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

Matthew Backus† Christopher Conlon‡ Michael Sinkinson§

February 10, 2021

Abstract

Models of firm conduct are the cornerstone of both theoretical and empirical work in industrial organization. A recent contribution (Berry and Haile, 2014) has suggested the use of exclusion restrictions to test alternative conduct models. We propose a pairwise testing procedure based on this idea and show that the power of the test to discriminate between models is tied to the formulation of those restrictions as moments and how they reflect the nonlinearity of equilibrium markups. We apply this test to the ready-to-eat cereal market using detailed scanner and consumer data to evaluate the "common ownership" hypothesis, which has received significant attention. Although we show that the potential magnitude of common ownership effects would be large, our test finds that standard own-firm profit maximization is more consistent with the data.

*Thanks to conference and seminar attendees and especially Steven Berry, Jeremy Fox, Amit Gandhi, Philip Haile, Jean-Francois Houde, Robin Lee, Nathan Miller, Aviv Nevo, and Ariel Pakes for thoughtful comments. We thank Jett Pettus and Alphonse Simon for excellent research assistance. 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, NBER, and CEPR, matthew.backus@columbia.edu

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

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

1. Introduction

The strategic behavior of firms is central to the study of industrial organization. Going back to at least Bresnahan (1982), economists have tried to determine which models of firm conduct are consistent with observed market outcomes. These models of conduct often have very different implications for consumer welfare, market efficiency, antitrust policy, and regulation, and therefore being able to distinguish between them is vital to the science of economics.

One particular model of conduct, the "common ownership" hypothesis, says that as firms are increasingly owned by overlapping shareholders, the best way to maximize the value of shareholders' entire portfolios might be for firms to unilaterally relax competition in product markets. If manifested in firm conduct as the theory predicts – an open question – the implications of this observation would be vast, affecting prices in almost every sector of the US economy. By 2017, the three largest institutional asset managers (Vanguard, Blackrock, and State Street) held a combined 21% of the average S&P 500 firm (Backus et al., 2021). However, most attempts at measuring price effects of common ownership have relied on regressions of prices on (augmented) concentration indices using market level data, reminiscent of the Structure-Conduct-Performance (SCP) literature in industrial organization (Schmalensee, 1989). The challenges of these regressions are well-documented, and have led them to fall out of favor with most economists studying strategic interactions among firms (Berry et al., 2019).

In this paper, we develop a testing procedure for comparing non-nested models of conduct that predict different markups in equilibrium, such as own-profit maximization or the common ownership hypothesis. We follow the identification argument of Berry and Haile (2014) and build our test on exclusion restrictions, exploiting variables that affect markups but not marginal costs. We use a moment-based version of the Rivers and Vuong (2002) test statistic to detect violation of these restrictions: under the true conduct assumption, recovered marginal cost shocks should not be correlated with those variables but under an incorrect conduct assumption they will be. The challenge is that the relationship between the excluded variables and the difference between the markups under the two models is likely to be nonlinear. We propose a simple scalar moment condition that uses the model to capture that nonlinearity.

Our testing procedure has several advantages. By estimating the nonlinear relationship between the excluded variables and the difference in markups, we are able to generate a more powerful test. This idea is motivated by the Chamberlain (1987) optimal instruments for nonlinear GMM, an analogy that guides us at several points in the analysis. Moreover, since our formulation of the exclusion restrictions reduces to a scalar moment condition, we avoid many of the problems associated with weighting matrices in moment-based non-nested model comparisons (Hall and Pelletier, 2011).

We apply this test to alternative models of conduct in the ready-to-eat (RTE) cereal market. The chosen industry is an ideal one for examining the "common ownership" hypothesis 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 Kellogg's, 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. This variation generates meaningful differences both in the cross-section and in the time series between pricing and markups under common ownership and pricing under own-firm profit maximization.

We first estimate demand in a standard random coefficients logit model using detailed scanner data from 2007–2017. We supplement this with demographic data of shoppers collected from panelist data on individual consumers and micro-moments of covariances between purchased product characteristics and consumer demographics. We find that category demand is inelastic while the median product faces an own-price elasticity of -2.67. We simulate price increases under a set of hypothetical mergers to understand the magnitudes implied by the substitution patterns identified in the demand data.

With demand in hand, we compute markups under a variety of conduct models including perfect competition, own-profit maximization, common ownership pricing, and monopoly pricing. We apply our testing procedure using a variety of alternative exclusion restrictions. The results strongly favor own-firm profit maximization. However, we show that capturing the nonlinearity of the model using our formulation of the exclusion restriction is critical. Alternative specifications that are predicated on functional form restrictions – linearity of the marginal cost function and linearity of the relationship between the exclusion restrictions and the difference in markups – offer inconsistent results across different sets of excluded instruments. Alternatively, when specified as we propose, the different sets of exclusion restrictions agree, satisfying the specification test critique of Hausman (1978).

Finally, we consider an "internalization parameter" approach that posits that some fraction τ of the common ownership incentives are transmitted to managers. We compute our test statistic comparing own-profit maximization to variations of the common ownership incentives at different levels of τ. We show that our test is able to reject that 30% or more of the common ownership incentives are reflected in firm pricing decisions. At lower levels of internalization, markups between the two models become indistinguishable.

Our contribution is both methodological and empirical. In terms of methodology, we describe a method for testing conduct with a number of advantages over previous approaches. Our approach uses the (known) nonlinearity of the model of equilibrium markups to improve power. Moreover, it frees the researcher from specifying a weighting matrix and allows for fully flexible functional forms in the testing procedure. Empirically, we provide evidence in the ongoing debate around the

common ownership hypothesis. Our analysis is one of few in a differentiated product setting and the first to our knowledge to undertake a pairwise model comparison test of the common ownership conduct model. We show how to evaluate the common ownership hypothesis in a conduct testing approach, but our framework is quite general and it can be used to test other models or hypotheses that predict equilibrium markups. We hope that others will find this a fruitful approach in new empirical contexts going forward.

1.1. Related Literature

Empirical testing of firm conduct is an endeavor with a storied history in industrial organization (IO) going back to at least Bresnahan (1982). Examples of conduct testing and estimation include Bresnahan (1987), Nevo (2001), Miller and Weinberg (2017), and Duarte et al. (2020).[1] Berry and Haile (2014) show nonparametric identification of conduct and suggest that models of conduct are testable in the presence of appropriate exclusion restrictions. 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) treat the same question, characterizing "partial mergers," where diversified ownership by shareholders and cross-ownership among firms can have anticompetitive effects. In particular, Bresnahan and Salop (1986) introduced the modified Herfindahl-Hirschman index (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 most notable 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 addresses 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. Besides the obvious concerns for identification (since this puts quantity sold on the right-hand side), these concentration

[1] 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 2007–2017), more stores (6 versus thousands), more detailed characteristics data, store-specific measurement of market size and, most importantly, an alternative conduct hypothesis.

measures have no monotonic relationship to the main outcome of interest, product market prices (O'Brien, 2017). Notwithstanding these concerns about the empirical results, this burgeoning literature has attracted the interest of the legal community (Elhauge, 2016) and antitrust authorities,[2] and there is already interest in regulatory remedies (Posner et al., 2017). The literature is further reviewed in 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 papers that endeavor a structural analysis of the common ownership hypothesis are Kennedy et al. (2017) and Park and Seo (2019), which take a different, nested approach using an internalization parameter (discussed later) and study the airline industry.[3]

2. A Test of Conduct

Throughout this article, we use $j \in J_T$ to denote products, $t \in T$ to denote markets, and n to denote all products and markets suitably stacked. We also use bold typeface to denote all of the products within a particular market (e.g.: $\mathbf{p_t}$ denotes the prices of all products in market t). We index by $-j$ to indicate exclusion of product j; for example, $\mathbf{p_{-j,t}}$ denotes the prices of all products in market t except for product j.

2.1. Setup and Testing Environment

Conduct testing is a classic problem in empirical industrial organization and antitrust economics. The idea, stemming from Bresnahan (1982), was to exploit "rotations of demand" to distinguish among different models of price-setting behavior (perfect competition, monopoly, Cournot, Bertrand). Later work followed Bresnahan (1987) in developing statistical tests for additional price-setting behavior such as double marginalization and two-part tariffs (Bonnet and Dubois, 2010; Villas-Boas, 2007). Much of this work relies on instruments excluded from a known parametric specification for marginal costs and compares the fit of the marginal cost specification in a likelihood ratio testing framework.

More recent work by Berry and Haile (2014) shows how a combination of conditions on excluded instruments (much broader than "rotations of demand") and a set of reasonable technical restrictions can be used to non-parametrically identify marginal costs, and thus discriminate among

[2]The Federal Trade Commission (FTC) held a 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 was released in September of 2020."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

[3]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.

alternative models of price setting behavior. Under some additional simplifying assumptions, we show how to take these ideas to data in a way that is computationally tractable while making as few parametric assumptions as possible.

In their analysis of marginal costs and firm conduct, Berry and Haile (2014) start with an assumption (Assumption 7a) which equates marginal costs to (generalized residual) marginal revenue $\psi_j(\cdot)$, the idea being that given prices $\mathbf{p_t}$, market shares $\mathbf{s_t}$, and a (known) demand system $D(\mathbf{z_t})$, where $\mathbf{z_t}$ is the full set of exogenous product characteristics, marginal cost can be recovered from the (known) first-order conditions of firms' price-setting behavior:[4]

$$\begin{aligned} mc_{jt} &= \psi_j(\mathbf{s_t}, \mathbf{p_t}, D(\mathbf{z_t})) \\ &= p_{jt} - \eta_j^m(\mathbf{s_t}, \mathbf{p_t}, D(\mathbf{z_t})). \end{aligned} \tag{1}$$

We modify this slightly and equate, in the second line, the generalized marginal revenue $\psi_j(\cdot)$ with the difference between observed prices p_{jt} and an additive markup $\eta_{jt}^m \equiv \eta_j^m(\mathbf{s_t}, \mathbf{p_t}, D(\mathbf{z_t}))$ which are subscripted by m to denote a particular model of conduct (such as Cournot, Bertrand, monopoly, perfect competition, fixed markups, or others). We use the notation η_{jt}^m without an argument, because given a conduct assumption and the fact that demand is assumed known, η_{jt}^m is also known and may be treated as data.[5]

We also need to characterize the marginal cost specification. At this point it is helpful to partition the set of exogenous product characteristics $\mathbf{z_t}$ into $\mathbf{x_t}$, variables that affect both demand and marginal costs; $\mathbf{v_t}$, variables that affect demand but are excluded from marginal costs, and $\mathbf{w_t}$, variables that affect marginal costs but are excluded from demand. Now,

$$mc_{jt} = c_j(x_{jt}, \mathrm{w}_{jt}, \mathbf{Q_t}, \omega_{jt}) = h_s(x_{jt}, \mathrm{w}_{jt}) + \omega_{jt}, \quad \text{where} \quad \mathbb{E}[\omega_{jt}|\mathbf{z_t}] = 0. \tag{2}$$

Here we have made two additional assumptions: the first is that there are no returns to scale or scope so that the vector of sales (in levels) $\mathbf{Q_t}$ does not impact the marginal cost; the second is that the unobservable cost term ω_{jt} is additively separable. Though it is unlikely to be important in our empirical study of RTE cereal, we can extend our approach to account for economies of scale, although it would require additional instruments to address the endogeneity of $\mathbf{Q_t}$.[6]

What we have not done is restricted the functional form regarding how observable cost components $(x_{jt}, \mathrm{w}_{jt})$ affect the marginal cost through $h_s(\cdot)$. This is in contrast with nearly all of the prior

[4] There are many examples of this in the literature including Nevo (2000) who shows that for multi-product oligopoly price-setting $\mathbf{p_t} - \mathbf{mc}_t = \Omega_t(\mathbf{s_t}, \mathbf{p_t})^{-1}\mathbf{s_t} \equiv \boldsymbol{\eta_t}(\mathbf{s_t}, \mathbf{p_t})$ where Ω_t is the matrix of demand derivatives $\Omega_{(j,k)} = \frac{\partial q_j}{\partial p_k}$ for products with the same owner.

[5] In practice, demand must be estimated from data. Our empirical example follows the standard approach in Berry et al. (1995) or Nevo (2001). Bresnahan (1987) is somewhat of an exception and estimates demand and marginal costs simultaneously when testing conduct models, while more recent work mostly follows the "menu" approach of Nevo (1998) and estimates demand prior to marginal costs.

[6] The substance of both restrictions are explained in Sections 4 and 6 of Berry and Haile (2014).

literature which generally treats $h_s(\cdot)$ as linear in covariates, or exponential (so that $\log(mc_{jt})$ is linear in the covariates).[7] Maintaining this flexibility is an important contribution of our approach.

If one combines (1) and (2) with a particular conduct assumption m, we get:

$$p_{jt} - \eta^m_{jt} = h_s(x_{jt}, \mathrm{w}_{jt}) + \omega^m_{jt}. \tag{3}$$

As in (2), the model is defined by the conditional moment restriction (CMR) $\mathbb{E}[\omega_{jt}|\mathbf{z_t}] = 0$. The boldfaced $\mathbf{z_t}$ indicates that the unobservable cost term ω_{jt} is conditionally mean independent not only of j's own characteristics $(x_{jt}, \mathrm{w}_{jt})$ in $h_s(\cdot)$ but of the characteristics of other products in the same market.[8] In practice, most researchers don't work with the conditional moment restriction directly, but rather with (weaker) unconditional moment restrictions implied by the CMR: $\mathbb{E}[\omega_{jt}|\mathbf{z_t}] = 0 \Rightarrow \mathbb{E}[\omega'_{jt}\, A(\mathbf{z_t})] = 0$ where $A(\cdot)$ denotes some matrix-valued function of $\mathbf{z_t}$. A substantial literature in econometrics is concerned with choosing $A(\cdot)$ in order to preserve the largest amount of information from the CMR (Newey, 1990; Ai and Chen, 2003; Imbens et al., 2003) including the seminal work by Chamberlain (1987) on "optimal instruments" in a nonlinear IV setting, which guides us at several points.[9] In that literature, the choice of $A(\mathbf{z_t})$ is largely about efficiency, while in our setting the choice of $A(\mathbf{z_t})$ is largely about the power of a test statistic to distinguish among alternative models of markups.

Consistent with much of the prior literature, we consider a non-nested testing framework where the null hypothesis is that both markup assumptions satisfy the moment restrictions equally well against the two-sided alternative where either model 1 (η^1) satisfies the restrictions better or model 2 (η^2) satisfies the restrictions better. The model is non-nested in the sense that both markups are not included in (3) at the same time. In the event of a rejection in favor of η^1 or η^2 we say that the model that satisfies restrictions $\mathbb{E}[\omega^{m\prime}_{jt} A(\mathbf{z_t})] = 0$ better is "preferred" rather than "true" because it is likely that both markups are mis-specified. The non-nested testing framework under mis-specification is formalized in Vuong (1989) for likelihood ratio tests, and extended to a broad class of objective functions (including GMM) in Rivers and Vuong (2002).

Formally, the null hypothesis of a Vuong-type test is that both models are equally far from the truth, taking as a distance metric a criterion function, $Q(\cdot)$ such as the GMM objective. The Rivers-Vuong test statistic is given by:

$$T = \frac{\sqrt{n}(Q_1(\eta^1) - Q_2(\eta^2))}{\sigma}, \tag{4}$$

where $\sigma/\sqrt{n}$ is the asymptotic (in n) standard error of the difference $(Q_1(\eta^1) - Q_2(\eta^2))$. Rivers and Vuong (2002) show that, for a broad class of specifications of the criterion function including

[7] See for example Michel and Weiergraeber (2018); Bonnet and Dubois (2010); Villas-Boas (2007).

[8] This is the foundation of the well-known "BLP instruments" for demand $\mathbb{E}[\xi_{jt}|\mathbf{z_t}] = 0$.

[9] The only case in which $\mathbb{E}[\omega'_{jt}\mathbf{z_t}] = 0$ contains the relevant information from the CMR is if the relationship between markups and $\mathbf{z_t}$ is exactly linear.

moment-based objective functions, T has a standard normal distribution. For $\alpha = 0.05$, this means that we can reject the null hypothesis in favor of model 1 if T is smaller than -1.96, and we can reject the null hypothesis in favor of model 2 if T is larger than 1.96.

This test has some advantages. First, it does not require that either model is the truth (as is necessary in the alternative Cox framework).[10] This is particularly appealing for model-dependent exercises like ours, as all models are, at best, an approximation. Second, recent work by Duarte et al. (2020) explored the performance of the Rivers-Vuong test statistic in Monte Carlo simulations of a conduct testing procedure. Their results suggest that it performs better than alternative testing procedures, particularly when the underlying demand model is misspecified.[11]

2.2. Choosing Moment Restrictions

Given a pair of candidate models that predict different markups, how can we choose a function $A(\mathbf{z}_t)$ that will allow us to discriminate between them? Defining $\Delta\eta_{jt}^{1,2} \equiv \eta_{jt}^1 - \eta_{jt}^2$, we propose the form $A(\mathbf{z}_t) = \mathbb{E}[\Delta\eta_{jt}^{1,2}|\mathbf{z}_t]$. There are several reasons we believe that this is a sensible choice. The simplest justification is that if we difference (3) for models 1 and 2 we obtain:

$$\underbrace{\eta_{jt}^1 - \eta_{jt}^2}_{=\Delta\eta_{jt}^{1,2}} = (\omega_{jt}^2 - \omega_{jt}^1) + (h_s^2(x_{jt}, \mathrm{w}_{jt}) - h_s^1(x_{jt}, \mathrm{w}_{jt})).$$

The intuition is that any discrepancy between two models of conduct should arise from the differences in the implied markups (ignoring differences in the estimates for the observed portion of marginal costs $h_s(\cdot)$). The challenge is that the markup difference is endogenous and directly depends on both the observed characteristics of all products $\mathbf{z_t}$ and the unobservable costs $\boldsymbol{\omega_t}$. Therefore we suggest replacing it with its expectation conditional on the exogenous variables: $\mathbb{E}[\Delta\eta_{jt}|\mathbf{z}_t]$. We provide a corresponding formal result in Proposition 1.

