to develop a structural approach that can guide antitrust and regulatory authorities in evaluating both the problem and potential remedies. To our knowledge, the only other paper that endeavors a structural analysis of the common ownership hypothesis is Kennedy et al. (2017), which takes a different approach and studies the airline industry.[4]

## 2. Theory of Common Ownership

The theoretical literature on common ownership has its early origins in Rotemberg (1984) and Bresnahan and Salop (1986). It offers a model of the objective function of the firm in terms of shareholder interests. Our derivation follows O'Brien and Salop (2000), beginning with the assumptions from Backus et al. (2020):

**Assumption 1.** *Stakeholder portfolio values are given by:* $v_s \equiv \sum_s \beta_{fs}\pi_f$.

**Assumption 2.** *(Rotemberg, 1984) Managers maximize a* $\gamma_{fs}$*-weighted average of stakeholder portfolio values:* $Q_f \equiv \sum_s \gamma_{fs} v_s$.

Each stakeholder $s$ receives some share of firm $f$'s profits $\beta_{fs}$ (presumably by owning shares in the firm). Assumption 1 says that investors hold portfolios made up of investments of many firms. In strategic games, firms may exert externalities upon each other's payoffs. Because different investors hold different portfolios, they have different objectives for the firm. In order to aggregate across investors with different interests, Assumption 2 states that the manager of $f$ places some Pareto weight $\gamma_{fs}$ on the profits of each shareholder and maximizes a weighted sum of investor payoffs.

Now if we consider a strategic action $x_f$ taken by firm $f$ which affects their profits and the

[4]The papers are quite different in approach and, given the gravity of the question, we view this heterogeneity in method as complementary. In particular they study airlines, mirroring Azar et al. (2018b), while we have deliberately chosen an industry with simpler pricing practices; they estimate a nested logit model of demand while ours is a random coefficient logit; and finally they estimate a conduct parameter where we follow the "menu" approach of Nevo (2000) and study discrete, fully specified, conduct models.

profits of rival firms $\pi_f(x_f, x_{-f})$, we can re-arrange the objective function of the firm $Q_f(x_f, x_{-f})$:

$$\begin{aligned} Q_f(x_f, x_{-f}) &= \sum_s \gamma_{fs} \cdot \pi_s(x_f, x_{-f}) \\ &= \sum_s \gamma_{fs} \cdot \left( \sum_{\forall g} \beta_{gs}\, \pi_g(x_f, x_{-f}) \right) \\ &= \sum_s \gamma_{fs}\, \beta_{fs}\, \pi_f + \sum_s \gamma_{fs} \sum_{\forall f \neq g} \beta_{gs}\, \pi_g \\ &\propto \pi_f + \sum_{g \neq f} \underbrace{\frac{\sum_s \gamma_{fs}\beta_{gs}}{\sum_s \gamma_{fs}\beta_{fs}}}_{\equiv \kappa^{co}_{fg}(\gamma_f, \beta)} \pi_g \\ &= \pi_f + \sum_{g \neq f} \kappa^{co}_{fg}(\gamma_f, \beta)\pi_g. \end{aligned} \tag{1}$$

Therefore each firm $f$ will place some non-negative weight $\kappa^{co}_{fg}$ on the profits of rival firms. By construction, $\kappa^{co}_{ff} = 1$, so that all $\kappa^{co}_{fg}$ are defined relative to the weight that firm $f$ places on its own profits.[5]

These assumptions are not controversial on their own. In fact, the model above is sufficiently general that it generates almost no testable implications. For example, the objective can include the manager's private investment portfolio and have them place a high $\gamma$ weight on their private returns. Alternatively, the manager can place a weight of $\gamma_{fs} = 0$ on particular stakeholders (including passive or index investors, investors below some minimal blockholder threshold, etc.). The challenge for empirical work is that even if we observed $\kappa^{co}_{fg}$ for every pair of firms within an industry, it is unlikely we would be able to recover $\gamma_{fs}$ without additional (parametric) assumptions, because while their might be ten firms within an industry (one-hundred pairwise interactions) there are often thousands of investors in each firm.

The controversy arises because the researcher must make an assumption on $\gamma_{fs}$. This is akin to specifying a particular model of corporate governance or corporate control. The most common choice for empirical work is the so-called "proportional control" assumption that sets $\gamma_{fs} = \beta_{fs}$ for investors and excludes other stakeholders. This is intuitively appealing as "one-share, one-vote" though is not necessarily consistent with a particular model of social choice. Backus et al. (2021) demonstrate that the proportional control assumption yields average values for $\kappa^{co}_{fg}$ around 0.7 for large publicly-traded firms given the current distribution of ownership, implying a large degree of potential "cooperation."

Another way to interpret the proportional control assumption is as a specific frictionless benchmark. Absent agency conflicts, managers are perfect agents for shareholders and investment man-

[5]To avoid dividing by zero, we need that $\gamma_f \cdot \beta_f > 0$, which is guaranteed, for instance, if $\gamma(\beta)$ is a strictly increasing monotone function and $\gamma(0) = 0$.

agers are perfect agents for investors. One way to interpret critiques of the common ownership hypothesis is that the "true" model of corporate governance specifies some alternative form for $\gamma_{fs}$. Another interpretation of a "lack of cooperation" in the data is that agency conflicts distort the manager's objective away from (1).

**Application to Cournot:** Much attention in the common ownership literature has been paid to the Modified Herfindahl-Hirschman Index (MHHI) concentration measure, which is derived from a Cournot oligopoly model of competition.[6] MHHI extends the traditional concept of HHI to incorporate common ownership, and is defined from the following firm objective function:

$$\max_{q_f} \pi_f(q_f, q_{-f}) + \sum_g \kappa_{fg} \pi_g(q_f, q_{-f}).$$

We let $\pi_f(q_f, q_{-f}) = q_f \cdot (p(Q) - c_f)$ where $c_f, q_f$ denote the marginal cost and output for firm $f$ respectively and $p(Q)$ denotes the inverse demand (at total output $Q = \sum_f q_f$). After taking the FOC (where $\epsilon_d$ represents the elasticity of demand and the market share is given by $s_f = \frac{q_f}{Q}$) we get:[7]

$$\frac{p_f - c_f}{p_f} = \frac{1}{\epsilon_d} \sum_g \kappa_{fg} s_g.$$

Aggregating across firms in the market, we get the share-weighted average markup in the market:

$$\begin{aligned} \sum_f s_f \frac{p_f - c_f}{p_f} &= \frac{1}{\epsilon_d} \underbrace{\sum_f \sum_g \kappa_{fg} s_g s_f}_{MHHI} \\ \text{– where } MHHI &= \underbrace{\sum_f s_f^2}_{HHI} + \underbrace{\sum_f \sum_{g \neq f} \kappa_{fg} s_f s_g}_{MHHID}. \end{aligned} \tag{2}$$

Following this decomposition of MHHI, MHHID is sometimes interpreted as the additional pressure of common ownership, over and above market concentration measured by HHI. This equation, which relates share-weighted average markups to MHHID, was used as motivation for regressions of market-level prices on MHHID.

**Application to differentiated Bertrand:** A similar, if less parsimonious, result follows in the differentiated multi-product Bertrand setting. We define $p_f$ as a vector of prices $\{p_j\}_{j \in \mathcal{J}_f}$ for the products produced by firm $f$, $\mathcal{J}_f$, and $\mathbf{p}$ as the vector of prices for all products and all firms. Firms

[6] Originally the MHHI was derived by Bresnahan and Salop (1986) in the context of a joint-venture.

[7] This follows Bresnahan and Salop (1986) which in turn generalizes the well-known result by Cowling and Waterson (1976).

solve:

$$\max_{p_f} \pi_f(p_f, p_{-f}) + \sum_g \kappa_{fg} \pi_g(p_f, p_{-f}).$$

Product-level profits for product $j$ are given by $\pi_j(\mathbf{p}) = D_j(\mathbf{p}) \cdot (p_j - c_j)$. This yields a set of $|\mathcal{J}_f|$ first-order conditions for firm $f$:

$$\underbrace{D_j(\mathbf{p}) + (p_j - c_j)\frac{\partial D_j}{\partial p_j}(\mathbf{p})}_{\text{single product FOC}} + \underbrace{\sum_g \kappa_{fg} \cdot \left( \sum_{k' \in \mathcal{J}_g} (p_{k'} - c_{k'}) \frac{\partial D_{k'}}{\partial p_j}(\mathbf{p}) \right)}_{\text{portfolio effects}} = 0 \tag{3}$$

The second term generalizes the multi-product portfolio effects to allow for common ownership. Absent common ownership effects, when $g = f$, $\kappa_{ff} = 1$ and the single product first-order condition is augmented by recaptured substitution to other products in $f$'s portfolio. In the presence of common ownership $\kappa_{fg} > 0$ for $g \neq f$, and for products $k \in \mathcal{J}_g$ that $f$ does not control. This highlights the connection between multi-product pricing and common ownership. For the firm selling multiple substitute products, in addition to the marginal cost of each sale, they consider the opportunity cost of foregone sales of their other products. Common ownership introduces yet another opportunity cost: that of foregone sales at competing firms, which are weighted by $\kappa_{fg}$, the profit weight firm $f$ places on firm $g$.

## 3. A Structural Test of Horizontal Conduct

Testing models of conduct is a central question in industrial organization. Here we develop a test inspired by the identification result in Berry and Haile (2014), which builds upon the argument of Bresnahan (1982). We begin with an overview of the model in 3.1, then we introduce the test in 3.2, and delve into finer points, in particular selection of instruments, in 3.3.

### 3.1. Conduct and Equilibrium

Conduct is neither supply nor demand, it is the equilibrium restriction that joins them. To set ideas, we begin with notation. A general formulation of demand is

$$s_{jt} = \sigma_j(\mathbf{p_t}, \mathbf{x_t}, \mathbf{v_t}, \xi_\mathbf{t}), \tag{4}$$

where $s_{jt}$ is the market share of product $j$ in market $t$, $\mathbf{p_t}$ is a vector of prices, $\xi_\mathbf{t}$ is a vector of product-specific unobservables, and $\mathbf{x_t}$ and $\mathbf{v_t}$ are matrices of variables that affect demand such as product characteristics or market-level variables; the distinction between them is that while $\mathbf{x_t}$ may also appear in the supply equation, variables in $\mathbf{v_t}$ do not. Variables in bold are vectors stacked over all products and firms, and therefore omit the $j$ subscript.

Supply is given by:[8]

$$mc_{jt} = h_s(x_{jt}, w_{jt}) + \omega_{jt}. \tag{5}$$

Here, $x$ and $w$ are variables that affect marginal costs, partitioned into those that also enter the demand system ($x$) and those that are excluded from the demand system ($w$). The final term $\omega_{jt}$ captures unobserved cost factors.

Finally, we close the model with conduct:

$$mc_{jt} = \psi(\mathbf{s_j}, \sigma_j(\mathbf{p_t}, \mathbf{x_t}, \mathbf{v_t}, \xi_\mathbf{t}), \mathbf{p_t}). \tag{6}$$

A model of conduct, in the notation of Berry and Haile (2014), is a specification of the marginal revenue function $\psi$. This function joins supply, on the left-hand side of (6), with demand on the right-hand side of (6), conditional on observed prices and market shares. With this, we can nest a variety of models of equilibrium.

The simplest case is perfect competition, in which $\psi = p_{jt}$, so that prices are equal to marginal cost. With a bit more work, $\psi$ can also accommodate models based on Bertrand price setting. The key step of the argument is to characterize first-order conditions using Bresnahan (1987)'s "matrix of cooperation," which we will denote by a $J \times J$ matrix $\kappa$. Precisely, this is a matrix of profit weights. The weight $\kappa_{fg}$ captures the rate at which the firm $f$ internalizes the profits generated by firm $g$, and so on-diagonal terms, i.e. $\kappa_{ff}$, are normalized to 1.

The flexibility of this framework is in the off-diagonal terms. They may be used to formalize either coordinated or unilateral effects. As an example of the former, consider two firms that efficiently collude to maximize their joint profits; in that case, $\kappa_{fg} = 1$.[9] Alternatively, as an example of unilateral effects, Section 2 above uses this framework to model the unilateral incentives implied by the theory of common ownership, $\kappa^{co}$, by deriving the off-diagonal elements from the cross-ownership of the two firms. Given a matrix of cooperation $\kappa$, firm $f$, which owns products $\mathcal{J}_f$, sets prices $\{p_{jt}\}_{j \in \mathcal{J}_f}$ equal to[10]

$$arg \max_{p_{jt} \,:\, j \in \mathcal{J}_f} \sum_{j \in \mathcal{J}_f} (p_{jt} - mc_{jt}) \cdot s_{jt}(\mathbf{p_t}) + \sum_g \kappa_{fg} \sum_{j \in \mathcal{J}_g} (p_{kt} - mc_{kt}) \cdot s_{kt}(\mathbf{p_t}),$$

yielding first-order conditions as in (3)

With a bit more notation we can write the system of first-order conditions in matrix form, which

[8] That $\omega_{jt}$ is additive with respect to $h_s$ is essential both here and in Berry and Haile (2014), but the additivity of $\omega_{jt}$ with respect to marginal costs is not.

[9] Alternatively, collusion might be incomplete, as in Miller and Weinberg (2017), who estimate a an internalization coefficient in the neighborhood of 0.3 for the MillerCoors joint venture.

[10] Two notes about this expression: first, $s_{jt}(\mathbf{p_t})$ is given by (4), suppressing the dependence on $(\mathbf{x_t}, \mathbf{v_t}, \xi_\mathbf{t})$. Second, for exposition we have held the ownership structure $\{\mathcal{J}_f\}_{\forall f}$ and matrix of cooperation ($\kappa$) fixed, but in practice this might vary over time, and can be subscripted by $t$.

is helpful for what follows. First, let $\Omega(\mathbf{p_t})$ denote the $J \times J$ Jacobian of the demand system (4) multiplied by -1, so that the $(j,k)$th entry is given by $-\frac{\partial s_{jt}}{\partial p_{kt}}$. Second, we extend the firm-indexed matrix of cooperation to a product-level matrix of profit weights, allowing for portfolio effects, by introducing the matrix

$$\mathcal{H}_{jk}(\kappa) \equiv \begin{cases} 1 & \text{if } (j,k) \in \mathcal{J}_f \text{ for any } f, \\ \kappa_{fg} & \text{if } j \in \mathcal{J}_f \text{ and } k \in \mathcal{J}_g \text{ for any } (f,g), \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Then, for a single market $t$, where $\odot$ denotes the element-wise (Hadamard) product of the two $J \times J$ matrices, we can rewrite the entire system of first-order conditions as:

$$\mathbf{s_t} = (\mathcal{H}_t(\kappa) \odot \Omega(\mathbf{p_t})) \cdot (\mathbf{p_t} - \mathbf{mc_t}). \tag{8}$$

And, inverting to group observables on the right-hand side,

$$\mathbf{mc_t} = \underbrace{\mathbf{p_t} - \underbrace{(\mathcal{H}(\kappa) \odot \Omega(\mathbf{p_t}))^{-1}\mathbf{s_t}}_{=\eta_\mathbf{t}(\cdot)}}_{\psi(\cdot)}. \tag{9}$$

This final expression clarifies two things, indicated by the braces. First, it expresses the standard derivation of markups ($\eta$) for models of conduct based on Bertrand price-setting, including own-firm profit maximization, common ownership, or collusion, simply by using the relevant matrix of cooperation. Second, it maps these models of conduct back to the general Berry and Haile (2014) conduct framework in (6) by taking the entire right-hand side to be $\psi(\cdot)$.

As a final note, observe from the derivation that the resulting markups depend not only on $\kappa$ but also all everything else in the model (via $\mathbf{p_t}$ and $\mathbf{s_t}$), including the demand shocks $\xi_\mathbf{t}$ and cost shocks $\omega_\mathbf{t}$ from (4) and (5), respectively.

### 3.2. Overview of the Test

Testing for conduct would be straightforward if we observed true marginal costs $\mathbf{mc_t}$. We could compare the predicted markups, from the above, to the true markups, given by $\mathbf{p_t} - \mathbf{mc_{jt}}$. This is the approach taken in Nevo (2001), which uses accounting data to form a measure of average markups. Unfortunately cost data are rarely available, and even when they are, we worry that they conflate economic and accounting costs. This limitation of the available data explains why identification of conduct testing is built on exclusion restrictions (Berry and Haile, 2014). From now on we will assume that the researcher observes $\mathbf{p_t}$ and $\mathbf{s_t}$, and has already estimated the demand

system that can be used to predict markups under alternative conduct assumptions.

**Motivation:** The researcher is considering two different models of conduct, indexed by $m \in \{0, 1\}$, each associated with a matrix of conduct $\kappa^m$ and predicted markups $\eta^m$. For exposition, suppose that model 0 is the correct model. For this model, equilibrium behavior relates the demand side of pricing to the supply side, represented by:

$$p_{jt} - \eta^0_{jt} = h_s(x_{jt}, w_{jt}) + \omega^0_{jt}. \tag{10}$$

In our notation, $x_{jt}$ are product-specific characteristics, $w_{jt}$ are cost shifters (e.g., input prices), and both are exogenous to $\omega^0_{jt}$, the idiosyncratic cost shock. Estimation and identification follow from the conditional moment restriction $\mathbb{E}[\omega^0_{jt}|\mathbf{z_t}] = 0$, where $\mathbf{z_t}$ includes all exogenous variables.

If the researcher estimates (10) for model 1, they will instead actually be estimating the following:

$$p_{jt} - \eta_{jt}(\kappa^1) = h_s(x_{jt}, w_{jt}) + \overbrace{\underbrace{\eta_{jt}(\kappa^0) - \eta_{jt}(\kappa^1)}_{\equiv \Delta\eta_{jt}(\kappa^1, \kappa^0)} + \omega^0_{jt}}^{\omega^1_{jt}}. \tag{11}$$

To be precise, under the "correct" conduct model $\mathbb{E}[\omega^0_{jt}|\mathbf{z_t}] = 0$, while under the incorrect conduct model $\kappa^1$, we have: $\mathbb{E}[\omega^0_{jt} + \Delta\eta_{jt}(\kappa^1, \kappa^0)|\mathbf{z_t}] \neq 0$. In other words, the misspecification in the conduct assumption "contaminates" the supply unobservable $\omega^1_{jt}$ with $\Delta\eta_{jt}(\kappa^1, \kappa^0)$. Much like the markup $\eta_{jt}(\cdot)$ itself, the difference in markups is a nonlinear function of all exogenous observables $\mathbf{z_t}$ as well as unobservables $(\boldsymbol{\xi_t}, \boldsymbol{\omega_t})$. This motivates our test to examine potential correlation between recovered cost shocks and differences in markups between conduct models.

**Previous Approaches:** Nearly all tests of conduct exploit the same conditional moment restriction $\mathbb{E}[\omega^0_{jt}|\mathbf{z_t}] = 0$ to form unconditional moment restrictions $\mathbb{E}[\omega^0_{jt} \cdot A(\mathbf{z_t})] = 0$. With a few exceptions, most of the literature compares goodness of fit measures (such as Likelihood Ratio statistics) for competing models in an overidentified GMM framework (Smith 1992). This approach is not without drawbacks: it requires a choice of instruments $A(\mathbf{z_t})$; it relies on a (fixed) weighting matrix and it is thus possible to reject model 0 in favor of model 1 and vice versa; and it amounts to a joint test of the specification of the markup (11) with a specification test for $h_s(x_{jt}, w_{jt})$.

Alternatively, one might consider a (Wald) test of whether the regression coefficient on $\Delta\eta_{jt}$ in (11) is equal to zero. Because both markups depend on $\boldsymbol{\omega_t}$, $\Delta\eta_{jt}$ is endogenous. The solution would be an IV estimator which replaces $\Delta\eta_{jt}$ with its "first stage" estimate: $\mathbb{E}[\Delta\eta_{jt}|\mathbf{z_t}]$. However, this approach does not directly allow for model comparisons and may reject all models.

**Our approach:** In our case, we use the model to both specify a scalar moment restriction – solving the weighting matrix issue – and choose the form of $A(\cdot)$. We adopt the non-nested testing framework of Rivers and Vuong (2002) to compare multiple models. This no longer requires that either model under consideration be the correct model, and instead lets us compare among two (potentially misspecified) models, as described further below.

Our test employs a criterion function that is based on detecting a correlation between the difference in markups between two models and the predicted cost shocks. In order to construct a test that the true model will pass, we require exclusion restrictions for our instruments $\mathbf{z_t}$. The motivating exercise above frames how we think about those exclusion restrictions, and where we should look for them. They must be exogeneous to $\omega_{jt}^0$, but they must also be relevant to the difference in markups $\Delta\eta_{jt}^{0,1}$ induced by the two models, which is the sense in which they "rotate marginal revenue" (Berry and Haile, 2014).

We see in (11) that the misspecification takes the form of the difference in markups between the two models, and so our test will exploit that information; in particular, the known nonlinearity of the model. For this reason, in testing between models 0 and 1 we build our exclusions following Newey (1990), using a first-stage prediction of $\Delta\eta_{jt}^{0,1}$. Our test is based on detecting violations of the following moment restriction:

$$\mathbb{E}[\omega_{jt}^m \cdot \mathbb{E}[\Delta\eta_{jt}^{0,1}|\mathbf{z_t}]] = 0. \tag{12}$$

An added advantage is that, by reducing the moments of interest to a singleton, it moots the problem of selecting a weighting matrix.

One last note: our motivation is based on the omitted variable $\Delta\eta^{0,1}$, which assumes that model 0 is correct, which may not be a model under consideration. Now assume we are comparing model 1 to an alternative model 2, that is also incorrect. To see why the logic of our example holds, suppose both models 1 and 2 are incorrect, but 0 is still correct. Then the omitted variable in (11) for model 1 is $\Delta\eta^{0,1} = \Delta\eta^{0,2} - \Delta\eta^{1,2}$. Similarly, the omitted variable for model 2 will be $\Delta\eta^{0,2} = \Delta\eta^{0,1} + \Delta\eta^{1,2}$. As model 1 approaches the truth, $\Delta\eta^{0,2}$ approaches $\Delta\eta^{1,2}$, relaxing the correlation between $\omega_{jt}^1$ and $\Delta\eta^{1,2}$. Meanwhile, as $\Delta\eta^{0,2}$, omitted variable in model 2, approaches $\Delta\eta^{1,2}$, the correlation between $\omega_{jt}^2$ and $\Delta\eta^{1,2}$ is exacerbated. A similar logic applies in the other direction as model 2 approaches the truth.

**Testing Procedure:** Let models $m \in \{1, 2\}$ be the two candidates, and $\mathbf{z_t}$ be a set of instruments satisfying $E[\omega_{jt}^0|\mathbf{z_t}] = 0$ (we will discuss finding such instruments in Section 3.3 below).

1. Estimate supply, as in (10), under models 1 and 2 to obtain residuals $\widehat{\omega_{jt}^1}$ and $\widehat{\omega_{jt}^2}$:

$$p_{jt} - \eta_{jt}^m = h_s(x_{jt}, w_{jt}) + \omega_{jt}^m.$$

2. Estimate $\Delta\eta_{jt}^{1,2} = g(\mathbf{z_t}) + \zeta_{jt}$ and predict $\widehat{\Delta\eta}_{jt}^{1,2}$.

3. For each candidate model, compute the value of the criterion function[11]

$$Q(\kappa^m) = (n^{-1} \sum_{j,t} \hat{\omega}_{jt}^m \cdot \widehat{\Delta\eta}_{jt}^{1,2})^2. \tag{13}$$

Finally, compute the test statistic,

$$T = \frac{\sqrt{n}(Q(\kappa^1) - Q(\kappa^2))}{\sigma} \tag{14}$$

where $\sigma^2$ is the asymptotic variance of the difference between the criterion function for the two models. We estimate the standard error of the difference $\hat{\sigma}/\sqrt{n}$ via a bootstrap procedure. Rivers and Vuong (2002) show that this test statistic is $N(0,1)$. The null hypothesis is that the two models are asymptotically equivalent with respect to the model selection criterion. A significant negative value is taken as a rejection in favor of model 1, and a significant positive value is taken as a rejection in favor of model 2.

To summarize, our procedure has some immediate advantages. The first is that steps (1) and (2) are functionally independent. The second is that because the resulting moment condition is scalar, it does not depend on how one specifies the weighting matrix. The third is that neither regression in steps 1 or 2 involve any endogenous variables, nor requires any interpretation of coefficients and can thus be done in a flexible semi-parametric (possibly machine-learning) manner. When we implement our test, we allow both $h_s(x_{jt}, w_{jt})$ and $g(\mathbf{z_t})$ to be random forest regressions, as discussed in more detail in Section 6.

### 3.3. Implementation of the Test

Here we consider the practical issues of implementation: what exclusion restrictions should be used in $z_{jt}$, prediction of intermediate objects $\{\widehat{\omega_{jt}^m}\}_{m=1,2}$ and $\widehat{\Delta\eta}_{jt}^{1,2}$, and the exact construction of the model comparison.

**Choosing $\mathbf{z_t}$.** The goal here is to ask: can we characterize all of the potential excluded variables from the marginal cost equation $\mathbf{z_t}$ that would be helpful in testing a model of conduct? While the set of potential excluded variables appears infinite, we can constrain our search by using the model itself to derive exclusion restrictions.

The obvious contents of $\mathbf{z_t}$ include the exogenous characteristics for demand (and excluded from supply) $\mathbf{v_t}$, those for supply (and excluded from demand) $\mathbf{w_t}$, and those regressors common

[11] At this step the researcher has some freedom; for instance, here we implement the criterion function directly as computation of the desired moment, but we might also have implemented it as the empirical likelihood that the moment holds. Our choice is informed mostly practical concerns for applied work, which are likely to involve both weighting and clustering (as ours does below).

to supply and demand $\mathbf{x_t}$. We use the bold notation to denote the matrix of characteristics for all $J_t$ products in the market to highlight the fact that we also utilize characteristics of other products $(\mathbf{x_{-j,t}}, \mathbf{v_{-j,t}}, \mathbf{w_{-j,t}})$.

Use of "demand-shifters" $v_{jt}$ forms the basis of the identification argument for supply in Berry and Haile (2014). These variables enter demand (4) but not supply (5), thus any effect $v_{jt}$ has on prices comes through markups (via higher residual demand) rather than marginal cost. Berry et al. (1995) exploited functional form differences so that miles-per-gallon shifted marginal cost $w_{jt}$, while miles-per-dollar shifted demand $v_{jt}$. A successful advertising campaign might generate a positive $v_{jt}$, while a product defect or recall might generate a negative $v_{jt}$.[12] If a promotion has an additional effect directly on demand beyond lower $p_{jt}$, then this might be valid as $v_{jt}$.[13] A special kind of "demand shifter" which enters the model nonlinearly rather than linearly are changes in the distribution of demographics across markets. That is, demand (and markups) for certain products may be higher in areas with high (low) income, but the marginal cost of serving consumers in these unlikely to be affected by demographics. In theory there are many such variables (more so than cost-shifters $w_{jt}$) and one could exploit seasonal or regional variations in demand as $v_{jt}$. Demand shifters for other products $\mathbf{v_{-j,t}}$ follow a similar argument, namely that demand and markups for $j$ are lower when $v_{kt}$ is high for close substitutes.

