THE RELATIONSHIP BETWEEN PRICE AND MARKET STRUCTURE: EVIDENCE FROM THE US FOOD RETAIL INDUSTRY

Vardges Hovhannisyan, University of Wyoming, vhovhann@uwyo.edu
Marin Bozic, University of Minnesota, mbozic@umn.edu

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Disclaimer

"Any opinions, findings, recommendations, or conclusions are those of the authors and do not necessarily reflect the views of the Economic Research Service, U.S. Department of Agriculture. The analysis, findings, and conclusions expressed in this paper also should not be attributed to either Nielsen or Information Resources, Inc. (IRI)."

"This research was conducted in collaboration with USDA under a Third Party Agreement with IRI."

THE RELATIONSHIP BETWEEN PRICE AND MARKET STRUCTURE: EVIDENCE FROM THE US FOOD RETAIL INDUSTRY

ABSTRACT

This study utilizes unique product barcode, store, and retail real estate data to calculate consistent estimates of the effects of retail market structure on food prices in the US. Our uniquely disaggregated data allow for identification strategy that corrects for the type of endogeneity that plagues many previous studies on price-concentration relationship. Empirical findings from an instrumental-variables fixed-effects model indicate that retail concentration is endogenous to price determination. Further, retail prices are found to rise with retail concentration. Importantly, ignoring endogeneity results in a severe downward bias in the estimated effects of concentration on food prices.

Keywords: *Retail concentration, retail food price, endogeneity of retail concentration, instrumental variables fixed-effects regression*.

THE RELATIONSHIP BETWEEN PRICE AND MARKET STRUCTURE: EVIDENCE FROM THE US FOOD RETAIL INDUSTRY

1. Introduction

The US food marketing system has been undergoing significant structural changes recently. One such change has been the increasing consolidation and coordination in the food retailing sector (Balagtas, 2010). Specifically, the four largest grocery chains have seen their market shares increase from 16 to 32 % over the period 1982-2005 (Hovhannisyan, Stiegert, and Bozic, 2014). This change has the potential to reshape the horizontal competitive landscape among food retailers, as well as the vertical relationships between retailers and upstream food system participants. Hence, rising retail concentration has been at the center of heated debates among economists, policymakers, lawyers, and industry stakeholders alike. For example, the US Departments of Justice and Agriculture organized a series of public workshops with the goal of providing policymakers with an improved understanding of market conditions that determine farm and consumer prices (US DOJ, 2011).

Our study aims at informing this discussion by using unique data to calculate consistent estimates of the effects of retail market structure on retail-level food prices in the US. It builds upon a long stream of economics and marketing literature that investigate the relationship between market structure and market performance. Early studies in this line of literature investigate the effects of market concentration on firm profitability based on industry-level cross-section data from a wide range of industries. A major finding emerging from this literature is that market concentration and firm market shares are positively related to firm profitability (Schmalensee, 1989). More recent studies have shifted the focus from cross-industry to single-industry profit-concentration analyses to sidestep issues that may stem from fundamental industry differences, profit measurement, and the "efficiency" critique put forth by Demsetz (1973) regarding firm superiority driving the positive relationship between market concentration and profit. A general finding emerging from this literature is that higher market concentration goes hand-in-hand with higher prices (e.g., Weiss, 1989; Newmark, 2004).[1]

Early studies of the relationship between price and market structure treat concentration as an exogenous variable. However, there are multiple reasons to believe that market concentration is endogenous to firm performance and price determination, as pointed out by Froeb and Werden (1991), Berry (1992), and Evans, Froeb, and Werden (1993), just to name a few. Unless accounted for, the endogeneity of concentration can result in a severe bias in the estimate of the concentration parameter (e.g., Evans, Froeb, and Werden, 1993; Manuszak and Moul, 2008). A variety of approaches have been taken in the literature to address the endogeneity of market concentration. For example, Evans, Froeb, and Werden (1993) adopt a fixed-effects instrumental variables (IV) technique to study concentration-price relationship in the US airline industry. A major challenge plaguing this approach is the lack of instrumental variables that are both relevant and valid in a specific application, given inherently limited economic data. Hence, Evans, Froeb, and Werden (1993) utilize lagged values of the Herfindahl-Hirschman Index (HHI) estimates for market concentration to construct an instrumental variable. In the same vein, Dafny, Duggan, and Ramanarayanan (2009) rely on lagged HHI estimates to account for the endogeneity of

[1] See Ellickson (2015) for an excellent review of the more recent studies in this literature that also include the bounds approach and structural approach.

concentration. Some of the more recent studies employ a two-stage approach, where the first stage is used to derive a correction term for the endogeneity of market structure that is incorporated in the price-concentration regressions in the second stage (e.g., Manuszak and Moul, 2008; Zhu, Singh, and Manuszak, 2009).