[10] The analogous Cox test would consider $H_0 : \lambda = [1, 0]$ vs $H_a : \lambda = [0, 1]$ using:

$$p_{jt} = h_s(x_{jt}, \mathrm{w}_{jt}) + [\lambda_1, \lambda_2] \cdot [\eta_{jt}^1, \eta_{jt}^2]^T + \omega_{jt}.$$

This is useful because it highlights the problems caused by the endogeneity of the markup η_{jt} and how moving it to the left hand side in (3) avoids the endogeneity problem via the Anderson-Rubin approach (imposing $\lambda_m = 1$ rather than estimating the parameter). See Duarte et al. (2020) for a more thorough discussion.

[11] The Vuong and Rivers-Vuong testing approach has been applied in prior work including Gasmi et al. (1992), Bonnet and Dubois (2010), and Duarte et al. (2020). The alternative, Cox-based tests, are employed in Bresnahan (1987) in an MLE context, and are extended to moment-based objective functions by Smith (1992), and employed by Villas-Boas (2007). Both test statistics were developed in a MLE framework (Cox, 1962; Vuong, 1989), and in both cases took the form of a likelihood ratio statistic. The main difference is the formulation of the null hypothesis. For a Cox test, one model is taken to be the null, and the other the alternative. This can lead to difficulties in testing using multiple moment restrictions because the weighting matrix now depends on which model is taken to be the null, generating an asymmetry in the test. Both in principle and in practice, this can lead to the scenario in which the researcher rejects model 1 (null) against model 2 (alternative), and model 2 (null) against model 1 (alternative). This problem does not arise for Rivers-Vuong tests, which are always symmetric, although Hall and Pelletier (2011) raise other issues with weighting matrices in that setting when the researcher uses multiple moment restrictions.

Proposition 1. *Under the following assumptions:*
(i) A fixed $k \times k$ positive semi-definite weighting matrix W;
(ii) A $(n \times k)$ matrix of instruments Z of full column rank k;
(iii) A vector of n unobservable cost components $\omega^m = p - h(z) - \eta^m$ and moment restrictions such that $E[\omega^{m\prime} Z] = 0 \in \mathbb{R}^k$ is satisfied at the true η^0 and not at $\eta^m \neq \eta^0$, and
(iv) The function $h(z)$ is known (rather than estimated).

(a) As $n \to \infty$ and under standard regularity conditions, the difference in GMM objective functions ($Q_W(\eta) \in \mathbb{R}^+$) under weighting matrix W, for any two markup models (η^1, η^2), can be expressed as:

$$Q_W(\eta^1) - Q_W(\eta^2) \xrightarrow{p} -\mathbb{E}[Z'\,\omega^1]'\,W\;\mathbb{E}[Z'\Delta\eta^{1,2}] - \mathbb{E}[Z'\,\omega^2]'\,W\,\mathbb{E}[Z'\Delta\eta^{1,2}] \text{ where } \Delta\eta^{1,2} = \eta^1 - \eta^2.$$

With an additional assumption, a stronger result obtains.
(v) Model 1 is the correctly specified model, i.e. $\mathbb{E}[Z'\,\omega_1] \xrightarrow{a.s} 0$.
(b) Under the additional condition (v), then:

$$Q_W(\eta^1) - Q_W(\eta^2) \xrightarrow{p} -\mathbb{E}[Z'\,\Delta\eta^{1,2}]\,W\,\mathbb{E}[Z'\,\Delta\eta^{1,2}].$$

Proof in Appendix A.1

The numerator of the Rivers-Vuong test statistic in (4) depends on the difference between GMM objective functions under the assumed markups (η^1, η^2). Proposition 1(a) highlights two features of this difference. First, it makes precise the relevance of instruments $A(\mathbf{z_t})$. If we choose instruments for each market $\mathbf{Z_t} = A(\mathbf{z_t})$ that are uncorrelated with $\Delta\eta^{1,2}_{jt}$, then $\mathbb{E}[A(\mathbf{z_t})'\Delta\eta^{1,2}_{jt}] = 0$ and the difference $Q_W(\eta_1) - Q_W(\eta_2) \xrightarrow{p} 0$ under both (a) and (b). For two competing models of markups to be testable, we require that they generate different objective functions, and therefore require that $\mathbb{E}[A(\mathbf{z_t})'\Delta\eta^{1,2}_{jt}] \neq 0$.[12]

Second, given a set of markups (η^1, η^2), a more powerful test will generate larger differences in $Q_W(\eta^1) - Q_W(\eta^2)$. Since both expressions in (a) are post-multiplied by $\mathbb{E}[Z'\,\Delta\eta^{1,2}]$, this suggests choosing $A(\mathbf{z_t})$ in such a way that $\mathbb{E}[A(\mathbf{z_t})'\Delta\eta^{1,2}_{jt}]$ is large. This alone isn't sufficient, because we would still need to know how the k moment restrictions in $\mathbb{E}[Z'\,\omega^m]$ covary with those in $\mathbb{E}[Z'\,\Delta\eta^{1,2}]$ (under some weighting matrix W). This motivates Proposition 1(b) which says that, under the true model, $\mathbb{E}[A(\mathbf{z}_t)'\,\omega_1] \xrightarrow{a.s} 0$, and the difference in the objective functions reduces to a quadratic form in $\mathbb{E}[Z'\,\Delta\eta^{1,2}]$. This suggests, in order to maximize the power of the test, we choose $A(\mathbf{z}_t)$ to be as close to $\Delta\eta^{1,2}_t$ as possible.[13]

[12] Likewise, if the markups themselves are indistinguishable $\eta^1_{jt} \approx \eta^2_{jt}$ it will become impossible to tell the models apart.

[13] Although $\Delta\eta^{1,2}_{jt}$ itself would be powered, it is not typically a valid exclusion restriction because markups are

A second argument for $A(\mathbf{z_t}) = \mathbb{E}[\Delta\eta_{jt}^{1,2}|\mathbf{z_t}]$ is that, for local comparisons, it appears organically as the (feasible approximation to the) optimal instrument for estimation of so-called "internalization parameters." Internalization parameters are a way to nest a family of objective functions of the firm. In this framework, markups can be written $\tilde{\eta}_{jt}(\tau)$, which we assume to be a continuous, but potentially nonlinear function of τ. Model selection amounts to estimating $\hat{\tau}$ and testing the hypotheses: $H_0 : \tau = \tau_1$ against $H_a : \tau = \tau_2$.[14] Consider estimating τ in:

$$p_{jt} = h_s(x_{jt}, \mathrm{w}_{jt}) + \tilde{\eta}_{jt}(\tau) + \omega_{jt} \quad \text{where} \quad \mathbb{E}[\omega_{jt}|\mathbf{z_t}] = 0. \tag{5}$$

The Chamberlain (1987) optimal instrument for the corresponding nonlinear GMM exercise is $\mathbb{E}\left[\frac{\partial\omega_{jt}}{\partial\tau}|\mathbf{z_t}\right] = \mathbb{E}\left[\frac{\partial\tilde{\eta}_{jt}(\tau)}{\partial\tau}|\mathbf{z_t}\right]$. Holding that in mind for a moment, next consider the non-nested model comparison of two models, at τ and $\tau + \epsilon$. Our proposed instrument $A(\mathbf{z_t}) = \mathbb{E}[\Delta\eta_{jt}^{1,2}|\mathbf{z_t}]$ is proportional to $\frac{\mathbb{E}[\Delta\eta_{jt}^{1,2}|\mathbf{z_t}]}{\epsilon} = \frac{\mathbb{E}[\tilde{\eta}_{jt}(\tau+\epsilon)-\tilde{\eta}_{jt}(\tau)|\mathbf{z_t}]}{\epsilon}$. This expression converges, as $\epsilon \to 0$, to $\mathbb{E}\left[\frac{\partial\tilde{\eta}_{jt}(\tau)}{\partial\tau}|\mathbf{z_t}\right]$, which is the (feasible approximation to the) optimal instrument for nonlinear GMM exercise described above.[15] In this sense we think of our instrument as a discrete analogue, in a non-nested model comparison exercise, of the Chamberlain (1987) optimal instrument for the corresponding estimation exercise. Moreover, this equivalence can be extended to non-local comparisons if $\tilde{\eta}_{jt}(\tau)$ is linear in τ.[16]

The third and final argument is that, since $k = 1$ for $A(\mathbf{z}_t) = \mathbb{E}[\Delta\eta^{1,2}|\mathbf{z}_t]$, i.e. we test using a single moment, so the need for a weighting matrix is obviated and we can write $\tilde{Q}(\eta)$ in place of $Q_W(\eta)$. Hall and Pelletier (2011) have shown how choice of a weighting matrix can be determinative in moment-based implementations of the Rivers and Vuong (2002) testing environment like our own. Moreover, the use of a scalar moment allows us to focus our computational efforts on the flexibility of our fit of the supply function $h_s(\cdot)$ from (3) as well as flexibility in the estimation of $\mathbb{E}[\Delta\eta^{1,2}|\mathbf{z}_t]$, which we detail next. Flexibility in the former is important because conduct tests based on (3) always jointly test both the exclusion restriction $\mathbb{E}[\omega'_{jt}A(\mathbf{z_t})]$=0 with the functional form of the cost function $h_s(x_{jt}, \mathrm{w}_{jt})$. Flexibility in the latter is important because the relationship between $\mathbf{z}_t$ and

usually modeled as endogenous functions of everything in the model, in particular ω_{jt}, and would violate the original moment restriction $\mathbb{E}[\Delta\eta_{jt}^{1,2}\omega_{jt}] \neq 0$.

[14] A number of recent papers have adopted this approach. In Miller and Weinberg (2017), that parameter captures the internalization of strategic externalities in pricing across Miller-Coors and ABI. Alternatively, in Crawford et al. (2018), the parameter characterizes the extent to which firms internalize the strategic consequences of their choices across divisions. Closer to our own question, Kennedy et al. (2017) and Park and Seo (2019) have estimate internalization parameters that convexify the space between own-profit maximization and common ownership incentives. Nevo (1998) offers a discussion of the contrast between conduct parameter estimation and our approach here, pairwise, non-nested model comparisons.

[15] The optimal but infeasible instrument for τ would require knowledge of the true parameter values including τ_0 itself.

[16] Consider a parameterized model of markups $\tilde{\eta}_{jt}(\tau)$ such that $\tilde{\eta}_{jt}(\tau_1) = \eta_{jt}^1$ and $\tilde{\eta}_{jt}(\tau_2) = \eta_{jt}^2$. The simplest parameterization would be $\tilde{\eta}_{jt}(\tau) = \tau \cdot \eta_{jt}^1 + (1-\tau) \cdot \eta_{jt}^2$. The approximation to the nonlinear optimal IV would be $\mathbb{E}\left[\frac{\partial\omega_{jt}}{\partial\tau}|\mathbf{z_t}\right] = \mathbb{E}\left[\frac{\partial\tilde{\eta}_{jt}(\tau)}{\partial\tau}|\mathbf{z_t}\right] = \mathbb{E}\left[\eta_{jt}^1 - \eta_{jt}^2|\mathbf{z_t}\right]$, which corresponds exactly to our choice of $A(\mathbf{z_t})$.

Algorithm 1 Testing Procedure

(a) Estimate the marginal cost function from (3), under models 1 and 2 to obtain residuals $\widehat{\omega_{jt}^1}$ and $\widehat{\omega_{jt}^2}$:

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

(b) Estimate the "first stage" regression, and compute the fitted values $\widehat{\Delta\eta}_{jt}^{1,2} = \hat{g}(\mathbf{z_t})$ of:

$$\Delta\eta_{jt}^{1,2} = g(\mathbf{z_t}) + \zeta_{jt}.$$

(c) For each candidate model, compute the value of the scalar moment:[17]

$$\tilde{Q}(\eta^m) = \left(n^{-1} \sum_{j,t} \hat{\omega}_{jt}^m \cdot \hat{g}(\mathbf{z_t}) \right)^2. \tag{6}$$

(d) Repeat steps (a)-(c) on bootstrapped samples and estimate $\hat{\sigma}/\sqrt{n}$ the standard error of the difference $\tilde{Q}(\eta^1) - \tilde{Q}(\eta^2)$.

(e) Compute the test statistic

$$T = \frac{\sqrt{n}(\tilde{Q}(\eta^1) - \tilde{Q}(\eta^2))}{\widehat{\sigma}} \sim \mathcal{N}(0,1). \tag{7}$$

Note: Steps (a) and (b) can be done in any order via non-parametric regression. Our preferred method is random forest regression (Breiman, 2001) which scales well as n becomes large and is well-suited to capturing nonlinear relationships.

$\Delta\eta^{1,2}$ will often be nonlinear.

All three of these arguments hinge on the unifying observation that the power of conduct tests based on Berry and Haile (2014) depends on capturing the nonlinear relationship between $\mathbf{z}_t$ and $\Delta\eta_{jt}^{1,2}$. So far, it's a theoretical observation. In our application to competition and common ownership in the ready-to-eat cereal market below, we will show that it has significant practical importance. Results from using forms of $A(\mathbf{z_t})$ which do not exploit the difference in markups suffer a loss of power relative to our proposed single moment. Examining this loss across different choices of $\mathbf{z_t}$, some of which better capture the nonlinearity of the model than others, the results confirm our observation. For now, however, we turn to the implementation of the test.

2.3. Our Testing Procedure

Our testing procedure is described in Algorithm 1. We adapt the non-nested test of Rivers and Vuong (2002) from (4) and use the expected difference in markups $A(\mathbf{z_t}) = \mathbb{E}[\Delta\eta_{jt}^{1,2}|\mathbf{z_t}]$ to formulate the sole moment restriction: $\mathbb{E}[\omega_{jt}' A(\mathbf{z_t})] = 0$.

An advantage of our procedure, which we discussed in Section 2.1 above, is that it allows

[17] 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).

substantial flexibility in the estimation of $h_s(\cdot)$ and $g(\cdot)$. One note, however: since $(x_{jt}, \mathrm{w}_{jt})$ are elements of $\mathbf{z_t}$ and appear in both equations, it is important to be as flexible in the way they enter $g(\cdot)$ as in the way they enter $h_s(\cdot)$. Not doing so may introduce implicit exclusion restrictions based on functional form. For instance, if x_{jt} enters $h_s(\cdot)$ linearly but with both a linear and a quadratic term in $g(\cdot)$, then the quadratic part of x_{jt} effectively behaves as an additional exclusion restriction.[18]

In our application, we use a random forest to fit both (Breiman, 2001). The main advantage of random forests over other flexible semi-parametric estimators is that they scale well when n is large, and they are good at capturing nonlinear relationships.

At this point we can highlight a second point of comparison with the recent proposal of Duarte et al. (2020). In place of $A(\mathbf{z_t}) = \mathbb{E}[\Delta\eta_{jt}^{1,2}|\mathbf{z_t}]$, they propose to take $A(\mathbf{z_t})$ to be a flexible sieve. Asymptotically, as the sieve becomes flexible, they show that their test statistic is equivalent to ours. This insight is analogous to Proposition 2 of Chamberlain (1987), which showed that for nonlinear GMM problems, taking $A(\cdot)$ to be a sufficiently flexible sieve allows the researcher to approximate the optimal instrument. Nonetheless, in applied exercises, when taking a sieve expansion of many instruments may prove infeasible,[19] we posit that it may be useful to exploit the known structure of the model and compute the conditional expectation of $\Delta\eta_{j,}^{1,2}$ given $\mathbf{z_t}$.

Continuing that analogy, observe that the problem of estimating $\mathbb{E}[\Delta\eta_{jt}^{1,2}|\mathbf{z_t}]$ mirrors the problem of estimating the infeasible Chamberlain (1987) optimal instruments. Relative to that literature, our approach follows most closely the suggestion of Newey (1990), who shows that the optimal instruments can be approximated by a first-stage nonparametric regression. An alternative approach, suggested in the appendix to Berry et al. (1999), would be to compute the expectation by simulating the entire model and integrating out over the unobserved cost and demand shocks. This is prohibitive for us because estimating and simulating from the marginal cost function requires us to take a stand, ex ante, on the correct conduct model, which begs the question we are trying to answer.[20]

As a final note, we have not said anything about how to choose the broader set of potential instruments $\mathbf{z_t}$ other than that they must include $(x_{jt}, \mathrm{w}_{jt})$. This will depend both on the models of conduct being considered and the specification of the demand system. In Section 6 we offer some comments on the full set of potential instruments when the researcher has estimated a demand system like the one we write down in Section 5.

[18] See Hartford et al. (2017) for further discussion on this point.

[19] For example, if there are many instruments in $\mathbf{z_t}$ (as when we use the BLP instruments in our exercise in Section 6, of which there are 143).