Another set of exclusion restrictions are $\mathbf{w_{-j,t}}$, the "cost shifters" of other products. In our cereal application, this might be the price of rice for Corn Flakes, or the price of corn for Rice Krispies. The most obvious explanation would be that a lower rival cost $w_{kt}$ increases the price-cost margin (and/or reduces the price) of the substitute. At the same time, the estimated marginal cost of Corn Flakes should not respond to the input prices for rice.[14]

The third set are the characteristics of other goods common to supply and demand, $\mathbf{x_{-j,t}}$. These are often referred to as "BLP instruments". The nested logit version of these instruments from Berry (1994) involve counting the number of competitors in each nest. Ghandi and Houde (2019) extend this idea to measure "closeness of competition" more generally. All things being equal, as the product space around $j$ becomes more crowded (more cereals with similar characteristics) we expect the markups for product $j$ to fall, while the characteristics of rival products should not directly impact the marginal costs of $j$.

A final set of potential instruments come from exogenous changes in conduct $\kappa_{fgt}$ itself. As

[12] On June 25, 2010, Kellogg's issued recalls 28 million boxes of Corn Pops, Honey Smacks, Froot Loops and Apple Jacks after customers complained of what the cereal maker called an "off-taste and smell" that caused "nausea and diarrhea" in some people. See `https://www.foodsafetynews.com/2010/06/kelloggs-recalls-some-popular-cereals/`. A more serious recall in September 2018 involved Honey Smacks and a potential Salmonella outbreak. `https://www.fda.gov/food/outbreaks-foodborne-illness/fda-investigated-multistate-outbreak-salmonella-mbandaka-infections-linked-kelloggs-honey-smacks`

[13] In contemporaneous work, Michel and Weiergraeber (2018) try to exploit the non-price demand shift effects of in store promotions to identify conduct.

[14] This mimics the well-known reduced form test of collusion. Conditional on a competitor's prices, own prices should not respond to changes in rival marginal costs which makes it a valid exclusion restriction.

an example, Miller and Weinberg (2017) use a binary instrument for "treated firms" in the post-joint venture period as an instrument for conduct $\kappa$. Another (somewhat obvious) example is that mergers may change $\kappa_{fgt}$ such as Post's acquisition of Malt-O-Meal in 2015.

Exogenous changes in financial ownership may also serve as an instrument for $\kappa_{fgt}$. This is the instrumental variables strategy in several reduced-form studies of common ownership. Inclusion in a stock market index is likely to increase overlap among investors; and serves as the basis for the identification strategy in Boller and Scott Morton (2019). Azar et al. (2018a) look before and after the BlackRock-Barclays acquisition with the implied assumption that it shifts $\kappa_{fgt}$. The challenge with using $\kappa_{fgt}$ restrictions is that they are often "event studies." We could use less discrete changes in $\kappa_{fgt}$ since it continuously varies period by period, however there is a question whether the variation is substantial enough and whether managers are fully aware of small changes in common ownership.

We should also mention what are *not* valid exclusion restrictions. The variables $(x_{jt}, w_{jt})$ are not *excluded*, but *included* and help to reduce the residual variance in $\omega_{jt}$. An example of a potentially excluded variable is the unobservable demand shifter $\xi_{jt}$. If we do exclude $\xi_{jt}$ we are implicitly assuming $Cov(\xi_{jt}, \omega_{jt}) = 0$.[15] Another example are the so-called "Hausman instruments" which are prices of the same product in other markets $p_{j,-t}$. The motivation for these instruments is that they capture unobserved variation in marginal costs that is common across markets such as unobserved shocks to input prices. The challenge is that it is difficult to identify changes in conduct $\kappa_{fg}$ separately from unobservable common shocks to marginal costs $\omega_{jt}$.[16] A stronger (than necessary) view is that anything not explicitly in the model of supply and demand $(\mathbf{x_t}, \mathbf{v_t}, \mathbf{w_t}, y_t)$ is not a valid instrument, thus any variable which shifts neither demand nor supply nor conduct $\kappa$, is not relevant, and thus is not a valid instrument.

**Obtaining $\widehat{\omega_{jt}^m}$ and $\widehat{\Delta\eta}_{jt}^{1,2}$.** If marginal costs are constant, then $\widehat{\omega_{jt}^m}$ can be obtained directly via nonparametric regression. Otherwise, the supply equation 5 becomes a function of quantity, or $h(x_{jt}, w_{jt}, q_{jt})$, and we require an additional exclusion restriction (above and beyond what we include in $z_{jt}$).

The researcher has two options for obtaining $\widehat{\Delta\eta}_{jt}^{1,2}$. Our approach is to estimate the relationship nonparametrically. An alternative would be to estimate the entire model—supply and demand—and simulate unobservables in order to construct the expectation. This choice is analogous to the problem of constructing optimal instruments for nonlinear models. Berry et al. (1995) propose the simulation method in demand estimation, while Newey (1990) and Ai and Chen (2003) develop

[15] This forms the basis of the identification argument in Miller and MacKay (2018), though it may be a stronger assumption than we are willing to impose in our context. Indeed, even $Cov(\xi_{jt}, \omega_{kt}) = 0$ may be too strong an assumption and we may not want to use $\xi_{-j}$ as instruments either.

[16] What looks to authorities like collusive behavior is explained by defendants as a "correlated shock to marginal costs". The "Hausman instruments" can be perfectly valid as instruments for demand $E[\xi_{jt}|x_{jt}, w_{jt}, v_{jt}, p_{j,-t}] = 0$ but are invalid as instruments for supply $E[\omega_{jt}|x_{jt}, w_{jt}, v_{jt}, p_{j,-t}] \neq 0$.

methods for nonparametrically approximating the optimal instrument. In our setting the nonparametric approach is rather more direct, and avoids the imposition of distributional assumptions on the unobserved demand and cost shocks.

**Comparing Models.** The main choice is whether to use a Cox or a Vuong testing framework. The differences, which stem from the definition of the null hypothesis, are mostly philosophic but have some practical consequences.

In the Cox testing framework (and variants), one model is taken to be the null hypothesis, and the the other model the alternative. The drawback of this setting is that the results are only well-defined when the consideration set—the two models—includes the truth.

In the Vuong testing framework, by contrast, the null hypothesis is that both models are equally distant from the truth. The alternative hypotheses are that one model (or the other) are closer to the truth. The advantage of this approach is that it does not require either candidate to be the true model (see Duarte et al. (2020) for further discussion of this point in the context of conduct testing). The drawback of this setting is a degeneracy issue when the models are not functionally independent; this required, in the original Vuong (1989), the use of a pre-test, with all of the attendant implications for validity. Recent work by Shi (2015) and Shennach and Wilhelm (2017) offers workarounds for this problem, which can also be directly avoided by sample-splitting (Yatchew, 1992; Whang and Andrews, 1993).

We adopt our testing framework from Vuong (1989), as extended by Rivers and Vuong (2002). Here the null hypothesis is that both conduct models are equally distant from the truth, as measured by the criterion function above.

### 3.4. Relationship to Internalization Parameters and Optimal IV

An alternative approach is to incorporate an "internalization parameter" $\tau \in [0, 1]$ which provides a scalar measure of the degree to which cooperative incentives are reflected in the data. In this case, $\tau = 1$ would imply that incentives are "fully internalized" while $\tau = 0$ would suggest that incentives are not internalized (or not present) at all. As an example of this approach in a vertically integrated setting, Crawford et al. (2018) estimate the extent to which a commonly owned regional sports network accounts for its effects on the profitability of a cable company to be $\tau = 0.79$ (where $\tau = 1$ would represent full joint profit maximization of the vertical channel).

We can extend the cooperation matrix from (15) to incorporate an internalization parameter $\tau$:

$$\mathcal{H}_{jk}(\kappa, \tau) \equiv \begin{cases} 1 & \text{if } (j,k) \in \mathcal{J}_f \text{ for any } f, \\ \tau \cdot \kappa_{fg} & \text{if } j \in \mathcal{J}_f \text{ and } k \in \mathcal{J}_g \text{ for any } (f,g), \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

This allows us to "convexify" the cooperation matrix so that $\tau = 0$ corresponds to the multi-product

differentiated Bertrand conduct, and $\tau = 1$ corresponds to conduct under the common ownership hypothesis.[17]

One can estimate $\tau$ as an additional parameter in (10) via nonlinear instrumental variables using the conditional moment restriction:[18]

$$\mathbb{E}[\underbrace{p_{jt} - \eta_{jt}(\kappa, \tau) - h_s(w_{jt}, x_{jt})}_{\omega^*_{jt}} | \mathbf{z_t}] = 0. \tag{16}$$

The optimal instrument for $\tau$ in (16) in the sense of Chamberlain (1987) is given by $\mathbb{E}\left[\frac{\partial \eta_{jt}}{\partial \tau} | \mathbf{z_t}\right]$ and the optimal instrument for $\kappa_{fg}$ (when treated as a free parameter) is given by $\mathbb{E}\left[\frac{\partial \eta_{jt}}{\partial \kappa_{fg}} | \mathbf{z_t}\right]$. We provide derivations and analytic expressions for both of these objects in Appendix A.3. The moment condition that forms the basis of our test in (12) uses the expectation of the discrete markup difference between two conduct models $\mathbb{E}\left[\Delta\eta^{1,2}_{jt} | \mathbf{z_t}\right]$. This highlights the connection between the "menu approach" to conduct testing (Nevo, 1998) and the "conduct parameter" approach (Bresnahan, 1982). When the difference between the two markups under consideration becomes small (and can be parametrized by $\tau$ or $\kappa_{fg}$), then the two expressions will coincide $\frac{\partial \eta_{jt}}{\partial \tau} \approx \frac{\eta_{jt}(\tau) - \eta_{jt}(\tau + \epsilon)}{\epsilon}$.

In this sense, for local comparisons, our testing procedure constructs instruments that are equivalent to the Chamberlain (1987) optimal instrument. Moreover, Step 2 of our testing procedure constructs the feasible approximation by nonparametric regression of $\Delta\eta_{jt}$ on the excluded instruments, just as Newey (1990) suggests for the estimation of the computation of optimal instruments. In this sense we think of our approach to conduct testing in Section 3.2 above as being a discrete analogue to the optimal IV in conduct parameter estimation.[19]

Generally speaking, the "conduct parameter" or "internalization" approach from (16) is not our preferred approach. One could specify $h_s(x_{jt}, w_{jt})$ as a partially linear model, but the resulting nonlinear IV problem is still challenging.

### 3.5. Relationship to Prior Testing Procedures

Conduct testing with discrete hypotheses, such as in Nevo (1998), has a long history in industrial organization. Much of the work in this literature has has employed criteria based on minimizing the errors in the fit of the supply function from (10) Bresnahan (1987); Gasmi et al. (1992); Bonnet

[17] Applying this approach to the airline industry (similar to Azar et al. (2018a)) using a nested logit demand system, Kennedy et al. (2017) estimate $\hat{\tau}$ to be less than zero, while Park and Seo (2019) estimate $\hat{\tau}$ to be close to 1 using the demand system of Berry and Jia (2010) which differentiates between business and leisure travelers.

[18] Note that $\eta_{jt}(\tau)$ is based on the same inversion expressed in (9).

[19] In private correspondence, the authors of Duarte et al. (2020) have made the superb observation that for a sufficiently flexible sieve in the set of instruments, denoted $\tilde{z}_{jt}$, their preferred test, based on the moments $\mathbb{E}[\omega_{jt} \cdot \tilde{a}(\mathbf{z_t})] = 0$ is equivalent to our own. We thank them for noting this point, which further highlights the relationship of our test to the optimal instruments. In the optimal IV setting, this argument corresponds to Proposition 2 of Chamberlain (1987), which observes that one can approximate the optimal instruments by including a sufficiently flexible polynomial expansion of ones instruments.

and Dubois (2010). However, the observation that markups are endogenous to cost shocks led subsequent work to take an exclusion restriction approach. Our work is closest to two prior testing procedures in that vein.

The first is the framework of Villas-Boas (2007). It uses the reduced form version of our exclusion, $\mathbb{E}[\omega_{jt}^k | z_{jt}] = 0$. Models are compared using a Cox testing procedure which is extended to GMM objective functions by Smith (1992).[20] The main difference between our procedure and that of Villas-Boas (2007) is that we make use of the known form of the misspecification introduced by using the wrong conduct model, as in (11). This has two advantages: First, it makes transparent what kinds of instruments are likely to be useful in distinguishing between models: we require instruments that manipulate the difference between the markups between two conduct models. Second, we expect it to be more powerful because it uses the model; in particular, it exploits the nonlinearity in the relationship between the excluded variables and markups.

The second point of comparison is what we refer to as the "Pakes-Wollmann regression", introduced in Pakes (2017). The idea here is to run the following regression:

$$p_{jt} = h_s(x_{jt}, w_{jt}) + \gamma \eta_{jt}^m + \omega_0, \tag{17}$$

and, since markups ($\eta_{jt}^m$) are endogenous, to use BLP instruments as exclusions to estimate $\gamma$. This last point is the novel insight that we borrow for our testing procedure. Pakes (2017) predicts that, for the true conduct model, we should obtain $\hat{\gamma} = 1$, and indeed he does using data and the demand system from Wollmann (2018). However, this framework does not permit the comparison of alternative models. In fact, if we interpret $\gamma$ as a LATE, it is straightforward to construct perverse examples that a) yield an estimate of 1, despite being far from Bertrand pricing, b) yield an estimate other than 1 at the correct markup (e.g. in the case of constant fractional markups), and c) the distance from 1 does not have a clear interpretation.[21] We replicate the Pakes-Wollmann regression in our setting and discuss this point further in Appendix E.2.

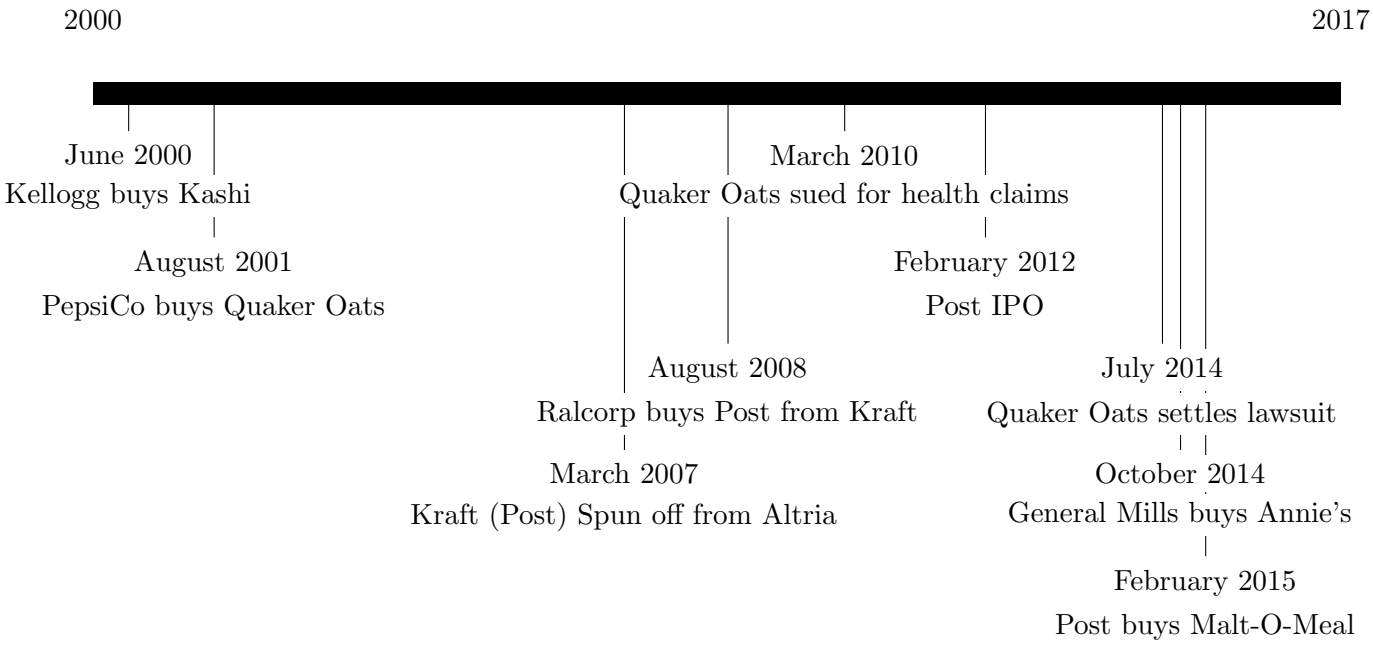
## 4. Ready-to-Eat Cereal

We focus our empirical exercise on the Ready-To-Eat (RTE) cereal industry for the period 2007–2016 in five major designated market areas. We chose this industry for a number of reasons. The first reason is that the industry is highly concentrated with four major players: Kellogg's, General Mills, Quaker Oats, and Post are responsible for approximately 85% of the overall market share.

[20] Duarte et al. (2020) also propose to implement the exclusion restrictions directly, using a Rivers-Vuong test instead of a Cox test, and shows that it performs well in Monte Carlo simulations

[21] For a particularly perverse example, consider the constant fractional markup case. If we compute markups, at fraction $x$ of marginal costs, to be $p \cdot x/(1+x)$ and the instruments for $p$ are relevant, then we will mechanically obtain $\hat{\gamma} = (1+x)/x$. For example, if $x = 0.5$, the correct estimate is $\gamma = 3$. Sending the hypothesized $x \to \infty$, we can get arbitrarily close to 1. However, if the model of constant fractional markups is correct, then BLP instruments should be weak in the first stage. Therefore, we require that the constant fractional markups model is *incorrect* to obtain the result that the $\hat{\gamma} \to 1$ as the hypothesis $x \to \infty$.

Figure 1: Chronology of RTE: 2000 – 2017




The second reason is that there are substantial differences in ownership patterns across firms. For historic reasons, Kellogg's has large undiversified shareholders while the other firms generally do not. In addition there are a substantial number of transactions both in the ownership space and the product space, particularly involving Post Brands which at various times is a component of the S&P 500 Index, the S&P 400 Midcap Index, and no index at all. The final reason is that there is substantial prior work on the RTE cereal industry which indicates that static Bertrand-Nash in differentiated products appears to be a reasonable empirical framework. Figure 1 highlights some important events in the RTE cereal market for the years 2000-2017.

### 4.1. The Big Four

**Kellogg's:** Kellogg's was founded in 1906 by brothers W.K. and J.H. Kellogg, who developed a technique for producing corn flakes that was originally meant for in-house use at the elder brother's Battle Creek Sanitarium. They believed that such a bland and tasteless food, which they called "granula," would reduce impure desires.

Over our sample period the holdings of Kellogg's are relatively stable, however there is one feature that differentiates them from the other three: approximately 20-30% of Kellogg's shares are held by the (entirely undiversified) W.K. Kellogg Foundation Trust. W.K. Kellogg, the longer-lived brother, established the foundation to ensure that "all children — regardless of race or income — have opportunities to reach their full potential," a marked departure from his elder brother's views, embodied by the Race Betterment Foundation, a now–defunct pillar of the also–defunct eugenics

movement. The second largest shareholder is the Gund family which acquired its stake in Kellogg's after selling the decaffeinated coffee brand Sanka to Kellogg's in 1927.

Kellogg's was a member of the S&P 500 over our entire sample and has around 30% market share. Well-known products include Kellogg's Corn Flakes, Fruit Loops, Rice Krispies, Raisin Bran, and Special K. In addition to RTE cereals, Kellogg's also sells other morning foods (e.g., Eggo Waffles, Pop Tarts) and snack foods (e.g., Pringles, Cheez–Its).

**General Mills:** General Mills' history goes back to the Minneapolis Milling company, founded in 1856. It's main RTE cereal innovation came in 1930, when it developed a technology for "puffing" cereal, as in Kix and Cheerios.

Holdings of General Mills are stable over our sample period, but they did acquire Annie's, a health-conscious brand, in October of 2014. They were included in the S&P 500 for the duration of our sample and have around 30% market share.

Well-known products include Cheerios, Chex, Lucky Charms, Total, and Wheaties. Outside of morning foods, they also own Betty Crocker brands, Pillsbury, Nature Valley, Hamburger Helper, Yoplait, and a variety of other food products.

**Post:** Like Kellogg's, Post also traces its history to the Battle Creek Sanitarium, where founder C.W. Post was a patient. The inaugural cereal, Grape–Nuts, was developed in 1897 under the pretense that it would cure appendicitis. The Postum Cereals company changed its name to the General Foods company in 1929, and was purchased by Phillip Morris (later rebranded as Altria) in 1985 and merged with Kraft in 1989, where it stayed until the beginning of our sample.

Post underwent three major ownership changes during our sample. First, on March 31, 2007, Kraft (which held Post) was spun off from Altria. Next, announced in November 2007, Kraft spun off Post Cereals and the resulting company was sold to Ralcorp Holdings on August 4 of 2008. This transition meant that Post left an S&P 500 company and was now owned by a non-S&P 500 company. Ralcorp Holdings is also a major producer of private label cereals as well as other food products. Finally, announced in July 2011, Ralcorp holdings announced an IPO for the Post Foods Unit, which was successfully spun off on February 7, 2012. The resulting company was not a member of the S&P 500.

Well-known products include Grape–Nuts, Honey Bunches of Oats, and Raisin Bran. In February 2015, Post purchased Malt–O–Meal, a major producer of private label cereals which comprised about 8–9% of the overall market.

**Quaker Oats:** Quaker Oats is the result of a four–way merger of Midwestern oat mills in 1901. The brand has no affiliation with the Religious Society of Friends (actual Quakers), who regularly display their annoyance at the representation with letter-writing campaigns.

In August of 2001 Quaker Oats was purchased by PepsiCo, an S&P 500 company, and it remained in their portfolio for the duration of our sample. From March of 2010 until a settlement in July of 2014, Quaker Oats was subject to a long and public legal battle over the veracity of their health claims. They did not claim that their products cured appendicitis or moral impurity; none the less, the legal battle may have contributed to a decline in sales.

Well-known brands include Cap'n Crunch, and Life and Quaker Oats has around 8–9% of the cereal market. The company also produces other morning foods (Oatmeals, and Aunt Jemima branded foods) as well as other food products, but should be considered in the larger PepsiCo setting, where it makes up a relatively small fraction of sales (2-3%).

### 4.2. Data Sources: Ownership Data

We are interested in the following firms, which at some point in 2004-2017 offered products in the ready-to-eat cereal market: General Mills (GIS), Kellogg (K), Kraft (KRFT: Q1 2013 - Q2 2015), Mondelez (MDLZ), Altria Group (MO), PepsiCo (PEP), Philip Morris (PM: Q4 2008 - Q2 2017), Post Holdings (Q1 2012 - Q2 2017), and Ralcorp (RAH: Q1 2008 - Q4 2011). We use the novel dataset of institutional holdings developed in Backus et al. (2021), and make a small number of corrections to address potential double counting of large private holdings, as described in Data Appendix C.

Table 1 provides summary statistics on the common ownership data. There are some important patterns to point out. The first is that Vanguard appears to be increasing its holdings across all firms over time. In part this is driven by the growing share of Vanguard within the index fund market. The second is that between 2004 and 2010 there is a reallocation from Barclays Global Investors (BGI) and Blackrock which acquired the BGI exchange-traded-fund (ETF) business in June of 2009. This made Blackrock the largest player in the ETF market.[22] State Street is another large player in the ETF market and also sees their ownership stakes increasing over time. Another large player is FMR LLC, which is better known as the financial entity behind Fidelity which is a major player in both actively managed and index funds. Capital Research, the parent of company of the American Funds family (primarily actively managed) is another major player particularly in the early periods. We provide a more detailed accounting of ownership stakes over time by major investors in the Appendix.

We continue by computing the common ownership profit weights as described in (1). These are depicted in Figure 2 for each of the four major firms in our sample. We make the proportional control assumption of $\gamma_{fs} = \beta_{fs}$ for this figure, although alternatives are possible. For example, the top right pane shows the implied weight that Kellogg's puts on the profits of their competitors. Notice that the weight Kellogg's puts on its own profit is normalized to one and constant over time. The weights are similar across competitors and slowly growing over time from around 8%

[22] Azar et al. (2018a) use this event as an instrument for changes in ownership as it substantially increases the holdings of Blackrock.

Table 1: Top 5 Owners of Major Firms, 2004-2016

**General Mills (GIS)**

| 2004 | | 2010 | | 2016 | |
|---|---|---|---|---|---|
| Capital Research and Management | 7.28% | BlackRock, Inc | 8.70% | BlackRock, Inc | 7.36% |
| Barclays Global Investors | 3.24% | State Street Global Advisors | 5.92% | The Vanguard Group | 6.92% |
| Wellington Management Group | 3.06% | The Vanguard Group | 3.56% | State Street Global Advisors | 6.14% |
| State Street Global Advisors | 2.48% | MFS | 2.65% | MFS | 3.37% |
| The Vanguard Group | 1.95% | Capital Research and Management | 2.43% | Capital Research and Management | 2.12% |

**Kellogg (K)**

| 2004 | | 2010 | | 2016 | |
|---|---|---|---|---|---|
| W.K. Kellogg Foundation | 29.87% | W.K. Kellogg Foundation | 22.94% | W.K. Kellogg Foundation | 19.75% |
| Gund Family | 7.26% | Gund Family | 8.65% | Gund Family | 7.68% |
| Capital Research and Management | 2.83% | Capital Research and Management | 3.54% | The Vanguard Group | 4.97% |
| Barclays Global Investors | 2.81% | BlackRock, Inc | 2.97% | BlackRock, Inc | 4.64% |
| W.P. Stewart & Co. | 2.63% | The Vanguard Group | 2.42% | MFS | 3.51% |

**Quaker Oats, a Unit of PepsiCo (PEP)**

| 2004 | | 2010 | | 2016 | |
|---|---|---|---|---|---|
| Barclays Global Investors | 4.40% | BlackRock, Inc | 4.64% | The Vanguard Group | 6.72% |
| State Street Global Advisors | 2.81% | Capital Research and Management | 4.37% | BlackRock, Inc | 5.63% |
| FMR LLC | 2.74% | The Vanguard Group | 3.64% | State Street Global Advisors | 3.98% |
| The Vanguard Group | 2.08% | State Street Global Advisors | 3.19% | Wellington Management Group | 1.48% |
| Capital Research and Management | 1.82% | Bank of America | 1.63% | Northern Trust | 1.37% |

**Post Brands, a Unit of Altria (2004, MO), Ralcorp (2010, RAH), and Post Holdings (2016, POST)**

| 2004 | | 2010 | | 2016 | |
|---|---|---|---|---|---|
| Capital Research and Management | 7.37% | FMR LLC | 10.18% | Wellington Management Group | 9.63% |
| State Street Global Advisors | 3.61% | BlackRock, Inc | 8.35% | BlackRock, Inc | 8.42% |
| Barclays Global Investors | 3.51% | The Vanguard Group | 3.57% | FMR LLC | 7.24% |
| FMR LLC | 2.60% | Baron Capital Group | 3.39% | The Vanguard Group | 6.93% |
| AllianceBernstein L.P. | 2.25% | Steinberg Asset Management | 2.68% | Tourbillon Capital Partners | 6.89% |

to 20%. These relatively small weights are due to the large undiversified Kellogg's shareholders (Kellogg Family Foundation and Gund Family). Contrast this with General Mills in the top left. General Mills places between 60-80% weight on the profits of Quaker Oats and Post as it does on its own profits, with substantial variation across time. It places slightly less weight on the profits of Kellogg's because of less overlapping ownership, though still more weight (40-60%) than Kellogg's places on the profits of General Mills. Quaker Oats (a division of PepsiCo) occasionally places more weight $\kappa > 1$ on competitor's (General Mills and Post) profits than it does on its own profits.[23] Quaker Oats puts somewhat less weight (though still $\kappa > 0.6$ on the profits of Kellogg's which has less overlap in ownership. Post generally puts less weight on each of its competitor's profits over time as Post is transitions from an S&P 100/500 component, to an S&P 400 Midcap Index Component, and briefly after its 2012 IPO is not included in any index, before rejoining the S&P 400 Midcap Index.