The contribution of this study is three-fold. First, it adopts a combination of fixed-effects and instrumental-variables (IV) techniques that account for store-level unobserved heterogeneity to investigate the relationship between prices and market structure in the US food retailing sector. This unobserved heterogeneity reflects time-invariant unobserved store characteristics such as quality of service and management, location, amenities, transition zone (*i.e.*, displays and other décor placed in front of stores), etc., which are important considerations when setting retail prices (Biscourp, Boutin, and Verge, 2013). Second, our empirical analysis is performed based on unique product, store, and retail real estate data that allow for identification strategy that corrects for the endogeneity of retail concentration. Retail food prices for an exhaustive list of food products are obtained from a novel Information Resources Inc. Infoscan (IRI) data that contain detailed store and product barcode-level price information from across the US. Further, HHI estimates are computed using Nielsen TDLinx store characteristics data that cover an exhaustive list of food retailers from the retail markets under study. Third, and most importantly, our approach accounts for the endogeneity of retail concentration by exploiting unique retail real estate data on newly constructed retail space provided by Marcus & Millichap.[2] Newly

[2] Marcus & Millichap is a commercial real estate brokerage firm and one of the largest US companies specializing in real estate investment services. It also conducts research on commercial real estate markets. See the "Data Description" section for additional details.

constructed retail space, through its effect on retail rent, is an important determinant of retail concentration.[3] Retail rent accounts for 8.3% of the retail total operating expenses in the US, third only to payroll and employer cost for fringe benefits (Annual Retail Trade Survey, US Census Bureau, 2012). Hence, retail rent is an important consideration in deciding whether or not to operate in a particular market (Newman and Cullen, 2002). To the extent that short-run retail food supply function is not related to retail rent, the latter constitutes a fixed cost. In practice, retail rent is a fixed payment that remains constant in the duration of rental contracts (Benjamin and Chinloy, 2004).[4]

One might expect that retail food prices might feed back into new retail construction through a change in retail rent and demand for retail space. Nonetheless, supply adjustments tend to be sluggish, given informational inefficiencies, investment-decision lags, and inherently long construction lags. This is further influenced by government policies regarding building permits, growth regulations, local zoning and other institutional controls that may hamper retail real estate development. As a result, it usually takes about two years to complete construction from the time of inception and one to two years for lease-up (Sivitanidou and Sivitanides, 2000). These

---

[3] Retail rents are determined in retail property markets by the demand and supply of retail space. The demand for space is derived from the demand for retail goods and services, while new construction and depreciation essentially determine how retail real estate inventory evolves over (Gyourko and Voith, 1993).

[4] To be more general, we allow for the possibility that retail rent comprises a fixed base payment and a variable component that is a percentage of retail sales volume, as discussed below.

institutional features of the retail real estate market are essential for new construction to be a valid instrument for market concentration.[5]

2. A Panel Data Model for Food Price and Market Structure

2 (*i*). *A Price Regression*

We empirically investigate the relationship between retail food prices and market structure by adopting a combination of IV and fixed-effects econometric techniques. Retail prices are expressed as a function of food supply and demand factors, as well as market structure that is represented by retail concentration. Discussions of the data and data sources are detailed in the next sections. The model follows:

$$p_{i,j}^{y,m} = \lambda^y X_{c(j)}^y + \delta^y HHI_{c(j)}^y + \alpha_j + \psi_i^{y,m} + \gamma_{\varphi(j)}^y + \varepsilon_{i,j}^{y,m} \tag{1}$$

where

$p_{i,j}^{y,m}$ = price (in logarithm) of product *i* in store *j* in month *m* of year *y*;

$X_{c(j)}^y$ = vector of market characteristics (other than concentration such as population and income) relating to market *c*, where store *j* is located in year *y*;

$HHI_{c(j)}^y$ = Herfindahl-Hirschman Index of market concentration;

α_j = store fixed effects;

$\psi_i^{y,m}$ = dummy variables representing interactions between product, year, and month;

[5] In contrast, we expect retail rent to be prone to endogeneity since the rent may contain a variable component that is determined by retail prices and sales volumes, as shown below. This would invalidate the use of retail rent as an instrument for market concentration.