[20] A third approach, would be to extend the idea of Reynaert and Verboven (2014) where the authors assume perfect competition and predict $\mathbb{E}[p_{jt}|x_{jt}, \mathrm{w}_{jt}]$ using a linear regression on the variables in the marginal cost relationship (3) and plug this into the model $\mathbf{s_t}(\hat{\mathbf{p}}_\mathbf{t})$ to derive markups. As they suggest, we can also predict $\mathbb{E}[p_{jt}|\mathbf{z_t}]$ nonparametrically using the full set of potential instruments in $\mathbf{z_t}$ to more flexibly account for oligopoly behavior. The interested reader will find we use this method to estimate the optimal instruments for the demand estimation exercise below in Section 5, and we discuss the alternative approaches to approximating them for testing further in Appendix A.3.

3. 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. (2020a):

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

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

Each shareholder s receives some share β_{fs} of firm f's profits π_f. 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 γ_{fs} on the profits of each shareholder and maximizes a weighted sum of their payoffs.

Now if we consider a strategic action x_f taken by firm f which affects their profits and the profits of rival firms $\pi_f(x_f, x_{-f})$, we can write the objective function of the firm $Q_f(x_f, x_{-f})$, with some rearrangement, as:

$$Q_f(x_f, x_{-f}) \propto \pi_f(x_f, x_{-f}) + \sum_{g \neq f} \underbrace{\frac{\sum_s \gamma_{fs}\beta_{gs}}{\sum_s \gamma_{fs}\beta_{fs}}}_{\equiv \kappa^{co}_{fg}(\gamma_f, \beta)} \pi_g(x_f, x_{-f}). \tag{8}$$

Therefore each firm f will act as if it places a non-negative weight κ^{co}_{fg} on the profits of rival firms. By construction, $\kappa^{co}_{ff} = 1$, so that all κ^{co}_{fg} are defined relative to the weight that firm f places on its own profits.[21]

This model is quite flexible. For example, the objective can include the manager's private investment portfolio and have them place a high γ weight on their private returns. Alternatively, the manager can place a weight of $\gamma_{fs} = 0$ on particular shareholders (including passive or index investors, investors below some minimal blockholder threshold, etc.). The problem for empirical work is that even if we observed κ^{co}_{fg} for every pair of firms within an industry, it is unlikely we would be able to recover γ_{fs} without additional (parametric) assumptions, because while there might be ten firms within an industry (one-hundred pairwise interactions), there are often thousands of investors in each firm.

The researcher must make an assumption on γ_{fs}, since they do not have a directly observable empirical counterpart. This is akin to specifying a particular model of corporate governance or

[21] 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 $\gamma(0) = 0$.

corporate control. The most common choice for empirical work is the so-called "proportional control" assumption that sets $\gamma_{fs} = \beta_{fs}$. 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 κ_{fg}^{co} 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 managers 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 γ_{fs}. Another interpretation of a "lack of cooperation" in the data is that agency conflicts distort the manager's objective away from (8).

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. MHHI extends the traditional Herfindahl-Hirschman index (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 ϵ_d represents the elasticity of demand and the market share is given by $s_f = \frac{q_f}{Q}$) and aggregating across firms in the market, we get the share-weighted average markup in the market:[22]

$$\begin{aligned} \sum_f s_f \frac{p_f - c_f}{p_f} &= \frac{1}{\epsilon_d} \underbrace{\sum_f \sum_g \kappa_{fg} s_g s_f}_{MHHI} \\ \text{– where } MHHI &= \underbrace{\sum_f s_f^2}_{HHI} + \underbrace{\sum_f \sum_{g \neq f} \kappa_{fg} s_f s_g}_{MHHID}. \end{aligned} \tag{9}$$

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 to evaluate whether common ownership

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

incentives could be detected in prices. However, such regressions faced a number of criticisms: first, since share-weighted average markups are not observed, prices were used in their place, although that would require that markups and costs be uniform across firms to maintain the above relationship. Second, outside of Cournot, one might expect to find spurious relationships between an object such as $MHHID$, since it interacts a measure of common ownership with (endogenous) market shares. Third, price and $MHHID$ are both simultaneously determined equilibrium outcomes and so a causal interpretation of an "effect" of one on the other is problematic. Section 4.5 replicates the $MHHI$ regressions in our empirical setting and finds negative and statistically significant effects of common ownership on prices; Appendix E discusses the issue further and shows simulation evidence of the spurious correlations that can result from such regressions.

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 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:

$$\underbrace{D_j(\mathbf{p}) + (p_j - c_j)\frac{\partial D_j}{\partial p_j}(\mathbf{p})}_{\text{single product FOC}} + \underbrace{\sum_g \kappa_{fg} \cdot \left(\sum_{k'\in\mathcal{J}_g} (p_{k'} - c_{k'})\frac{\partial D_{k'}}{\partial p_j}(\mathbf{p}) \right)}_{\text{portfolio effects}} = 0. \tag{10}$$

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 κ_{fg}, the profit weight firm f places on firm g. This leads to different predicted markups under common ownership compared to own-profit maximization.

4. Ready-to-Eat Cereal

We focus our empirical exercise on the Ready-To-Eat (RTE) cereal industry for the period 2007–2016. 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. 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 at least during the 1990's (Nevo, 2000, 2001).

4.1. The Big Four

Kellogg's: 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. 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: 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: 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 (MOM), 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).

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; nevertheless, 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's (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 Appendix B.

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 December of 2009. This made BlackRock the largest player in the ETF market.[23] 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 the financial entity behind Fidelity, a major player in both actively managed and index funds. Capital Research, the parent 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 (8). These are depicted in Figure 1 for each of the four major firms in our sample.[24] For example, the top

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

[24] Here and throughout the analysis we make the proportional control assumption $\gamma_{fs} = \beta_{fs}$. Our framework can accommodate fully-specified alternative assumptions on γ, some of which we consider in our companion piece Backus et al. (2021).

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's (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%

Notes: This table documents the five largest institutional investors with holdings in each of the four largest RTE cereal companies for 2004, 2010, and 2016. Source: Backus et al. (2020b)

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% 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.[25] 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 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 κ. In Backus

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

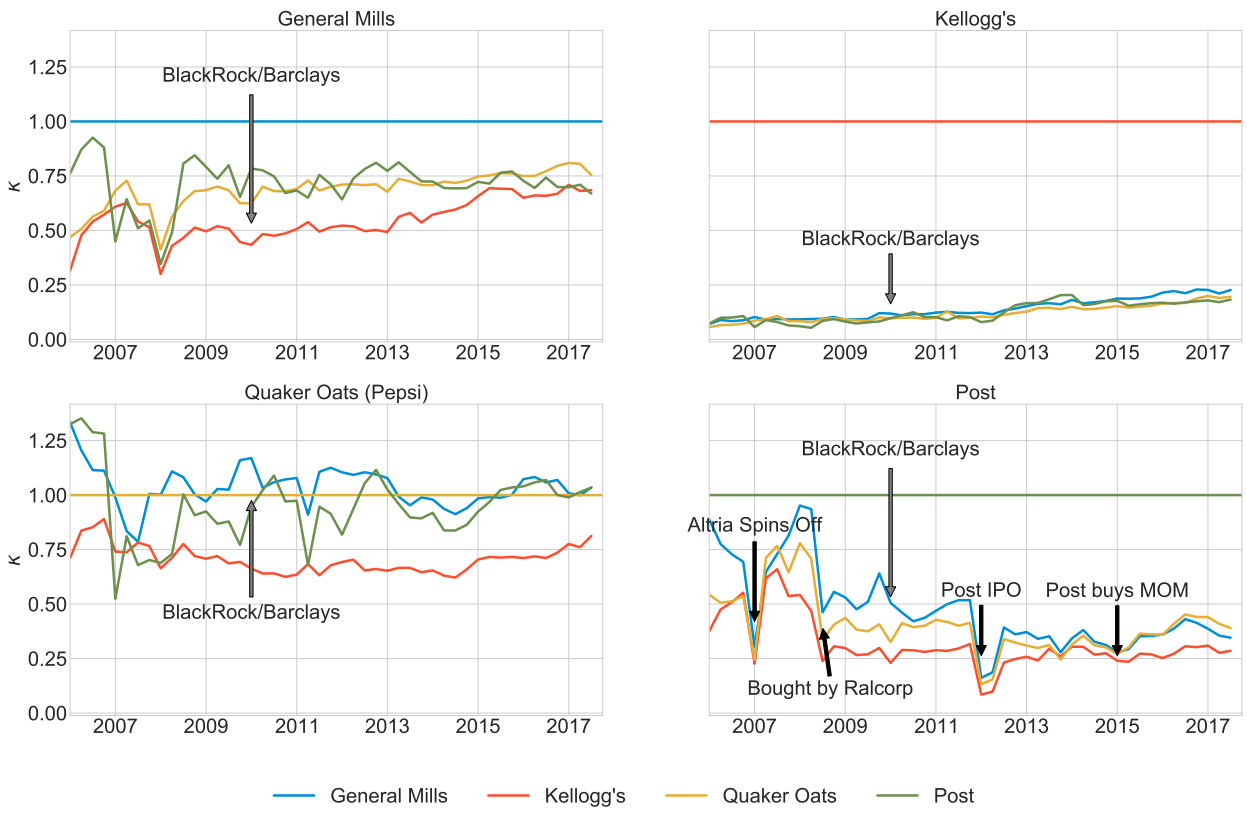


Figure 1: κ Profit Weights for Ready-To-Eat Cereal (Proportional Control)

Notes: This figure depicts the common ownership profit weights for each of the four largest RTE cereal companies between 2007 and 2017. Horizontal lines at 1 represent the normalization of κ_{ff}, i.e. the weight on a firm's own profits, to one. Arrows indicate major financial events that affect the ownership of the firms as described in the text. Source: Backus et al. (2020b) augmented by insider holdings data as detailed in Appendix B.1

et al. (2021) we showed that $\kappa_{fg} \neq \kappa_{gf}$ when firms f and g differ in their *investor concentration*, as measured by $IHHI \equiv \sum_s \beta_{fs}^2$. This asymmetry is on display in Figure 1. 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 Kellogg's has concentrated ownership, driven by a modest retail share and the presence of the Kellogg's Foundation. One disadvantage of $MHHI$ is that, by 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.[26] Because there is little price variation across stores within the same chain (see DellaVigna and Gentzkow (2019)), we aggregate unit sales and revenues to the DMA-chain level. We focus on the years 2007-2016.[27] 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.[28] 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).[29]

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).[30] Consistent with Nevo (2000), we convert boxes of cereal into serving equivalents, and report prices and quantities on a per-serving basis.[31] 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

[26] 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-Nielsen documentation, retailer coupons or loyalty card discounts are included in the "average price" calculation while manufacturer coupons are not.

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

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

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

[30] 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.

[31] 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. Consumer reports conducted a survey (https://www.consumerreports.org/cro/news/2014/12/cereal-portion-control-matters/index.htm) 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 Muesli or Granola where the average overpour was 282%.

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.[32] We assign each product to its ultimate owner in each period, taking to account mergers and acquisitions.[33]

	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

Notes: This table depicts summary statistics for our dataset of RTE cereal consumption in six DMAs (by column) between 2007 and 2017. Source: Nielsen Retailer Scanner Data: Stores and Sales, Nielsen Consumer Panelist Data: Income and Presence of Children, Nutritionix: Product Characteristics, Serving Size, FRED: Unemployment.

Figure 2 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 from Figure 1 in the figure.

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.[34] Instead, we must match these products to 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

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

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

[34]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*.

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 C.1. 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 C-1. 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.[35] 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.[36] To smooth the income distribution, we estimate two lognormal distributions of income for each DMA-chain-year, one for households with children and one for households without.[37] 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).[38] We report these demographic variables in Table 2.

Finally, we also gather data on input prices of different commodities. In particular, we gather prices for rice, wheat, corn, and oat from quandl.com, while we gather sugar prices from Global Financial Data. Changes in these input prices 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

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

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

[37] For more details please consult Appendix C.6.

[38] 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.

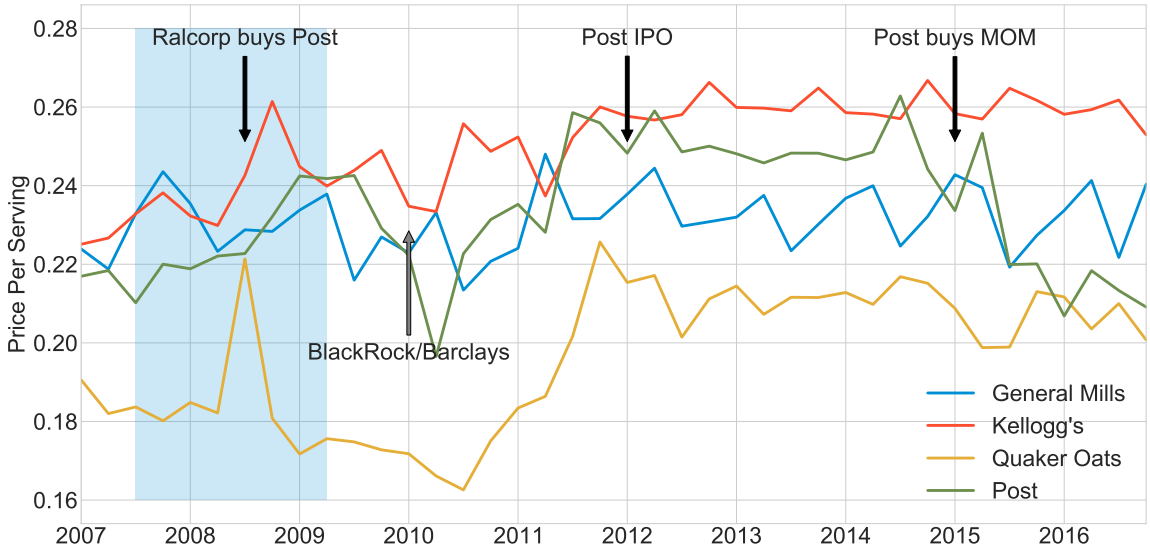


Figure 2: Average Serving Price by Firm

Notes: This figure depicts the average serving price, given by the ratio of quarterly firm revenue to total servings, for the period 2007-2017. The Great Recession is indicated by the shaded area. The decline in Post's average price after buying M-O-M, is largely compositional due to M-O-M's low prices. Source: Nielsen.

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 C-3. 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 3 and across markets in Appendix Figure E-10. Because we do not necessarily know which manufacturers produce the private label products (around 50% of the private label volume 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.

We break out concentration by DMA in Appendix Figure E-10. There is substantial cross market variation in HHI with Denver and Phoenix being the least concentrated (HHI typically below 2200) while Chicago is the most concentrated (HHI in excess of 2500 for most of the period). The concentration approximately 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 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.

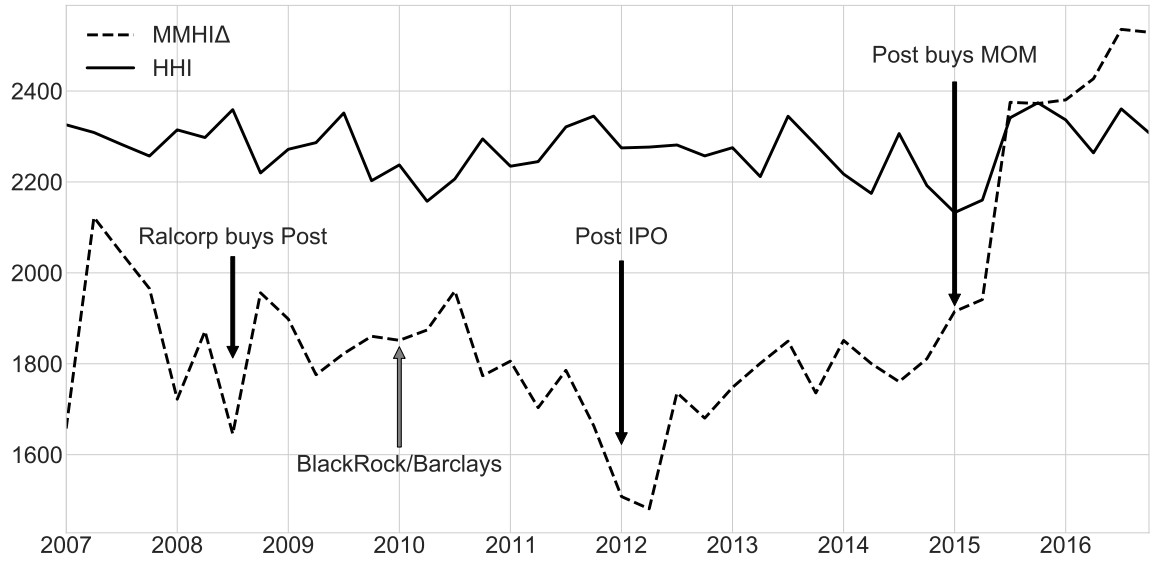


Figure 3: HHI and $MHHI\Delta$ over time (Six Markets)

Notes: This figure depicts the time series of the HHI and MHHIΔ concentration measures over the time period 2007-2017. For the purposes of the computation, private Label products are treated as single unified firm and "Other" independent sellers are treated as atomistic. $MHHI\Delta$ assumes proportional control. Source: Authors' computations.