One advantage of RTE cereal is that there is a large amount of useful variation in $\kappa$. In Backus et al. (2021) we showed that $\kappa_{fg} \neq \kappa_{gf}$ when firms $f$ and $g$ differ in their *investor concentration*, as measured by $IHHI \equiv \sum_s \beta_{fs}$. This is on display in Figure 2. For instance, Kellogg's puts much lower weight on General Mills than vice versa, and this is in large part due to the fact that General mills has more unconcentrated ownership, driven in part by a relatively high retail share, while Kelloggs has concentrated ownership, driven by a modest retail share and the presence of the

[23]This is consistent with the observation of Backus et al. (2021) that common ownership weights are higher in firms with a greater retail share. Both General Mills and PepsiCo, but particularly the latter, have high retail shares.

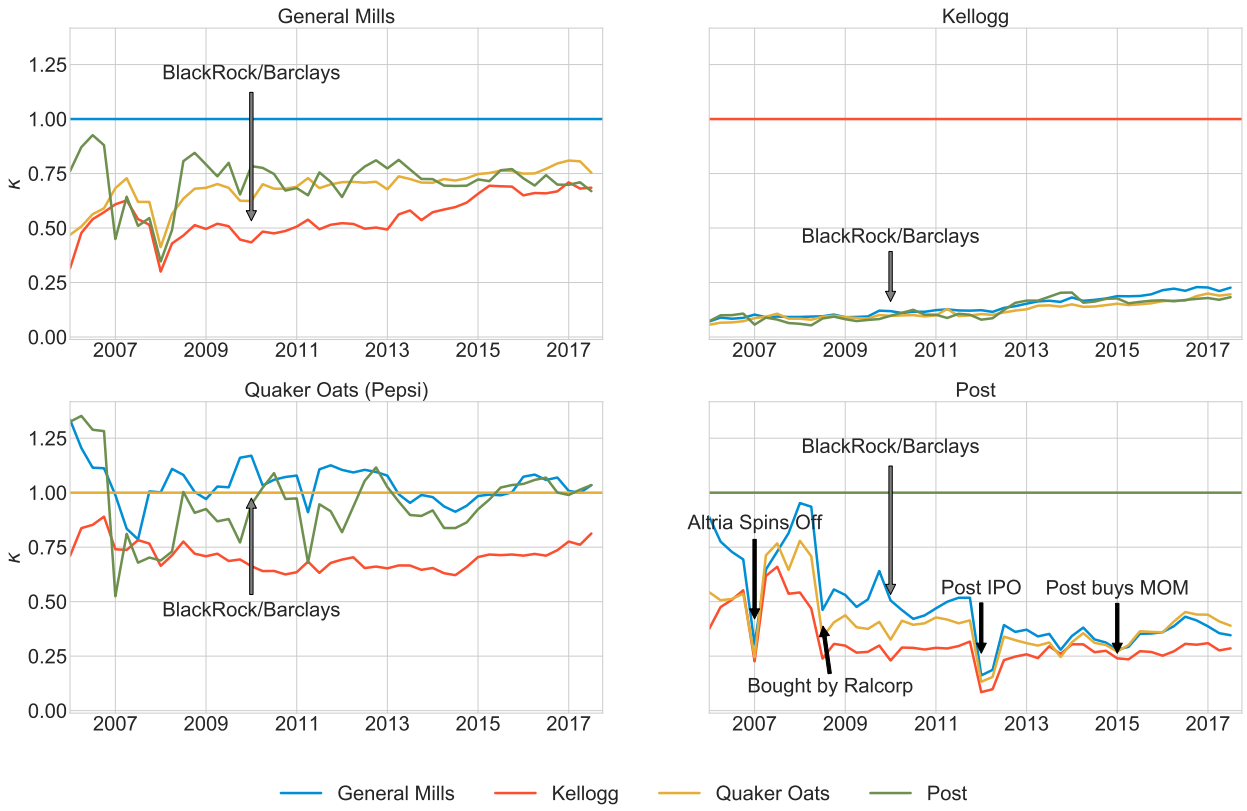



Figure 2: $\kappa$ Profit Weights for Ready-To-Eat Cereal (Proportional Control)

Kelloggs Foundation. One disadvantage of $MHHI$, is that providing a single aggregate measure for all firms in the market, it obscures these asymmetric relationships.

### 4.3. Data Sources: Sales, Product, and Input Price Data

Our primary data source for unit sales and prices of ready-to-eat (RTE) cereal comes from the Kilts Nielsen Scanner Dataset. The data are organized by store, week, and UPC code. For each store–week–upc we observe unit sales as well as a measure of "average price" which is revenue divided by sales.[24] Because there is little price variation across stores within the same chain (DellaVigna and Gentzkow, 2019), we aggregate unit sales and revenues to the DMA-chain level.[25] We focus on the years 2007-2016.[26] We further restrict our attention to a set of six DMAs (cities): Boston, Chicago, Charlotte, Denver, Phoenix, and Richmond. We choose these six cities to provide some geographic diversity, and also because Nielsen reports high coverage of "all-commodities-volume" for supermarket sales in these DMAs. Within these cities, we focus exclusively on the set of conventional supermarket sales (which Nielsen labels as "F" stores) and exclude convenience stores and pharmacies which sometimes sell RTE cereal ("C" or "D" stores) and mass- merchandise ("M" stores). We further restrict our attention to a set of DMA-chains where we observe a sufficiently high volume of individuals from the Nielsen Panelist Dataset as customers.[27] This allows us to construct chain-specific demographic profiles of customers at the DMA-chain-year level rather than just by geography. We focus on two key demographics (income and presence of children) that were previously shown to be important in demand for RTE cereal (Nevo, 2000).[28]

The data are recorded at the level of a universal product code (UPC) with around 3500 unique UPCs, we consolidate multiple package sizes at the "brand level", where our definition of brand also includes flavor (Honey Nut Cheerios or Blueberry Frosted Mini-Wheats).[29] Consistent with Nevo (2000), we convert boxes of cereal into serving equivalents, and report prices and quantities on a per-serving basis.[30] Given this definition, we find that in a DMA-Chain-Week a typical consumer

[24]Prices may vary within a UPC–store–week for a number of reasons: the first is that price changes may occur within the middle of the reporting week, the second is that some consumers may use coupons; according to the Kilts-Nilsen documentation, retailer coupons or loyalty card discounts are included in the "average price" calculation while manufacturer coupons are not.

[25]We provide supporting evidence in favor of chain-level pricing in Appendix B.9.

[26]We exclude the year 2006 because the set of stores observed in that year does not sufficiently overlap with stores observed in subsequent years.

[27]This causes us to lose less than 4.3% of overall sales (as measured in servings).

[28]We've examined other characteristics such as race, but this often required slicing the Panelist data too thinly to observe differential patterns in cereal preferences.

[29]This is especially important because a large number of individual UPCs are associated with a single "brand" and new UPCs might represent new packaging for a movie-tie-in or small changes in product size. If a manufacturer changes the package size from 14oz to 12.5oz and keeps the price fixed, we don't want to misinterpret this as "new product" but rather we want to interpret this as a serving price change for an existing product. This causes us to "miss" some nonlinear pricing where "Family Packs" are often less expensive on a per serving basis than smaller boxes. Focusing exclusively on conventional supermarkets mitigates this somewhat, but not completely.

[30]Nutrient dense cereals generally display smaller serving sizes (by weight) which lead to much smaller serving sizes by volume. There is some evidence that serving sizes are chosen so that caloric content falls in the (100-150)

chooses among 100-250 unique "products".

We report some basic summary statistics for our main dataset in Table 2. We observe 1590 stores which we consolidate into 26 DMA-chains, over 522 weeks between 2007-2016. The average price per serving to be around 20 cents across our sample, though it varies substantially across products, chain-DMAs, and time. We find that at the DMA/city level, aggregate sales are between 1.5 and 8.5 million servings per week, with a typical box containing around 12-20 servings. We find that between 80%-90% of sales (as measured in servings) are to branded cereals, with the remainder being private label or store-brands.[31] We assign each product to its ultimate owner in each period, taking to account mergers and acquisitions.[32]

| | Boston | Charlotte | Richmond | Chicago | Denver | Phoenix |
|---|---|---|---|---|---|---|
| # Chains | 4 | 4 | 2 | 7 | 3 | 6 |
| # Stores | 257 | 296 | 98 | 366 | 221 | 352 |
| Nielsen ACV Coverage | 83% | 86% | 81% | 65% | 86% | 84% |
| # Products (chain-week) | 216 | 161.6 | 179.2 | 179.7 | 205.6 | 166.9 |
| Servings/ Week (millions) | 7.61 | 3.26 | 1.52 | 8.47 | 5.5 | 6.92 |
| Price Per Serving (cents) | 20.89 | 20.7 | 19.4 | 20.14 | 20.13 | 18.59 |
| Servings Per Box | 13.81 | 14.13 | 14.28 | 14.09 | 14.92 | 16.41 |
| % Branded | 0.82 | 0.85 | 0.81 | 0.92 | 0.81 | 0.81 |
| $PCA_0$ | -0.26 | 0.13 | 0.18 | 0.19 | -0.21 | 0.14 |
| $PCA_1$ | -0.07 | 0.04 | 0.1 | 0.03 | -0.01 | 0.05 |
| $PCA_2$ | 0.15 | -0.06 | -0.07 | -0.02 | -0.04 | -0.09 |
| % of HH with kids | 0.19 | 0.18 | 0.22 | 0.23 | 0.19 | 0.16 |
| Mar 2009 Unemployment | 7.02 | 11.65 | 7.09 | 9.58 | 6.97 | 8.52 |
| Mar 2016 Unemployment | 3.77 | 4.85 | 4.1 | 6.12 | 3.29 | 4.84 |
| Median Income (07-09) | 56,650 | 49,905 | 67,206 | 62,681 | 57,730 | 49,885 |
| Median Income (10-16) | 58,080 | 51,373 | 61,595 | 63,796 | 58,774 | 50,669 |

Table 2: Summary Statistics of Sales and Demographic Data by City

Source: Nielsen Retailer Scanner Data: Stores and Sales, Nielsen Consumer Panelist Data: Income and Presence of Children, Nutritionix: Product Characteristics, Serving Size, FRED: Unemployment.

We augment the Nielsen dataset with nutritional information from the Nutritionix Database. This database is organized by UPC code and was designed to provide API access for various fitness tracking mobile apps. It encodes the nutritional label on the product packaging (serving size, calories, sugar, fat, vitamin content, ingredient lists). We merged the Nielsen UPC information with the nutritional label information from Nutritionix. A large number of products in our dataset (and 9-10% by volume) are private-label brands. For these products, we do not have UPC codes which we can match to the Nutritionix database.[33] Instead, we must match these products to

range rather than measuring typical serving sizes by consumers. [CITE:CR] conducted a survey and found that 92% of consumers exceed the posted serving size when pouring bowls of cereal. The average overpour on Cheerios was 30%-130% while it was even greater for denser cereals like Museli or Granola where the average overpour was 282%.

[31] We provide additional detail on how the private label share varies across chains, DMAs, and with the business cycle in Appendix B.4, see Figure B-3.

[32] For example, prior to September 2014 we assign Annie's Homegrown products as belonging to Annie's Homegrown and independent firm, and after September 2014, we assign them to General Mills.

[33] UPCs are obfuscated by Nielsen to prevent researchers from de-identifying stores. The true UPC code could identify the product as *Chain X Brand Honey Toasted O's*.

the most similar branded product (for example, *HNY TSTD O'S* to *Honey-Nut Cheerios*) and use the nutritional information from the branded product. For some private label products we cannot identify the most similar brand (e.g. *CTL BR C-M-C RTE*), rather than dropping these products we impute product characteristics using averages.

Because there are a large number of product characteristics (23), and we aren't interested in nutritional aspects of products per se, we consolidate the nutritional information into a number of principal components. The idea is to reduce the dimension of the characteristic space while preserving the variation across products. This lets us measure which products are more (or less) similar based on nutritional content. We elaborate on this process in Appendix B.2. Each component is designed to have a mean of zero, and a variance of one. An additional advantage is that these principal components form an orthogonal basis which aids in estimation of our random coefficients model. We provide a map of the product space as defined by the first two principal components in Figure 3. Using only nutritional information (and nothing about substitution), it appears to do a good job of separating the "kids" cereals from "adult" cereals and "healthy" cereals from "less healthy" cereals. As Table 2 indicates, there appears to be more variation in product assortment across DMAs in the "kids vs. adults" dimension than the others. If our goal in estimating counterfactuals were to alter these product characteristics, we would not be able to with a demand system based on these principal components, but that is clearly not the goal of this analysis.

We construct the distribution of consumer demographics to be representative of the consumers shopping at a particular DMA-chain-year which is somewhat different than in the previous literature.[34] We use the Nielsen HMS (household) dataset and construct a sample of all households who visited a particular DMA-chain in that year. For each panelist we measure their income and whether or not they have children. We weight each panelist by their annual number of trips to each DMA-chain, with the idea being that more frequent shoppers at a store receive more weight.[35] To smooth things out, we estimate two lognormal distributions of income for each DMA-chain-year, one for households with children and one for households without.[36] We report these overall averages in Table 2 where 18-25% of households report children at home, and the median income varies from around $50,000 in Phoenix to around $67,000 in Chicago. In addition to consumer demographics, we also record the unemployment rate for each MSA (which we match to the corresponding DMA).[37] We report these demographic variables in Table 2.

Figure 4 plots a simple firm-level price index over time across all markets. This price index is based on the price per serving of cereal. We highlight a few financial transactions in the figure.

Finally, we also gather data on input prices of different commodities. In particular, we gather

[34] For example, Nevo (2000) samples demographics from the Current Population Survey (CPS) for the corresponding city.

[35] We impose a one visit per week maximum in the weighting.

[36] For more details please consult Appendix B.8.

[37] We use data provided by the St. Louis Federal Reserve (FRED) database, which is reported at the monthly level. We impute to the weekly level using linear interpolation.

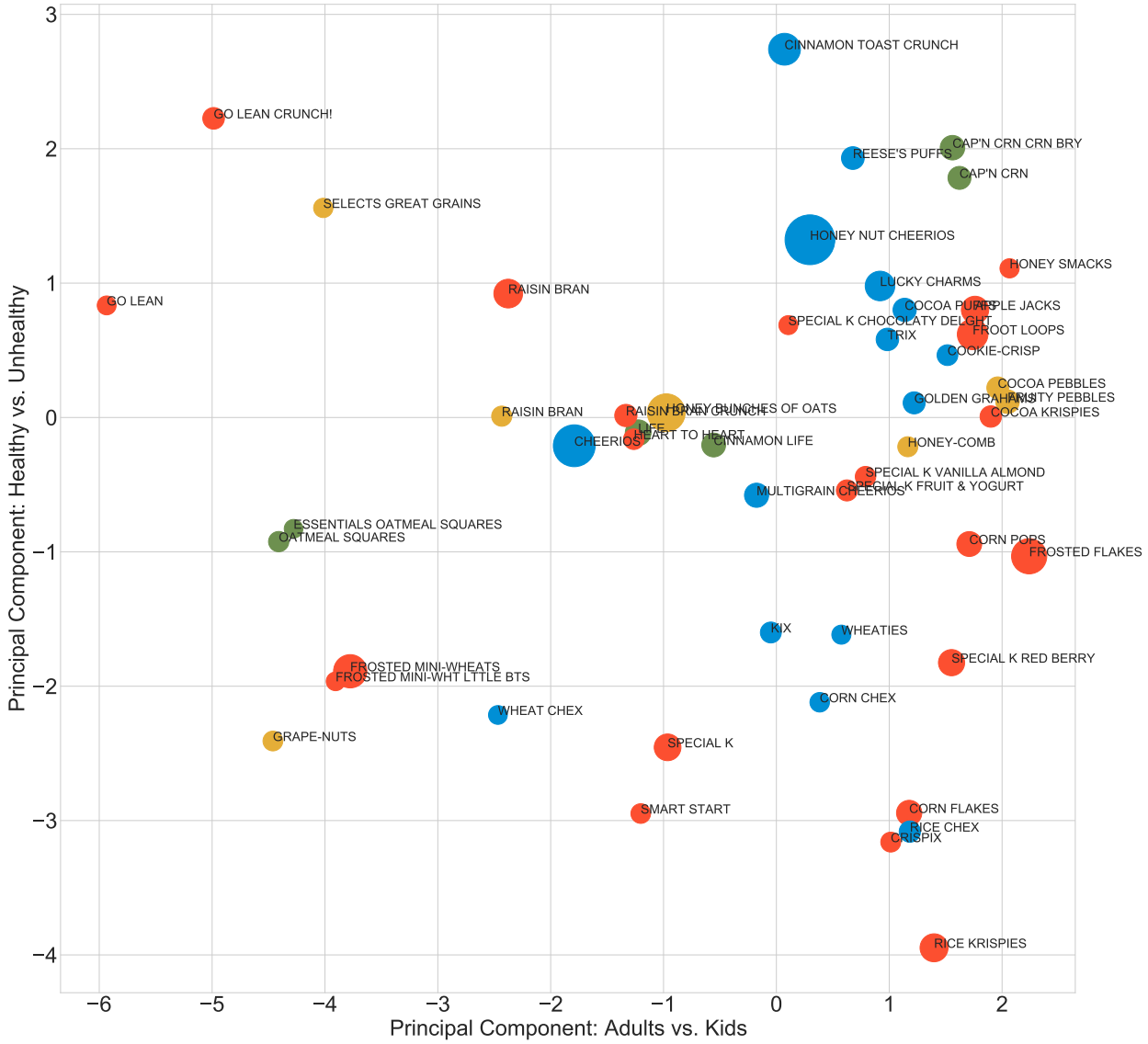



Figure 3: Principal Components of Nutritional Data: Top 50 Brands
Circle size is proportional to the inverse of national sales rank. Colors indicate different manufacturers.

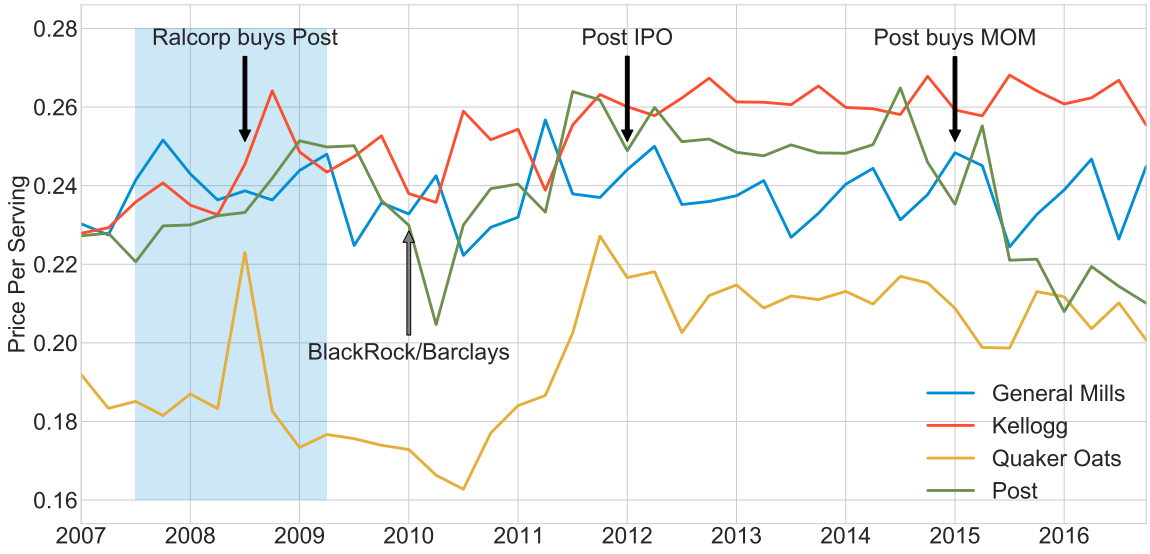



Figure 4: Average Serving Price by Firm

Total quarterly firm revenue (not inflation adjusted) divided by total firm servings for all six markets. Great Recession in shaded area.

prices for rice, wheat, corn, and oat from quandl.com, while we gather sugar prices from Global Financial Data. Changes in these input prices are represent cost shocks for the production of a given product, depending on the main ingredient of that product. That is, if a product's main ingredient is rice (e.g. Rice Krispies), it is more exposed to a price increase in rice than a price increase in corn, when compared to a corn-based cereal (e.g. Corn Pops). We identify the main ingredient for each product from the Nutritionix database discussed above. We plot the commodity prices in Figure 5. There is significant variation both over time and across ingredients in prices during our sample period.

### 4.4. Prices and Market Concentration

We can also use market shares to construct quarterly concentration measures (such as the HHI and $MHHI\Delta$). We present plots of these measures over time in Figure 6 and across markets in Figure B-4. Because we do not necessarily know which manufacturers produce the private label products (around 50% of the private label market is produced by Malt-O-Meal but we do not observe which products), we instead assume that each privately label product is produced by a different manufacturer. As we see in Appendix Figure B-4 there is substantial cross market concentration in HHI with Denver and Phoenix being relatively unconcentrated (HHI $<$ 1800) while Chicago is highly concentrated (HHI $>$ 2500). The concentration more less mirrors the inverse of the private label share (Chicago and Charlotte are more concentrated and have a lower private label share). When we look at HHI averaged across all markets we see relatively little

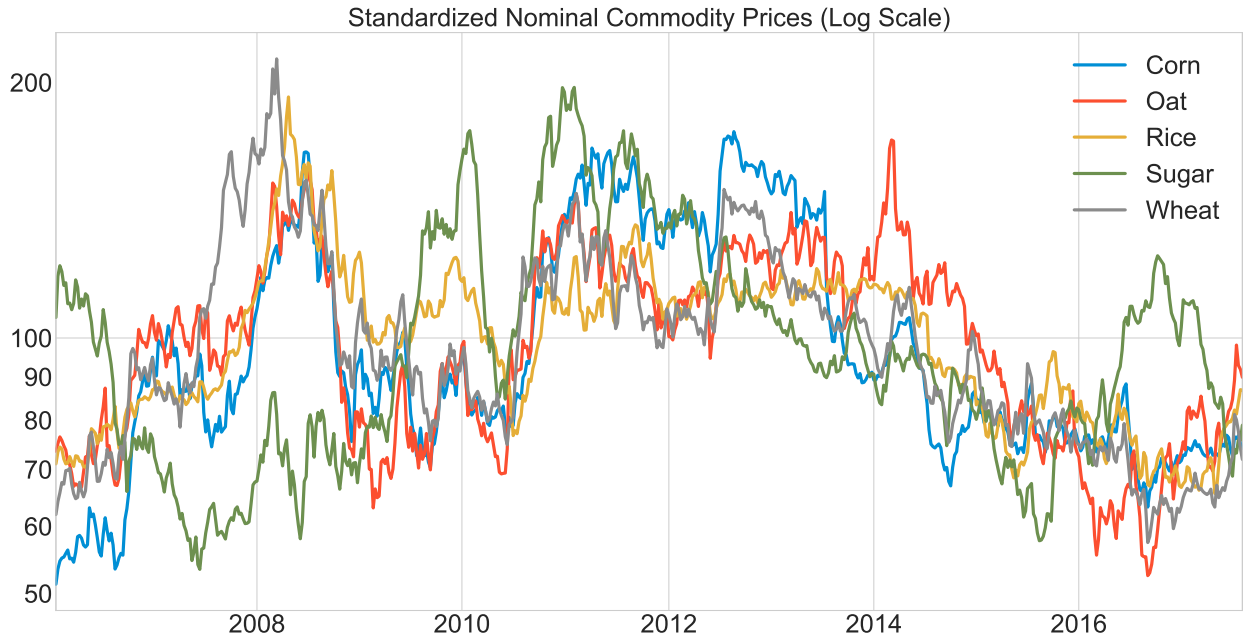



Figure 5: Input Prices

response to the Blackrock/BGI event (as we would expect), we see a substantial increase in HHI after the Post/Malt-O-Meal and General Mills/Annie's Homegrown acquisitions towards the end of the sample. We also see a substantial decline in HHI around the same time as Kraft sold Post to Ralcorp. Across time we rarely see more than a 150 point change in the national aggregate HHI.

The main point of Figures B-4 and 6 is to demonstrate that other than the spike in private label sales during 2009, there isn't much variation within a DMA over time, but there is much larger variation across DMAs in market concentration. This cross market variation in $HHI$ is likely to drown any time series variation in $\kappa$ when we construct $\Delta MHHI$. Indeed when we construct the $\Delta MHHI$ and plot it across DMAs in Figure B-4 we find the majority of the variation is across market rather than within market over time. We find that the $\Delta MHHI$ is approximately the same magnitude as the HHI: around 1500 for Phoenix and around 2400 for Chicago.