$\gamma^{y}_{\varphi(j)}$ = store-type effects with $\varphi(j)$ denoting the retail format to which store j belongs (*i.e.*, convenience store, mass merchandiser, discount store, etc.);

$\varepsilon^{y,m}_{i,j}$ = standard *i.i.d.* disturbance for product i in store j in month m of year y.

2 (*ii*). *Endogeneity of Retail Concentration and Identification Strategy*

Endogeneity of market concentration may arise through a correlation between the concentration $\left(HHI^{y}_{c(j)}\right)$ and either or both of the disturbances in the price-concentration regression, *i.e.* store-specific unobservable variables (α_j) and the *i.i.d.* disturbance ($\varepsilon^{y,m}_{i,j}$), as can be seen from equation (1). Our use of the fixed-effects estimation procedure yields a consistent estimator even when $Cov(\alpha_j, HHI^{y}_{c(j)}) \neq 0$, assuming that the condition $Cov(\varepsilon^{y,m}_{i,j}, HHI^{y}_{c(j)}) = 0$ is satisfied. In reality, however, the assumption of $Cov(\varepsilon^{y,m}_{i,j}, HHI^{y}_{c(j)}) = 0$ may not hold owing to a variety of reasons, as laid out in detail by Evans, Froeb, and Werden (1993). First and foremost, industry performance and structure may be interrelated through a feedback effect. More specifically, market structure may affect firm performance through changes in firm efficiency and/or market power. On the other hand, industry structures evolve over time with firm conduct (e.g., various promotional campaigns, and decisions regarding entry, exit, and investment in new capacity, etc.) playing a major role in this dynamics. Firm conduct, in turn, responds to performance with the latter usually represented by firm profitability and price levels in empirical applications (Berry, 1992). Another reason why concentration is endogenous is that *HHI* and specifically its revenue-based estimates are a function of firm outputs and/or revenues, while outputs and revenues are determined with prices in a simultaneous fashion.[6]

Endogeneity of market concentration results in biased OLS estimates. The magnitude of the bias may be larger than the true coefficient as illustrated by Froeb and Werden (1991) based on an analytical framework with the underlying assumptions of a linear demand, constant marginal cost, and Cournot competition. However, it is also possible for the bias to be small in certain applications when various effects cancel each other out. As regards the sign of the bias, there seems to be a general consensus that the OLS estimator underestimates the association between price and concentration (e.g., Evans, Froeb, and Werden, 1993; Manuszak and Moul, 2008). In general, though the asymptotic bias may also be positive depending on the modelling assumptions and the specificities of markets under study (Froeb and Werden, 1991).

This study combines fixed-effects and IV procedures based on disaggregate store and product-level panel data on retail food prices to address the endogeneity of retail concentration. A major consideration concerns the choice of instrumental variables. Ideally, these instruments are excluded from the reduced-form price equation, affect food prices indirectly through their effects on market concentration, and are uncorrelated with the error term in this outcome equation. Retail fixed costs present one such example. Nevertheless, limited retail cost data has led researchers, as previously presented, to rely on other instruments such as lagged values of *HHI* estimates or various transformations thereof (e.g., Evans, Froeb, and Werden, 1993; Dafny, Duggan, and Ramanarayanan, 2009).

---

[6] Biscourp, Boutin, and Verge (2013) use an alternative measure of market concentration, *i.e.* capacity-based *HHI* estimates, which sidesteps this particular source of endogeneity.