The main point of Figure E-10 and Figure 3 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 κ when we construct $\Delta MHHI$. Indeed when we construct the $\Delta MHHI$ and plot it across DMAs in Figure E-10, the most salient feature is common, cross-market shocks in the time series. We find that the $\Delta MHHI$ is approximately the same magnitude as the HHI: around 1500 for Phoenix and around 2250 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 Section 3, although they do not correspond to a true reduced form if any of the assumptions of Section 3 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 –

	(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

Notes: This table reports results for a regression of log(average serving price) on concentration measures such as HHI and MHHIΔ. An observation is a manufacturer-chain-DMA-quarter. Prices and shares are computed based on servings. Source: Authors' computations.

an 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 E.4 contains additional specifications for this type of regression.

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 demographic 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 and therefore we leave the micro-foundation of the discrete choice multinomial logit demand system to Appendix A.2.

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

$$\sigma_j^{-1}(\mathcal{S}_t, \mathbf{p_t}, y_t; \theta_2) = h_d(x_{jt}, v_{jt}; \theta_1) + \xi_{jt} \quad \text{with} \quad \mathbb{E}[\xi_{jt}|\mathbf{z_t}] = 0. \tag{11}$$

We follow the notation of Conlon and Gortmaker (2020) and partition the parameter space into $[\theta_1, \theta_2]$ where the first set of parameters govern exogenous variables which enter (11) linearly, while the second set enter (11) 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.[39] We draw y_{it} from the (DMA-chain specific) joint distribution of income for households with and without children. This is meant to better reflect the demographics (particularly income) of shoppers at a particular chain beyond just the demographics of the area.[40]

Each micro-moment is meant to target a different demographic interaction parameter. We use two types of micro-moments: the first are 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.[41] The second type are covariances between purchaser demographics and product characteristics including price conditional on purchase.

We can use (11) 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

[39] 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, and we don't find much impact of including age.

[40] This allows us to address the fact that shoppers at a premium, natural, and organic supermarkets in Chicago are wealthier or have different preferences than shoppers at discount supermarkets in Chicago.

[41] This is substantially lower than the unconditional share of households with children, suggesting childless households are less likely to consume cereal overall and leads to a positive interaction coefficient $\pi_{const,kids} > 0$.

moments and micro-moments $g^M(\theta_1, \theta_2)$ to form a GMM estimator for $[\theta_1, \theta_2]$:[42]

$$\arg\min_{\theta_1,\theta_2} q(\theta_1, \theta_2) \equiv g(\theta_1, \theta_2)' W g(\theta_1, \theta_2), \quad g(\theta_1, \theta_2) = \begin{bmatrix} g^D(\theta_1, \theta_2) \\ g^M(\theta_1, \theta_2) \end{bmatrix}. \tag{12}$$

We estimate (12) via two-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 (2020), and update the weighting matrix W.[43]

We use three kinds of instruments in the initial $b(\mathbf{z_t})$. The first set are the w_{jt} instruments, or own-cost-shifters, which we plot in Figure C-3. 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 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 are observable variation in demographics y_t. We use the share of households with children, as well as the 10%, 50%, and 90% quantiles of the income distribution calculated separately for households with and without children.

The third type of instruments are a variant of the BLP instruments. Here we follow Gandhi 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.[44] We then obtain an estimate of expected price: $E[p_{jt}|x_{jt}, \mathrm{w}_{jt}, y_t, z_{jt}^{GH}, \mathrm{promo}_{jt}] = \hat{p}_{jt}$ via random forest regression.[45] We include this expected price measure in our set of characteristics and repeat the procedure above (to construct "differentiation instruments" for expected price). Following the recommendation in Gandhi 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 number of

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

[43] 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 C.4.

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

[45] We also estimated this regression with linear fixed effects: chain (26), product (987), and week (522). The R^2 of the linear model is 0.87 and the within $R^2 = 0.22$. The random forest regression does not include the fixed effects and does not interact the (44) instruments and exogenous regressors x_{jt} or with the demographic moments. The random forest has an $R^2 = 0.97$ and reduces MSE by 75% compared to the linear regression. Correlation in predicted prices across both specifications is high: $\rho = 0.957$. Both specifications include a dummy for promotional activity as recorded in the scanner dataset though this flag is often missing. This flag is also included in other studies of RTE cereal (Michel and Weiergraeber, 2018).

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 and are left with 143 instruments.

Parameter	Variable	Prices Only		Correlated Normal	
β	Prices	-13.369		-9.531	
		(0.096)		(0.113)	
	$\text{Unemp}_t \times \text{Branded}_j$	-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)	
Σ^2		Cons	Prices	Cons	Prices
	Cons			2.214	-1.000
				(0.040)	
	Prices		1.945		5.013
			(0.049)		(0.043)
Π		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

Notes: This table reports demand estimates for RTE cereal between 2007 and 2017 in our six DMAs on a sample of 2 weeks per quarter. We estimate parameters using the Cholesky root of the correlated normal $LL' = \Sigma^2$. The transformed parameter estimates we report are computed using the Delta method in `pyBLP`. The correlated normal estimates the components relating to prices and the constant as nearly perfectly negatively correlated, and the component on prices independent of the constant to be nearly zero. This implies that the marginal consumer (controlling for demographics) tends to prefer the least expensive products. Source: Authors' computations.

We estimate our demand model using `pyblp` (Conlon and Gortmaker, 2020). We estimate demand on a random subset of 2 weeks from each quarter (16% sample) in our data. We do this to alleviate concerns around potential storage (Hendel and Nevo, 2006), as well as to mitigate the computational burden, and provide some out-of-sample diagnostics. We report the second-stage GMM estimates of $[\theta_1, \theta_2]$ in Table 4 after updating both the weighting matrix and the (feasible

approximation to the) optimal instruments. 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 experience high unemployment, consumers substitute away from branded products.[46]

Both sets of estimates include demographic interactions Π 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 expected, high income households are substantially less price-sensitive, but have a lower overall preference for cereal.[47] 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 Σ. 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 goods.[48] In general our estimates of own price elasticities are around $\epsilon_{jj} = -2.6$ which are in line with Nevo (2000) and other studies of RTE cereal Michel and Weiergraeber (2018) using data from the 1990's.

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. We report both the fit on the random sample of 2 weeks from each quarter (Estimation Sample) and the overall dataset (Full Sample). The latter is an out-of-sample evaluation of model fit, and in general performs similarly to the in-sample fit.[49]

In Table 6, we report the firm 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

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

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

[48] 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.

[49] The GMM weighting matrix determines how much priority is given to satisfying the micro-moments versus the aggregate data $\mathbb{E}[\xi'_{jt} Z_{jt}] = 0$ moments. As noted by Berry et al. (2004) and Petrin (2002), there is no covariance between micro-moments and aggregate data moments as they are expectations computed over different measures. If one fails to account for the fact that the micro-moments have sampling uncertainty from the panelist data, the naive update to the optimal weighting matrix places a very high weight on the micro-moments relative to the aggregate data moments (and can fit them to 6-7 decimals). This would effectively ignore the aggregate data moments, and makes it difficult to estimate the parameters other than Π. For this reason we scale the weight on the micro-moment block so that it fits to roughly the third decimal place.

	Data		Estimation Sample		Full Sample	
	Income	Kids	Income	Kids	Income	Kids
$E[y_{it}$|Outside Good]	0.1785	0.6593	0.1820	0.6372	0.1823	0.6367
$Cov(y_{it}, p_{jt})$	-0.0052	0.0027	-0.0086	0.0011	-0.0086	0.0011
$Cov(y_{it}, \text{PCA0}_{jt})$	0.1626	-0.0152	0.1364	-0.0207	0.1365	-0.0206
$Cov(y_{it}, \text{PCA1}_{jt})$	0.0790	0.0066	0.0726	-0.0016	0.0724	-0.0015
$Cov(y_{it}, \text{PCA2}_{jt})$	-0.0500	-0.0003	-0.0432	0.0042	-0.0435	0.0042

Table 5: Micro Moment Values

Notes: This table reports values for the micro moments for the fitted model, the estimation sample, and the full sample. Estimation Sample is T = 1,973 markets and N = 307,675 observations. Full Sample is T = 12,733 markets and N = 1,981,930 observations. Source: Authors' computations.

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, while General Mills products (such as Cheerios) are more popular in markets with more children. Likewise, private label products are more popular in markets (defined as DMA-chain) with low incomes. Average prices (sales-weighted) in low-income markets tend to be 5-10% lower than prices in high-income markets. We also report the average share across markets by firm $\sum_{j \in \mathcal{J}_f} s_{jt}$ (inclusive of the outside good share). This represents the weekly probability that a consumer, having walked into the supermarket, purchases a product. The plain IIA logit model would predict diversion ratios proportional to share (around 7% for Kellogg's or General Mills products) while our estimate model predicts much higher diversion ratios (closer to 20%). Likewise, though the outside good share is nearly 74% in the data (roughly 26% of consumers purchase cereal each week), our estimated diversion ratios are closer to 35% suggesting that most cereal customers choose another brand as their preferred product becomes more expensive.[50]

6. Non-Nested Model Comparisons

6.1. Predicted Markups

We begin by recovering markups and marginal costs under different conduct assumptions.[51] For our application to ready-to-eat cereal sold in food outlets, this exercise requires an additional as-

[50] This should be unsurprising given the relatively large variance estimated on the random coefficient for the constant in Table 4, as well as the large π value estimated for households with children for the constant.

[51] Recovering markups, which is a standard exercise in the literature, requires prices and market shares $(\mathbf{p_t}, \mathbf{s_t})$; a matrix of demand derivatives a observed prices $\mathbf{p_t}$, which we take as known from the demand estimation exercise and write $\Omega(\mathbf{p_t})$, and a product-level ownership matrix, $\mathcal{H}(\kappa_t)$, which is an extension of the typical firm-level matrix of profit weights. This latter object embeds the model of conduct. Then, markups are computed as

$$\eta_t(\mathbf{p_t}, \mathbf{s_t}, \kappa_t) = (\mathcal{H}(\kappa_t) \odot \Omega(\mathbf{p_t}))^{-1} \mathbf{s_t}.$$

See Appendix A.3 for a complete derivation from the first order conditions of Nash pricing.

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)

Notes: This table reports quantity-weighted average diversion ratios by firm pair (row and column) for different demographic groups (by panel). 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 $> 21\%$). Diversion ratios are reported as percentages (sum to 100%). Prices are reported in cents. Share is reported inclusive of the outside good. Source: Authors' computations.

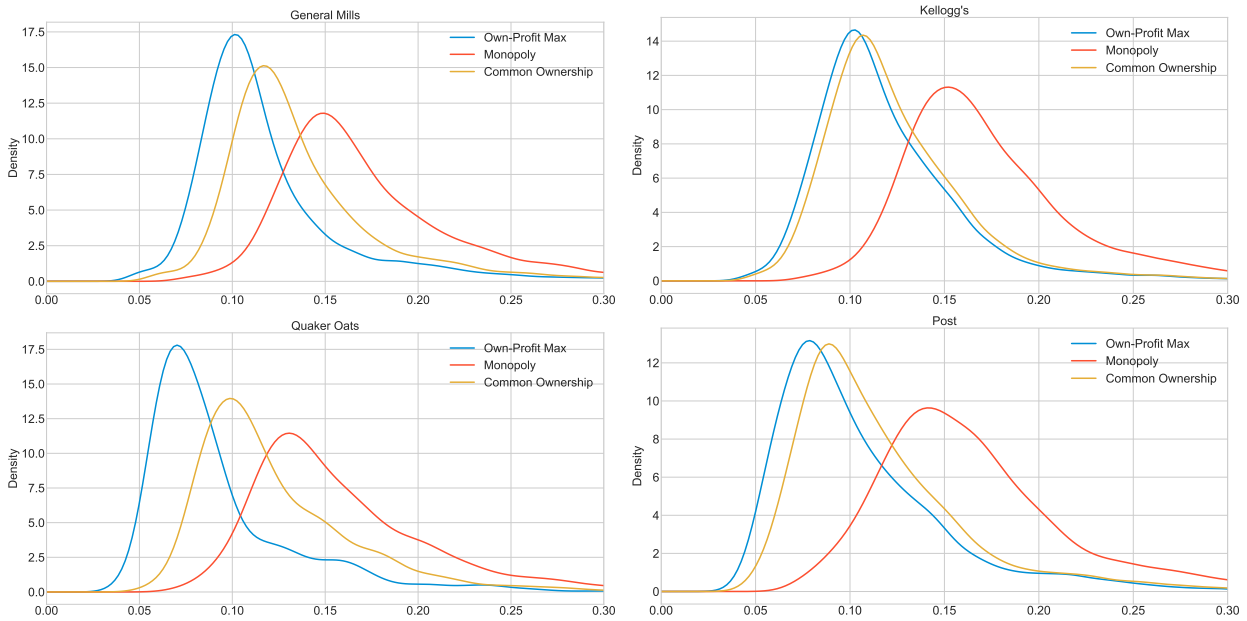


Figure 4: Distribution of Implied Price-Cost Margins by Conduct Assumption and Firm

Notes: This figure depicts the distributions of price-cost margins for the four largest RTE cereal firms in the fourth quarter of 2016 under three different conduct hypotheses: own-profit maximization, a hypothetical monopolist, and common ownership. The unit of observation is a product-chain-DMA-week. Source: Authors' computations.

sumption that manufacturers directly set retail prices.[52] The set of conduct hypotheses we consider includes own-profit maximization, perfect competition, a hypothetical monopolist composed of the four largest firms, and common ownership (and variants). Note that everything that follows uses the full sample.

We depict the price-cost margins for all of these conduct hypotheses in Figure 4. We focus on the fourth quarter of 2016, the last period in the data; we also exclude perfect competition, for which price-cost margins would be trivially zero. Under own-profit maximization ($\kappa_{fg} = 0$), larger firms (here, Kellogg's and General Mills) have larger implied markups as they recapture more diverted sales with their own products. We also see that the difference between markups under own-profit maximization and common ownership maps neatly to the profit weights depicted in Figure 1. Kellogg's large undiversified owner depresses their profit weights, and so the difference between own-profit maximization and common ownership is slight. In contrast, Quaker Oats' (PepsiCo's) large retail share inflates their profit weights, and so the difference is much larger.

To frame the economic magnitude of these differences, we compare the effect of common ownership on prices to the effect of hypothetical pairwise mergers. To be precise, we derive marginal

[52]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 using an ownership matrix which also 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 own-profit maximization.

costs under the assumption of own-profit maximization, 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 pricing of a hypothetical monopolist composed of 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 both firms raising prices around 5% and the price index rising by 3.3%. Mergers of the two smaller firms (Quaker Oats and Post) and one of the two larger firms, lead to smaller increases in the overall price index but often large changes (5-6%) for the smaller firm.

The effects of any pairwise predicted merger, however, are dwarfed by the 5.42% increase in the price index 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, a hypothetical monopolist composed of the four largest firms yields a price increase of 10.25%, and so common ownership would get us $5.42/10.25 \approx 53\%$ of the way to monopoly pricing. In Appendix B and Table B-1, we show how common ownership can lead to smaller or larger price effects of a merger, and those effects can depend on the financial structure of the merger (an all cash deal or a share-swap).

The question remains, however, whether patterns of common ownership are reflected in pricing. For this we implement the conduct testing procedure developed in Section 2.

Firm	GM-KEL	GM-QKR	GM-POST	KEL-QKR	KEL-POST	QKR-POST	Monopoly	κ^{CO}
General Mills	4.69	1.10	3.00	-0.10	-0.16	-0.11	9.42	3.97
Kellogg's	5.13	-0.13	-0.19	1.04	2.90	-0.11	9.30	5.34
Quaker Oats	-0.37	6.39	-0.36	5.60	-0.37	3.93	14.87	7.75
Post	-0.15	-0.11	6.16	-0.10	5.19	1.45	12.76	7.06
Price Index	3.32	0.91	2.22	0.78	1.96	0.58	10.25	5.42

Table 7: Price Increases under Counterfactual Scenarios

Notes: This table reports counterfactual percent price increases from the observed baseline in the fourth quarter of 2016. We compute the serving-weighted percentage change in price at the market and firm level (with the last row containing all firms) and report the median price change across markets. The market-level mean price per serving observed in the data is \$0.2255. Columns indicate the counterfactual scenario: the first six are pairwise mergers, the seventh is a hypothetical monopolist composed of the four largest firms, and the last is under common ownership. In some cases, prices of non-merging firms decline in response to the merger. This is because with random coefficients and demographics (or any mixture of logits), prices need not be strategic complements as is often assumed. Source: Authors' computations.

6.2. Testing: Inclusions and Exclusions

Our test employs a criterion function that is based on detecting a correlation between the difference in markups between two models and the recovered cost shocks. This means that useful instruments are those which are able to explain the difference in markups $\Delta\eta_{jt}^{1,2}$ and are excluded from the marginal cost equation in (3).

Since markups are fully determined by the model of marginal costs, demand, and conduct, we can also fully characterize the set of relevant variables. The contents of $\mathbf{z_t}$ are five sets of instruments. The first set of instruments are the exogenous own-product characteristics $(x_{jt}, \mathrm{w}_{jt})$ included in equation (3). In our tests, product characteristics x_{jt} (which also appear in the demand system) are captured by a full set of product, retailer-DMA, and quarter fixed effects. This set also includes the fraction of servings that are on sale, as well as dummies for periods in which a product was on recall (see Appendix C.7 for further discussion of recalls). Variables in w_{jt} (which do not appear in the demand system) include the input costs for designated "main" ingredients. The four remaining instruments in $\mathbf{z}_t$ are all excluded from (3).