### 4.5. MHHI Regressions

To place our work in the context of the existing work on the effect of common ownership on pricing, we next perform some regressions of the style used in many early papers in this literature. These regressions are motivated by Equation 2, although do not correspond to a true reduced form if any of the assumptions of equation 2 are violated. Table 3 shows regressions of log(price) on $HHI$ and $MHHI\Delta$. An observation is at the manufacturer-DMA-retailer-quarter level. Prices and shares are computed at the level of a serving in this analysis. The different specifications across columns add additional fixed effects or controls. The regressions show that – if anything – an

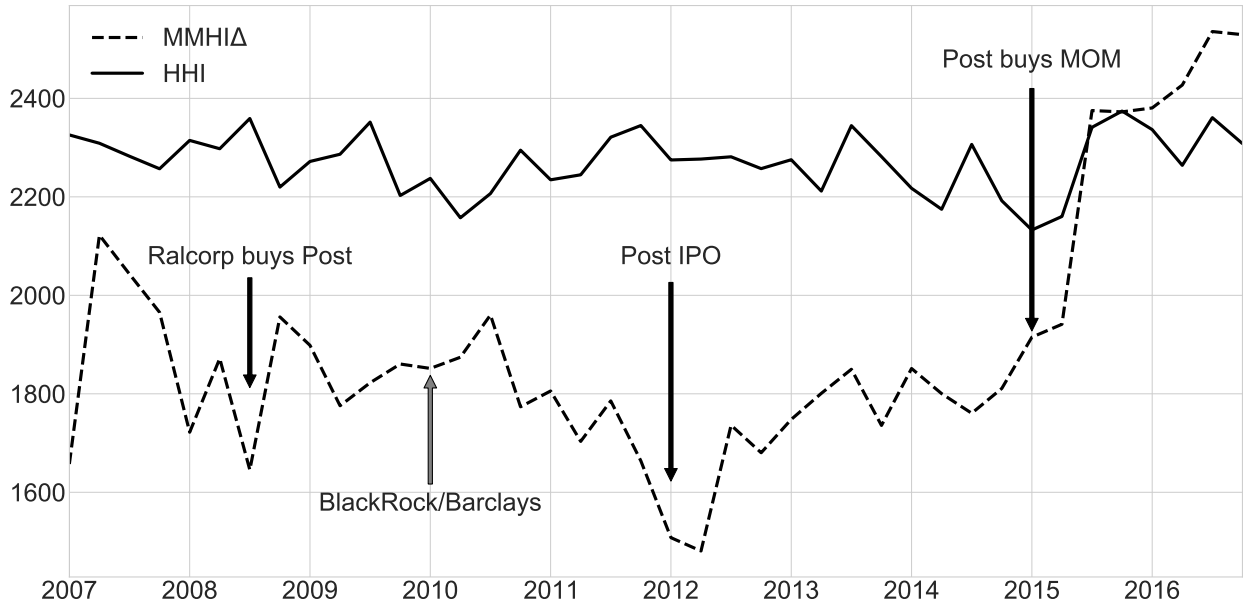



Figure 6: $HHI$ and $MHHI\Delta$ over time (Six Markets)

Private Label products treated as single unified firm. "Other" independent sellers treated as atomistic. $MHHI\Delta$ assumes proportional control.

increase in $MHHI\Delta$ is associated with lower prices in this context. However, in a differentiated products setting, there is no interpretation of these regressions and so they are only included for completeness. Appendix D.3 contains additional specifications for this type of regression.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| HHI | 0.106*** | 0.164*** | 0.0681*** | 0.0647*** | 0.0707*** | 0.525 |
| | (13.04) | (16.58) | (5.73) | (5.43) | (6.64) | (0.58) |
| MHHI-Δ | -0.0366*** | 0.00585 | -0.0427*** | -0.0429*** | -0.0370*** | -0.0367*** |
| | (-8.50) | (1.32) | (-5.86) | (-5.90) | (-5.69) | (-5.59) |
| Share | | | | | -1.135*** | -1.135*** |
| | | | | | (-40.53) | (-40.53) |
| DMA FE | No | Yes | No | Yes | Yes | Yes |
| Retailer FE | No | No | Yes | Yes | Yes | Yes |
| Quadratic Time Trend | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Cubic HHI | No | No | No | No | No | Yes |
| Observations | 6538 | 6538 | 6538 | 6538 | 6538 | 6538 |
| $R^2$ | 0.753 | 0.786 | 0.826 | 0.828 | 0.862 | 0.862 |

$t$ statistics in parentheses
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 3: MHHI Regressions

log(average serving price) is dependent variable.
An observation is a manufacturer-chain-DMA-quarter. Prices and shares are computed based on servings.

## 5. Demand Specification and Results

One main advantage of studying the RTE cereal market is that Nevo (2001, 2000) estimate a BLP-style demand model with random coefficients, product fixed effects, and demographc interactions which provides a roadmap for developing our model of consumer demand. Readers familiar with Berry et al. (1995) or Nevo (2000) should find the setup familiar. We leave the micro-foundation of the discrete choice multinomial logit demand system to the Appendix.

In short, one can derive estimating equations from (4) where $\sigma_j^{-1}(\cdot)$ represents the inverse share equation (generalized Berry inverse):

$$\sigma_j^{-1}(\mathcal{S}_t, \mathbf{p_t}, y_t; \theta_2) = h_d(x_{jt}, v_{jt}; \theta_1) + \xi_{jt} \quad \text{with} \quad \mathbb{E}[\xi_{jt}|\mathbf{z_t}] = 0 \tag{18}$$

We follow the notation of Conlon and Gortmaker (Forthcoming), summarized for convenience in Table A-1 and partition the parameter space into $[\theta_1, \theta_2]$ where the first set of parameters govern exogenous variables which enter (18) linearly, while the second set enter (18) nonlinearly and also affect the markups $\eta_{jt}(\theta_2, \kappa)$. The inverse share equation depends on the observed shares $\mathcal{S}_t$, the observed prices $\mathbf{p_t}$ and the distribution of demographics $y_t$.

We augment our aggregate scanner data on price and quantity with demographics $y_t$ and micro-moments (see Petrin (2002) or Berry et al. (2004)) formed from the Nielsen Panelist dataset. We use two demographic variables (inspired by Nevo (2000)): household income (in \$100,000s), and an

indicator for the presence of children.[38] For each DMA-chain-year in our dataset, we estimate the (conditional) log-normal distribution of income for households with and without children, as well as the fraction of households with children. This allows us to draw $y_{it}$ for each "household" from the (DMA-chain specific) joint distribution of demographics. This is meant to better reflect the demographics (particularly income) of shoppers at a particular chain beyond just the demographics of the area.[39]

Each micro-moment is meant to target a different demographic interaction parameter. We use two types of micro-moments: the first are simple conditional expectations of demographic variables conditional on purchasing the outside good. For example, conditional on purchasing the outside good, the expectation of the variable "kids", which is an indicator for the presence of children in a household, is 17.85% and the expected income is around $66,000.[40] The second type are covariances between purchaser demographics and product characteristics including price conditional on purchase.

We can use (18) to construct aggregate moments $g^D(\theta) = \frac{1}{N}\sum_{j,t} \xi'_{jt} \cdot b(\mathbf{z_t})$, where $b(\mathbf{z_t})$ is some (matrix-valued) function of the exogenous variables. We combine the aggregate data demand moments and micro-moments $g^M(\theta_1, \theta_2)$ to form a GMM estimator for $[\theta_1, \theta_2]$:[41]

$$\arg\min_{\theta_1,\theta_2} q(\theta_1, \theta_2) \equiv g(\theta_1, \theta_2)' W g(\theta_1, \theta_2), \quad g(\theta_1, \theta_2) = \begin{bmatrix} g^D(\theta_1, \theta_2) \\ g^M(\theta_1, \theta_2) \end{bmatrix} \tag{19}$$

We estimate (19) via 2-step GMM. For the second-step we replace $b(\mathbf{z_t})$ with the feasible approximation to the optimal instruments following the recipe in Conlon and Gortmaker (Forthcoming), and update the weighting matrix $W$.[42]

We use two kinds of instruments in $b(\mathbf{z_t})$. The first set are the $w_{jt}$ instruments, or own-cost-shifters, which we plot in Figure 5. We measure commodity prices of the main ingredient over time. This has advantages and disadvantages. Fortunately, these series do not covary as much as one might expect. The advantage is that commodity prices of corn are different from commodity prices of wheat and we can compare retail prices of corn-based cereals to those of wheat-based

[38] Nevo (2000) also includes the age of the head of the household although it appears to be insignificant. While Nielsen reports the age of the head of household for panelists, we face a bias-variance trade off when slicing the demographic bins too finely.

[39] One might assume that shoppers at a premium, natural, and organic supermarkets in Chicago are wealthier or have different preferences than shoppers at discount supermarkets in Chicago.

[40] This is substantially lower than the unconditional share of households with children, suggesting childless households are less likely to consume cereal overall.

[41] The covariance between aggregate data moments and micro moments is assumed to be zero in $W$. One can not estimate this covariance because the expectations are taken over different measures. We do estimate the covariance between micro-moments in $W$.

[42] For optimal instruments using only the demand side, we need an estimate of "expected price" $E[p_{jt}|\mathbf{z_t}]$ which we describe below, and evaluate shares at this "expected price" and an initial guess of $[\widehat{\theta}_1, \widehat{\theta}_2]$. This is similar to the procedure described in (Amemiya, 1977; Chamberlain, 1987; Berry et al., 1995, 1999; Reynaert and Verboven, 2014) and additional details are provided in B.6.

cereals. The disadvantage is that this instrument provides no geographic variation to explain prices in different stores for the same product at the same time.

The second type of instruments in are a variant of the BLP instruments. Here we follow Ghandi and Houde (2019). This measures the distance between each pair of products in characteristic space for each characteristic $l$: $d_{j,k}^{(l)} = |x_{kt}^{(l)} - x_{jt}^{(l)}|$. We use the quadratic version of their differentiation measure which contains all quadratic interactions of $d_{j,k}^{l}$ summed over $k$.[43] Following the recommendation in Ghandi and Houde (2019), we further interact these instruments with moments of the distribution of demographics (income and children) for each market $y_t$. This leads to a large nubmer of such instruments, many of which are highly correlated with one another. We perform dimension reduction following Carrasco (2012) or Conlon (2017) by projecting these instruments down onto principal components which explain 99% of the variance in the original set. We then obtain an estimate of "expected price": $E[p_{jt}|x_{jt}, w_{jt}, z_{jt}^{GH}] = \hat{p}_{jt}$ via linear regression on these instruments.[44] We include this "expected price" measure in our set of characteristics and repeat the procedure above leaving us with 143 instruments altogether.

We estimate our demand model using `pyblp` (Conlon and Gortmaker, Forthcoming) and report the second-stage GMM estimates of $[\theta_1, \theta_2]$ in Table 4. Both sets of estimates include product (brand) and market (DMA-chain-week) fixed effects in $h_d(x_t, w_t)$ as well as indicators for the weeks following three product recall events (interacted with the relevant products), and the DMA-week unemployment rate interacted with a dummy for branded products. Parameters have the expected sign: the recall events have a small but negative effect on demand, and when areas or periods when unemployment is high, consumers substitute away from branded products.[45]

Both sets of estimates include demographic interactions $\Pi$ between the presence of children and income with: prices, the first three principal components of product characteristics, and a constant. These parameters are largely related to the corresponding micro-moments. As is often true, high income households are substantially less price-sensitive, but have a lower overall preference for cereal.[46] Households with children are more price sensitive but have a higher overall preference for cereal. Tastes for the principal components are small but precisely estimated from the micro-moments.

Finally, we estimate two distributions of unobserved (normally distributed) heterogeneity $\Sigma$. The first estimates a normally distributed taste for price, while the second estimates a bivariate normal distribution for price and the constant. We find a strong negative correlation which suggests that as households become more price sensitive, they also have a higher mean utility for all inside

[43] We found this led to a stronger "first stage" than the local version of the Ghandi and Houde (2019) variant of the BLP instruments.

[44] We could have estimated this via some more complicated non-parametric regression but the $R^2$ is high.

[45] See Figure B-2 for cross-sectional and time series variation in unemployment.

[46] With a large outside good share, we can end up with the richest (and least price sensitive) households purchasing all of the cereal if we fail to include interactions with the constant.

| Parameter | Variable | Prices Only | | Correlated Normal | |
|---|---|---|---|---|---|
| $\beta$ | Prices | -13.369 | | -9.531 | |
| | | (0.096) | | (0.113) | |
| | Unemp × Branded | -0.024 | | -0.028 | |
| | | (0.002) | | (0.002) | |
| | Recall 1 | -0.243 | | -0.209 | |
| | | (0.068) | | (0.069) | |
| | Recall 2 | -0.033 | | -0.034 | |
| | | (0.039) | | (0.042) | |
| | Recall 3 | -0.063 | | -0.053 | |
| | | (0.045) | | (0.047) | |
| $\Sigma$ | | Cons | Prices | Cons | Prices |
| | Cons | | | 2.214 | -1.000 |
| | | | | (0.040) | |
| | Prices | | 1.945 | | 5.013 |
| | | | (0.049) | | (0.043) |
| $\Pi$ | | Kids | Inc | Kids | Inc |
| | Cons | 3.822 | -2.554 | 2.876 | -0.518 |
| | | (0.012) | (0.008) | (0.028) | (0.007) |
| | Prices | -17.665 | 7.815 | -15.429 | 3.102 |
| | | (0.072) | (0.031) | (0.184) | (0.030) |
| | $PCA_0$ | 0.064 | 0.066 | 0.052 | -0.027 |
| | | (0.000) | (0.000) | (0.000) | (0.000) |
| | $PCA_1$ | 0.182 | -0.011 | 0.206 | -0.033 |
| | | (0.000) | (0.000) | (0.001) | (0.000) |
| | $PCA_2$ | -0.244 | 0.089 | -0.287 | 0.068 |
| | | (0.000) | (0.000) | (0.001) | (0.000) |
| | Observations | 361,105 | | 361,105 | |
| | Product Fixed Effects | 946 | | 946 | |
| | Market Fixed Effects | 1,970 | | 1,970 | |
| | Median Own-Elasticity | -2.420 | | -2.665 | |
| | Median Aggregate Elasticity | -0.395 | | -0.344 | |
| | Mean Outside Good Diversion | 0.621 | | 0.482 | |

Table 4: Demand Parameter Estimates

goods.[47]

In Table 5, we report the values of the micro-moments used to estimate our parameters and the violations at $[\widehat{\theta}_1, \widehat{\theta}_2]$. For the most part, we fit these moments well.

| Panel A: Conditional Expectation Moments: | Kids | Income |
|---|---|---|
| $E[d_{it}$\|Outside Good] | 0.1785 | 0.6593 |
| Panel B: Covariance Moments: | | |
| $Cov(d_{it}, p_{jt})$ | -0.0052 | 0.0027 |
| $Cov(d_{it}, \text{PCA0}_{jt})$ | 0.1626 | -0.0152 |
| $Cov(d_{it}, \text{PCA1}_{jt})$ | 0.0790 | 0.0066 |
| $Cov(d_{it}, \text{PCA2}_{jt})$ | -0.0500 | -0.0003 |

Table 5: Micro Moment Values [UPDATE: ADD FIT]

In general our estimates of own price elasticities are around $\epsilon_{jj} = -2.6$ which is line with Nevo (2000) and other studies of RTE cereal Michel and Weiergraeber (2018). Our estimate of substitution to the outside good says that around half of consumers 0.48 substitute to the outside good when prices increase (compared to an outside good share of 74%).

In Section 5, we report the brand level (sales-weighted) diversion ratios, as well as the overall purchase probabilities for different subsets of markets. We report the diversion ratios for the markets in the top 25% of household income and the bottom 25% of household income, and the top 25% of children at home and the bottom 25% of children at home. We find substantial heterogeneity in the substitution patterns across markets. For example, Post products (such as Honey Bunches of Oats) are much more popular in markets with fewer children. Likewise, private label products are more popular in markets with low incomes.

## 6. Non-Nested Model Comparisons

### 6.1. Predicted Markups

We begin by recovering markups and marginal costs under different conduct assumptions using (9): In our setting, this requires an additional assumption that manufacturers directly set retail prices.[48] We reporting distributions of implied markups under various conduct assumptions in Figure 7. The results suggest that implied markups under profit maximization ($\kappa_{fg} = 0$) are more in line with those reported in Nevo (2000) or gross margins reported by Post (30-45%) in their

[47]These estimates are derived from the estimates of the Cholesky root of the bi-variate normal $(l_{11}, l_{21}, l_{22})$, and the lower triangular term $l_{22} \approx 0$, while the other two terms have opposite signs.

[48]There is some evidence that this is true for consumer products in supermarkets (Villas-Boas, 2007; Bonnet and Dubois, 2010). Recent work suggests that retailers extract surplus via slotting fees Hristakeva (2019) rather than markups above wholesale prices. Under double marginalization, the retailer internalizes substitution across brands from competing manufacturers. This softening of competition can be implemented an ownership matrix which sets $\kappa_{fg} > 0$. Therefore, if one has a strong prior that double marginalization is salient, we should expect the results to be biased against profit maximization.

| High Income | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Firm | Gen Mills | Kellogg | Quaker | Post | Private Label | Outside | Price | Share |
| General Mills | 19.8 | 19.2 | 5.5 | 11.9 | 8.7 | 34.2 | 20.2 | 7.5 |
| Kellogg | 20.3 | 18.4 | 5.1 | 11.7 | 8.4 | 35.2 | 21.5 | 7.1 |
| Quaker Oats | 21.4 | 18.8 | 6.5 | 12.7 | 9.2 | 30.6 | 17.0 | 1.8 |
| Post | 20.5 | 19.1 | 5.6 | 13.0 | 9.9 | 31.1 | 17.3 | 4.2 |
| Private Label | 21.1 | 19.5 | 5.8 | 14.1 | 11.2 | 27.2 | 13.4 | 2.7 |
| Other | 17.7 | 16.5 | 4.4 | 10.9 | 8.7 | 40.8 | 26.4 | 0.4 |
| **Low Income** | | | | | | | | |
| Firm | Gen Mills | Kellogg | Quaker | Post | Private Label | Outside | Price | Share |
| General Mills | 20.2 | 19.7 | 5.1 | 10.6 | 10.7 | 33.6 | 18.2 | 7.3 |
| Kellogg | 20.7 | 18.0 | 5.0 | 10.4 | 10.3 | 35.4 | 20.5 | 7.1 |
| Quaker Oats | 22.3 | 20.8 | 5.1 | 11.5 | 11.6 | 28.4 | 15.3 | 1.6 |
| Post | 21.2 | 19.9 | 5.3 | 10.7 | 11.4 | 31.3 | 16.6 | 3.5 |
| Private Label | 21.9 | 20.1 | 5.4 | 11.6 | 12.1 | 28.7 | 14.0 | 3.3 |
| Other | 16.5 | 15.3 | 4.1 | 8.9 | 9.9 | 44.9 | 27.8 | 0.1 |
| **High Children Share** | | | | | | | | |
| Firm | Gen Mills | Kellogg | Quaker | Post | Private Label | Outside | Price | Share |
| General Mills | 22.0 | 17.1 | 5.3 | 9.1 | 9.3 | 36.0 | 20.4 | 7.3 |
| Kellogg | 19.9 | 17.5 | 5.0 | 9.1 | 8.9 | 38.5 | 23.3 | 6.4 |
| Quaker Oats | 22.6 | 18.2 | 6.3 | 9.6 | 9.4 | 32.8 | 18.0 | 1.7 |
| Post | 21.8 | 18.9 | 5.4 | 10.0 | 9.2 | 33.7 | 18.9 | 3.0 |
| Private Label | 22.7 | 19.0 | 5.4 | 9.5 | 13.3 | 28.8 | 13.6 | 2.7 |
| Other | 20.5 | 16.3 | 4.5 | 7.6 | 9.2 | 40.8 | 25.3 | 0.4 |
| **Low Children Share** | | | | | | | | |
| Firm | Gen Mills | Kellogg | Quaker | Post | Private Label | Outside | Price | Share |
| General Mills | 18.2 | 17.9 | 6.9 | 14.1 | 6.8 | 36.0 | 19.8 | 4.9 |
| Kellogg | 13.9 | 27.2 | 7.1 | 14.0 | 7.2 | 30.3 | 17.3 | 6.7 |
| Quaker Oats | 16.1 | 21.1 | 11.8 | 14.8 | 7.0 | 29.0 | 15.3 | 2.1 |
| Post | 14.4 | 18.5 | 6.6 | 17.6 | 9.9 | 32.8 | 16.4 | 4.7 |
| Private Label | 13.7 | 18.4 | 6.2 | 19.6 | 12.6 | 29.3 | 12.1 | 2.2 |
| Other | 14.7 | 10.2 | 5.6 | 11.9 | 6.3 | 51.2 | 29.3 | 0.1 |

Table 6: Diversion Ratio Estimates by Income and Presence of Children (2016 Q4)

High and low represent the top quartile and bottom quartile of all markets in 2016 Q4 (after Post's acquisition of Malt-O-Meal) by median household income (below \$70k or above \$88k), and fraction of households with children ($< 14\%$ or $> 21\%$).

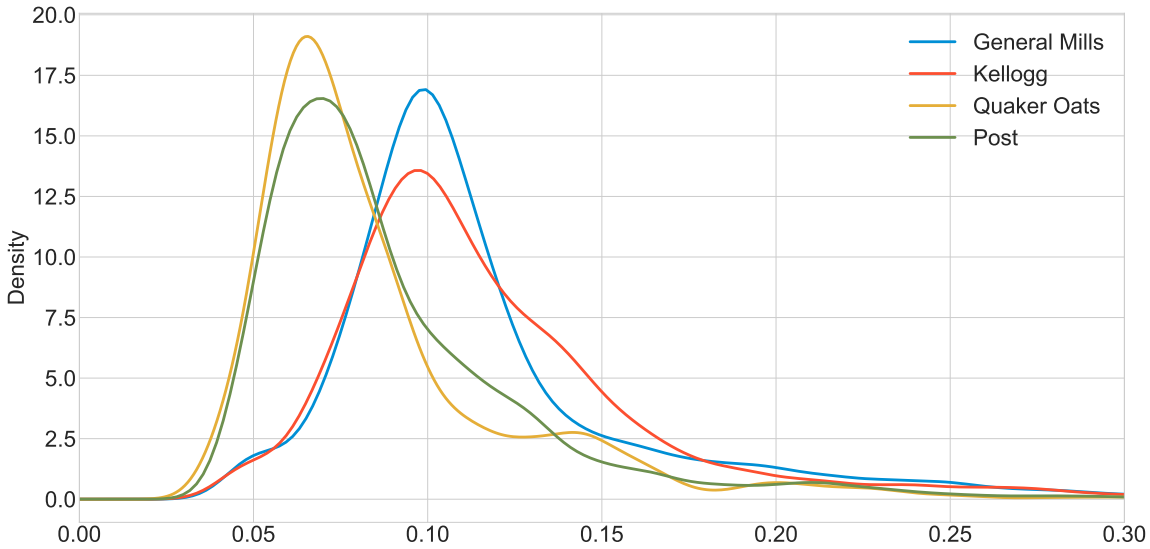


Figure 7: Distribution of Implied Markups by Firm (for Multiproduct Bertrand 2016)

financial filings.[49] We take evidence on the level of markups as largely descriptive, but it suggests that the common ownership hypothesis implies margins which are likely to be too large.

We further break out the sales weighted average implied markups and marginal costs over time by firm in Figure 8. As is often the case, under profit maximization $\kappa_{fg} = 0$, larger firms have larger implied markups as they recapture more diverted sales with their own products.

As a second way to evaluate implied profit margins, we compare the effect of common ownership on prices to the effect of any of the power set of potential pairwise mergers, as well as full collusion among the four largest firms. To be precise, we derive marginal costs under the assumption of profit maximization ($\kappa_{fg} = 0$), and then compute counterfactual equilibrium prices using those marginal costs under all of the alternative scenarios. Results are presented in Table 7. Each column represents a counterfactual scenario: the first six are pairwise merger counterfactuals, while the seventh represents "turning on" common ownership incentives, and the eight represents optimal collusion among the four largest firms. Price effects, weighted by quantity of servings sold, are represented in the table. The first four rows represent prices for the four largest firms in the market, while the fifth row represents an aggregate servings-weighted price index.

We see, predictably, that the most anticompetitive merger in the marketplace would be that between General Mills and Kellogg's, resulting in a 5.50% price increase. Mergers of Quaker Oats and any other firm are the least anticompetitive, all yielding less than a 1.5% price increase, presumably because the products sold by Quaker Oats are more differentiated. In contrast, GM-

[49] We use the gross margins reported by Post because they are the closest to a "pure play" in RTE Cereal. Quaker Oats is a small division of PepsiCo, Kellogg's produces a number of other snack foods: Keebler Cookies, Pringles Potato Chips, and Pop-Tarts, while General Mills also owns Yoplait yogurt, Pillsbury baking products, and Häagen-Dazs ice cream.

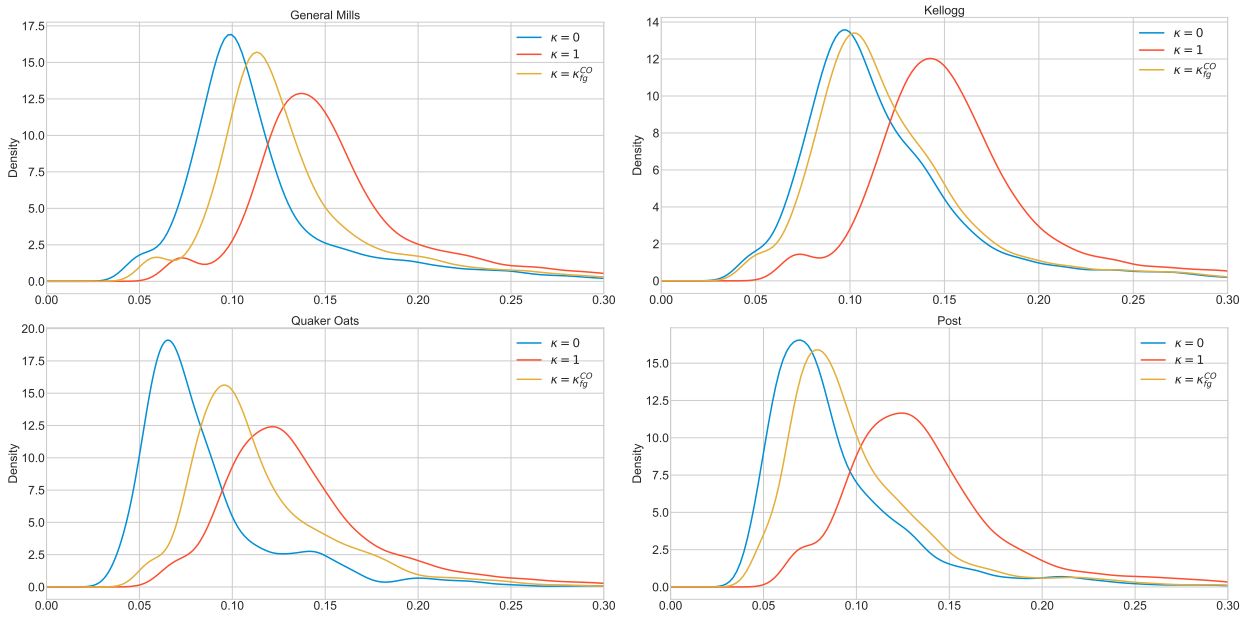



Figure 8: Distribution of Implied Markups by Conduct Assumption and Firm

Post and Kelloggs-Post mergers both yield price increases in excess of 3%.

The effects of any pairwise predicted merger, however, are dwarfed by the 8.34% price increase predicted under common ownership. This highlights the anticompetitive potential of the hypothesis: it yields substantially larger price increases than a hypothetical merger between the two dominant firms in the market, which we would expect antitrust authorities to block. By way of comparison, full collusion among the four largest firms yields a price increase of 17.28%, and so common ownership would get us 8.34/17.28 = 48.25% of the way to collusive pricing.

The question remains, however, whether patterns of common ownership are reflected in pricing. For this we implement the conduct testing procedure developed in Section 3.

| Firm | GM-KEL | GM-QKR | GM-POST | KEL-QKR | KEL-POST | QKR-POST | $\kappa^{CO}$ | $\kappa = 1$ |
|---|---|---|---|---|---|---|---|---|
| General Mills | 7.24 | 1.82 | 4.22 | 0.25 | 0.44 | 0.18 | 6.33 | 15.18 |
| Kellogg | 7.29 | 0.21 | 0.42 | 1.71 | 4.18 | 0.17 | 8.17 | 15.13 |
| Quaker Oats | 0.08 | 9.84 | 0.04 | 9.02 | 0.04 | 5.81 | 13.62 | 28.24 |
| Post | 0.50 | 0.19 | 8.61 | 0.20 | 8.46 | 2.19 | 10.65 | 21.66 |
| **Price Index** | 5.50 | 1.46 | 3.33 | 1.39 | 3.32 | 0.94 | 8.34 | 17.28 |

Table 7: Price Increases under Counterfactual Scenarios

Entries are percent price increases from the given baseline. The market-level mean price per serving observed in the data is \$0.2255.