Our approach corrects for the endogeneity of retail concentration by exploiting unique retail real estate data on newly constructed retail space offered by Marcus & Millichap. To put new retail space in a more meaningful context, we express it as a share of the total retail space that is available in a given market and time period. We formulate our identification strategy by making explicit the relationships that result in the endogeneity of retail concentration. These relationships highlight the role of new construction space as a source of exogenous identifying variation in the retail rents and therefore non-price determinants of HHI. In what follows, we offer a brief discussion concerning the relevance and the validity of our instrument. Consider the following simultaneous set of price and concentration equations:

(2) $$p_{i,j}^{y,m} = f\left(X_{c(j)}^{y}, HHI_{c(j)}^{y}, Y_{i,j}^{y,m}, \varepsilon_{i,j}^{y,m}\right)$$

(3) $$HHI_{c(j)}^{y} = g\left(p_{i,j}^{y,m}, F_{j}^{y}, \xi_{c(j)}^{y}\right)$$

In this system, equation (2) is the outcome equation with $p_{i,j}^{y,m}, X_{c(j)}^{y}, HHI_{c(j)}^{y}$, and $\varepsilon_{i,j}^{y,m}$ as previously defined, and $Y_{i,j}^{y,m}$ denotes other potentially important variables in equation (1) such as the retail format, store fixed-effects, etc. Equation (3) represents an equation for retail concentration with F_{j}^{y} denoting retail fixed costs (*i.e.*, retail real estate, equipment, rent, etc.), and $\xi_{c(j)}^{y}$ accounts for the remaining factors that have direct bearing on market concentration.

As discussed above, retail rent accounts for a considerable portion of total operating costs of the US food retailers (Annual Retail Trade Survey, US Census Bureau, 2012). As such, retail rent is a significant consideration in deciding whether or not to operate in a particular market (Newman and Cullen, 2002). In practice, retail rents are fixed payments that remain constant in the duration of rental contracts. Nevertheless, to be more general, we allow for the possibility that rents comprise a *fixed base payment* and a variable component that is a percentage of retail sales volume (Benjamin and Chinloy, 2004). Therefore, we decompose retail rents into the following components: (*i*) a *fixed component* that is exogenous to retail food price determination in the short-run, and (*ii*) a variable component that is responsive to food prices and sales volumes. In turn, these *fixed base rates* are determined as an equilibrium outcome in the retail rental market. Consider the following reduced-form equation for retail rents:

(4) $$R_{c(j)}^{y} = \Gamma\left(E_{c(j)}^{y}, D_{c(j)}^{y}, \upsilon_{c(j)}^{y}\right)$$

where $R_{c(j)}^{y}$ represents retail *base* rent, $E_{c(j)}^{y}$ represents observed retail-space supply-side factors such as retail real estate stock (S_{c}^{y}) and *newly constructed retail space* (C_{c}^{y}), $D_{c(j)}^{y}$ is observed retail-space demand-related factors, and $\upsilon_{c(j)}^{y}$ denotes unobserved rent determinants.

New construction is very important when analyzing real estate markets, given the long life of real estate assets. To see how new construction of retail space affects the inventory of retail real estate, it is useful to consider the *stock-flow identity*, which describes the dynamic evolution of total real estate stock (S_{c}^{y}) in market *c* in year *y* (Gyourko and Voith, 1993):

(5) $$S_{c}^{y} = S_{c}^{y-1}(1-d) + C_{c}^{y},$$

(6) $$S_{c}^{y} = S_{c}^{y-1}(1-d) + \omega PRM_{c}^{y-n},$$

where

S_{c}^{y} = retail real estate stock in market *c* in year *y*;

d = depreciation rate;

C_{c}^{y} = retail space completed in market *c* in year *y*;

PRM_{c}^{y-n} = retail space permitted in market *c* in year *y-n*;

ω = percent of permits completed;

n = time between permit issuance and project completion.

Equations (5) and (6) reveal that the amount of *new construction* and depreciation essentially determine how real estate markets and the retail real estate inventory evolve over time. Therefore, *new construction* is a key supply concept in real estate market analyses, and is a key determinant of fixed base component of retail rent. The term new construction refers to completions, or equivalently the total square footage in all new buildings that have passed the final inspection under the building permit during the period under consideration and are ready-to-use (Mourouzi-Sivitanidou, 2002). As such, completions reflect an important supply-side determinant of retail rents. Hence, we expect new completions to satisfy the relevance requirement for instruments.

Our major argument for instrument validity, which relies on the institutional characteristics of the retail real estate market, is presented in Section I. Specifically, retail real estate inventory adjustments tend to be sluggish, driven by informational inefficiencies, investment-decision lags, and inherently long construction lags. This process can be further prolonged by government policies regarding building permits, growth regulations, local zoning and other institutional controls that may hamper retail real estate development. As a result, it usually takes about two years to complete construction from the time of inception and one to two years for lease-up (Sivitanidou and Sivitanides, 2000).