The second set of instruments are the demand shifters $\mathbf{v_t}$ which affect demand but not marginal costs and form the basis of the identification argument in Berry and Haile (2014).[53] In theory there are many such variables (more so than cost-shifters w_{jt}) and one could exploit seasonal or regional variation in demand or promotional activity.[54] In our cereal example, we have variation in the local unemployment rate (interacted with branded products). Demand shifters for other goods $\mathbf{v_{-j,t}}$ are also valid instruments.

A special kind of demand shifter which enters demand non-linearly rather than linearly are changes in the distribution of demographics across markets. That is, demand (and markups) for certain products (sugary cereals) may be higher in areas with lots of children, but the marginal cost of serving consumers in these areas is unlikely to be affected by demographics.

A third 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.[55]

The fourth set are the characteristics of other goods $\mathbf{x_{-j,t}}$ or the "BLP instruments." The nested logit version of these instruments from Berry (1994) involve counting the number of competitors

[53] 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 $\mathbf{v_t} = [v_{j,t}, \mathbf{v_{-j,t}}]$.

[54] 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}. Michel and Weiergraeber (2018) exploit the non-price demand shift effects of in-store promotions.

[55] 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 (Genesove and Mullin, 1998).

in each nest. Gandhi 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 markup for product j to fall, while the characteristics of rival products should not directly impact the marginal costs c_{jt}.

The final set of potential instruments come from exogenous changes in conduct κ_{fgt} itself. As an example, Miller and Weinberg (2017) use a binary instrument for treated firms in the post-joint venture period as an instrument for conduct κ. Another (somewhat obvious) example is that mergers may change κ^{CO}_{fgt} such as Post's acquisition of Malt-O-Meal in 2015.

Exogenous changes in financial ownership may also serve as an instrument for common ownership profit weights κ^{CO}_{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 (2020). Azar et al. (2018a) look before and after the BlackRock-Barclays acquisition with the implied assumption that it shifts κ^{CO}_{fgt}. The challenge with using κ^{CO}_{fgt} restrictions is that they are often event studies. We could use less discrete changes in κ^{CO}_{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. To be conservative, we don't include these events in our set of instruments $\mathbf{z_t}$.

We should also mention what are not valid exclusion restrictions. The variables $(x_{jt}, \mathrm{w}_{jt})$ are not excluded, but included and help to reduce the residual variance in ω_{jt}. The so-called "Hausman instruments" which are prices of the same product in other markets $p_{j,-t}$ are not valid instruments for markup differences. 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. However, unobservable changes in marginal cost are potentially correlated with ω_{jt}.[56]

In light of these observations, we run our tests with four different specifications of the excluded variables in $g(\cdot)$. Comparing results across them will both shed light on the role of the nonlinear relationship between $\mathbf{z}_t$ and $\Delta\eta_{jt}$, and also offer us an informal overidentification test insofar as the results are similar or different. We report the number of variables in parentheses.

1. Other firms' input costs (5): $\mathbf{w}_{-jt}$ in the above, we use the price of ingredients (corn, rice, oat, wheat, and sugar) that are not designated as "main" ingredients in the Nutritionix dataset.

2. Demographic moments (5): these 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.

[56] What looks to authorities like collusive behavior (changes in κ) is explained by defendants as a "correlated shock to marginal costs". In this case, the "Hausman instruments" can be perfectly valid as instruments for demand $E[\xi_{jt}|\mathbf{z_t}, p_{j,-t}] = 0$ but are invalid as instruments for marginal costs $E[\omega_{jt}|\mathbf{z_t}, p_{j,-t}] \neq 0$.

3. BLP instruments (143): these are explained in detail above in Section 5 where we follow Gandhi and Houde (2019). These summarize variation in $\mathbf{x}_{-jt}$ and $\mathbf{v}_t$.

4. Demand-optimal instruments (18): these too are explained in detail above in Section 5. Note that these are constructed using the model, summarize all of the variation in the four sets of excluded instruments in $\mathbf{z}_t$.

6.3. Implementation

The specification of $g(\cdot)$ in Step (b) of our algorithm 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 (Ai and Chen, 2003; Newey and Powell, 2003), we fit both functions flexibly. We use a random forest (Breiman, 2001) to estimate both $h_s(\cdot)$ in Step (a) and $g(\cdot)$ in Step (b).[57] By way of comparison, we will also consider a version of our test that uses a linear fit for $h_s(\cdot)$ and $g(\cdot)$. The reason we do this is to highlight the importance of fitting the nonlinearity in the relationship between $\Delta\eta_{jt}$ and $\mathbf{z_t}$ in the Step (b) regression. We discuss results for both sets of regressions further and document the mean-squared error of the Step (b) regression in Appendix Section D.3.

In order to compute $\hat{\sigma}$, the standard deviation of $\tilde{Q}(\eta^1) - \tilde{Q}(\eta^2)$ 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 (a) and (b) regressions, as well as the computation of $\tilde{Q}(\eta^m)$—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 hypothesis, perfect competition, and monopoly pricing (by row) against own-firm profit maximization using different sets of exclusions (by column), for different specifications (by panel). To interpret an entry in the table, recall that the Rivers-Vuong test statistic is distributed $\mathcal{N}(0,1)$. The null hypothesis is that the two models perform equally well. A negative value can reject that null in favor of the alternative that model 1 better fits the data (here, own-profit maximization), while a

[57] Alternative approaches might include LASSO regression (Belloni et al., 2012) and deep neural nets (Hartford et al., 2017).

Table 8: Testing Results: Own-Profit Maximization vs Alternatives

	Others' Costs	Demographics	BLP Inst.	Dmd. Opt. Inst.
Own Profit Max vs.	Panel 1: $A(\mathbf{z}_t) = \mathbf{z}_t$, linear $h_s(\cdot)$			
Common Ownership	-2.4732	-0.0079	-1.2333	-4.9099
Common Ownership (MA)	-2.5918	0.0070	-1.2105	-4.9215
Common Ownership (Lag)	-2.5208	0.0075	-1.2125	-4.9351
Perfect Competition	0.8611	-2.3033	-3.1652	-10.9229
Monopolist	-2.4166	-0.8783	-3.5162	-6.0048
Own Profit Max vs.	Panel 2: $A(\mathbf{z}_t) = \mathbb{E}[\Delta\eta^{12}\|\mathbf{z}_t]$, linear $h_s(\cdot)$ and $g(\cdot)$			
Common Ownership	-1.2859	-0.2126	-0.8317	-5.2361
Common Ownership (MA)	-1.3993	-0.2071	-0.8340	-5.3019
Common Ownership (Lag)	-1.3506	-0.2093	-0.8367	-5.3271
Perfect Competition	1.1732	-0.8843	-1.4708	-10.7559
Monopolist	-1.4038	-0.3243	-1.0613	-5.3183
Own Profit Max vs.	Panel 3: $A(\mathbf{z}_t) = \mathbb{E}[\Delta\eta^{12}\|\mathbf{z}_t]$, random forest $h_s(\cdot)$ and $g(\cdot)$			
Common Ownership	-4.8893	-5.4460	-5.4412	-5.9585
Common Ownership (MA)	-5.4345	-6.1348	-5.8757	-6.4357
Common Ownership (Lag)	-5.1770	-5.9221	-5.7041	-6.2255
Perfect Competition	-7.7749	-8.7051	-8.9758	-10.0654
Monopolist	-5.2711	-6.7789	-5.9158	-6.5933

Notes: This table reports testing results for pairwise comparisons of own-firm profit maximization against various alternatives. Variations on common ownership include a one-quarter lagged κ (Lag), and a four-quarter backwards looking moving average (MA). The column label indices the set of exclusions used in the Step 2 regression. A negative value constitutes evidence for own-profit maximization and a positive value constitutes evidence for the model of that row. 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. The bootstrap is clustered at the retailer by DMA by year level and observations are weighted by servings volume.

positive value can do so for model 2 (specified by the row).

The three panels are contrasted to underscore the importance of, first, using the model to inform the design of moment restrictions (recall $A(\mathbf{z_t})$ from Section 2.3) and, second, loosening parametric restrictions in order to capture the nonlinear relationships between the excluded instruments and markups. Panel 1 depicts the conduct test that the literature has focused on, in which instruments are translated directly into unconditional moments of the form $\mathbb{E}[\omega_{jt}^m z_{jt}] = 0$. Since there are potentially many such moments, we weight them by $(Z'Z)^{-1}$ to form a scalar criterion function, where Z is the full $n \times k$ matrix of instruments stacked across markets.[58] We also follow them in adopting a linear form for $h_s(\cdot)$. Panel 2 uses Algorithm 1 but retains a linear functional form for $h_s(\cdot)$ and $g(\cdot)$ in Step (a) and Step (b), respectively.

In both Panels 1 and 2 we find substantial variation in results across different sets of instruments. The similarity of the test statistics in these two panels reflects the fact that when everything is linear, the Step (a) regression in our method accomplishes little more than to change the weighting matrix over the moments interacting the instruments with $\hat{\omega}_{jt}^m$. Using others' costs rejects common ownership against profit maximization, but fails to reject (and even weakly prefers) perfect competition. In contrast, demographic moments and BLP instruments are inconclusive. However, the demand optimal instruments decisively favor own-profit maximization against all other alternatives. This is particularly notable because the demand optimal instruments are a nonlinear aggregation of the variation captured by the prior three (see Appendix C.4). This suggests that the different results have more to do with functional form than the informational content of the exclusion restrictions.

Therefore, Panel 3, our preferred specification, uses a random forest to fit both $h_s(\cdot)$ and $g(\cdot)$. Here, all sets of instruments reject all alternative hypotheses in favor of own-profit maximization. The demand optimal instrument generally rejects most strongly, which reflects the fact that it aggregates variation from all of the other three. Demographic variables do particularly well, which is consistent with their prominence in explaining variation in demand, as we saw in Section 5. Finally, we also observe that the test statistics for the demand optimal instruments are not much stronger, and occasionally weaker, than those from Panel 2. This too is intuitive: since they already capture the nonlinearity of the model, using an approximately correct linear functional form restriction can offer greater power, like using the true functional form offers efficiency gains in estimation problems.[59]

In order to bridge the gap between our testing procedure and recent conduct parameter esti-

[58] We follow Duarte et al. (2020) in selecting the 2SLS weighting matrix $(Z'Z)^{-1}$ for the test with multiple moment restrictions, but caveat that the test in Panel 1 does not correspond to theirs. They propose using a sieve expansion of $\mathbf{z_t}$ which will, in the limit, fit the nonlinearity of the model.

[59] A reasonable question might be to ask why the demand optimal IV work so well independent of functional form. One possible explanation is in (10), the difference in markups between all of the alternative models can be written as proportional to $\sum_g \Delta\kappa_{fg}^{1,2} \cdot \sum_{k \in \mathcal{J}_g} \eta_k \cdot D_{jk}$ where $\Delta\kappa_{fg}^{1,2}$ represents the difference in profit weights between the two models. The demand optimal IV are good at explaining the diversion ratios D_{jk}.

mation exercises, we also use our procedure to evaluate own-profit maximization against "partial" common ownership, the latter of which we specify as $\tau \in \{0.1, \ldots, 0.9\}$, where τ is a multiplicative scalar on all of the off-diagonal terms of the common ownership profit weight matrix κ_t^{CO}. So, $\tau = 0$ implies own-profit maximization, and $\tau = 1$ implies common ownership pricing.[60] Results are depicted in Table 9. As before, a negative value less than 1.96 represents a rejection in favor of own-profit maximization at $\alpha = 0.05$. In brackets, under each test statistic, we write the numerator and the denominator, i.e. our estimates of $\tilde{Q}(\eta^1) - \tilde{Q}(\eta^2)$ and $\sigma/\sqrt{n}$.

Two features are salient. First, the test statistic is monotonically decreasing in τ, so that models closer to own-profit maximization are preferred. Second, however, for no set of excluded variables do we statistically reject the model for $\tau < 0.3$. There is a third, more subtle feature of the results evident in the brackets beneath the test statistics. While the difference in the criterion function (the numerator, $\tilde{Q}(\eta^1) - \tilde{Q}(\eta^2)$) is converging to zero as the models become more similar ($\tau \to 0$), the standard error of the difference (the denominator, $\sigma/\sqrt{n}$) is too. This reflects a degeneracy problem with the Rivers and Vuong (2002) statistic when testing sufficiently similar models. As $\tau \to 0$, that standard error shrinks to zero because the models become identical (they imply nearly identical markups), and the test statistic is degenerate. While it is not our primary interest, for applications in which the researcher is interested in local comparisons, this could be resolved by sample splitting (Yatchew, 1992; Whang and Andrews, 1993), partial sample splitting (Shennach and Wilhelm, 2017), or debiasing methods (Shi, 2015).

To summarize, our results consistently reject the common ownership hypothesis in favor of own-firm profit maximization. We emphasize that our results are not driven by comparisons of average markups. All of our regressions in the testing procedure include product dummies. Moreover, we would also stress that our result does not answer the ill-posed question of whether common ownership, broadly construed as the presence of common owners, has *any* price effects. Rather, we have shown that the pricing model implied by the common ownership hypothesis, as outlined in Section 3 above, is inconsistent with the pricing behavior we observe.[61]

7. Conclusion

In this paper we study how to test among competing assumptions for conduct in a differentiated products market by exploiting implied differences in additive markups. We show that exclusion restrictions can be used to differentiate between alternative models of firm conduct. Our testing procedure has advantages over previous approaches: it is easy to implement and performs well across different sets of instruments. We take this testing procedure to data from the RTE cereal market where we first estimate demand in a rich framework using detailed scanner data as well as demographic data of shoppers. We then apply our test against several alternative models of

[60] This internalization parameter design mirrors that in Kennedy et al. (2017).

[61] In Appendix Section D.1 we offer further pairwise comparisons. Our findings consistently favor own-firm profit maximization against alternative conduct models.

Table 9: Profit Maximization vs. τ

	Others' Costs	Demographics	BLP Inst.	Dmd. Opt. Inst.
$\tau = .1$	-0.4328	-0.321	-0.6851	-0.7026
	[-0.0176/0.0407]	[-0.0145/0.0452]	[-0.0372/0.0543]	[-0.0437/0.0622]
$\tau = .2$	-1.0734	-0.9217	-1.5237	-1.6487
	[-0.224/0.2087]	[-0.201/0.2181]	[-0.418/0.2743]	[-0.509/0.3087]
$\tau = .3$	-2.6522	-2.6837	-3.1898	-3.5225
	[-1.590/0.5995]	[-1.640/0.6111]	[-2.570/0.8057]	[-3.140/0.8914]
$\tau = .4$	-3.8962	-4.1062	-4.5616	-4.9736
	[-5.140/1.3192]	[-5.210/1.2688]	[-7.950/1.7428]	[-9.620/1.9342]
$\tau = .5$	-3.2687	-3.4082	-3.7689	-4.1914
	[-10.6435/3.2562]	[-10.5109/3.0840]	[-16.512/4.3811]	[-20.022/4.7769]
$\tau = .6$	-3.3722	-3.6771	-4.0833	-4.547
	[-17.6249/5.2266]	[-17.5416/4.7706]	[-28.7501/7.0409]	[-34.935/7.6831]
$\tau = .7$	-4.1593	-4.4611	-4.6689	-5.2397
	[-36.6078/8.8014]	[-35.6581/7.9931]	[-56.8988/12.1867]	[-69.4023/13.2456]
$\tau = .8$	-4.7044	-5.1992	-5.3273	-5.8925
	[-64.5747/13.7264]	[-61.8094/11.8882]	[-99.7635/18.7269]	[-119.7003/20.3141]
$\tau = .9$	-4.4131	-4.7389	-4.8305	-5.3131
	[-120.783/27.3691]	[-114.8521/24.2358]	[-179.9577/37.2543]	[-214.9648/40.4595]

Notes: This table reports testing results for pairwise comparisons of own-firm profit maximization against τ-adjusted common ownership (as indicated by row). The column label indicates the set of excluded variables used in the Step 2 regression. A negative value constitutes evidence for profit maximization and a positive value constitutes evidence for the τ-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. The bootstrap is clustered at the retailer by DMA by year level and observations are weighted by servings volume.

conduct. We find support for the classical model of own-profit maximization; however, simulated effects of common ownership incentives imply large potential effects. While the common ownership hypothesis is of interest in this paper, the testing procedure we develop is more general and applicable elsewhere. We hope our method makes it possible for other researchers to examine firm price-setting behavior in a variety of other settings.

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.

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

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

Backus, M., C. T. Conlon, and M. Sinkinson (2020b): "Common Ownership Data: Scraped SEC form 13F filings for 1999-2017," Https://doi.org/10.7910/DVN/ZRH3EU, Harvard Dataverse, V1.

Belloni, A., D. Chen, V. Chernozukhov, and C. Hansen (2012): "Sparse Models and Methods for Optimal Isntruments with an Application to Eminent Domain," *Econometrica*, 80, 2369–2429.

Berry, S., M. Gaynor, and F. Scott Morton (2019): "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization," *Journal of Economic Perspectives*, 33, 44–68.

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 (2020): "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.

Breiman, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32.

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," Https://arxiv.org/abs/2011.06158.

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

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

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

Cox, D. R. (1962): "Further Results on Tests of Separate Families of Hypotheses," *Journal of the Royal Statistical Society. Series B (Methodological)*, 24, 406–424.

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.