### 6.2. Testing: Inclusions and Exclusions

Recall that in order to implement our test, we need to specify, first, the "inclusions" which appear both in Step 1 supply and Step 2 markup regressions and, second, the the exclusions which appear in the Step 2 markup regression but *not* in the Step 1 supply regression.

For the Step 1 supply regressions, we take the marginal costs predicted by a particular model on the left-hand side, and on the right-hand side we have $x_{jt}$ and $w_{jt}$. Variables in $w_{jt}$ include the input costs for designated "main" ingredients, the fraction of servings on sale, and finally dummies for periods in which a product was on recall (see Footnote 12 above). We also include product fixed effects, retailer-DMA fixed effects, and quarter fixed effects.

For the Step 2 markup regression to predict $\widehat{\Delta\eta}_{jt}^{1,2}$, we include $x_{jt}$ and $w_{jt}$ as well as four different sets of exclusions, which all follow from the discussion in Section 3.3. We use other firms' input costs; demographic moments at the market level; the Ghandi and Houde (2019) variation on BLP instruments, and the demand-optimal instruments.

For other firms' input costs we use the price of ingredients (corn, rice, oat, wheat, and sugar) that are not designated as "main" ingredients in the Nutrionix dataset. So, for instance, if oats are the main ingredient for Honey Bunches of Oats and corn is the main ingredient for Corn Flakes, then the price of oats will be included for the former and excluded for the latter, and vice-versa for the price of rice.

The demographic moments we use correspond to those we use in predicting the distribution of the nonlinear coefficients in Section 5. In particular, we use the fraction of households with children, and the mean and standard deviation of income for households with and without children.

Finally, the construction of BLP instruments and the demand-optimal instruments is explained in detail above in Section 5. We emphasize that the latter are optimal from the perspective of estimating the parameters of the demand system, not necessarily for testing. Nonetheless, since they capture the nonlinearity of markups, we conjecture that they will be effective for testing supply restrictions as well.

### 6.3. Implementation

The specification of $g(\cdot)$ in the Step 2 regression is important because markups are a nonlinear function of everything in the model. The power of the test, therefore, depends on the fit of this regression. However, if we allow arguments of $g(\cdot)$ to enter flexibly while using a restrictive functional form for $h_s(\cdot)$, then any misspecification of the $h_s$ will functionally enter as additional exclusion restrictions in our test (this is a known problem when using flexible forms to estimate the first-stage regression of a linear IV problem; see Chen et al. (2020) for a recent discussion). Therefore, as in nonparametric instrumental variables regression (**??**), we fit both functions flexibly.

Table 8: Testing Results: Profit Maximization vs Variations of Common Ownership

| | Others' Costs | Demographics | BLP Inst. | Dmd. Opt. Inst. |
|---|---|---|---|---|
| Common Ownership | -2.5571 | -2.6003 | -2.9191 | -3.1160 |
| Common Ownership (MA) | -2.5761 | -2.5903 | -2.9201 | -3.1240 |
| Common Ownership (Lag) | -2.5357 | -2.5683 | -2.8908 | -3.0888 |

Notes: This table reports testing results for pairwise comparisons of own-firm profit maximization against variations on common ownership. The column label indices the set of exclusions used in the Step 2 regression. A negative value constitutes evidence for profit maximization and a positive value constitutes evidence for common ownership. The null hypothesis is that they satisfy the criterion function equally well, and the critical values for rejection at $\alpha = 0.5$ are –1.96 and 1.96.

We use a random forest to estimate both $h_s(\cdot)$ in Step 1 and $g(\cdot)$ in Step 2.[50] We discuss results for both sets of regressions further in Appendix Section E.3.

In order to compute $\hat{\sigma}$, the standard deviation of $Q(\kappa^1) - Q(\kappa^2)$ which appears in the denominator of our test statistic, we bootstrap the entire procedure. Bootstrap samples, and standard errors in all of the regressions reported, are clustered at the retailer-DMA-year level of aggregation.

Finally, note that at every step—Step 1 and 2 regressions, as well as the computation of $Q(\kappa^m)$—we weight observations (which are product-chain-DMA-weeks) by the total quantity of servings for that product across all markets. We do this to guarantee that our results are not driven by the large number of niche products in our sample.

## 6.4. Results: Testing Common Ownership

Our main results are presented in Table 8. Here we test variations on the common ownership (by row) against own-firm profit maximization using different sets of exclusions (by column). To read this table, recall that the Rivers-Vuong test statistic is distributed $N(0, 1)$. The null hypothesis is that the two models perform equally well. A negative value may reject that null in favor of the alternative that model 1 better fits the data (here, profit maximization), while a positive value may do so for model 2 (common ownership).

The results consistently reject the common ownership hypothesis in favor of own-firm profit maximization. We note two additional observations. First, the difference between the alternative models of common ownership considered (lagged ownership and a moving average) is insignificant. Therefore we shall focus on the baseline common ownership case. Second, demand-optimal instruments generate the strongest rejections. This is as expected, since they exploit the full set of conditional moment restrictions implied by the model, which is reflect in the fit of the Step 2 regressions, documented further in Appendix Section E.3. There we show that the demand optimal instruments yield the smallest mean-squared error.

We emphasize that our results are not driven by comparisons of average markups. All of our

[50] Alternative approaches considered by the literature in a linear IV setting include LASSO regression (**?**) and deep neural nets (**?**).

regressions in the testing procedure include product fixed effects. Therefore, we are using the within-product, across-market variation in markups as predicted by our exclusions. Moreover, we would also stress that our result does not answer the question of whether common ownership, broadly construed as the presence of common owners, has *any* price effects. That is an ill-defined, untestable question. Rather, we have shown that the pricing model implied by the common ownership hypothesis, as outlined in Section 2 above, is inconsistent with the pricing behavior we observe.

In Appendix Section E.1 we offer further pairwise comparisons, including to perfect competition and four-firm collusion; we consider within-firm conduct tests, and we test profit maximization against the conduct parameter model discussed in Section 3.4 for $\tau \in \{0.1, 0.2 \ldots, 0.9\}$. Our findings consistently favor own-firm profit maximization against these alternative conduct models.

## 7. Conclusion

In this paper we study how to test among competing assumptions for conduct in a differentiated products market. We show that exclusion restrictions can be used to differentiate between alternative models of firm conduct. Our application is to RTE cereal, where we first estimate demand in a rich framework using detailed scanner data as well as demographic data of shoppers. The literature on common ownership has for the most part focused on reduced-form regressions of prices on measures of concentration to make progress. Our test exploits exclusion restrictions to separate models of conduct. We find support for traditional Bertrand competition; however, simulated effects of common ownership incentives imply large potential effects. We hope that our illumination of a structural approach inspired more research into the topic.

## References


Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Amemiya, T. (1977): "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955–968.

Antón, M., F. Ederer, M. Giné, and M. C. Schmalz (2020): "Common ownership, Competition, and Top Management Incentives," Working Paper.

Azar, J., S. Raina, and M. C. Schmalz (2016): "Ultimate Ownership and Bank Competition," Https://ssrn.com/abstract=2710252. Accessed on 2020-01-08.

Azar, J., M. C. Schmalz, and I. Tecu (2018a): "Anti-Competitive Effects of Common Ownership," *Journal of Finance*, 73.

——— (2018b): "Reply to:"Common Ownership Does Not Have Anti-Competitive Effects in the Airline Industry"," .

Backus, M., C. Conlon, and M. Sinkinson (2019): "The Common Ownership Hypothesis: Theory and Evidence," Brookings Working Paper.

——— (2020): "Theory and Measurement of Common Ownership," *AEA: Papers and Proceedings*, 110.

——— (2021): "Common Ownership in America: 1980–2017," *American Economic Journal: Microeconomics*.

Berry, S. and P. Jia (2010): "Tracing the Woes: An Empirical Analysis of the Airline Industry," *American Economic Journal: Microeconomics*, 2, 1–43.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

——— (1999): "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," *American Economic Review*, 89, 400–430.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25, 242–262.

Berry, S. T. and P. A. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

Boller, L. and F. Scott Morton (2019): "Testing the Theory of Common Stock Ownership," Working Paper.

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

Bresnahan, T. F. (1982): "The Oligopoly Solution is Identified," *Economics Letters*, 10, 87–92.

——— (1987): "Competition and Collusion in the American Automobile Industry: The 1955 Price War," *The Journal of Industrial Economics*, 35, 457–482.

Bresnahan, T. F. and S. C. Salop (1986): "Quantifying the Competitive Effects of Production Joint Ventures," *International Journal of Industrial Organizatoin*, 4, 155–175.

Carrasco, M. (2012): "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170, 383 – 398, thirtieth Anniversary of Generalized Method of Moments.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

Chen, J., D. L. Chen, and G. Lewis (2020): "Mostly Harmless Machine Learning: Learning Optimal Instruments in Linear IV Models," .

Conlon, C. (2017): "The Empirical Likelihood MPEC Approach to Demand Estimation," Working Paper.

Conlon, C. and J. Gortmaker (Forthcoming): "Best Practices for Differentiated Products Demand Estimation with PyBLP," *Rand Journal of Economics*.

Cowling, K. and M. Waterson (1976): "Price-Cost Margins and Market Structure," *Economica*, 43, 267–274.

Crawford, G. S., R. S. Lee, M. D. Whinston, and A. Yurukoglu (2018): "The Welfare Effects of Vertical Integration in Multichannel Television Markets," *Econometrica*, 86.

DellaVigna, S. and M. Gentzkow (2019): "Uniform Pricing in U.S. Retail Chains*," *The Quarterly Journal of Economics*, 134, 2011–2084.

Duarte, M., L. Magnolfi, and C. Sullivan (2020): "Testing Firm Conduct," Unpublished Manuscript.

Elhauge, E. (2016): "Horizontal Shareholding," *Harvard Law Review*, 109, 1267–1317.

Farrell, J. and C. Shapiro (2010): "The B.E. Journal of Theoretical Economics,," *The B.E. Journal of Theoretical Economics*, 10.

Gasmi, F., J. Laffont, and Q. Vuong (1992): "Econometric Analysis of Collusive Behavior in a Soft-Drink Market," *Journal of Economics & Management Strategy*, 1, 277–311.

Ghandi, A. and J.-F. Houde (2019): "Measuring Substitution Patterns in Differentiated Products Industries," Working Paper.

Gramlich, J. and S. Grundl (2017): "Estimating the Competitive Effects of Common Ownership," FEDS Working Paper 2017-029.

Hristakeva, S. (2019): "How Do Vertical Contracts Affect Product Availability? An Empirical Study of the Grocery Industry," Working Paper.

Kennedy, P., D. P. O'Brien, M. Song, and K. Waehrer (2017): "The Competitive Effects of Common Ownership: Economic Foundations and Empirical Evidence," Working Paper.

Michel, C. and S. Weiergraeber (2018): "Estimating Industry Conduct in Differentiated Products Markets The Evolution of Pricing Behavior in the RTE Cereal Industry," Working Paper.

Miller, N. and A. MacKay (2018): "Instrument-Free Demand Estimation," Working paper.

Miller, N. and M. Weinberg (2017): "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica*, 85, 1763–1791.

Nevo, A. (1998): "Identification of the Oligopoly Solution Concept in a Differentiated-Products Industry," *Economics Letters*, 59, 391–395.

——— (2000): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *RAND Journal of Economics*, 31, 395–421.

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

Newey, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

O'Brien, D. and S. Salop (2000): "Competitive Effects of Partial Ownership: Financial Interest and Corporate Control," *Antitrust Law Journal*, 67, 559–614.

O'Brien, D. P. (2017): "Price-Concentration Analysis: Ending the Myth, and Moving Forward," Https://ssrn.com/abstract=3008326. Accessed on 2020-01-08.

Pakes, A. (2017): "Empirical tools and competition analysis: Past progress and current problems," *International Journal of Industrial Organization*, 53, 241 – 266.

Park, A. H. and K. Seo (2019): "Common Ownership and Product Market Competition: Evidence from the U.S. Airline Industry," *Korean Journal of Financial Studies*, 45, 617–640.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110, 705–729.

Posner, E. A., F. M. Scott Morton, and E. G. Weyl (2017): "A Proposal to Limit the Anti-Competitive Power of Institutional Investors," *Antitrust Law Journal*, 81.

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

Rivers, D. and Q. Vuong (2002): "Model selection tests for nonlinear dynamic models," *The Econometrics Journal*, 5, 1–39.

Rotemberg, J. (1984): "Financial Transactions Costs and Industrial Performance," Http://hdl.handle.net/1721.1/47993. Accessed on 2020-01-08.

Shennach, S. M. and D. Wilhelm (2017): "A Simple Parametric Model Selection Test," *Journal of the American Statistical Association*, 112, 1663–1674.

Shi, X. (2015): "A Nondegenerate Vuong Test," *Quantitative Economics*, 6, 85–121.

Smith, R. J. (1992): "Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments," *Econometrica*, 60, 973–980.

Villas-Boas, S. B. (2007): "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data," *The Review of Economic Studies*, 74, 625–652.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–333.

Whang, Y.-J. and D. W. K. Andrews (1993): "Tests of Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277–318.

Wollmann, T. G. (2018): "Trucks without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles," *American Economic Review*, 108, 1364–1406.

Yatchew, A. J. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435–451.

# Appendices

## A. Theory Appendix

Table A-1: Model Notation

| | | | |
|---|---|---|---|
| $j$ | Products | $p_{jt}$ | Price |
| $t$ | Markets | $c_{jt}$ | Marginal cost |
| $i$ | "Individuals" | $x_{jt}$ | Exogenous product characteristic |
| $(f, g)$ | Firms | $v_{jt}$ | Excluded demand-shifter |
| | | $w_{jt}$ | Excluded supply-shifter |
| $N$ | Set/number of products across all markets | $U_{ijt}$ | Indirect utility |
| $T$ | Set/number of markets | $\delta_{jt}$ | Mean utility |
| $J_t$ | Set/number of products in market $t$ | $\mu_{ijt}$ | Heterogeneous utility |
| $I_t$ | Set/number of individuals in market $t$ | $\varepsilon_{ijt}$ | Idiosyncratic preference |
| $F_t$ | Set/number of firms in market $t$ | | |
| $J_{ft}$ | Set/number of products of firm $f$ in market $t$ | $s_{ijt}$ | Choice probability |
| | | $s_{jt}$ | Calculated market share |
| | | $\mathcal{S}_{jt}$ | Observed market share |
| | | $\xi_{jt}$ | Demand-side structural error |
| $\theta$ | Parameters with dimension $K$ | | |
| $\theta_1$ | Demand-side parameters with dimension $K_1$ | $\mathcal{H}_t(\kappa)$ | Ownership or holdings matrix |
| $\theta_2$ | Nonlinear common parameters with dimension $K_2$ | $\Delta_t$ | $J_t$ matrix of demand derivatives |
| $\theta_3$ | Supply-side parameters with dimension $K_3$ | $\eta_{jt}$ | Multi-product Bertrand markup |
| $\alpha$ | Parameter in $\theta_2$ on price | $\omega_{jt}$ | Supply-side structural error |
| $\widetilde{\theta}_2$ | Parameters in $\theta_2$ governing heterogeneity | | |
| | | $Z$ | Instruments |
| $\kappa_{fg}$ | Profit weight that $f$ places on $g$'s profits | $W$ | Weighting matrix |
| $\gamma_{fs}$ | Pareto weight that $f$ places on $s$'s payoffs | $g$ | Sample moments |
| $\beta_{fs}$ | Cashflow rights of investor $s$ in firm $f$ | $q$ | GMM objective function |

### A.1. Microfoundation of Demand Model

We define the utility of consumer $i$ for product $j$ and market (chain-dma-week) $t$ as:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \tag{1}$$

Following the usual convention we decompose the utility into three parts: a mean preference which all consumers agree on $\delta_{jt}$, a consumer-specific taste shock $\mu_{ijt}$, and an idiosyncratic error $\varepsilon_{ijt}$. Consumers select the option among the set of $\mathcal{J}_t = \{0, 1, \ldots, J_t\}$ alternatives which provides the most utility. Following convention we include a no purchase alternative, denoted $j = 0$, which provides $u_{i0t} \equiv \varepsilon_{i0t}$. Aggregate market shares are given by integrating over heterogeneous consumer choices:

$$s_{jt}(\boldsymbol{\delta_t}) = \int \mathbb{I}[u_{ijt}(\boldsymbol{\delta_t}, \boldsymbol{\mu_{it}}) > u_{ikt}(\boldsymbol{\delta_t}, \boldsymbol{\mu_{it}})]\, d\mu_{it}\, d\varepsilon_{it}.$$

A frequent assumption is that $\varepsilon_{ijt}$ is distributed IID type I extreme value (Gumbel) so that market shares are given by:

$$s_{jt}(\boldsymbol{\delta_t}, \theta_2) = \int \frac{\exp[\delta_{jt} + \mu_{ijt}]}{\sum_{k \in J_t} \exp[\delta_{kt} + \mu_{ikt}]} f(\mu_{it} \mid \widetilde{\theta}_2)\, d\mu_{it}. \tag{2}$$

We follow standard practice and put an additive structure on the mean utilities:

$$\delta_{jt}(\mathcal{S}_t, \widetilde{\theta}_2) = h_d(x_{jt}, v_{jt}; \theta_1) + \xi_{jt}. \tag{3}$$

We specify the linear component of utility $h_d(x_{jt}, v_{jt}; \theta_1)$ as follows. We include four $v_{jt}$ variables which affect demand, but are excluded from supply. Three of those variables are dummies for the twelve weeks following a product recall for the affected products. For further discussion of the product recall events (including plots demonstrating the effect on demand) consult Appendix B.10. We also include an interaction between the DMA-week level unemployment rate and a dummy for branded products. This is meant to capture the fact that during the Great Recession there is an increase in the relative demand for private label as opposed to branded products. We provide additional support for this in Figure B-2 and Figure B-3. We also include a large number of two-way fixed effects: 946 product fixed effects and 1970 chain-week fixed effects.[51]

We specify the distribution of unobservable heterogeneity from (1) similar to Nevo (2000):

$$\mu_{ijct}(\widetilde{\theta}_2) = (\Sigma \cdot \nu_{ict} + \Pi \cdot D_{ict}) \cdot \begin{bmatrix} 1, & p_{jct} \end{bmatrix}^T$$

We define the nonlinear parameters as $\widetilde{\theta}_2 = [\Sigma, \Pi]$ where the $\Pi$ represent the parameters on interactions between demographic draws $D_{ict} = [\text{kids}_{ict}, \text{income}_{ict}]$ and observable characteristics $x_{jct}$, and $\Sigma$ represents the covariance of the (normally distributed) matrix of unobserved heterogeneity. We allow consumers to have heterogeneous preferences over prices and a constant (which captures inside vs. outside good substitution).[52] Demographics are drawn from the DMA-chain-year specific distribution calculated using the Nielsen Panelist data. Our demographics capture the joint distribution of income and presence of children in the household.[53]

### A.2. Relationship to UPP

To illustrate the relevance of various exclusion restrictions in discriminating among models of conduct, we can revisit the firm's first order condition for product $j$ in (**??**). The right side below in (4) represents marginal revenue, while the left represents marginal cost. We've followed the convention of Farrell and Shapiro (2010) and written this in terms of the *diversion ratio* $D_{jk}(\theta_2) =$

[51] Nevo (2000) and Nevo (2001) include one fixed effect for each of the 26 products in the dataset.
[52] See Section 4.3 or **??** for a discussion on how prices and product characteristics are constructed.
[53] See Section 4.3 and Appendix B.8.

$\frac{\frac{\partial q_k}{\partial p_j}(\theta_2)}{\frac{\partial q_j}{\partial p_j}(\theta_2)}$ and Upward Pricing Pressure (UPP) $UPP_{j,g} = \sum_{k' \in \mathcal{J}_g}(p_{k'} - mc_{k'})D_{jk'}(\theta_2)$:

$$p_j \underbrace{\left(1 + \frac{1}{\epsilon_j(\mathbf{p})}\right)}_{=\frac{1}{\mu_j(\mathbf{p})}} = mc_j + \sum_{k \in \mathcal{J}_f}(p_k - mc_k)D_{jk} + \sum_g \kappa_{fg} \underbrace{\left(\sum_{k' \in \mathcal{J}_g}(p_{k'} - mc_{k'})D_{jk'}\right)}_{UPP_{j,g}}$$

$$h(x_j, w_j; \theta_3) + \omega_j = f\left(\frac{p_j}{\mu_j(\mathbf{p}, \theta_2)} - \sum_{k \in \mathcal{J}_f}(p_k - mc_k)D_{jk}(\theta_2) - \sum_g \kappa_{fg} \cdot UPP_{j,g}(\theta_2)\right) \tag{4}$$

We've written $f(\cdot)$ in (4) above as three terms: the first is the single-product firm's marginal revenue; the second is the decrease in variable profits from cannibalizing other products in the multi-product firm's portfolio; the third is the cannibalization of "sympathetic" competitor variable profits. If we are interested in discriminating among competing models for conduct $\kappa$, we can difference (4) as we did in (**??**) for $(\kappa_0, \kappa_1)$:

$$\Delta\omega_j \approx p_j \left[\frac{1}{\mu_j(\mathbf{p}(\kappa_0, \theta_2))} - \frac{1}{\mu_j(\mathbf{p}(\kappa_1, \theta_2))}\right] - \sum_g \Delta\kappa_{fg} \cdot UPP_{j,g}(\mathbf{p}(\kappa_0, \theta_2)) \tag{5}$$

We should caution that (7) represents an approximation, and embeds several assumptions: that $f(x) = x$, the multi-product term differences out, and that $UPP_{j,g}$ does not vary with $\kappa$ or $\mathbf{p}$.[54] The first term above contains the difference in the proportional markups $\mu_j(\kappa, \mathbf{p}, \theta_2)$, which depends on how much the own price elasticities $\epsilon_j(\mathbf{p}, \kappa)$ vary in *equilibrium* as conduct changes.[55] The point of (7) is that the second term is linear in the *difference* in conduct $\Delta\kappa_{fg}$, and shows us that a relevant instrument $Z^e_{jt}$ is likely to shift $UPP_{j,g}$. This gives us an intuitive way to check whether an instrument is likely to be relevant or not. As in (**??**), it is tempting to treat $UPP_{j,g}$ as an exclusion restriction directly and test $\lambda_{fg} = 0$:

$$f(p_{jt} - \eta^0_{jt}) = h(x_{jt}, w_{jt}; \theta_3) + \sum_g \lambda_{fg} \cdot \Delta\kappa_{fg} \cdot UPP_{j,g}(\mathbf{p}(\kappa_0, \theta_2)) + \omega_{jt} \tag{6}$$

Just as $E[\omega_{jt}|x_{jt}, w_{jt}, \eta_{jt}] = 0$ as invalid, so is $E[\omega_{jt}|x_{jt}, w_{jt}, UPP_{jg}] = 0$, because both are functions of equilibrium objects $(\mathbf{p}, \mathbf{s})$ and most importantly $\omega_{jt}$.

[54] We assume $f(x) = x$ but this need not be true so long as $f(\cdot)$ is a monotone function of the argument. Otherwise we need to provide additional terms in the Taylor expansion as in (**??**). Similarly the assumption that $UPP_{j,g}$ is unchanged by conduct ignores the terms $\kappa_{fg} \cdot \Delta UPP_{j,g}(\kappa, \mathbf{p}, \theta_2)$ and $\Delta\kappa_{fg} \cdot \Delta UPP_{j,g}(\kappa, \mathbf{p}, \theta_2)$. The relation in (7) is meant only to provide intuition.

[55] The difference in markups is given by $\left[\frac{1}{\mu_j(\mathbf{p}(\kappa_0, \theta_2))} - \frac{1}{\mu_j(\mathbf{p}(\kappa_1, \theta_2))}\right] = \left[\frac{\epsilon_j(\mathbf{p}(\kappa_0)) - \epsilon_j(\mathbf{p}(\kappa_1))}{(1+\epsilon_j(\mathbf{p}(\kappa_0)))(1+\epsilon_j(\mathbf{p}(\kappa_1)))}\right]$. This depends on how the elasticities change from one equilibrium to the next. This is what Bresnahan (1982) called "rotations of demand" and what Berry and Haile (2014) called rotations of marginal revenue.

$$\Delta\omega_j \approx p_j \left[\frac{\Delta\epsilon_{jj}(\kappa_0,\kappa_1)}{(1+\epsilon_{jj}(\kappa_0))(1+\epsilon_{jj}(\kappa_1))}\right] - \sum_g \Delta\kappa_{fg} \cdot UPP_{j,g}(\mathbf{p},\kappa,\theta_2) \tag{7}$$

The first term above depends on how the *equilibrium own price elasticity* changes (compared to its geometric mean) from one conduct assumption to another.[56] The second term depends on Upward Pricing Pressure (UPP) $UPP_{j,g} = \sum_{k\in\mathcal{J}_g}(p_k - mc_k)D_{jk}$ which in turn depends on the diversion ratio $D_{jk} = \frac{\partial q_k}{\partial p_j} / \frac{\partial q_j}{\partial p_j}$. This approximation is useful in that the second term is linear in $\Delta\kappa_{fg}$. This tells us that in order to discriminate among any two models of conduct (without knowledge of the true model) excluded variables $Z^e_{jt}$ should shift $UPP_{j,g}$ and make $j$'s competitors stronger or weaker either by changing the diversion ratio $D_{jk}$ and/or the price cost margin $(p_k - mc_k)$.

### A.3. Derivative of Markup

We are interested in the Jacobian matrix $\frac{\partial \boldsymbol{\eta_t}}{\partial \kappa}$ and $\frac{\partial \boldsymbol{\eta_t}}{\partial \tau}$. We will consider a single market and omit the $t$ subscript. We can augment the original setup with an "internalization parameter" $\tau$. Here $\tau = 1$ denotes acting in accordance with the common ownership hypothesis while $\tau = 0$ represents own profit maximization. The downside is that $\tau \in (0,1)$ is not necessarily interpretable although intermediate values possibly represent some partial internalization.

$$\mathcal{H}_{jk}(\kappa,\tau) = \begin{cases} 1 & \text{if } (j,k) \in \mathcal{J}_f \text{ for any } f, \\ \tau \cdot \kappa_{fg} & \text{if } j \in \mathcal{J}_f \text{ and } k \in \mathcal{J}_g \text{ for any } (f,g), \\ 0 & \text{otherwise.} \end{cases}$$

The markup has derivative given by:

$$\begin{aligned} \frac{\partial \boldsymbol{\eta_t}}{\partial \tau} &= \frac{\partial}{\partial \tau}\left[(\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\mathbf{s}(\mathbf{p})\right] \\ &= (\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\left[\frac{\partial}{\partial\tau}(\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))\right](\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\mathbf{s}(\mathbf{p}) \\ &= (\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\left[I_{f\neq g}\odot\Omega(\mathbf{p})\right]\boldsymbol{\eta_t}(\kappa,\tau;\theta_2) \end{aligned}$$

[56] These are the "rotations of demand" in Bresnahan (1982) or "rotations in marginal revenue" of Berry and Haile (2014).