3. Data Description

We compile data on retail food prices, store and market characteristics, as well as newly constructed retail space from a variety of sources: (*i*) IRI scanner data that contain store and product barcode-level information on retail dollar sales and quantity for food and beverage products marketed throughout the US, (*ii*) Nielsen TDLinx provides annual store characteristics data for an exhaustive list of food retailers, (*iii*) US Department of Commerce, Bureau of Economic Analysis provides annual market characteristics data regarding population density and per capita income, (*iv*) US Bureau of Labor Statistics retail wage data from the respective US metropolitan statistical areas, (*v*) Marcus & Millichap real estate data on newly constructed retail space from 16 US metropolitan statistical areas. Below, we provide a detailed description of these datasets:

IRI Retail Food Price Data: We use novel and uniquely disaggregate IRI data that provide food price information from across the US. The analysis is performed over the period 2008-2012. IRI collects information on all items scanned at cash registers from more than 11,000 local grocery stores in the US on weekly basis. The data are then scaled up to reflect all sales from stores with annual revenues of $2 million and higher. The IRI dataset contains information on dollar sales and physical volumes for a large group of food products from five departments (dairy, deli, bakery, frozen food, fresh produce) at brand, UPC or item level (a total of 36 billion observations). Most stores in this dataset belong to a retail chain. The remaining non-chain/independent stores are chosen by the IRI based on the random stratified sampling method. Specifically, a fraction of stores is dropped each month and replaced by others using a rotating panel design (see Ward et al. (2002) for further details).

We base our empirical analysis on more than 1,000 food products that are sufficiently homogenous across stores and dates (e.g. water, milk, sugar, salt, etc.) to sidestep potential issues related to the effects of product differentiation on food prices. To ensure empirical tractability, we aggregate these food items into closely related food groups. Our final data set contains monthly prices for 129 food groups marketed in 16 US retail markets over the period 2008-

2012.[7] This results in unbalanced panel data that cover a wide range of retail formats (*i.e.*, 14.6 % convenience store, 20.6 % variety stores better known as dollar stores, 32.1 % drug stores, 26.1 % grocery stores, and 7.6 % mass merchandisers) and contain **7,755,166 observations**.

Nielsen TDLinx Store Characteristics Data: We confine our empirical analysis to 16 US metropolitan statistical areas for which we have data on retail new space. Despite retail grocery competition being limited to local geographic areas, market delineation remains a challenging task (Biscourp, Boutin, and Verge, 2013). In practice, markets are defined based on the competing stores located within a certain radius (Barros, Brito, de Lucena, 2006). In this study, we assume retail markets to be represented by metropolitan areas and/or cities located in different geographic areas of the US. We utilize Nielsen TDLinx data on store characteristics to calculate revenue-based HHI estimates of concentration for the respective markets (Table 1). Markets vary widely in terms of market concentration, however most markets appear to be low to moderately concentrated. The underlying reason may be the way the markets are defined. Further, spatial variation seems to outweigh the temporal variation with the coefficient of variation ranging from as low as 2.25 % for Minneapolis, MN to 10.35 % for San Diego, CA.

US Department of Commerce Market Characteristics Data: Two important descriptors are used to characterize the retail markets under scrutiny. Specifically, we compile population density and per capita income data for the period 2008-2012 from the US Department of Commerce, Bureau of Economic Analysis. Our goal with the inclusion of the population and income variables is to account for the effects of demand-related factors on retail food prices. Markets vary considerably

[7] A complete list of these product groups and composition thereof are available from the authors upon request.

in terms of population density. For example, while the average number of population per square mile is 2,543 in Jacksonville, FL the estimate for New York, NY is 5,684. Further, the coefficient of variation for population change over time varies from as low as 0.49 % for Chicago, IL to a high of 2.94 % for Houston, TX. Markets also manifest considerable heterogeneity in terms of consumer income. Specifically, average per capita income varies from as low as 36,800 in San Antonio, TX to as high as 56,900 in New York, NY. A general tendency that stands out is that per capita income has declined in a majority of markets following the great recession in 2008. Nevertheless, this effect is predominantly felt in 2009 and starting the following year income reverted back to a rising trend in most markets, eventually surpassing the pre-recession levels.[8]