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

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.

Genesove, D. and W. P. Mullin (1998): "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914," *The RAND Journal of Economics*, 29, 355–377.

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

Hall, A. R. and D. Pelletier (2011): "Nonnested Testing in Models Estimated via Generalized Method of Moments," *Econometric Theory*, 27, 443–456.

Hartford, J., G. Lewis, K. Leyton-Brown, and M. Taddy (2017): "Counterfactual Prediction using Deep Instrumental Variable Networks," Working Paper.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

Hendel, I. and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74, 1637–1673.

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

Imbens, G., S. Donald, and W. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55–93.

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 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.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variables Estimation of Nonparamatric Models," *Econometrica*, 71, 1565–1578.

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.

Schmalensee, R. (1989): "Chapter 16 Inter-industry studies of structure and performance," Elsevier, vol. 2 of *Handbook of Industrial Organization*, 951 – 1009.

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

A.1. Proofs of Propositions

Lemma 1: A Property of Quadratic Forms For a symmetric, positive semi-definite matrix $A \in \mathbb{R}^{k \times k}$ and two vectors $(x, y) \in \mathbb{R}^k$, the difference in quadratic forms is given by:

$$x^T Ax - y^T Ay = (x+y)^T A(x-y).$$

Proof of Lemma:

$$\begin{aligned}
(x+y)^T A(x-y) &= x^T A(x-y) + y^T A(x-y) \\
&= x^T Ax - x^T Ay + y^T Ax - y^T Ay \\
&= x^T Ax - \left(\left(x^T Ay\right)^T\right)^T + y^T Ax - y^T Ay \\
&= x^T Ax - \left(y^T A^T x\right)^T + y^T Ax - y^T Ay \\
&= x^T Ax - \left(y^T Ax\right)^T + y^T Ax - y^T Ay \\
&= x^T Ax - y^T Ax + y^T Ax - y^T Ay \\
&= x^T Ax - y^T Ay.
\end{aligned}$$

Statement of Proposition 1a:
Under the following assumptions:
(i) A fixed $k \times k$ positive semi-definite weighting matrix W;
(ii) A $(n \times k)$ matrix of instruments Z of full column rank k;
(iii) A vector of n unobservable cost components $\omega^m = p - h(z) - \eta^m$ and moment restrictions such that $E[\omega^{m\prime} Z] = 0 \in \mathbb{R}^k$ is satisfied at the true η^0 and not at $\eta^m \neq \eta^0$, and
(iv) The function $h(z)$ is known (rather than estimated).

(a) As $n \to \infty$ and under standard regularity conditions, the difference in GMM objective functions ($Q_W(\eta) \in \mathbb{R}^+$) under weighting matrix W, for any two markup models (η^1, η^2), can be expressed as:

$$Q_W(\eta^1) - Q_W(\eta^2) \xrightarrow{p} -\mathbb{E}[Z'\,\omega^1]'\,W\;\mathbb{E}[Z'\Delta\eta^{1,2}] - \mathbb{E}[Z'\,\omega^2]'\,W\,\mathbb{E}[Z'\Delta\eta^{1,2}] \text{ where } \Delta\eta^{1,2} = \eta^1 - \eta^2.$$

Proof of Proposition 1a:
Under (iii) and (iv) note we can write: $\omega_1 - \omega_2 = -\Delta\eta^{1,2}$ and $\omega_2 = \omega_1 + \Delta\eta^{1,2}$.

The difference in GMM objective functions is given by:

$$\begin{aligned} Q_W(\eta^1) - Q_W(\eta^2) &= \omega^{1T} Z\, W\, Z^T \omega^1 - \omega^{2T} Z\, W\, Z^T \omega^2 \\ &= (\omega^1 + \omega^2)^T Z\, W\, Z^T (\omega^1 - \omega^2) \text{ by Lemma 1} \\ &= -(\omega^1 + \omega^2)^T Z\, W\, Z^T \Delta\eta^{1,2} \\ &= -\omega^{1T} Z\, W\, Z^T \Delta\eta^{1,2} - \omega^{2T} Z\, W\, Z^T \Delta\eta^{1,2} \\ &\xrightarrow{p} -\mathbb{E}[Z'\,\omega^1]^T\, W\, \mathbb{E}[Z'\Delta\eta^{1,2}] - \mathbb{E}[Z'\,\omega^2]^T\, W\, \mathbb{E}[Z'\Delta\eta^{1,2}]. \end{aligned}$$

As $n \to \infty$ and under standard regularity conditions (first two sets of first moments are finite), the SLLN applies so $(Z^T\omega^m) \xrightarrow{a.s.} \mathbb{E}[Z^T\omega^m]$ and $(Z^T\Delta\eta^{1,2}) \xrightarrow{a.s.} \mathbb{E}[Z^T\Delta\eta^{1,2}]$, which delivers the final line.

Statement of Proposition 1b:
(v) Model 1 is the correctly specified model, i.e. $\mathbb{E}[Z'\,\omega_1] \xrightarrow{a.s} 0$.
(b) Under the additional condition (v), then:

$$Q_W(\eta^1) - Q_W(\eta^2) \xrightarrow{p} -\mathbb{E}[Z'\,\Delta\eta^{1,2}]\, W\, \mathbb{E}[Z'\,\Delta\eta^{1,2}].$$

Proof of Proposition 1b:
Using the fact that $\omega_2 = \omega_1 + \Delta\eta^{1,2}$ and the second to last line of the proof of Proposition 1(a)

$$\begin{aligned} Q_W(\eta^1) - Q_W(\eta^2) &= \omega^{1T} Z\, W\, Z^T \Delta\eta^{1,2} + \omega^{2T} Z\, W\, Z^T \Delta\eta^{1,2} \\ &= \omega^{1T} Z\, W\, Z^T \Delta\eta^{1,2} + (\omega^1 + \Delta\eta^{1,2})^T Z\, W\, Z^T \Delta\eta^{1,2} \\ &= 2 \cdot \omega^{1T} Z\, W\, Z^T \Delta\eta^{1,2} + \Delta\eta^{1,2T} Z\, W\, Z^T \Delta\eta^{1,2} \\ &\xrightarrow{p} \mathbb{E}[Z^T\Delta\eta^{1,2}]^T W\, \mathbb{E}[Z^T\Delta\eta^{1,2}] \text{ since } \mathbb{E}[Z^T\omega^1] \xrightarrow{a.s.} 0 \text{ by (v).} \end{aligned}$$

A.2. Microfoundation of Demand Model

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

$$U_{ijt} = \delta_{jt} - \alpha p_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \tag{1}$$

Following the usual convention we decompose the utility into three parts: a mean preference which all consumers agree on δ_{jt}, a consumer-specific taste shock μ_{ijt}, and an idiosyncratic error ε_{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 ε_{ijt} is distributed IID type I extreme value (Gumbel) so that market shares are given by:

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

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

$$\delta_{jt}(\mathcal{S}_t, \theta_2) = h_d(x_{jt}, v_{jt}; \theta_1) + \xi_{jt}. \tag{3}$$

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 C.7. 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 C-2 and Figure C-4. We also include a large number of two-way fixed effects: 946 product fixed effects and 1970 chain-week fixed effects.[62]

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

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

We define the nonlinear parameters as $\widetilde{\theta}_2 = [L, \Pi]$ where the Π represent the parameters on interactions between demographic draws $y_{ict} = [\text{kids}_{ict}, \text{income}_{ict}]$ and observable characteristics x_{jct}, and L represents the Cholesky root of covariance matrix ($LL' = \Sigma^2$) for the (multivariate normally distributed) unobserved heterogeneity. We allow consumers to have heterogeneous preferences over prices and a constant (which captures inside vs. outside good substitution).[63] 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.[64]

A.3. Markups and their Derivatives

We can augment the usual static multi-product Bertrand Nash pricing game to allow for common ownership (or "internalization" effects). In this case a firm chooses the price for product j but now treats \$1 of rival g's profit as κ_{fg} of their own. We further augment this setup with an "internalization parameter" τ. 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.

The first-order condition of firm f which controls several products $\mathcal{J}_f$ with respect to p_j is given by the following:

$$\arg \max_{p_j \,:\, j \in \mathcal{J}_f} \sum_{j \in \mathcal{J}_f} (p_j - mc_j) \cdot s_j(\mathbf{p}) + \tau \cdot \left(\sum_g \kappa_{fg} \sum_{j \in \mathcal{J}_g} (p_k - mc_k) \cdot s_k(\mathbf{p}) \right),$$

$$s_j(p) + \frac{\partial s_j}{\partial p_j}(\mathbf{p}) \cdot (p_j - mc_j) + \sum_{k \in \mathcal{J}_f} \frac{\partial s_k}{\partial p_j}(\mathbf{p}) \cdot (p_k - mc_k) + \tau \cdot \left(\sum_g \kappa_{fg} \sum_{k \in \mathcal{J}_g} \frac{\partial s_k}{\partial p_j}(\mathbf{p}) \cdot (p_k - mc_k) \right) = 0. \tag{4}$$

[62] Nevo (2000) and Nevo (2001) include one fixed effect for each of the 26 products in the dataset.
[63] See Section 4.3 or Appendix C.1 for a discussion on how prices and product characteristics are constructed.
[64] See Section 4.3 and Appendix C.6.

With a bit more notation we can write the system of first–order conditions in matrix form, which is helpful for what follows. First, let $\Omega(\mathbf{p})$ denote the $J \times J$ matrix of demand derivatives so that each (j,k) entry is given by $\frac{\partial s_j}{\partial p_k}$. Second, we introduce the product level "cooperation matrix". We can augment the setup in Nevo (2001) with κ_{fg} for potential common ownership effects, and τ as the "internalization parameter":

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

Then, for a single market t, where $\odot$ denotes the element-wise (Hadamard) product of the two $J_t \times J_t$ matrices:

$$\mathbf{s_t}(\mathbf{p_t}) = (\mathcal{H}_t(\kappa,\tau) \odot \Omega(\mathbf{p_t})) \cdot (\mathbf{p_t} - \mathbf{mc_t}). \tag{5}$$

The augmented markup in matrix form can be written as:

$$\eta_t(\mathbf{p_t}, \mathbf{s_t}, \kappa_t, \tau) = (\mathcal{H}(\kappa_t,\tau) \odot \Omega(\mathbf{p_t}))^{-1}\mathbf{s_t}.$$

This gives us a multi-product, multi-firm markup which incorporates potential conduct effects: $\eta(\mathbf{p}, \mathbf{s}, \kappa, \tau)$. Note that this depends on the conduct among all firms κ as well as all of the endogenous variables.

We are also 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.

The markup has a derivative given by:

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

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

We also consider the derivative with respect to a scalar element of κ, κ_{fg}:

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

Where $I_{f=g}$ 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.

The Jacobians of the markup with respect to the internalization parameters (τ, κ_{fg}) are important because they represent the (infeasible) optimal instruments (in the sense of Amemiya (1977);

Chamberlain (1987)) for the conduct parameters themselves. They are also the continuous analogue of the discrete markup difference that forms the basis for our test $E[\Delta\eta_{jt}|\mathbf{z_t}]$.

In order to compute the feasible approximations $E\left[\frac{\partial\eta_{jt}}{\partial\kappa_{fg}}|\mathbf{z_t}\right]$ or $E\left[\frac{\partial\eta_{jt}}{\partial\tau}|\mathbf{z_t}\right]$, we have two choices which parallel the discussion of optimal instruments in Berry et al. (1995, 1999); Conlon and Gortmaker (2020). The Newey (1990) approach is to compute the expectations by calculating the Jacobian at the true values of $(\mathbf{p_t}, \mathbf{s_t})$ (and the conjectured τ and θ_2) and then projecting those values onto exogenous variables $\mathbf{z_t}$ via (nonlinear) regression. The Chamberlain (1987) approach as discussed in Berry et al. (1999); Reynaert and Verboven (2014) instead would allow us to calculate $E\left[p_{jt}|\mathbf{z_t}\right]$ with a "first-stage" estimate $\hat{\mathbf{p}}_\mathbf{t}$ from a (potentially nonlinear) regression, draw $\boldsymbol{\xi}_\boldsymbol{t}^*$ and calculate $\mathbf{s_t}(\hat{\mathbf{p}}_\mathbf{t}, \boldsymbol{\xi}_\boldsymbol{t}^*; \widehat{\theta}_2) = \hat{\mathbf{s}}_\mathbf{t}$. Given $(\hat{\mathbf{s}}_\mathbf{t}, \hat{\mathbf{p}}_\mathbf{t})$ we can calculate $\Omega(\hat{\mathbf{s}}_\mathbf{t}, \hat{\mathbf{p}}_\mathbf{t}; \widehat{\theta}_2)$ and $\boldsymbol{\eta_t}(\hat{\mathbf{s}}_\mathbf{t}, \hat{\mathbf{p}}_\mathbf{t}; \widehat{\theta}_2)$ now purged of the endogeneity problem of $\mathbf{p_t}$ and $\boldsymbol{\xi_t}$, and use these to calculate the Jacobian. The conventional approach (Berry et al., 1999) replaces $\boldsymbol{\xi_t}^* = 0$, and Conlon and Gortmaker (2020) show that it works well in Monte Carlo analysis.

B. Financial Market Data Appendix

B.1. Ownership Data

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 Backus et al. (2019).

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.

B.2. Share Swap Mergers

One important aspect of common ownership is that post-merger incentives may be affected by the financial structure of the transaction. For example, in an all-cash deal, the equity of the target is purchased by the acquirer and the merged firm retains the ownership structure of the acquirer β_{fs}. We report the ownership matrix for Q4 2016 below. In practice, many transactions are realized as share-swaps where the acquirer issues additional shares and uses those to acquire the shares of the target.

Absent any acquisition, we have that the ownership stake of s in f is given by:

$$\beta_{fs} = \frac{\text{shares}_{fs}}{\text{outstanding}_f}.$$

After a share swap, we need to adjust both the numerator and denominator. First we calculate what is known as the *exchange ratio*

$$ER_{target,acquirer} = \frac{\text{price}_{target}}{\text{price}_{acquirer}}.$$

Then we update the shares held in the acquirer:

$$\text{shares}_{s,merged} = \text{shares}_{s,acquirer} + \text{shares}_{s,target} \cdot ER_{target,acquirer}$$
$$\text{outstanding}_{s,merged} = \text{outstanding}_{s,acquirer} + \text{outstanding}_{s,target} \cdot ER_{target,acquirer}.$$

Because κ_{fg} depends on relative investor concentration, the effects of a share-swap merger on κ can be difficult to predict and need not be a convex combination of pre-merger κ and the merged ownership under an all-cash deal. We provide several examples below. The most striking feature is that acquisitions involving Quaker Oats (a division of PepsiCo) can lead to huge changes in profit weights for the merged entity. The extremely large market capitalization of PepsiCo tends to swamp the other cereal manufacturer's ownership structure, and the relative generosity $\kappa > 1$ can be even more pronounced under a share-swap acquisition.

	GM	*KEL*	*POST*	*QKR*
GM	1.00	0.67	0.70	0.80
KEL	0.23	1.00	0.17	0.19
POST	0.41	0.30	1.00	0.44
QKR	1.07	0.74	1.00	1.00

	GM	*KEL/QKR*	*POST*
GM	1.00	0.78	0.78
KEL/QKR	1.08	1.00	0.99
POST	0.37	0.34	1.00

	GM/KEL	*POST*	*QKR*
GM/KEL	1.00	0.85	0.85
POST	0.33	1.00	0.39
QKR	0.93	1.12	1.00

	GM/QKR	*KEL*	*POST*
GM/QKR	1.00	0.75	1.0
KEL	0.19	1.00	0.2
POST	0.36	0.27	1.0

	GM	*KEL*	*QKR/POST*
GM	1.00	0.67	0.80
KEL	0.23	1.00	0.19
QKR/POST	1.06	0.73	1.00

	GM/POST	*KEL*	*QKR*
GM/POST	1.00	0.69	0.83
KEL	0.22	1.00	0.19
QKR	1.06	0.74	1.00

	GM	*KEL/POST*	*QKR*
GM	1.00	0.69	0.80
KEL/POST	0.31	1.00	0.26
QKR	1.07	0.78	1.00

C. Product Market Data Appendix

C.1. 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 collinearity 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 C-2

Share-Swap Mergers under Common Ownership							
Firm	GM-KEL	GM-QKR	GM-POST	KEL-QKR	KEL-POST	QKR-POST	Monopoly
General Mills	0.68	-3.43	-0.58	1.59	-2.73	-0.29	3.12
Kellogg	6.75	4.88	5.69	5.55	7.03	5.76	8.80
Quaker Oats	-4.36	-7.30	-6.73	-4.41	-9.46	-5.79	0.40
Post	8.24	5.65	5.89	8.11	6.97	3.89	9.51
Price Index	3.74	0.92	2.33	3.66	1.89	2.10	6.09
Cash Mergers under Common Ownership							
Firm	GM-KEL	GM-QKR	GM-POST	KEL-QKR	KEL-POST	QKR-POST	Monopoly
General Mills	1.31	-3.00	-0.55	-4.86	-3.52	-0.22	3.12
Kellogg	7.29	4.25	5.49	4.88	6.82	5.79	8.80
Quaker Oats	-5.03	-6.81	-7.17	-5.46	-10.38	-5.79	0.40
Post	5.96	2.39	5.93	1.03	6.74	3.89	9.51
Price Index	3.60	0.21	2.26	-0.28	1.37	2.14	6.09
Base Case: Mergers under Own Profit Maximization							
Firm	GM-KEL	GM-QKR	GM-POST	KEL-QKR	KEL-POST	QKR-POST	Monopoly
General Mills	4.69	1.10	3.00	-0.10	-0.16	-0.11	9.42
Kellogg	5.13	-0.13	-0.19	1.04	2.90	-0.11	9.30
Quaker Oats	-0.37	6.39	-0.36	5.60	-0.37	3.93	14.87
Post	-0.15	-0.11	6.16	-0.10	5.19	1.45	12.76
Price Index	3.32	0.91	2.22	0.78	1.96	0.58	10.25

Table B-1: Price Increases under Counterfactual Scenarios

Notes: This table reports counterfactual percent price increases from the observed baseline in the fourth quarter of 2016. We compute the serving-weighted percentage change in price at the market and firm level (with the last row containing all firms) and report the median price change across markets. The market-level mean price per serving observed in the data is \$0.2255. Columns indicate the counterfactual scenario: the first six are pairwise mergers, the seventh is a hypothetical monopolist composed of the four largest firms, and the last is under common ownership. Source: Authors' computations.