We consider the derivative with respect to a scalar element of $\kappa$, $\kappa_{fg}$

$$\begin{aligned}\frac{\partial \boldsymbol{\eta_t}}{\partial \kappa_{fg}} &= \frac{\partial}{\partial \kappa_{fg}}\left[(\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\mathbf{s}(\mathbf{p})\right] \\ &= (\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\left[\frac{\partial}{\partial \kappa_{fg}}(\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p})\right](\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\mathbf{s}(\mathbf{p}) \\ &= (\mathcal{H}(\kappa,\tau)\odot\Omega(\mathbf{p}))^{-1}\left[\tau\times I_{fg}\odot\Omega(\mathbf{p})\right]\boldsymbol{\eta_t}(\kappa;\theta_2)\end{aligned}$$

Where $I_{fg}$ is a $J_t \times J_t$ matrix of zeros and ones, where the ones occur if and only if the row product is owned by $f$ and the column by $g$. The result are two $J \times J$ matrices multiplying a $J \times 1$ vector which results in a $J \times 1$ vector.

where $I_{f=g}$ represents a $\mathcal{J}_t \times \mathcal{J}_t$ matrix of zeros and ones, with ones if $\kappa_{fg} > 0$ and $f \neq g$.

In order to compute $E[\frac{\partial \eta_{jt}}{\partial \kappa_{fg}}|\mathbf{x_t}, \mathbf{w_t}, \mathbf{v_t}, \mathbf{z^e_{jt}}]$ or $E[\frac{\partial \eta_{jt}}{\partial \tau}|\mathbf{x_t}, \mathbf{w_t}, \mathbf{v_t}, \mathbf{z^e_{jt}}]$ we must replace $\Omega(\mathbf{p})$ and $\boldsymbol{\eta_t}$ with their exogenous simulated analogues $(\Omega(\mathbf{p})^*, \boldsymbol{\eta_t^*})$ as in the optimal instruments approach of Berry et al. (1999).

## B. Product Market Data Appendix

### B.1. Sample Selection

Our primary data source for unit sales and prices of ready-to-eat (RTE) cereal comes from the Kilts Nielsen Scanner Dataset. The data are organized by store, week, and UPC code. For each store–week–upc we observe unit sales as well as a measure of "average price" which is revenue divided by sales.[57] Because there is little price variation across stores within the same chain (DellaVigna and Gentzkow, 2019), we aggregate unit sales and revenues to the DMA-chain level.[58] We focus on the years 2007-2016.[59] We further restrict our attention to a set of six DMAs (cities): Boston, Chicago, Charlotte, Denver, Phoenix, and Richmond. We choose these six cities to provide some geographic diversity, and also because Nielsen reports high coverage of "all-commodities-volume" for supermarket sales in these DMAs.

Within these cities, we focus exclusively on the set of conventional supermarket sales (which Nielsen labels as "F" stores) and exclude convenience stores and pharmacies which sometimes sell RTE cereal ("C" or "D" stores) and mass- merchandise ("M" stores). We further restrict our attention to a set of DMA-chains where we observe a sufficiently high volume of individuals from the Nielsen Panelist Dataset as customers.[60] This allows us to construct chain-specific demographic

[57] Prices may vary within a UPC–store–week for a number of reasons: the first is that price changes may occur within the middle of the reporting week, the second is that some consumers may use coupons; according to the Kilts-Nilsen documentation, retailer coupons or loyalty card discounts are included in the "average price" calculation while manufacturer coupons are not.

[58] We provide supporting evidence in favor of chain-level pricing in Appendix B.9.

[59] We exclude the year 2006 because the set of stores observed in that year does not sufficiently overlap with stores observed in subsequent years.

[60] This causes us to lose less than 4.3% of overall sales (as measured in servings).

profiles of customers at the DMA-chain-year level rather than just by geography. We focus on two key demographics (income and presence of children) that were previously shown to be important in demand for RTE cereal (Nevo, 2000).[61]

The data are recorded at the level of a universal product code (UPC) with around 3500 unique UPCs, we consolidate multiple package sizes at the "brand level", where our definition of brand also includes flavor (Honey Nut Cheerios or Blueberry Frosted Mini-Wheats).[62] Consistent with Nevo (2000), we convert boxes of cereal into serving equivalents, and report prices and quantities on a per-serving basis.[63] Given this definition, we find that in a DMA-Chain-Week a typical consumer chooses among 100-250 unique "products".

We report some basic summary statistics for our main dataset in Table 2. We observe 1590 stores which we consolidate into 26 DMA-chains, over 522 weeks between 2007-2016. The average price per serving to be around 20 cents across our sample, though it varies substantially across products, chain-DMAs, and time. We find that at the DMA/city level, aggregate sales are between 1.5 and 8.5 million servings per week, with a typical box containing around 12-20 servings. We find that between 80%-90% of sales (as measured in servings) are to branded cereals, with the remainder being private label or store-brands.[64] We assign each product to its ultimate owner in each period, taking to account mergers and acquisitions.[65]

### B.2. Product Characteristics

Because we are not interested in the relationship between demand and the 19 nutritional characteristics *per se*, and because the characteristics are often correlated with one another we consolidate our product characteristics into a smaller number of principal components. This is particularly helpful because nutritional information often creates colinearity problems. For example total calories are a linear function of fat, carbohydrates, and protein; or sugar and dietary fiber are a source of carbohydrates. We report average nutritional information on a per-serving basis in Table B-2 We use a principal components approach to reduce the dimension of the product space and provide

[61] We've examined other characteristics such as race, but this often required slicing the Panelist data too thinly to observe differential patterns in cereal preferences.

[62] This is especially important because a large number of individual UPCs are associated with a single "brand" and new UPCs might represent new packaging for a movie-tie-in or small changes in product size. If a manufacturer changes the package size from 14oz to 12.5oz and keeps the price fixed, we don't want to misinterpret this as "new product" but rather we want to interpret this as a serving price change for an existing product. This causes us to "miss" some nonlinear pricing where "Family Packs" are often less expensive on a per serving basis than smaller boxes. Focusing exclusively on conventional supermarkets mitigates this somewhat, but not completely.

[63] Nutrient dense cereals generally display smaller serving sizes (by weight) which lead to much smaller serving sizes by volume. There is some evidence that serving sizes are chosen so that caloric content falls in the (100-150) range rather than measuring typical serving sizes by consumers. [CITE:CR] conducted a survey and found that 92% of consumers exceed the posted serving size when pouring bowls of cereal. The average overpour on Cheerios was 30%-130% while it was even greater for denser cereals like Museli or Granola where the average overpour was 282%.

[64] We provide additional detail on how the private label share varies across chains, DMAs, and with the business cycle in Appendix B.4, see Figure B-3.

[65] For example, prior to September 2014 we assign Annie's Homegrown products as belonging to Annie's Homegrown and independent firm, and after September 2014, we assign them to General Mills.

| | mean | standard deviation |
|---|---|---|
| calories | 133.636 | 43.111 |
| calories_from_fat | 12.785 | 11.659 |
| vitamin_a_dv | 7.715 | 6.637 |
| vitamin_c_dv | 9.910 | 15.105 |
| iron_dv | 30.718 | 24.454 |
| calcium_dv | 4.844 | 10.456 |
| serving_weight (g) | 36.111 | 11.604 |
| As a Percentage of Weight | | |
| cholesterol | 0.080 | 0.515 |
| dietary_fiber | 0.075 | 0.053 |
| monounsaturated_fat | 0.007 | 0.015 |
| polyunsaturated_fat | 0.006 | 0.012 |
| protein | 0.077 | 0.042 |
| saturated_fat | 0.006 | 0.015 |
| sodium $\times 1000$ | 4.403 | 2.186 |
| sugars | 0.264 | 0.125 |
| total_carbohydrate | 0.811 | 0.076 |
| total_fat | 0.041 | 0.035 |
| other_carbs | 0.472 | 0.114 |

Table B-2: Summary Statistics of Product Characteristics (3478 Products)

a lower dimensional basis of *orthogonal* components. We describe our procedure in detail below.

Consider a $(J \times K)$ matrix $X$ with rows denoting products and columns denoting nutritional characteristics. Principal components constructs an eigen-decomposition of the covariance matrix $(X'X) = W\Lambda W'$ where $W$ is the matrix of eigenvectors and $\Lambda$ is a diagonal matrix of eigenvalues. We can approximate the covariance matrix by considering the largest $k < K$ eigenvalues and their corresponding eigenvectors and construct a basis in dimension $k$ rather than $K$. This yields the approximation $(X'X) \approx W\Lambda_k W'$ where $\Lambda_k$ is the same diagonal matrix as $\Lambda$ but with zeros for all eigenvalues except the $k$ largest. This lets us write $\tilde{X} = XW_k'$ where $W_k$ is the $J \times k$ matrix of eigenvectors which correspond to the $k$ largest eigenvalues in $\Lambda$. $\tilde{X}$ is an orthonormal approximation in dimension $k$ to the basis spanned by the covariance matrix $(X'X)$ in dimension $K$.

This has some advantages: for one we have reduced the number of dimensions of nutritional information that we need to keep track of. This is especially helpful when we want consumers to have heterogeneous preferences (random coefficients) across these characteristics. The second advantage is that the columns of $\tilde{X}$ are now orthogonal. In general this makes estimating parameters much easier because we no longer have to worry about columns of $X$ being highly correlated with one another (such as sugar and carbohydrate content). The disadvantage is that the transformed components of $\tilde{X}$ can be difficult to interpret.

When we apply this to our dataset, instead of working with the covariance matrix $(X'X)$ we work with the weighted covariance matrix $(X'\Omega X)$ where $\Omega$ is an $J \times J$ diagonal matrix of weights where the weights correspond to the brand-level market shares for each row of $X$. This prevents less popular brands with strange product characteristics from becoming outliers and distorting our transformed product space. It is also helpful to pre-standardize and scale the nutritional informa-

tion so that each column of $X$ (and $\tilde{X}$) has mean zero and variance one. The first five components comprise around 75% of the overall variation of the 19 nutritional characteristics. In the paper Figure 3 plots the first two principal components, and Figure B-1 plots the next two below.

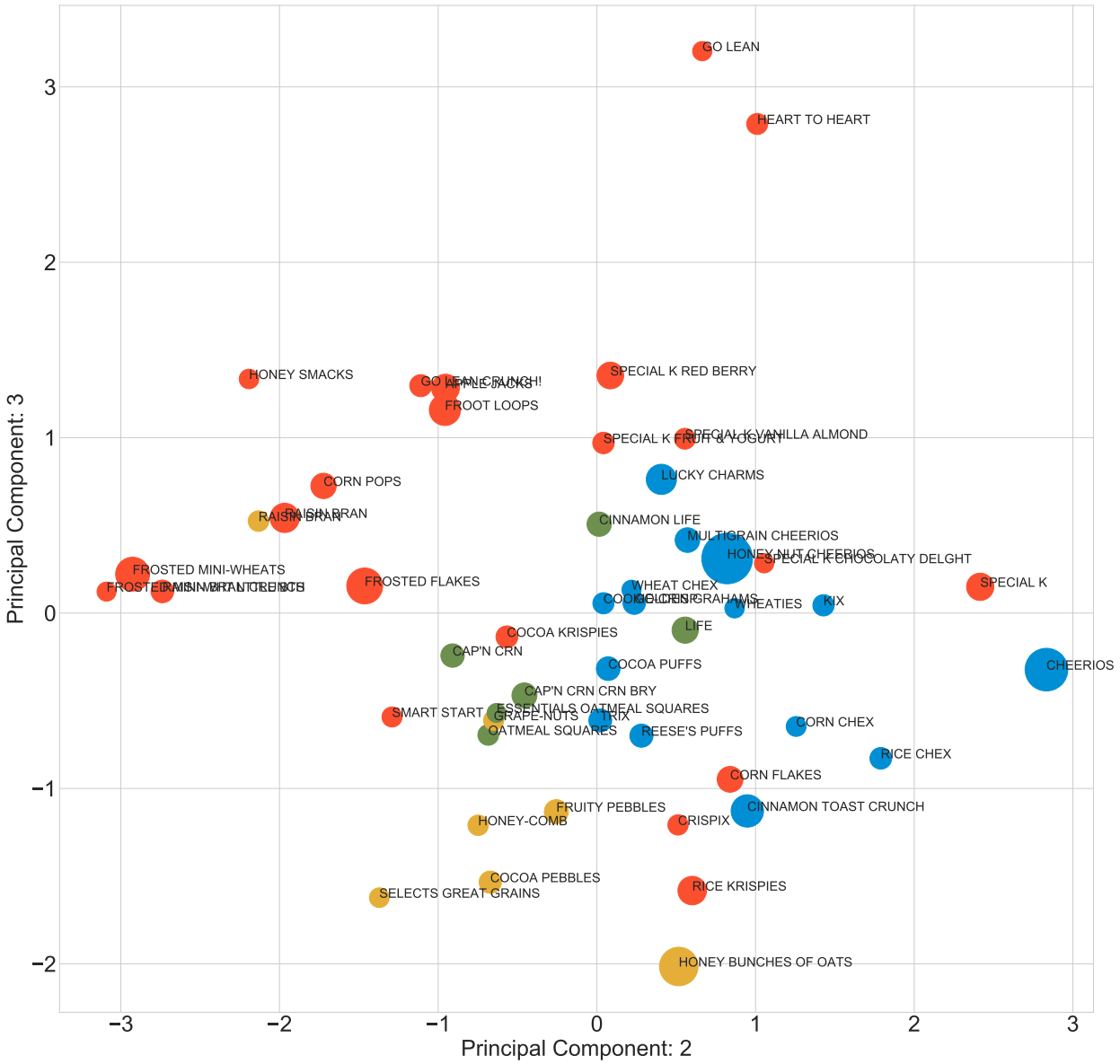



Figure B-1: Principal Components of Nutritional Data: Top 50 Brands
Circle size is proportional to the inverse of national sales rank. Colors indicate different manufacturers.

## B.3. Macro Variables

Here we document some additional macroeconomic varibles that we use in our demand estimates. One important variable is the unemployment rate. This is obtained from the St. Louis Federal Reserve (FRED) database. For each DMA, we find the corresponding (monthly) unemployment series for the corresponding Metropolitan Statistical Area (MSA). For most MSAs, the unemployment rate peaks in early 2010, but the magnitude of the swing is much larger in Charlotte than in Boston. The rate of recovery also varies across MSA, and is slower in Chicago, for example, than in other cities. We do simple linear interpolation to construct a weekly estimate of the unemployment rate.

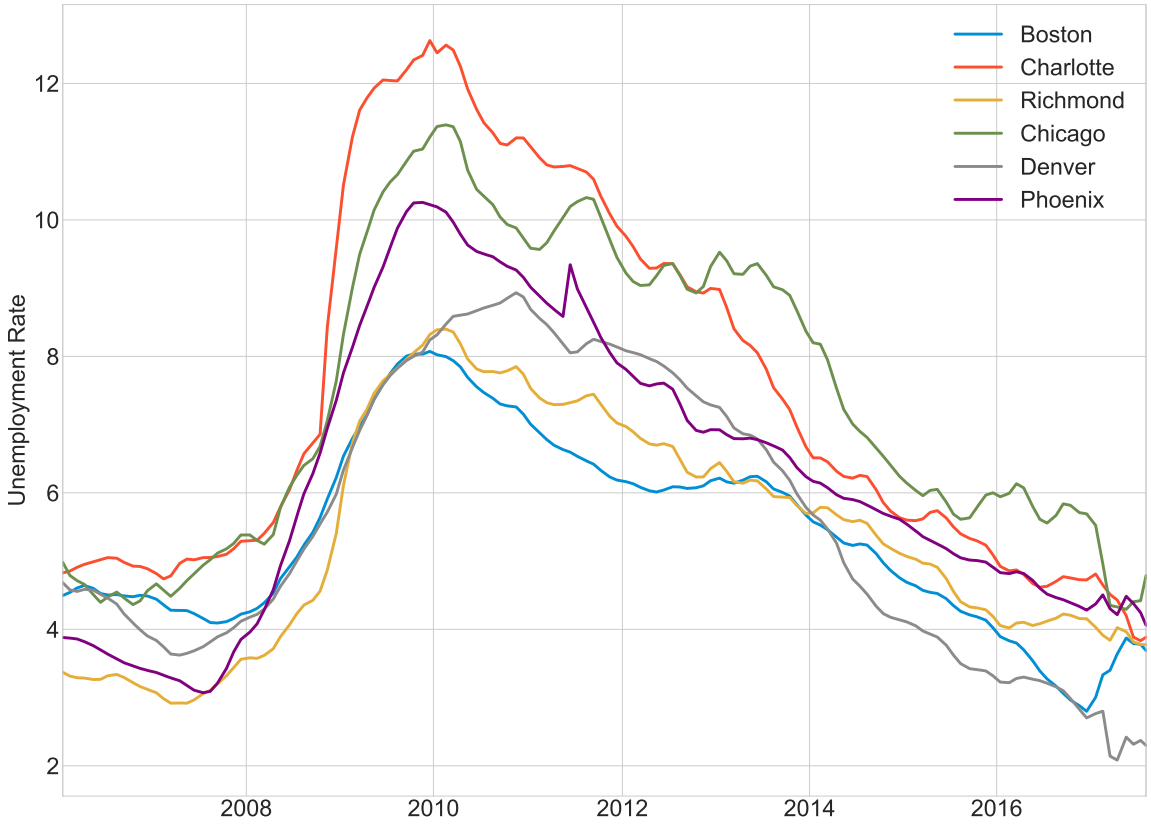



Figure B-2: Unemployment Rate by MSA/DMA

## B.4. Private Label

Below, in Figure B-3 we show the share captured by private label sales over time in our sample. The private label share grows during the Great Recession. It begins to decline immediately after the BlackRock acquisition of iShares from Barclays. During this period the four major brands all cut prices substantially, and then begin to increase prices in 2012. From 2013 onwards, the private

label share continues to decline, press accounts often attribute this to rising incomes during the recovery rather than pricing changes by the four major manufacturers.

We also plot the private label share for each DMA/city. There is a lot of heterogeneity across cities, for example Chicago (and to a lesser extent Charlotte) have a much lower private label share than other cities. This may be due to a number of factors, the most likely of which are the particular stores and chains that operate in each DMA/city and their own focus on private label brands. It may also be due to persistent geographic preferences, or heterogeneity in incomes across cities. For example, the private label share in Phoenix follows the business cycle more closely than in other cities. The cross-sectional variation in private label share is an important driver in the cross-sectional variation in the concentration measures ($HHI$ and $MHHI\Delta$).

### B.5. Concentration and Modified Concentration Measures

Below we plot the Herfindahl-Hirschman Index (HHI) and common-ownership modified HHI Bresnahan and Salop (1986) for each of our six DMAs over time. When constructing the concentration measures, we compute marketshares for each of the four major manufacturers (General Mills, Kellogg's, Quaker (PepsiCo), and Post) making sure to incorporate any brands they acquire during the sample. For example in September 2014, General Mills acquires Annie's Homegrown and independent organic focused brand. The largest remaining manufacturers include Malt-O-Meal which produces both its own products and also produces approximately 50% of the store-brand and private label products in the United States. We treat the Malt-O-Meal branded products as a distinct firm. We also treat private label products as a single monolithic firm. This makes sense because our unit of observation is a DMA-chain, so we assume that prices are being set by the retail chain.

### B.6. Instruments

The so-called BLP instruments are functions of characteristics of other goods within the same market $f(\tilde{x}_t)$. We follow the "best practices" of Conlon and Gortmaker (Forthcoming) and begin with the recipe in Ghandi and Houde (2019). First we construct a product-characteristic specific differentiation measure. We suppress the market subscript $t$ to economize on notation, and consider the distance between products $(j,k)$ in the dimension of characteristic $l$, and we let $x_j^{(l)}$ represent the suitably standardized $l$th principal component for product $j$ in matrix $\tilde{X}$:

$$d_{jk,(l)} = \tilde{x}_k^{(l)} - \tilde{x}_j^{(l)}$$

We form a second-order basis using $d_{jk,(l)}$ including interactions with other characteristics $d_{jk,(m)}$:

$$z_{jt}^{\text{quad}} = \left\{ \sum_k d_{jk,(l)}^2, \sum_k d_{jk,(l)} \times d_{jk,(m)} \right\} \quad \forall (l,m)$$

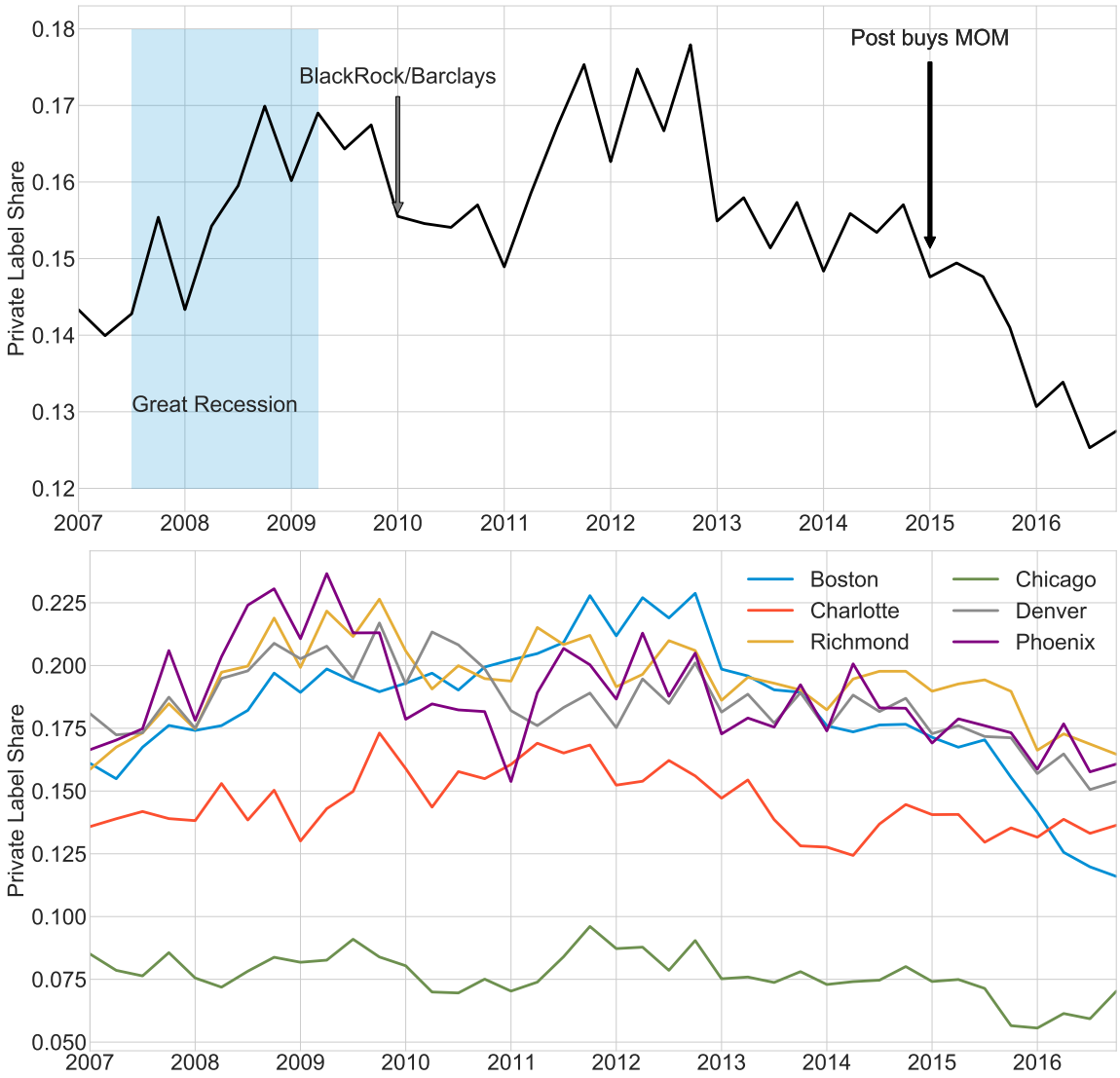



Figure B-3: Private Label Market share

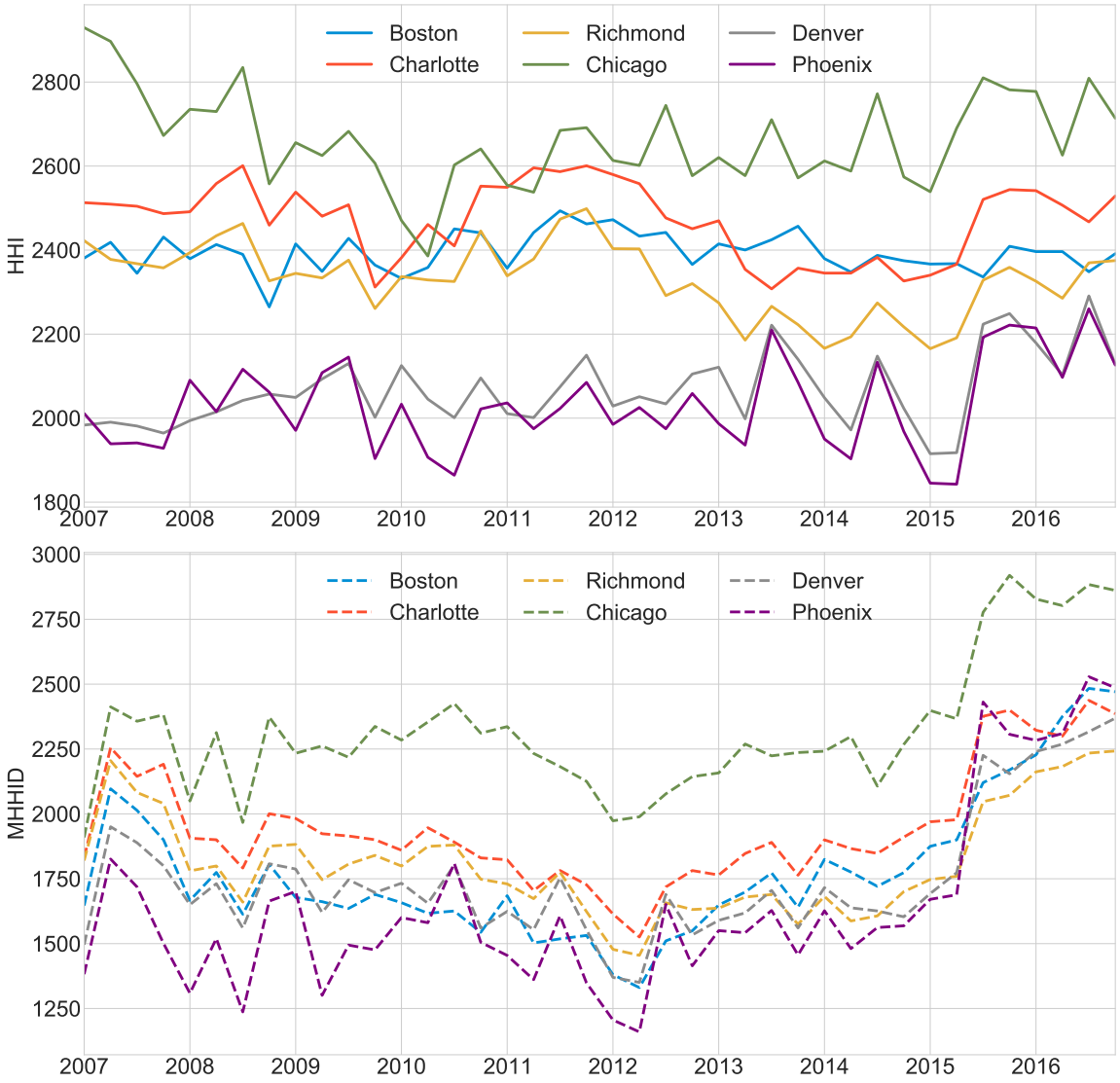



Figure B-4: HHI and $MHHI\Delta$ over time by City/DMA

We then use that second-order basis, along with the observable cost shifters $w_{jk}$ to form a measure of expected price (via a linear or polynomial regression). As in the demand model we allow $x_{jt}$ to contain product specific intercepts:

$$E[p_{jt}|x_{jt}, w_{jt}, z_{jt}] = \hat{p}_{jt}$$

We compute the differentiation measure in terms of "expected price":

$$d_{jk,(p)} = \hat{p}_k - \hat{p}_j$$

This allows us to extend $z_{jt}^{\text{quad}}$ as:

$$z_{jt}^{\text{quad}} = \left\{ \sum_k d_{jk,(l)}^2, \sum_k d_{jk,(p)}^2, \sum_k d_{jk,(l)} \times d_{jk,(m)}, \sum_k d_{jk,(l)} \times d_{jk,(p)} \right\} \quad \forall (l,m)$$

Ghandi and Houde (2019) show these instruments are most useful in the estimation of the standard deviations and covariances of the random coefficients $\Sigma$. Since we are also interested in estimating demographic interactions $\Pi$, we follow their suggestion and interact these instruments with moments of the distribution of demographics. In our case the moments of the demographic distribution we use are: mean of income, the standard deviation of income, and the fraction of households with children.

$$z_{jt}^D = z_{jt}^{\text{quad}} \otimes \{1, \overline{y_{it}}, \text{sd}(y_{it}), \%kids_t\}$$

After obtaining consistent initial estimates of $[\widehat{\theta}_1, \widehat{\theta}_2]$ we construct the feasible approximation to the optimal instruments (Amemiya, 1977; Chamberlain, 1987; Berry et al., 1999; Reynaert and Verboven, 2014). Because we estimate the demand side separate from any supply side, we are unable to exploit the additional cross-equation restrictions (see Conlon and Gortmaker (Forthcoming)). The simplified version of the expression in Conlon and Gortmaker (Forthcoming) becomes:

$$\frac{\partial \xi_j}{\partial \theta_1} = \begin{bmatrix} x_j, & v_j \end{bmatrix}, \quad \frac{\partial \xi_j}{\partial \theta_2} = \left[ E\left[ \frac{\partial \xi_j}{\partial \alpha} | x_t, v_t, w_t, z_t^D \right], \quad E\left[ \frac{\partial \xi_j}{\partial \widetilde{\theta}_2} | x_t, v_t, w_t, z_t^D \right] \right]$$

The first expectation is simply the expected price from above $E[p_{jt}|x_{jt}, w_{jt}, z_{jt}] = \hat{p}_{jt}$, while the second expectation is the Jacobian $\frac{\partial \xi_j}{\partial \widetilde{\theta}_2}$ evaluated at $\hat{p}_{jt}$ and $\xi_t = 0$. For each set of instruments we construct, we standardize each column to mean zero and variance one, and then perform the principal components dimension reduction outlined above.[66]. This has the advantage that it produces a well-scaled and orthogonal set of instruments, and has the added benefit that the identity matrix is often close the the optimal weighting matrix.