Marcus & Millichap Real Estate Data: Marcus & Millichap is a commercial real estate brokerage firm and one of the largest US companies specializing in real estate investment services. Marcus & Millichap also conducts research on commercial real estate markets. We exploit unique Marcus & Millichap data on newly constructed retail space (square footage) to instrument for market concentration (Table 2, left panel). Retail new space completion manifests considerable spatial and temporal variability during our sample period with the coefficient of variation ranging from 41.36 % for Milwaukee, WI to 123.75 % for Phoenix, AZ. To put in a meaningful context, we construct the ratio of new retail space to the total retail space, which is

[8] Further details are not presented to preserve space, but are available from the authors upon request.

used to instrument for concentration (Table 2, right panel).[9] Sampling variability of this ratio is also noteworthy with the mean share of the new retail space varying from 2.45 % for Louisville, KY to 39.41 % for Dallas, TX. Further, the respective coefficient of variation extends from 33.42 % for San Diego, CA to 110.93 % for Jacksonville, FL. This variability is vital from the identification perspective.

4. Empirical Relationship between Retail Price and Retail Concentration

Our empirical results are provided in Tables 3 and 4. For benchmark comparison, we first report the estimates from the cross-section/OLS regressions, which disregard both unobserved store heterogeneity and the endogeneity of concentration (Table 3, left panel). These results confirm that mass merchandisers offer the lowest prices on a wide spectrum of food products among the retail formats in our sample. Moreover, the price gap between mass merchandisers and the remaining formats has been on a rise over the 5-year span. These findings are also supported by the IV-OLS results (Table 3, right panel). We also find that food prices generally rise with population density and per capita income. Finally, the OLS estimates indicate that prices go hand-in-hand with market concentration with the *HHI* estimated coefficient varying from 0.051 to 0.067 (*i.e.*, 10 % rise in retail concentration is associated with 0.51-0.67 % price increase). By contrast, the IV-OLS estimates provide a mixed evidence with the majority of coefficients exceeding their OLS counterparts in magnitude (0.091-0.440), while the *HHI* estimated coefficient for year 2010 is found to be negative (-0.107). This may well be a result of omitted

[9] We calculate total retail space from the respective markets based on the Nielsen TDLinx data on store characteristics.

variables such as unobserved store heterogeneity confounding our estimation of the relationship between price and concentration.

As has been found in previous studies, unobserved store heterogeneity plays a significant role in retail price formation (e.g., Evans, Froeb, and Werden, 1993; Biscourp, Boutin, and Verge, 2013; Lin and Wang, 2015). Therefore, the fixed-effects regression is our preferred specifications. Based on a Hausman test, we may reject the null hypothesis that all of the coefficients from the fixed-effects and random-effects panel data models are the same (the value of the associated test statistic is $\chi^2(3245)=2{,}342$, or equivalently, p-value=0.00).[10] This implies that $Cov(\alpha_j, HHI_{c(j)}^{y}) \neq 0$ and that the OLS and random-effects models yield inconsistent estimates. Estimation results from a fixed-effects regression are provided in Table 4 (left panel). As before, we find that food prices at convenience, dollar (also known as a variety store), drug, and grocery stores exceeds those at mass merchandisers, however these differences appear to be moderate relative to the findings from the OLS and IV-OLS specifications. The estimated coefficients associated with population density (0.078-0.100) and per capita income (0.009-0.036) suggest direct relationship with price. Finally, a majority of the *HHI* estimated coefficients are statistically significant (2010 onward) and fall in the range of 0.013 to 0.037. Interestingly, the *HHI* coefficients obtained via the OLS regression exceed those from the fixed-effects model in magnitude. This may be a result of, for example, unobserved variation in factor

---

[10] We acknowledge that Hausman test may be neither necessary, nor sufficient statistic for deciding between fixed and random-effects estimators, as illustrated by Clark and Linzer (2015) based on a series of simulation experiments.

prices exceeding the unobserved variation in demand in the OLS specification, as illustrated by Froeb and Werden (1991), with the net bias being positive.