	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 ×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 C-2: Summary Statistics of Product Characteristics (3478 Products)

We use a principal components approach to reduce the dimension of the product space and provide 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 Λ 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 Λ_k is the same diagonal matrix as Λ 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 Λ. $\tilde{X}$ 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 Ω 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 information 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. The first four principal components are plotted in pairs in Figure C-1

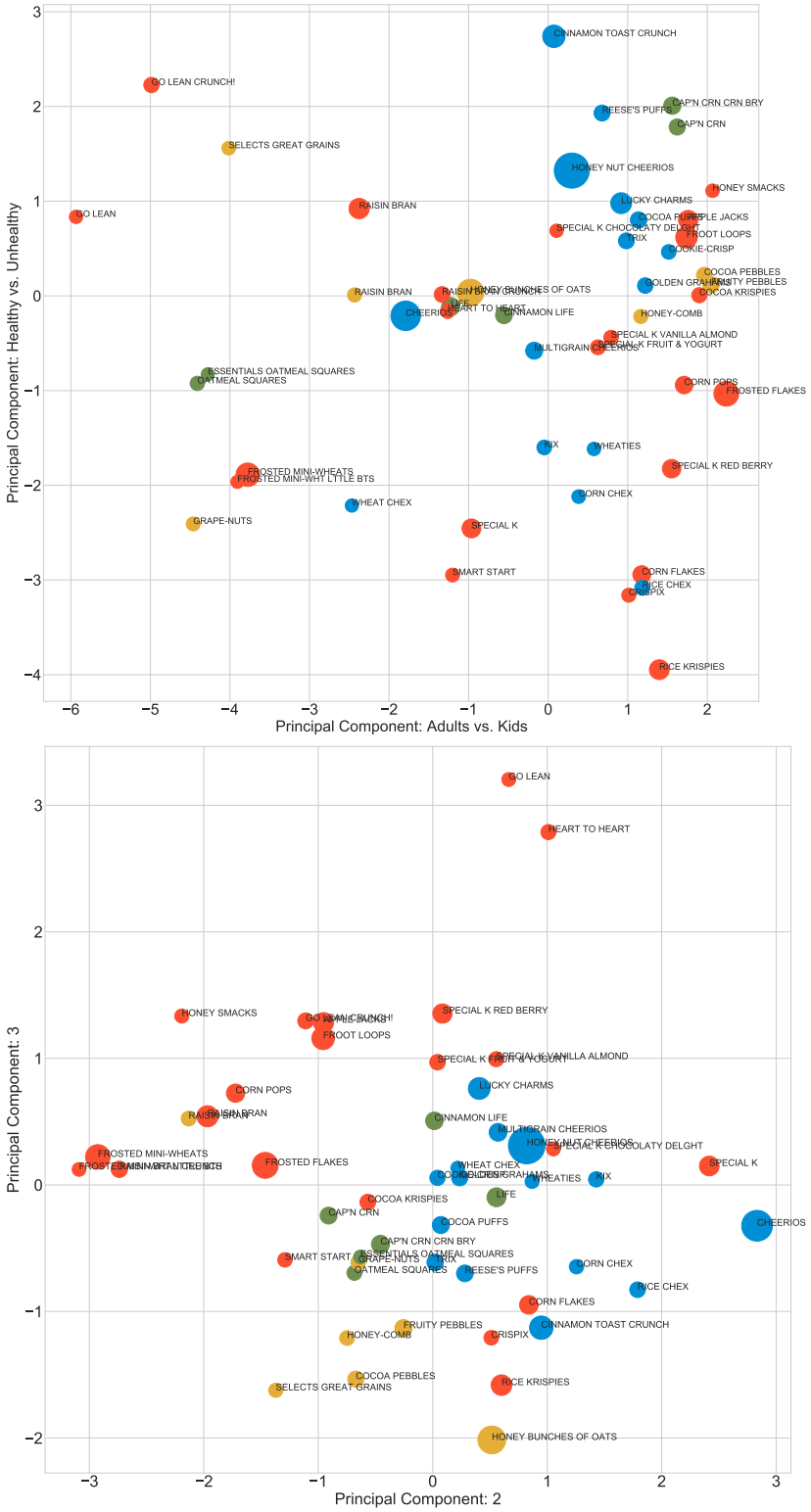


Figure C-1: Principal Components of Nutritional Data: Top 50 Brands

Notes: This figure depicts the PCA location of the top 50 brands of cereal. Circle size is proportional to the inverse of national sales rank. Colors indicate different manufacturers. Source: Authors' computations.

C.2. Macro Variables

Here we document some additional macroeconomic variables 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). We plot MSA level unemployment in Figure C-2.

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.

We also acquire data on commodities prices from Global Financial Data on a weekly basis. We track the prices of: corn, oats, rice, wheat, and sugar and compute an index value relative to the middle of our sample (2011). We report those input prices in Figure C-3.

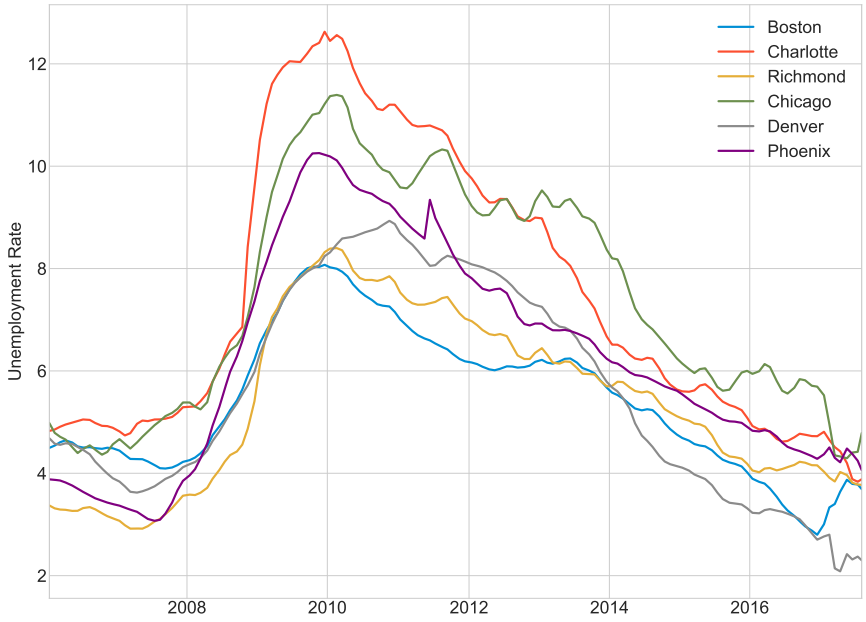


Figure C-2: Unemployment Rate by MSA/DMA

C.3. Private Label

Below, in Figure C-4 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

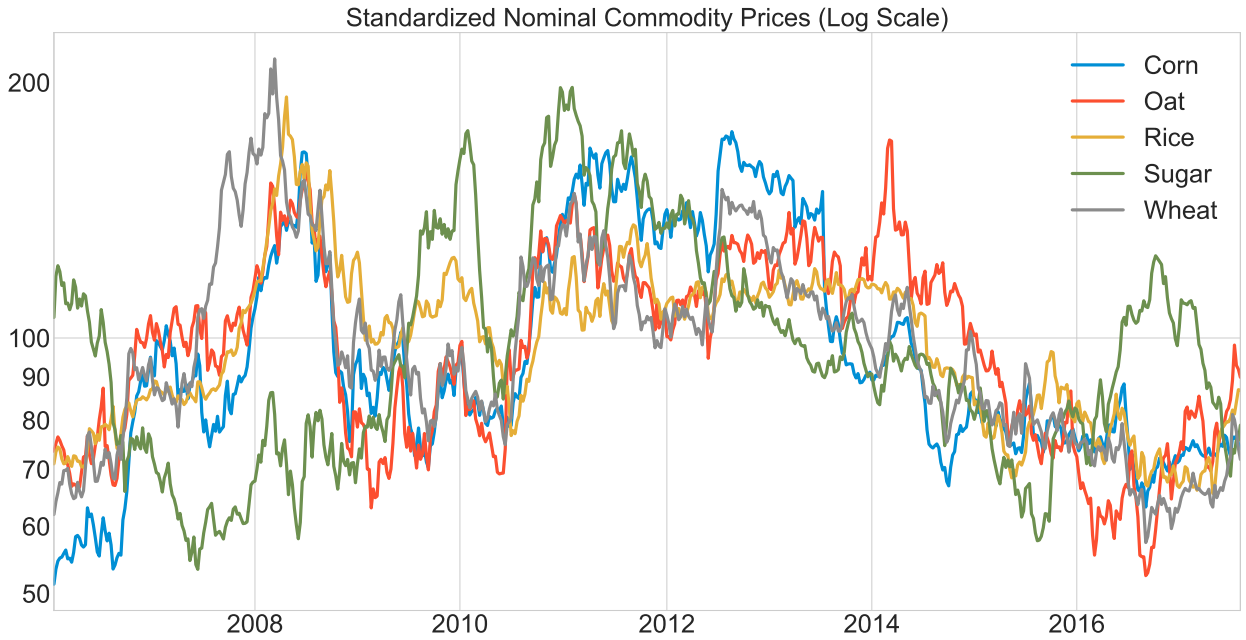


Figure C-3: Input Prices

Notes: This figure depicts the time series of major RTE cereal ingredient input prices for the period 2007–2017. Source: Global Financial Data.

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 $MHHI\Delta$).

C.4. 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 (2020) and begin with the recipe in Gandhi 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 lth 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 $d_{jk,(m)}$:

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

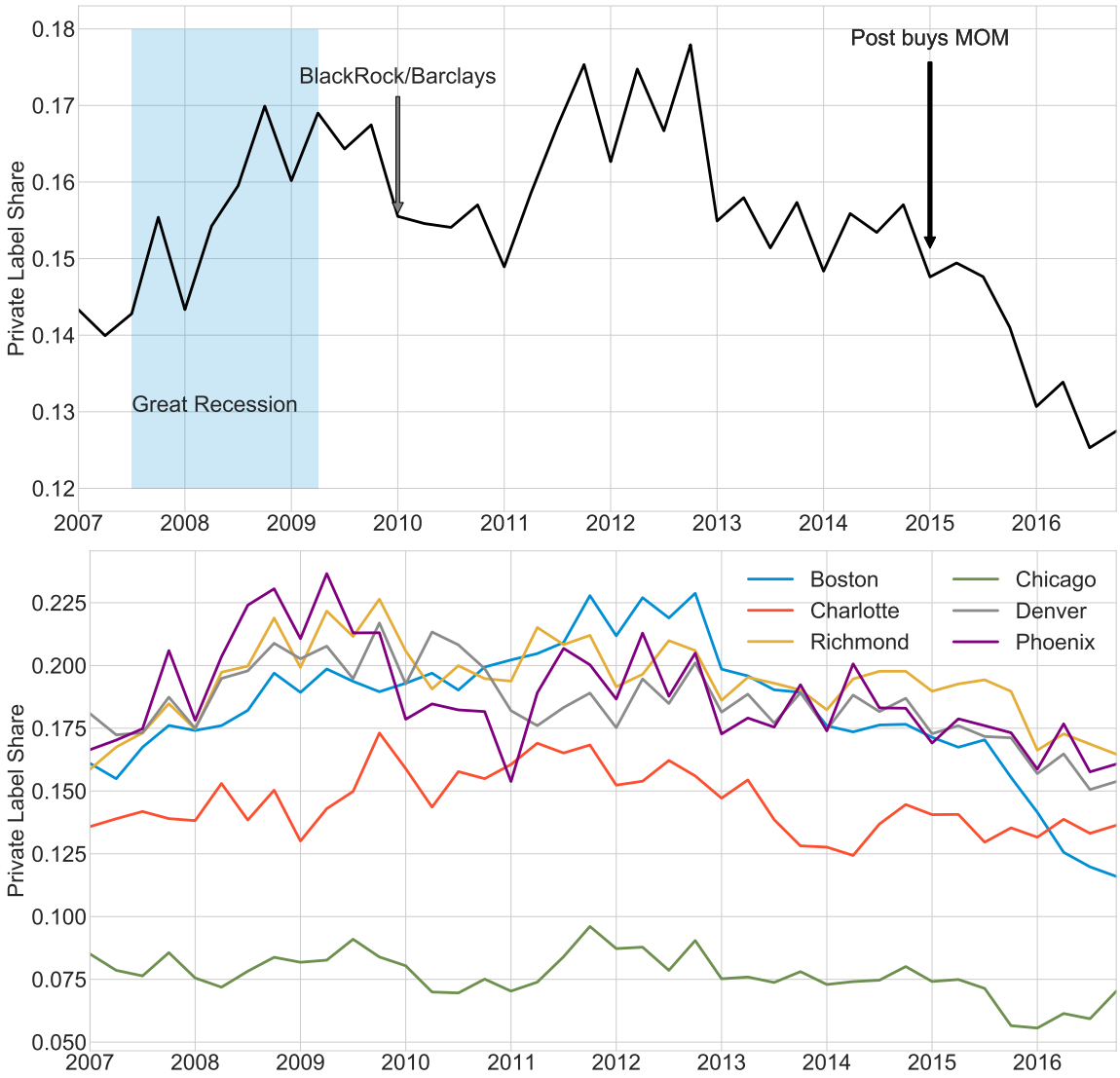


Figure C-4: Private Label Market share

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:

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

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}^{quad} as:

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

Gandhi and Houde (2019) show these instruments are most useful in the estimation of the standard deviations and covariances of the random coefficients Σ. Since we are also interested in estimating demographic interactions Π, 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: with quantiles of the income distribution (conditional on children), and the fraction of households with children.

$$z_{jt}^D = z_{jt}^{\text{quad}} \otimes \left\{ 1, \%kids_t, inc_{it}^{10\%,k}, inc_{it}^{50\%,k}, inc_{it}^{90\%,k}, inc_{it}^{10\%,nk}, inc_{it}^{50\%,nk}, inc_{it}^{90\%,nk} \right\}.$$

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 (2020)). The simplified version of the expression in Conlon and Gortmaker (2020) 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 $E[p_{jt}|x_{jt}, \mathrm{w}_{jt}, z_{jt}] = \hat{p}_{jt}$, 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.[65] 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 optimal weighting matrix.

C.5. 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

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

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:

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

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 C-3. 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 C-3: Predicted Market Size Using Milk and Eggs Sales

We then define our market size as follows, where λ 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 C-5, 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.

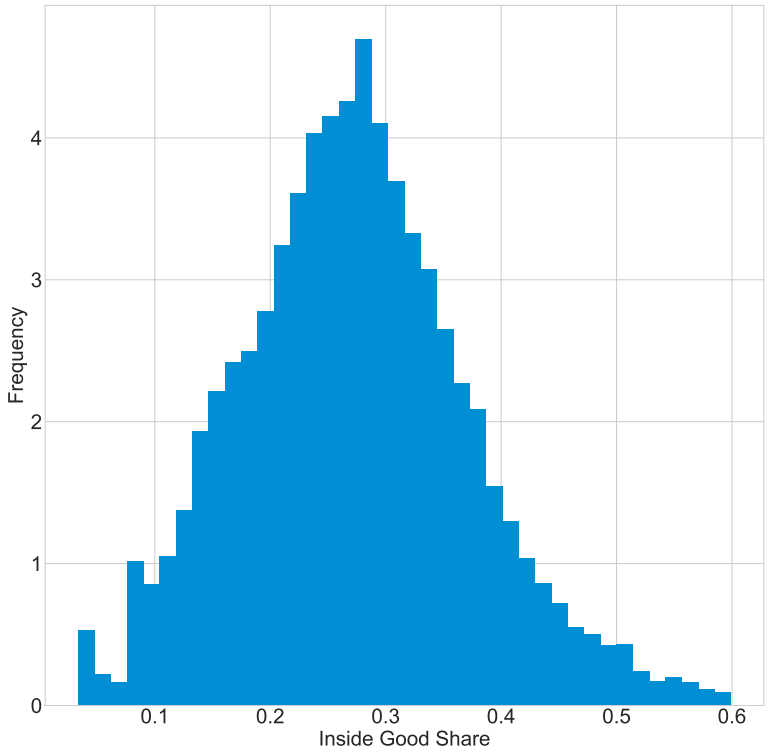


Figure C-5: Distribution of Inside Good Shares

C.6. Demographics and Micro Moments

We construct the distribution of demographics for each DMA-chain-year as follows. We are interested in estimating the joint distribution of:

$$f(\text{inc}_{it}, \text{kids}_{it} | i \in \text{DMA-Chain}_{it}) = f(\text{inc}_{it} | \text{kids}_{it}, i \in \text{DMA-Chain}_{it}) \cdot \mathbb{P}(\text{kids}_{it} | i \in \text{DMA-Chain}_{it}).$$

First we estimate $\mathbb{P}(\text{kids}_{it} | i \in \text{DMA-Chain}_{it})$ from a weighted sample of households in the Nielsen Panelist (HMS) data. 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.[66] We then split the sample into households with children at home (roughly 15-25%) and those without and separately estimate a parametric lognormal distribution by method of moments. The moments we estimate are of the form:

$$\mathbb{P}(a_s \leq \text{income}_{ht} \leq a_{s+1}).$$

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:

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

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, conditional on not purchasing cereal, we compute the expected value of income and the expected probability of children in the home.