[66] We make sure our principal components always span at least 99% of the variance of the initial set of instruments

Figure B-5: Average Package Size Over Time: REDACTED

To measure the relevance of the instruments, we report the fit of the two different first-stage estimates below.The first panel looks at the first stage F-statistic in a regression of prices on covariates (with and without fixed effects) $E[p_{jt}|x_{jt}, w_{jt}, z_{jt}]$, the second mimics the IIA test of Ghandi and Houde (2019) which regresses:

$$\log s_{jt} - \log s_{0t} - \widehat{\alpha} p_{jt} = d_j + \lambda_1[x_{jt}, v_{jt}] + \lambda_2[z^D_{jt}, w_{jt}] + \epsilon_{jt}$$

and reports the F-statistic of the hypothesis that $\lambda_2 = 0$ (with various sets of fixed effects). We find evidence both that our instruments for price are strong $F > 10$ and that we can clearly reject the null hypothesis that the IIA model is correctly specified.

Table B-3: First Stage for Price [UPDATE]

| | | | | | |
|---|---|---|---|---|---|
| Product FE | No | Yes | No | No | No |
| Product-Chain FE | No | No | Yes | No | Yes |
| Time FE | No | No | No | Yes | Yes |
| $N$ | 912386 | 912310 | 909901 | 912386 | 909901 |
| adj. $R^2$ | 0.080 | 0.837 | 0.893 | 0.097 | 0.894 |
| F | 15864.5 | 270.0 | 264.1 | 17979.7 | 15.60 |

$t$ statistics in parentheses

## B.7. Market Size

Recall that our market is defined as a DMA-chain-week and we have 21,833 markets. In other words we combine weekly sales for stores within the same chain, so long as those chains belong to the same Nielsen DMA. Our market definition is: the number of people who walk into the store during a particular week (summed over stores within the same chain and DMA).

Since we don't directly observe the number of consumers to walk into a store, we estimate this number using two other commonly purchased categories: milk and eggs. We run the following fixed effects regression:

$$\log q^{RTEServings}_{c,t} = \beta_1 \log q^{milk}_{c,t} + \beta_2 \log q^{eggs}_{c,t} + \gamma_c + \varepsilon_{ct}$$

We predict the weekly number of RTE Cereal servings purchased as a function of milk sales (in gallons) and egg sales (in dozens) along with fixed effects at the DMA-Chain level. We don't include any time varying fixed effects because those might be correlated with the overall price level for RTE cereal. Instead we rely on the milk and eggs purchases to capture the time varying traffic of individual chains. We report the results of those regressions in Table B-4. We find that on its own 1% increase in milk sales increases the sales of RTE Cereal (measured in servings) by 0.88%

while a 1% increase in egg purchases increases cereal purchases by 0.81%. While not displayed in the table, chain fixed effects alone explain around 94% of the variation in the cereal sales, while the addition of milk and egg sales increase the $R^2$ to almost 99% and the within $R^2 = 0.7742$.

| | log(Servings) | | |
|---|---|---|---|
| log(milk) | 0.886*** | | 0.535*** |
| | (0.0666) | | (0.0572) |
| log(eggs) | | 0.809*** | 0.407*** |
| | | (0.0587) | (0.0648) |
| Observations | 21,836 | 21,833 | 21,833 |
| R-squared | 0.983 | 0.982 | 0.987 |
| Chain FE | Yes | Yes | Yes |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table B-4: Predicted Market Size Using Milk and Eggs Sales

We then define our market size as follows, where $\lambda$ is a constant chosen to match the (average) outside good share of $s_0 = 1 - 0.276 = 0.723$:

$$\hat{M}_{c,t} = \lambda \cdot \left( \hat{\beta}_1 \log q_{c,t}^{milk} + \hat{\beta}_2 \log q_{c,t}^{eggs} + \hat{\gamma}_c \right)$$

In this process we find a very small number of outliers and drop 249 out of 21,833 observations (1.1% of the sample by observations, 0.8% by revenue) by trimming the largest and smallest values by the outside good share. We report the resulting distribution of the inside good share $1 - s_{0ct} = \frac{\sum_j s_{jct}}{M_{ct}}$ in Figure B-6, which we have standardized around the 27.6% purchase probability. This is a bit different from the usual approach. Most papers, including Nevo (2001), construct market size using population demographics for the surrounding area (the population of the DMA, the population of the zipcode, etc.). That approach generally does not capture the fact that some stores face more competition than other stores (ie: we often don't know if there is a major competitor nearby or not). One disadvantage of our approach is that it can become potentially contaminated if sales of milk or eggs respond to the price for cereal or if consumers go to the supermarket just to buy cereal. Even if cereal and milk are complements (or Leontief in breakfast production) as long as the price of cereal doesn't causally impact the sales of milk and eggs we should be fine.

## B.8. Demographics and Micro Moments

The first form of micro-moment are covariances between the observed demographics and the observed characteristics conditional on the purchase of some product. We then take the average of this market-level covariance across markets:

$$\frac{1}{I_t \cdot T} \sum_{t=1}^{T} \sum_{i=1}^{I_t} w_i (d_{it} - \overline{d}_t)(x_{it} - \overline{x}_t) - b$$

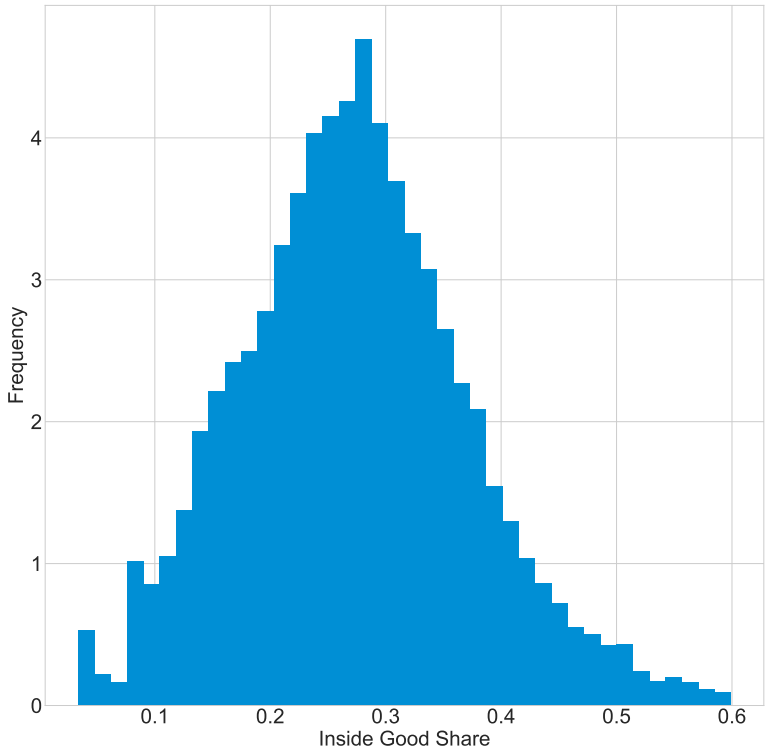



Figure B-6: Distribution of Inside Good Shares

We construct our estimate of the true covariance $b$ using the Nielsen Panelist data. That is we compute the covariance between the panelists income and the price per serving (or principal component) for the product they purchase. We construct this covariance moment conditional on purchase (for the set of households purchasing cereal). We construct similar micro moments for the covariance between the indicator for presence of children in the household and the product characteristics (including price per serving).

We use the panelist data to construct two additional moments, we compute the expected value of income, and the expected probability of children in the home conditional on not-purchasing cereal.

## B.9. Chain Pricing

As in DellaVigna and Gentzkow (2019), we find that for the most part prices do not vary across chains. We provide a plot in Figure B-7 in which each pixel represents a store-week price for a single UPC within a large chain. Each row tracks prices within a single store over time. We arrange the rows by their three-digit zipcode so that stores closer together in physical space are closer together within the matrix. Each column represents a weekly observation. A dark vertical bar represents a temporary sale. We can see visually from the heatmap plot that within the chain stores have very common prices and even time temporary sales simultaneously. There is some geographic dispersion. The colors in the top and bottom half of the graph are somewhat different from one another. This split corresponds to two separate DMAs (Denver and Phoenix) with prices being slightly higher in the Phoenix DMA than the Denver DMA. The white pixels are store-weeks where we are missing data.

This lets us aggregate sales (and prices) within a Chain-DMA pair. We do not aggregate across multiple DMAs as our figure indicates that there may be some DMA-level variation in chain pricing.

Figure B-7: Chain Pricing Heatmap : Redacted

[BREAK] The easiest estimator to work with is the EL estimator because it provides likelihood units.[67] We can consider the ELR test statistic which is just $ELR = (\sum_i \hat{\pi}_i) - N \cdot \log N$. We can now use likelihood ratio style test statistics. This has the advantage that unlike say the pairwise Cox testing procedure used in Villas-Boas (2007) we can test and rank all models at once by their ELR statistics rather than using the optimal weighting matrix associated with model $A$ to test model $B$ and the optimal weighting matrix associated with model $B$ to test model $A$ (which could, in practice, lead to such hard-to-interpret results as each model rejecting the other). We can also compare optimized CUE objective functions if we wanted something more in line with traditional

[67] This is often called a non-parametric maximum likelihood estimator NPMLE. See (**?**).

GMM estimators but without the problem of multiple optimal weighting matrices.

## B.10. Product Recalls

We use a number of product recalls as demand shifting instruments $v_{jt}$ in the notation of the model. We report a list of recalls in Table B-5 and report the effect of the recall on sales in Figure B-8. We only use the three Kellogg's recalls in our data. For the Malt-O-Meal Puffed Wheat and Puffed Rice Recall and the Post Great Grains Protein Blend recall we don't observe any sales after the recall period so it is as if those products are removed from the choice set. Variation in the choice set enters through $x_{-j,t}$ rather than $v_{jt}$.

| Date | Affected Brands | Reason |
|---|---|---|
| 6/25/10 | Frosted Mini-Wheats | Metal Shards |
| 10/10/12 | Honey Smacks, Froot Loops, Corn Pops, and Apple Jacks | GI Issues |
| 2/20/13 | Special K Red Berries | Glass Fragments |
| 04/04/08 | M-O-M Puffed Wheat and Rice* | Salmonella |
| 06/10/16 | Post Great Grains Protein* | Listeria |

Table B-5: List of Product Recalls

* denotes products which were permanently withdrawn from the market after recall.

## B.11. Demand Estimation

Below is the problem definition in `pyblp`.

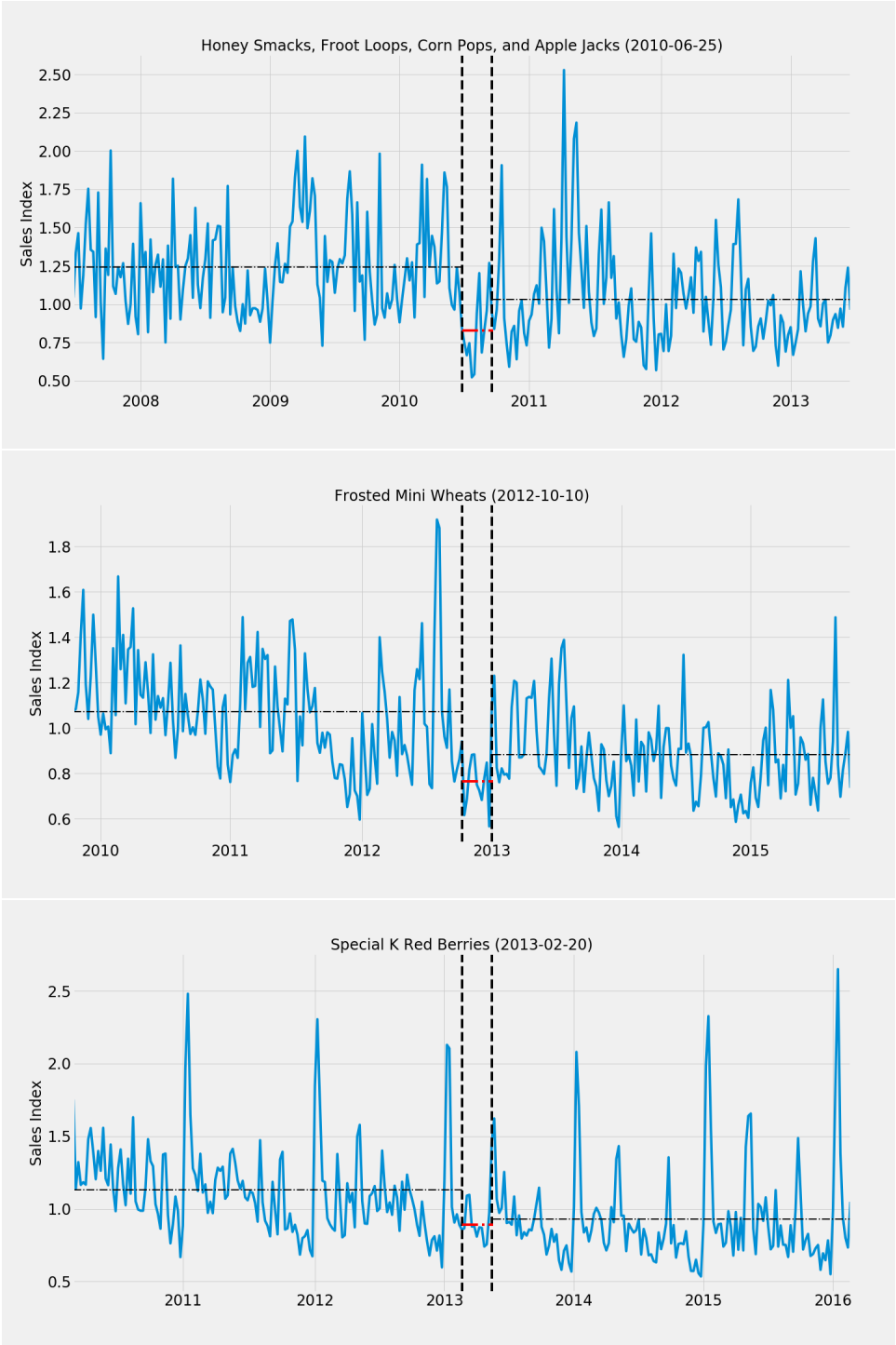



Figure B-8: Effects of Product Recals

Vertical lines denote date of recall and 12 weeks later (the period we consider the demand shock $v_{jt}$ applies to).
Horizontal lines denote the mean of each period (before, during, after) the recall.

## C. Ownership Data Appendix

While the main data source on holdings used in this literature is the Thomson Reuters S34 database, we instead use our hand-collected dataset described in Backus et al. (2021) due to the problems with the Thomson dataset discussed in **?**.

Additionally, we gather data on insider holdings from the Thompson Reuters Insider Holdings database accessible via WRDS. This compiles SEC forms 3,4,5, and 144 to get a picture of insider holdings that we may be missing from the 13f filings — in particular, holdings of executives and board members. We find that these holdings are always less than five percent of shares outstanding (and typically substantially so) and so, for the RTE cereal industry, have a minor effect on the computed common ownership profit weights.

## D. Problems with the MHHI Regression Approach

### D.1. Discussion of MHHI-based Approaches to Testing [Move this?]

As summarized above, the literature on common ownership so far has largely relied on MHHI measures of market concentration; these papers regress market-level prices on HHI and MHHID, where the coefficient on the latter, if positive and statistically significant, is taken as evidence for common ownership effects on pricing.

The problem is that symmetric Cournot is a restrictive framework.[68] If the world is actually characterized by product differentiation, as workhorse empirical models in IO typically assume, MHHI-based measures fare particularly poorly. This is because they are functions of market shares, or quantity. So, any misspecification in demand means that the misspecification bias will be correlated with quantities, and quantities are, as we know, endogenous to prices. In short, misspecification pursuant to the assumption of symmetric Cournot pricing will introduce a non-classical measurement error problem.

This problem is generic to HHI and MHHI measures, and has led to the wholesale abandonment of their use in modern empirical IO. In Appendix D.1 we discuss these problems in detail, and offer a numerical example to show how the misspecification bias could yield both positive and negative correlations between MHHID and price, even when MHHID is "instrumented" with upstream financial shocks in an (unsuccessful, we show) attempt to isolate good variation in common ownership incentives. We also replicate this approach for our own setting, RTE cereal, and estimate a negative coefficient — a stark and theoretically uninterpretable result which we contend is most easily reconciled not as a causal relationship, but a critique of the MHHI measure more broadly.

There are several other reasons we prefer a direct focus on the common ownership weights $\kappa$ to the use of MHHI concentration measures. Many of these are related to measurement and market

[68] Symmetry is a necessary condition for interpretation. O'Brien (2017) shows that the mapping from MHHI to prices is no longer one-to-one when that assumption alone is relaxed; therefore under asymmetric Cournot, a regression of prices on MHHI is not a "reduced form" of anything.

Table D-6: Example Ownership Structure

| | Firm 1 | Firm 2 |
|---|---|---|
| Investor 1 | 20% | - |
| Investor 2 | - | 20% |
| Investor 3 | 20% | 20% |
| Retail Share | 60% | 60% |

Notes: This table presents investor holdings in two firms for the example of Section D.2.

definition, and we discuss these in Backus et al. (2021) and **?**. More importantly for testing, however, MHHI reduces the entire market-level cooperation matrix to a single index. In doing so, it throws away what we view as the most useful source of variation: the within-market asymmetries in $\kappa$. For this reason we adopt a structural approach based on the Bertrand first-order conditions and a series of exclusion restrictions.

Here we offer an overview of some of the objections that have been raised to the MHHI approach to testing for common ownership effects in pricing. The approach involves regressing prices on HHI and MHHID using panel data, i.e. with multiple markets (e.g., geographic markets for local banking services, or airline routes) over time. As a robustness check, AST propose using the 2010 Blackrock-Barclays merger as an instrument. We consider both arguments in what follows. First, we offer an analytic example to show how the classical problem of price-quantity endogeneity introduces non-classical measurement error into the usual MHHI regresisons when applied to data generated from a Bertrand pricing game. Next we run MHHI regressions ourselves for the RTE cereal industry and use these observations to guide our interpretation.

## D.2. Analytic Example

We begin with two single-product firms playing a static (Betrand) pricing game. Suppose that the ownership structure of the two firms is as follows, based on Example 1 from Backus et al. (2021): 60 percent of each firm is held by small, undiversified retail investors. 20 percent of each is held, respectively, by two large, undiversified investors. The final 20 percent of each is held by a single, diversified investor. This ownership pattern is summarized in Table D-6. From these numbers we would compute $\kappa_{12} = \kappa_{21} = 0.5$, and MHHID $= 2 \cdot 0.5 \cdot s_1(a_1)\, s_2(a_1)$.

Despite the ownership structure assumed above, we will maintain the null hypothesis that firms are profit maximizers, and ignore rival profits. Our question is, under this null hypothesis, can we none the less find (spurious) evidence for common ownership using the MHHI approach?

Demand for both products is linear and given by:

$$\begin{aligned} q_1 &= a_1 - \alpha_1 \cdot p_1 + \beta_1 \cdot p_2 \\ q_2 &= a_2 - \alpha_2 \cdot p_2 + \beta_2 \cdot p_1. \end{aligned}$$

To further simplify we assume symmetric marginal costs of zero so that $mc_1 = mc_2 = 0$. Then, under the assumption of own-firm profit maximization, best replies are given by:

$$\begin{aligned} p_1(p_2) &= \frac{a_1 + \beta_1 \cdot p_2}{2\alpha_1} \\ p_2(p_1) &= \frac{a_2 + \beta_2 \cdot p_1}{2\alpha_2}, \end{aligned}$$

Which solves to:

$$p_i = \frac{2a_i\alpha_j + \beta_i a_j}{4\alpha_i\alpha_j - \beta_i\beta_j},$$

**The panel regression:** Suppose that across markets there is variation in $a_1$. We set $(a_2, \alpha_1, \alpha_2, \beta_1, \beta_2) = (100, 1, 1, 0.5, 0.5)$ and parametrize each "market" by a different demand intercept $a_1$ which we vary over $[80, 100]$ and so that:

$$\begin{aligned} q_1 &= a_1 - p_1 + \frac{1}{2} \cdot p_2 \quad \text{for } a_1 \in [80, 100] \\ q_2 &= 100 - p_2 + \frac{1}{2} \cdot p_1 \end{aligned}$$

We can calculate $p_1^*(a_1), p_2^*(a_1)$ as well as $HHI(a_1)$. These are depicted in Figure D-10, where we plot both prices and $\Delta MHHI$ as a function of the demand intercept $a_1$ holding all other quantities fixed.

For $a_1 < 100$ we find that there is a positive correlation between prices ($p_1$ in blue and $p_2$ in red) and $\Delta MHHI$, while for $a_1 > 100$ we find that there is a negative correlation between prices and $\Delta MHHI$. Therefore if the researcher observes sample 1, they will conclude that common ownership raises prices, and if they observe sample 2, they will find that common ownership lowers prices. Both results are spurious.[69]

**The Barclays acquisition instrument:** The response to this concern has been to use the Blackrock-Barclays merger in 2010, and show that results are robust, quoting from the abstract of AST, which introduced this strategy, "even when we only use variation in ownership due to the combination of two large asset managers." To this end they instrument using the within-market variation alone, i.e. the change in MHHID due to the merger.

Does this differencing exercise "clean out" the endogenous variation driven by demand shocks? We continue with the same paramterization of demand and variation in $a_1$.

An obvious objection would be that variation in $a_1$ may be correlated with exposure to the Blackrock-Barclays merger. The following argument does not rely on such a claim.

[69] The astute observer will note that the expression for MHHID here is log-linear, and so running the regression in logs would solve the problem. This is not generically true for $n > 2$ firms.

Continuing with our example, suppose that that in the ownership structure of Table D-6, the holdings of Investor 3 are the result of a merger of two symmetric investors, call them "Blackpebble" and "Pubclays." Therefore prior to the merger both investors held a 10% stake in both Firm 1 and Firm 2. This means that prior to the merger, we would compute $\kappa_{12} = \kappa_{21} = 1/3$. And so the differenced MHHID (pre-merger minus post-merger) would be equal to: $\Delta MHHID = 2 \cdot (1/3 - 1/2) \cdot s_1(a_1)\, s_2(a_1)$.

We plot this difference in Figure D-11. What is immediately apparent is that the differencing exercise has not removed the endogenity of market shares. If the researcher observes sample 1, then in markets where there was a small increase in $a_1$, there would be a rise in prices and a decrease in MHHID. Therefore they would conclude that MHHID and prices are negatively correlated. Alternatively, if the researcher observes sample 2, then in markets where there was an increase in $a_1$, there would be a rise in prices and a rise in MHHID, and they would conclude the opposite. The key insight is that the merger may have affected $\kappa$, but since the MHHID measure interacts with market shares, the endogeneity problem remains. As a result, both results would again be spurious.

### D.3. MHHI Regressions in RTE Cereal

We now perform the same set of price-concentration regressions as the previous literature: Azar et al. (2018a), Azar et al. (2016), Kennedy et al. (2017), Gramlich and Grundl (2017). We run the following regression:

$$\log p_{jmt} = \gamma_{mj} + \beta_1 HHI_{mt} + \beta_2 \Delta MHHI_{mt} + \beta_3 s_{jmt} + c_1 t + c_2 t^2 + \ldots + \epsilon_{jt}.$$

We report our regression results in Tables D-7 - D-10. We perform these regressions at two levels of aggregation. The first is at the manufacturer-chain-DMA-quarter level, and the second is at the manufacturer-DMA-quarter level. While we might want to use higher frequency time series data, because we only observe 13-F filings once per quarter such variation would be dubious to use to identify common ownership effects. We use both prices and log prices as our left hand side variable. When we construct prices we use the $P_V(t)$ version of the price index (just total revenue divided by total servings). We include fixed effects at the dma-level, chain-dma-level, manufacturer level, and a quadratic time trend. In some specifications we also include a cubic polynomial in HHI, in case $\Delta MHHI$ is picking up nonlinearities in the price-concentration relationship. In all of our specifications there is a negative and significant relationship between the $\Delta MHHI$ and price. This suggests that additional overlapping ownership is correlated with *lower prices* rather than higher prices. We don't interpret this as causal mechanism. That is, it is inappropriate to conclude that common ownership leads to lower prices. Instead we offer our numerical example from Section D.2 as an explanation: cross sectional variation in demand provides variation in market shares. When we interact these market shares with $\kappa_{fg}$ we can produce spurious positive or negative correlations

between $\Delta MHHI$ and price.