C.7. 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 C-4. 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}.

C.8. Demand Estimation

Below is the problem definition in `pyBLP`.

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

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 C-4: List of Product Recalls

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

Figure C-6: 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-5: Full Pairwise Comparisons

	Others' Costs			
	Profit Max	Common Ownership	Perfect	Monopolist
Profit Max	-	-4.8318	-7.7749	-5.2711
Common Ownership		-	-8.0656	-5.237
Perfect Competition			-	-4.0458
Monopolist				-
Demographics				
Profit Max	-	-5.5819	-8.7051	-6.7789
Common Ownership		-	-9.223	-6.6265
Perfect Competition			-	-4.6365
Monopolist				-
BLP Inst.				
Profit Max	-	-5.4286	-8.9758	-5.9158
Common Ownership		-	-9.1621	-5.8744
Perfect Competition			-	-4.6767
Monopolist				-
Dmd. Opt. Inst.				
Profit Max	-	-5.9627	-10.0654	-6.5933
Common Ownership		-	-10.4858	-6.5205
Perfect Competition			-	-5.1268
Monopolist				-

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 (b) 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.

D. Testing Appendix

D.1. Additional Pairwise Tests

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

Results for our four sets of exclusions are presented in Table D-5. 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 monopoly pricing as well.

We can break these same results out by firm, as we do in Table D-6. We see that the evidence against common ownership is somewhat weaker for Kellogg's, but this is unsurprising because, as

we saw in Figure 1, they have the smallest common ownership weights (κ^{CO}), and so it is relatively more difficult to distinguish own-profit maximization from common ownership.

D.2. Pakes-Wollmann Regressions

A separate innovation in the conduct testing literature is what we refer to as the "Pakes-Wollmann regression," introduced in Pakes (2017).[67] The author is interested in whether or not own-profit maximization is an appropriate model of firm behavior given an estimated demand system. The idea is to run the following regression:

$$p_{jt} = h_s(x_{jt}, \mathrm{w}_{jt}) + \lambda \cdot \eta^m_{jt} + \omega_{jt}, \tag{6}$$

and, since markups (η^m_{jt}) are endogenous, to use BLP instruments to estimate λ. This last point is the novel insight that we borrow for our testing procedure. Pakes (2017) predicts that, if own-profit maximization is what firms engage in, we should obtain $\hat{\lambda} = 1$ for markups c, and indeed he does using data and the demand system from Wollmann (2018). However, this approach is less suited to comparisons between alternative models. In general, distance of $\hat{\lambda}$ from 1 does not correspond to an economically interpretable measure of the similarity of firm behavior to the candidate model.[68]

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

Results are presented in Table D-7. 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 for profit maximization.

D.3. Step (a) and Step (b) Regressions

Here we document the intermediate marginal cost and markup difference regressions that make up Step (a) and Step (b) 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 D-8. 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 monopolist 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 (b) regressions in Figure D-7. 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

[67] This test is closely related to the "informal" test of Villas-Boas (2007).

[68] If firms behave heterogeneously, then the estimate $\hat{\lambda}$ will represent a weighted average effect, with weights that depend on the variation induced by the instruments. In this way it is possible to construct examples that will yield $\hat{\lambda}$ close to 1 for a candidate model that may be far from any particular firm's behavior. Even when firm behavior is homogenous, however, it is possible to construct examples in which $\hat{\lambda} = 1$ for a false model, and $\hat{\lambda}$ is arbitrarily far from 1 for a true model. 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 $\hat{\lambda} = (1+x)/x$. For example, if $x = 0.5$, the correct estimate is $\lambda = 3$. 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 $\hat{\lambda} \to 1$ as the hypothesis $x \to \infty$.

Table D-6: Pairwise Comparisons by Firm

	General Mills				Kellogg's			
Others' Costs	Profit Max	Common Ownership	Perfect	Monopolist	Profit Max	Common Ownership	Perfect	Monopolist
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 Competition			-	-1.2465			-	-0.9356
Monopolist				-				-
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 Competition			-	-1.1029			-	-0.7526
Monopolist				-				-
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 Competition			-	-1.4759			-	-1.1429
Monopolist				-				-
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 Competition			-	-1.5415			-	-1.2211
Monopolist				-				-
	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
Monopolist				-				-
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
Monopolist				-				-
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
Monopolist				-				-
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
Monopolist				-				-

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 (b) 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 D-7: 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
Monopolist	-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.

substantively better for all but the BLP instruments, and we believe this has to do with the fact that there are 143 of them.

E. Problems with the MHHI Regression Approach

E.1. Discussion of MHHI-based Approaches to Testing

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.[69] 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 the next section 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 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

[69] 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.

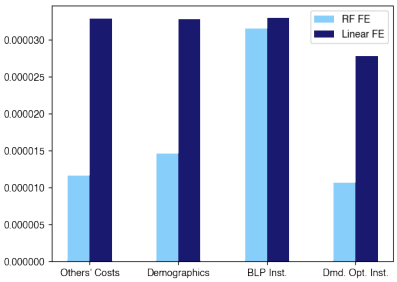
Table D-8: Linear Supply Regressions

	Profit Max (1)	Common Ownership (2)	Perfect (3)	Monopolist (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

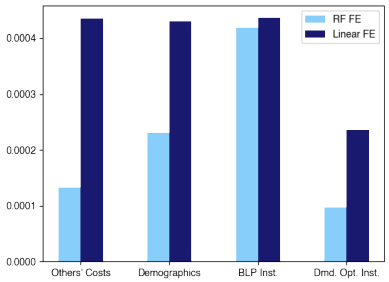
Notes: This table reports regressions of marginal costs, given by the column-wise conduct model, on included variables (x_{jt} and w_{jt}). Standard errors are clustered at the retailer by DMA by year level and observations are weighted by servings volume. * denotes statistical significance at $\alpha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

Figure D-7: Step (b) Mean-Squared Error

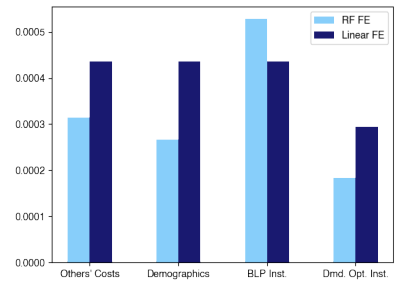
(a) Profit Max. vs. CO



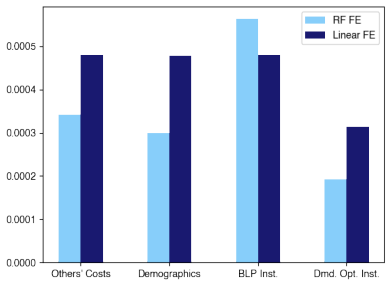
(b) Profit Max. vs. Monopoly



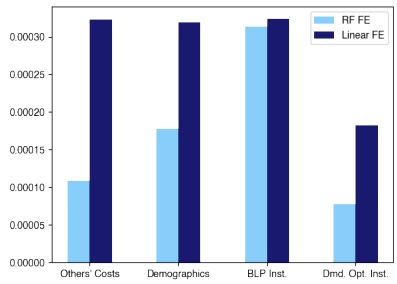
(c) Profit Max. vs. Perfect Comp.



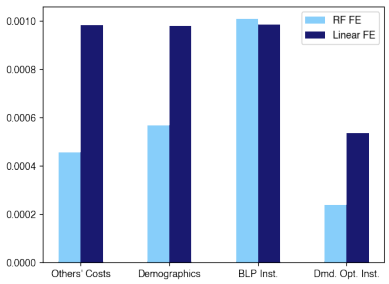
(d) CO vs. Perfect Comp.



(e) CO vs. Monopoly



(f) Perfect Comp. vs. Monopoly



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

as a causal relationship, but an expression of the misspecification of the regression.

There are several other reasons we eschew MHHI concentration measures. Many of these are related to measurement and market definition, and we discuss these in Backus et al. (2021) and Backus et al. (2019). 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 κ^{CO}. For this reason we adopt a structural approach based on the Bertrand first-order conditions and a series of exclusion restrictions.

Table E-9: 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 E.2.

E.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 E-9. From these numbers we would compute $\kappa_{12}^{CO} = \kappa_{21}^{CO} = 0.5$, 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 still find (spurious) evidence for common ownership using the MHHI approach?

Demand for both products is linear and given by:

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

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:

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

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:

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

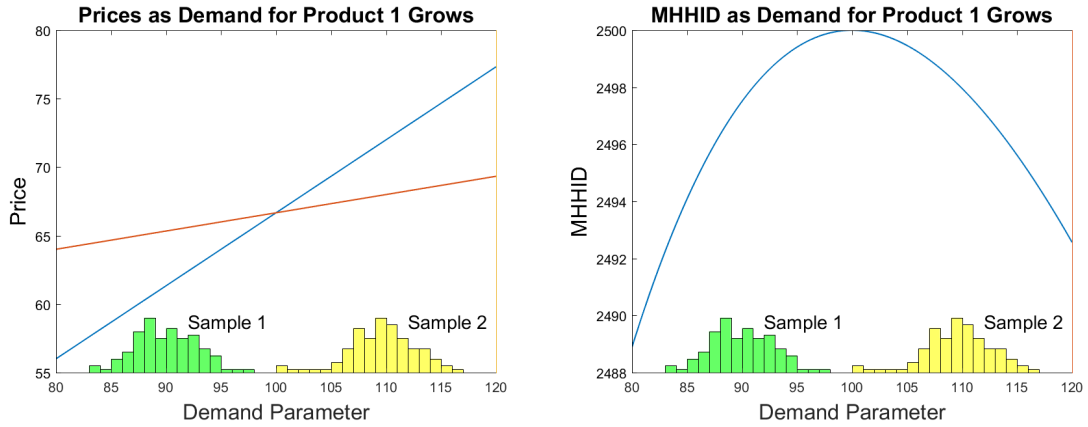


Figure E-8: Analytic/Numerical Example: Varying Demand Intercept

We can calculate $p_1^*(a_1), p_2^*(a_1)$ as well as $HHI(a_1)$. These are depicted in Figure E-8, 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.[70]

The Barclays acquisition instrument: One response to this concern has been to use the BlackRock-Barclays merger in 2010, and show that results are robust. To this end Azar et al. (2018a) instrument using the within-market variation alone, i.e. the change in MHHID due to the merger. This may help address the endogeneity of κ^{CO}, but does this differencing exercise also "clean out" the endogenous variation driven by demand shocks? We continue with the same parameterization 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.

Continuing with our example, suppose that in the ownership structure of Table E-9, 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 $\kappa_{12}^{CO} = \kappa_{21}^{CO} = 1/3$. 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 E-9. 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

[70] 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.

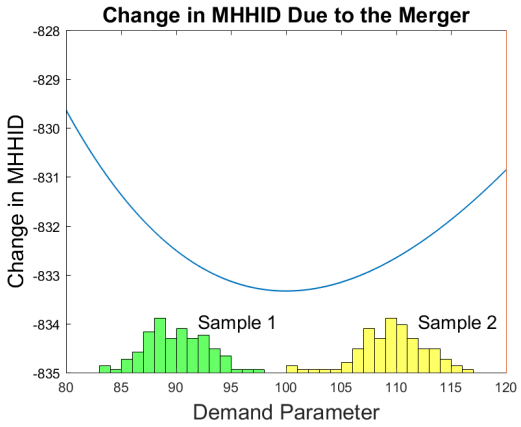


Figure E-9: Analytic/Numerical Example: Varying Demand Intercept

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 κ^{CO}, but since the MHHID measure interacts with market shares, the endogeneity problem remains. As a result, both results would again be spurious.

E.3. Concentration and Modified Concentration Measures in RTE Cereal

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.

E.4. 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 E-10 - E-13. 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

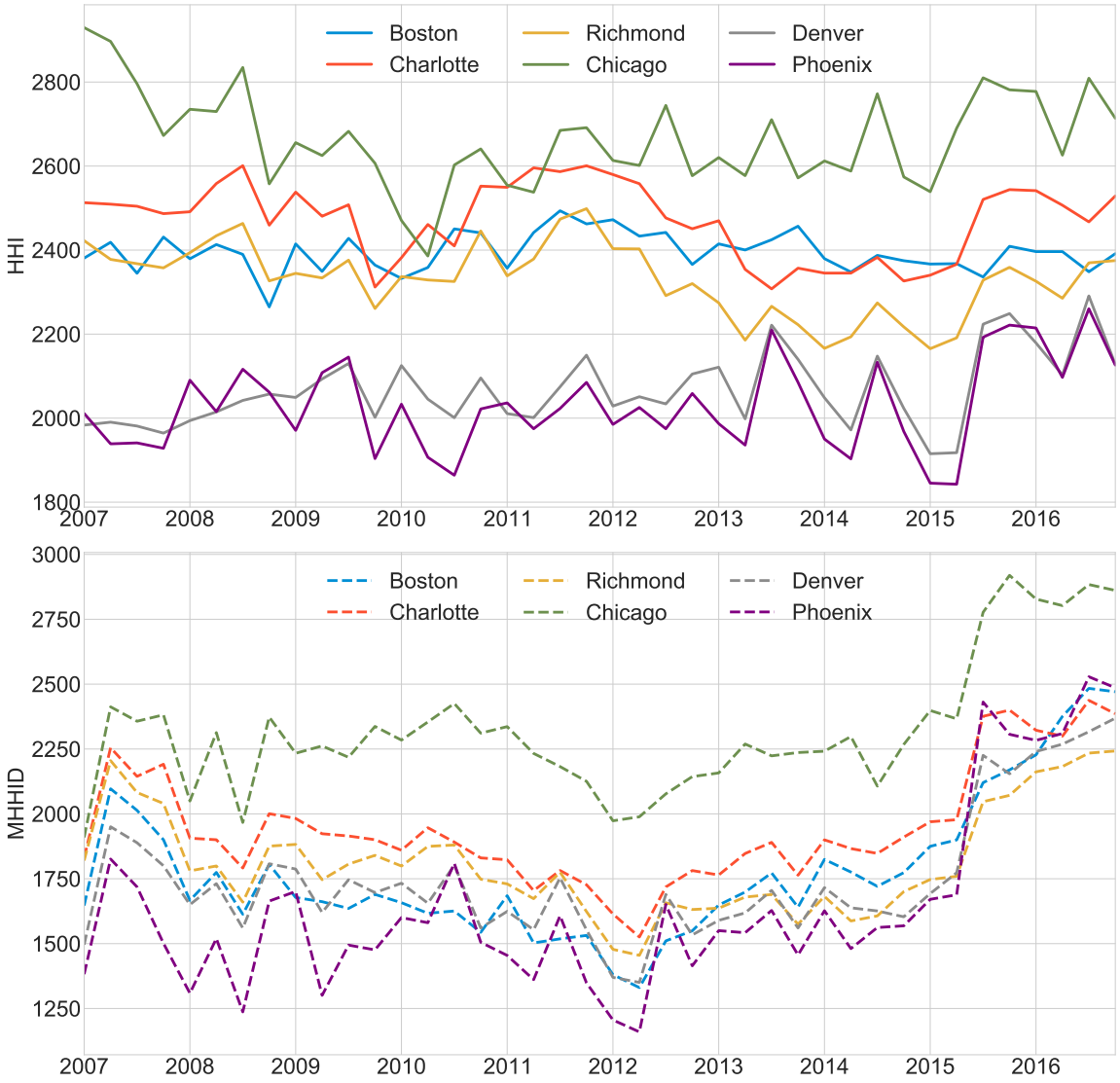


Figure E-10: HHI and $MHHI\Delta$ over time by City/DMA

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 E.2 as an explanation: cross sectional variation in demand provides variation in market shares. When we interact these market shares with κ_{fg}^{CO} we can produce spurious positive or negative correlations between $\Delta MHHI$ and price.

Table E-10: 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 E-11: 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

As a final exercise we run a "true reduced form" regression as suggested by O'Brien (2017). Here an observation is a manufacturer-chain-dma-quarter. The variable of interest is κ. Theoretically it should be the entire matrix K which we summarize by using the arithmetic or geometric mean of the off-diagonal κ^{CO} entries. We report the results of the regression of average prices on these functions of K in Table E-14. Once again we observe a negative (and significant) relationship between the common ownership profit weights κ and the prices p_{jmt}. For some specifications the results are barely (in)significant.

Table E-12: 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

Table E-13: 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

* p¡0.10, ** p¡0.05, *** p¡0.01

Table E-14: $\log p$ on κ regression

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