Table D-7: DMA Level MHHI Regressions $p$

| | (1) Price b/se | (2) Price b/se | (3) Price b/se | (4) Price b/se | (5) Price b/se |
|---|---|---|---|---|---|
| hhi_servings | 0.0123*** | 0.0221*** | -0.0062*** | 0.0021 | 0.0028* |
| | (0.0029) | (0.0031) | (0.0019) | (0.0015) | (0.0016) |
| mhhid_servings | -0.0261*** | -0.0254*** | -0.0351*** | -0.0232*** | -0.0243*** |
| | (0.0034) | (0.0034) | (0.0020) | (0.0016) | (0.0016) |
| DMA FEs | No | Yes | No | Yes | Yes |
| Manufacturer FEs | No | No | Yes | Yes | Yes |
| Quadratic Time Trend | Yes | Yes | Yes | Yes | Yes |
| Cubic HHI | No | No | No | No | Yes |
| $R^2$ | 0.110 | 0.204 | 0.831 | 0.903 | 0.904 |
| N | 1294 | 1294 | 1294 | 1294 | 1294 |

* p¡0.10, ** p¡0.05, *** p¡0.01

Table D-8: DMA Level MHHI Regressions $\log p$

| | (1) Log(Price) b/se | (2) Log(Price) b/se | (3) Log(Price) b/se | (4) Log(Price) b/se | (5) Log(Price) b/se |
|---|---|---|---|---|---|
| hhi_servings | 0.0396*** | 0.0730*** | -0.0141* | 0.0121* | 0.0028* |
| | (0.0115) | (0.0123) | (0.0074) | (0.0062) | (0.0016) |
| mhhid_servings | -0.1092*** | -0.1059*** | -0.1388*** | -0.0959*** | -0.0243*** |
| | (0.0134) | (0.0136) | (0.0075) | (0.0064) | (0.0016) |
| DMA FEs | No | Yes | No | Yes | Yes |
| Manufacturer FEs | No | No | Yes | Yes | Yes |
| Quadratic Time Trend | Yes | Yes | Yes | Yes | Yes |
| Cubic HHI | No | No | No | No | Yes |
| $R^2$ | 0.124 | 0.193 | 0.846 | 0.899 | 0.904 |
| N | 1294 | 1294 | 1294 | 1294 | 1294 |

* p¡0.10, ** p¡0.05, *** p¡0.01

Table D-9: DMA-Chain Level MHHI Regressions $p$

| | (1) Price b/se | (2) Price b/se | (3) Price b/se | (4) Price b/se | (5) Price b/se | (6) Price b/se |
|---|---|---|---|---|---|---|
| hhi_servings | 0.0130*** | 0.0234*** | -0.0067*** | 0.0021 | 0.0110*** | 0.0105*** |
| | (0.0017) | (0.0018) | (0.0016) | (0.0015) | (0.0013) | (0.0014) |
| mhhid_servings | -0.0269*** | -0.0265*** | -0.0358*** | -0.0240*** | -0.0181*** | -0.0192*** |
| | (0.0019) | (0.0019) | (0.0016) | (0.0015) | (0.0013) | (0.0014) |
| share_servings | | | | | -0.0026*** | -0.0026*** |
| | | | | | (0.0001) | (0.0001) |
| DMA FEs | No | Yes | No | Yes | Yes | Yes |
| Retailer FEs | No | No | Yes | Yes | Yes | Yes |
| Quadratic Time Trend | Yes | Yes | Yes | Yes | Yes | Yes |
| Parent FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Cubic HHI | No | No | No | No | No | Yes |
| r2_a | 0.229 | 0.295 | 0.720 | 0.767 | 0.820 | 0.820 |
| N | 5173 | 5173 | 5173 | 5173 | 5173 | 5173 |

* p¡0.10, ** p¡0.05, *** p¡0.01

As a final exercise we run a "true reduced form" regression as suggested by O'Brien (2017). Here

Figure B-9: Demand Estimation Code

```
>>> problem = pyblp.Problem(
...     product_formulations=(
            # Linear demand
...         pyblp.Formulation('0 + prices+ unemp_branded +v_0+v_1+v_2',
...         absorb='C(product_ids)+C(market_ids)'),
...         pyblp.Formulation('1+ prices + pca_0 + pca_1+pca_2')    # Nonlinear demand
...     ),
...     agent_formulation=pyblp.Formulation('0+kids+inc'),           # Demographics
...     product_data=df,
...     agent_data=agent_df
... )
>>> print(problem)

Dimensions:
=================================================
 T      N      F     I      K1    K2    D    MD    ED
----  ------  ---  ------  ----  ----  ---  ----  ----
1970  361105   7   563420   5     5     2    18    2
=================================================

Formulations:
================================================================================
       Column Indices:             0             1          2      3      4
------------------------------  ----------  -------------  -----  -----  -----
 X1: Linear Characteristics      prices    unemp_branded   v_0    v_1    v_2
X2: Nonlinear Characteristics      1          prices      pca_0  pca_1  pca_2
       d: Demographics         demog_kids    demog_inc
================================================================================
```

This is the definition of our demand model in `pyblp`.

Table D-10: DMA-Chain Level MHHI Regressions $\log p$

| | (1) Log(Price) b/se | (2) Log(Price) b/se | (3) Log(Price) b/se | (4) Log(Price) b/se | (5) Log(Price) b/se | (6) Log(Price) b/se |
|---|---|---|---|---|---|---|
| hhi_servings | 0.0495*** | 0.0857*** | -0.0105 | 0.0181*** | 0.0578*** | 0.0559*** |
| | (0.0068) | (0.0074) | (0.0066) | (0.0064) | (0.0057) | (0.0060) |
| mhhid_servings | -0.1216*** | -0.1204*** | -0.1523*** | -0.1097*** | -0.0833*** | -0.0892*** |
| | (0.0079) | (0.0081) | (0.0065) | (0.0064) | (0.0056) | (0.0059) |
| share_servings | | | | | -0.0116*** | -0.0117*** |
| | | | | | (0.0003) | (0.0003) |
| DMA FEs | No | Yes | No | Yes | Yes | Yes |
| Retailer FEs | No | No | Yes | Yes | Yes | Yes |
| Quadratic Time Trend | Yes | Yes | Yes | Yes | Yes | Yes |
| Parent FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Cubic HHI | No | No | No | No | No | Yes |
| r2_a | 0.267 | 0.312 | 0.722 | 0.754 | 0.814 | 0.814 |
| N | 5173 | 5173 | 5173 | 5173 | 5173 | 5173 |

* pj0.10, ** pj0.05, *** pj0.01

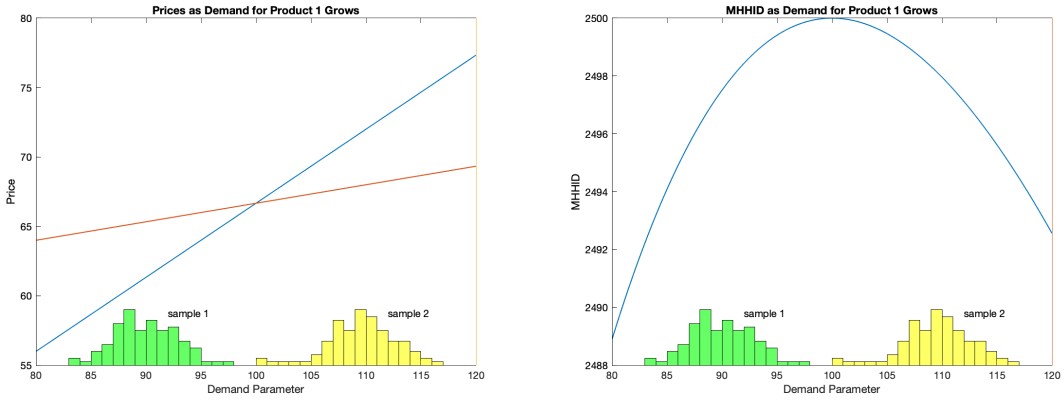



Figure D-10: Analytic/Numerical Example: Varying Demand Intercept

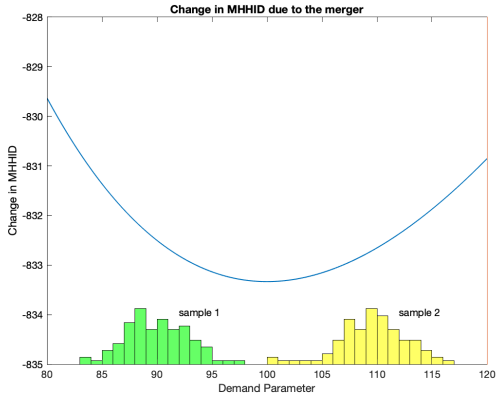



Figure D-11: Analytic/Numerical Example: Varying Demand Intercept

an observation is a manufacturer-chain-dma-quarter. The variable of interest is $\kappa$. Theoretically it should be the entire matrix $K$ which we summarize by using the arithmetic or geometric mean of the off-diagonal $\kappa$ entries. We report the results of the regression of average prices on these functions of $K$ in Table D-11. Once again we observe a negative (and significant) relationship between the common ownership profit weights $\kappa$ and the prices $p_{jmt}$. For some specifications the results are barely (in)significant.

## E. Testing Appendix

### E.1. Additional Pairwise Tests

Here we consider additional pairwise hypothesis tests using out framework. We begin by including perfect competition and perfect collusion as two extremal cases, and consider all pairwise combinations of conduct tests including these, profit maximization, and common ownership.

Table D-11: $\log p$ on $\kappa$ regression

| | Mean | Geometric Mean |
|---|---|---|
| | (1) | (2) |
| Market Share | -0.0026* | -0.0026* |
| | (0.0001) | (0.0001) |
| $\kappa$ | -0.0771* | |
| | (0.0115) | |
| $\kappa$ | | -0.0951* |
| | | (0.0120) |
| Retailer, Parent, DMA FE | ✓ | ✓ |
| Quadratic Time Trend | ✓ | ✓ |
| N | 5173 | 5173 |

Results for our four sets of exclusions are presented in Table E-12. To read this table, negative numbers favor the row-wise conduct model, while positive numbers favor the column-wise conduct model. The diagonal is undefined, and the lower triangle is symmetric to the upper triangle (by construction; the test is symmetric). We see that profit maximization is preferred to not only common ownership, but perfect competition and collusion as well.

We can break these same results out by firm, as we do in Table E-13. We see that the evidence against common ownership is somewhat weaker for Kellogg's, but this is unsurprising because, as we saw in Figure 2, they have the smallest common ownership weights ($\kappa^{CO}$), and so it is relatively more difficult to distinguish profit maximization from common ownership.

Next, we consider tests of profit maximization against the $\tau$-parameterized variations of common ownership. Results are presented in Table E-14. On the one hand, it is clear that others' costs and the demand optimal instruments offer clear rejections of any value of $\tau > 0$. Intuitively, we would expect the test statistic to converge to zero as $\tau \to 0$. Instead, the test statistic is stable in $\tau$. This reflects an underlying degeneracy of the Rivers-Vuong statistic when testing sufficiently similar models. In brackets, under each test statistic, we write the numerator and the denominator, the latter of which is the asymptotic standard error of the difference between the two models. As $\tau \to 0$, that standard error shrinks to zero because the models become equipollent, and the test statistic is degenerate. Our intuition is restored if we focus on the numerator, which is clearly monotone and diverging from zero as $\tau$ increases.

## E.2. Pakes-Wollmann Regressions

Here we use our four sets of exclusion restrictions to estimate the Pakes-Wollmann regressions detailed in Section 3.5. Recall that for this test, we say that the model is confirmed if $\hat{\gamma} = 1$.

Results are presented in Table E-15. The four sets of exclusion differ substantively in their implications. Others' costs are inconclusive; demographic exclusions and BLP instruments are most consistent with common ownership, and finally the demand optimal instruments cast a vote

Table E-12: Additional Hypotheses

| | Profit Max | Common Ownership | Perfect | Collusion |
|---|---|---|---|---|
| **Others' Costs** | | | | |
| Profit Max | - | -2.5571 | -2.4681 | -2.7059 |
| Common Ownership | | - | -1.9066 | -2.7775 |
| Perfect | | | - | -1.1028 |
| Collusion | | | | - |
| **Demographics** | | | | |
| Profit Max | - | -2.6003 | -2.6847 | -2.5516 |
| Common Ownership | | - | -2.0872 | -2.6642 |
| Perfect | | | - | -0.879 |
| Collusion | | | | - |
| **BLP Inst.** | | | | |
| Profit Max | - | -2.9191 | -2.9192 | -3.0706 |
| Common Ownership | | - | -2.2646 | -3.219 |
| Perfect | | | - | -1.2443 |
| Collusion | | | | - |
| **Dmd. Opt. Inst.** | | | | |
| Profit Max | - | -3.116 | -3.3885 | -3.2555 |
| Common Ownership | | - | -2.6742 | -3.3907 |
| Perfect | | | - | -1.3054 |
| Collusion | | | | - |

Notes: This table reports testing results for pairwise comparisons of different conduct models, as indicated by the row and column. The panel labels indicate the set of exclusions used in the Step 2 regression. A negative value constitutes evidence for row model and a positive value constitutes evidence for the column model. The null hypothesis is that they satisfy the criterion function equally well, and the critical values for rejection at $\alpha = 0.5$ are –1.96 and 1.96.

Table E-13: Additional Hypotheses by Firm

| | General Mills | | | | Kellogg's | | | |
|---|---|---|---|---|---|---|---|---|
| **Others' Costs** | Profit Max | Common Ownership | Perfect | Collusion | Profit Max | Common Ownership | Perfect | Collusion |
| Profit Max | - | -2.2732 | -1.611 | -2.5548 | - | -1.45 | -1.9016 | -2.6387 |
| Common Ownership | | - | -0.8248 | -2.6282 | | - | -1.8277 | -2.7001 |
| Perfect | | | - | -1.2465 | | | - | -0.9356 |
| Collusion | | | | - | | | | - |
| **Demographics** | | | | | | | | |
| Profit Max | - | -2.1394 | -1.7664 | -2.4701 | - | -1.3899 | -2.1431 | -2.5216 |
| Common Ownership | | - | -0.9534 | -2.5632 | | - | -2.0555 | -2.5802 |
| Perfect | | | - | -1.1029 | | | - | -0.7526 |
| Collusion | | | | - | | | | - |
| **BLP Inst.** | | | | | | | | |
| Profit Max | - | -2.4684 | -1.8368 | -2.9054 | - | -1.641 | -2.3678 | -3.0363 |
| Common Ownership | | - | -0.8616 | -2.9978 | | - | -2.2599 | -3.0911 |
| Perfect | | | - | -1.4759 | | | - | -1.1429 |
| Collusion | | | | - | | | | - |
| **Dmd. Opt. Inst.** | | | | | | | | |
| Profit Max | - | -2.6369 | -2.1915 | -3.0621 | - | -1.8304 | -2.6658 | -3.2037 |
| Common Ownership | | - | -1.1207 | -3.1487 | | - | -2.54 | -3.2578 |
| Perfect | | | - | -1.5415 | | | - | -1.2211 |
| Collusion | | | | - | | | | - |
| | **Quaker Oats** | | | | **Post** | | | |
| **Others' Costs** | | | | | | | | |
| Profit Max | - | -1.749 | -2.2883 | -2.1783 | - | -2.1508 | -1.9202 | -2.4236 |
| Common Ownership | | - | -1.3747 | -2.3534 | | - | -1.6392 | -2.3983 |
| Perfect | | | - | -0.721 | | | - | -0.8617 |
| Collusion | | | | - | | | | - |
| **Demographics** | | | | | | | | |
| Profit Max | - | -1.6486 | -2.6564 | -2.1565 | - | -2.0544 | -2.1703 | -2.3658 |
| Common Ownership | | - | -1.4841 | -2.4233 | | - | -1.8503 | -2.3596 |
| Perfect | | | - | -0.6289 | | | - | -0.7302 |
| Collusion | | | | - | | | | - |
| **BLP Inst.** | | | | | | | | |
| Profit Max | - | -1.9596 | -2.6735 | -2.5427 | - | -2.2828 | -2.1713 | -2.7349 |
| Common Ownership | | - | -1.3429 | -2.8117 | | - | -1.7999 | -2.7258 |
| Perfect | | | - | -1.0085 | | | - | -1.0921 |
| Collusion | | | | - | | | | - |
| **Dmd. Opt. Inst.** | | | | | | | | |
| Profit Max | - | -2.1256 | -3.041 | -2.7207 | - | -2.4638 | -2.5659 | -2.8826 |
| Common Ownership | | - | -1.5703 | -2.9736 | | - | -2.1574 | -2.8657 |
| Perfect | | | - | -1.0371 | | | - | -1.0983 |
| Collusion | | | | - | | | | - |

Notes: This table reports testing results for pairwise comparisons of different conduct models, as indicated by the row and column. The panel labels indicate the set of exclusions used in the Step 2 regression. A negative value constitutes evidence for row model and a positive value constitutes evidence for the column model. The null hypothesis is that they satisfy the criterion function equally well, and the critical values for rejection at $\alpha = 0.5$ are –1.96 and 1.96.

Table E-14: Profit Maximization vs. $\tau$

| | Others' Costs | Demographics | BLP Inst. | Dmd. Opt. Inst. |
|---|---|---|---|---|
| $\tau = .1$ | -1.8554<br>[-0.1776/0.0957] | -2.0084<br>[-0.2035/0.1013] | -2.1622<br>[-0.2383/0.1102] | -2.4412<br>[-0.2977/0.1219] |
| $\tau = .2$ | -1.6799<br>[-1.2163/0.7240] | -1.8167<br>[-1.3942/0.7674] | -2.0053<br>[-1.7181/0.8568] | -2.2366<br>[-2.1368/0.9554] |
| $\tau = .3$ | -1.7847<br>[-3.7215/2.0852] | -1.9202<br>[-4.2205/2.1980] | -2.1379<br>[-5.4462/2.5475] | -2.3772<br>[-6.7613/2.8443] |
| $\tau = .4$ | -1.9656<br>[-9.5633/4.8654] | -2.0758<br>[-10.4122/5.0160] | -2.3122<br>[-13.7536/5.9483] | -2.559<br>[-17.0692/6.6703] |
| $\tau = .5$ | -2.0145<br>[-20.3951/10.1243] | -2.1113<br>[-22.0896/10.4625] | -2.3662<br>[-29.6962/12.5501] | -2.6006<br>[-36.7038/14.1134] |
| $\tau = .6$ | -2.1318<br>[-40.0954/18.8082] | -2.2209<br>[-43.1363/19.4227] | -2.4923<br>[-58.6461/23.5311] | -2.7196<br>[-71.9892/26.471] |
| $\tau = .7$ | -2.2062<br>[-71.4729/32.3969] | -2.2781<br>[-75.0284/32.9344] | -2.562<br>[-103.5226/40.4063] | -2.7797<br>[-126.3524/45.4555] |
| $\tau = .8$ | -2.3405<br>[-120.9867/51.693] | -2.401<br>[-125.2252/52.1558] | -2.7021<br>[-174.948/64.7449] | -2.9124<br>[-212.0695/72.8169] |
| $\tau = .9$ | -2.4842<br>[-199.5463/80.3261] | -2.5184<br>[-202.79/80.5224] | -2.8343<br>[-286.0827/100.9349] | -3.0386<br>[-345.2732/113.6282] |

Notes: This table reports testing results for pairwise comparisons of own-firm profit maximization against $\tau$-adjusted common ownership (as indicated by row). The column label indices the set of exclusions used in the Step 2 regression. A negative value constitutes evidence for profit maximization and a positive value constitutes evidence for the $\tau$-adjusted model of common ownership. The null hypothesis is that they satisfy the criterion function equally well, and the critical values for rejection at $\alpha = 0.5$ are –1.96 and 1.96.

Table E-15: Pakes-Wollmann Regression Results

| | Others' Costs | Demographics | BLP Inst. | Dmd. Opt. Inst. |
|---|---|---|---|---|
| **Profit Max** | -0.4233 | 1.5141 | 1.3701 | 1.0028 |
| SE | 0.1435 | 0.4794 | 0.1766 | 0.0456 |
| First Stage Fstat | 8.7038 | 9.9748 | 60.187 | 482.7362 |
| R-squared | 0.2594 | 0.6647 | 0.6651 | 0.644 |
| **Common Ownership** | -0.3104 | 1.013 | 0.9633 | 0.6704 |
| SE | 0.0855 | 0.3864 | 0.1856 | 0.0397 |
| First Stage Fstat | 13.1769 | 6.8737 | 26.9474 | 285.2434 |
| R-squared | 0.3094 | 0.5968 | 0.5955 | 0.5726 |
| **Collusion** | -0.1223 | 0.201 | 0.191 | 0.1242 |
| SE | 0.0268 | 0.168 | 0.1295 | 0.0172 |
| First Stage Fstat | 20.9002 | 1.4312 | 2.1762 | 52.1729 |
| R-squared | 0.39 | 0.4462 | 0.4461 | 0.4425 |

Notes: This table reports results for the Pakes-Wollmann regression setup using different exclusion restrictions (indicated by the column), for different conduct models (indicated by the row). We report the coefficient on the markup, the standard error below that, and then the first-stage F-statistic and the $R^2$.

for profit maximization.

## E.3. Step 1 and Step 2 Regressions

Here we document the intermediate supply and markup regressions that make up Step 1 and Step 2 of our proposed test. While our main results use a random forest to estimate the supply regression, for expositional purposes we report results here for a simple linear regression. Estimates are presented, for each conduct hypothesis, in Table E-16. Results are remarkably consistent across different conduct models, although we observe substantial variation in the $R^2$ and within-$R^2$, which are lowest for the collusive conduct model. Reassuringly, almost all of the input cost variables enter positively (rice enters negatively, but insignificantly). The sales variable enters negatively, as we would expect, and the recall dummies enter positively.

We report mean-squared error for the fit of the Step 2 regressions in Figure E-12. Here we compare the random forest approach, our preferred specification, to a linear fit with a third-order orthogonal polynomial expansion of own and others' input costs. The random forest performs substantively better for all but the BLP instruments, and we believe this has to do with the fact that there are 143 of them.

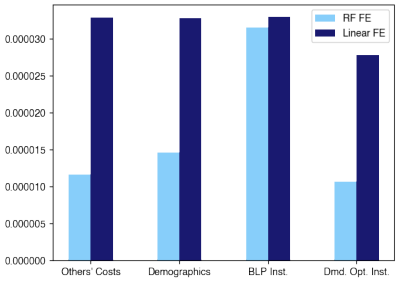
Table E-16: Linear Supply Regressions

| | Profit Max | Common Ownership | Perfect | Collusion |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Pct. Sales | -0.0594* | -0.0594* | -0.0740* | -0.0657* |
| | (0.0009) | (0.0009) | (0.0012) | (0.0010) |
| Corn | 0.0048* | 0.0050* | 0.0050* | 0.0051* |
| | (0.0011) | (0.0011) | (0.0013) | (0.0013) |
| Oat | 0.0040* | 0.0058* | 0.0049* | 0.0040* |
| | (0.0017) | (0.0019) | (0.0021) | (0.0019) |
| Rice | -0.0009 | -0.0016 | -0.0019 | -0.00005 |
| | (0.0023) | (0.0024) | (0.0029) | (0.0028) |
| Sugar | 0.0129* | 0.0127* | 0.0163* | 0.0140* |
| | (0.0023) | (0.0024) | (0.0032) | (0.0026) |
| Wheat | 0.0011 | 0.0015 | 0.0026 | 0.0015 |
| | (0.0015) | (0.0015) | (0.0018) | (0.0017) |
| Recall 1 | 0.0101* | 0.0117* | 0.0123* | 0.0081* |
| | (0.0033) | (0.0039) | (0.0045) | (0.0037) |
| Recall 2 | 0.0059* | 0.0043* | 0.0065* | 0.0049* |
| | (0.0021) | (0.0021) | (0.0030) | (0.0023) |
| Recall 3 | 0.0041* | 0.0033 | 0.0068* | 0.0029 |
| | (0.0020) | (0.0019) | (0.0025) | (0.0023) |
| Product | ✓ | ✓ | ✓ | ✓ |
| Quarter | ✓ | ✓ | ✓ | ✓ |
| Retailer × DMA | ✓ | ✓ | ✓ | ✓ |
| Observations | 1,404,106 | 1,404,106 | 1,404,106 | 1,404,106 |
| $R^2$ | 0.80179 | 0.80299 | 0.82532 | 0.74929 |
| Within $R^2$ | 0.34721 | 0.32784 | 0.34384 | 0.23974 |

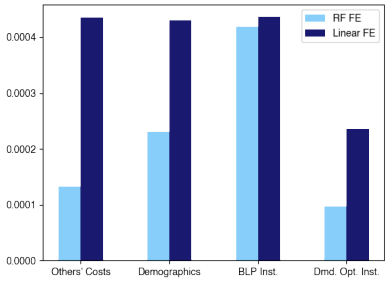
Notes: This table reports regressions of marginal costs, given by the column-wise conduct model, on included variables ($x_{jt}$ and $w_{jt}$). Standard errors are clustered at the retailer by DMA by year level. * denotes statistical significance at $\alpha = 0.05$.

Figure E-12: Step 2 Mean-Squared Error

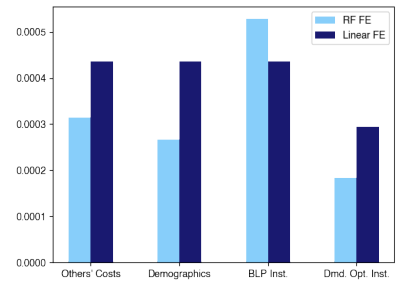
(a) Profit Max. vs. CO



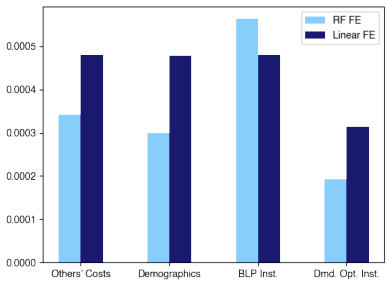

(b) Profit Max. vs. Collusion



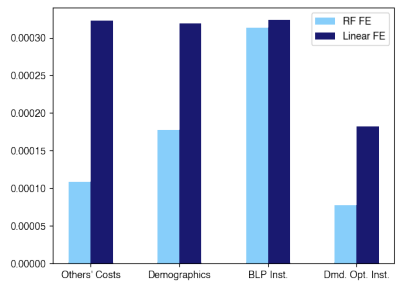

(c) Profit Max. vs. Perfect Comp.



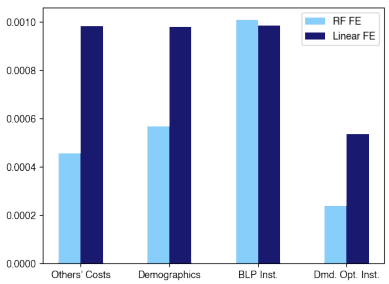

(d) CO vs. Perfect Comp.




(e) CO vs. Collusion




(f) Perfect Comp. vs. Collusion




Notes: This figure depicts mean-squared error for the Step 2 regressions. The panel name indicates the candidate model pair.