As has been discussed above, there are good reasons to believe that concentration is endogenous even after unobserved store heterogeneity has been accounted for. Hence, we further estimate an IV fixed-effects model using the instrumental variables that are discussed above. Indeed, based on a Durbin-Wu-Hausman test procedure, we are able to empirically confirm that concentration is endogenous to price formation.[11] Moreover, our instruments satisfy the relevance requirement as evidenced by the computed first-stage F-statistic values, which exceed 200 (Stock and Yogo, 2005). This finding is robust to the choice of the test, which include Andersen-Rubin Wald test ($\chi^2(10) = 60.59$), Stock-Wright LM test ($\chi^2(10) = 66.67$), Cragg-Donald Wald test ($\chi^2(10) = 54,452$), and Kleibergen-Paap Wald rk test that yields heteroskedasticity-robust results ($\chi^2(10) = 49.21$).[12]

---

[11] The Durbin-Wu-Hausman test is used to test whether the estimates from the IV fixed-effects model are statistically significantly different from the fixed-effects estimates that ignores the endogeneity of concentration. To perform the test, we first obtain the residuals from regressing concentration on the control variables and the excluded instruments. Retail price is then regressed on the control variables, concentration, and the residuals from the previous regression. A finding of statistically insignificant coefficient for the residual implies that we may not be able to reject the null that concentration is exogenous.

[12] We do not perform a test for the overidentifying restrictions since our equation is exactly identified. It is worth noting that overidentifying restrictions provide little information on the validity of instruments (Parente and Silva, 2012). Instead, the validity of instruments should be

Estimation results from the IV fixed-effects model with the underlying store-level cluster-robust standard errors are presented in Table 4 (right panel).[13] It can be observed that a great majority of the coefficients of the retail formats trace the fixed-effects estimates closely (Table 4, left panel). *HHI* estimates present an important exception in that they exceed in magnitude the respective estimates from the fixed-effects model and are all statistically significant for the entire sample period. Specifically, the estimates from the IV model range from 0.236 to 0.351 (e.g., the fixed-effects coefficient for year 2012 is about one/tenth of the respective IV fixed-effects coefficient). This implies that a 10 % increase in retail concentration is associated with 2.36-3.51 % rise in retail food prices. An interesting finding that emerges is that the *HHI* coefficients from the IV fixed-effects regression manifest a decline over the period 2008-2010 from 0.351 to 0.236, which is followed by a steady increase through 2012. This might be indicative of food demand becoming less inelastic in the aftermath of the 2008 recession thus intensifying retail competition in the short-run.

As a robustness check, we also estimate the Fixed-effects and IV-Fixed effects specifications that include three high-dimensional fixed effects. Specifically, following Biscourp, Boutin, and Verge (2013), we include interaction effects between products and retail formats in

---

based on the underlying model such as the one presented earlier. Further, overidentifying restriction tend to be rejected in presence of parameter heterogeneity.

[13] As illustrated by Angrist and Pischke (2009), the finite sample bias of the formula for homoscedastic errors is smaller than that of the robust sandwich estimator. Nonetheless, our very large sample size suggests using the robust estimator.

addition to the store fixed effects and the interaction of year and month variables.[14] This allows for the possibility of food products being priced differently across the retail formats (Biscourp, Boutin, and Verge, 2013). The results from this specification are reported in Table A.1. Overall, the parameter estimates are similar to those in Table 4 both in terms of the sign and magnitude.

4. Summary and Conclusions

This study exploits unique product, store, and retail real estate data to calculate consistent estimates of the effects of retail market structure on food prices in the US. Our uniquely disaggregated data allow for identification strategy that accounts for the type of endogeneity that plagues many previous studies on price-concentration relationship. The study has several distinguishing characteristics. First, it employs a combination of fixed-effects and IV techniques that account for store-level unobserved heterogeneity to empirically examine the relationship between prices and market structure in the US food retailing sector. Second, our empirical analysis is conducted using extremely detailed product, store, and retail real estate data. Third, and most importantly, our approach accounts for the endogeneity of retail concentration by utilizing unique retail real estate data on newly constructed retail space obtained from Marcus & Millichap, which conducts research on commercial real estate markets.

Our empirical results indicate that unobserved store heterogeneity plays a major role in price determination, therefore the OLS estimates cannot be relied on in the food price - retail concentration studies. Findings from the fixed-effects panel data model indicate that retail food

[14] We estimate this multiple high-dimensional fixed-effects model using the *REGHDFE* Stata program developed by Correia (2014), which is based on memory-saving techniques and requires considerably less run-time vis-à-vis the standard panel-data estimation programs.

prices rise with retail concentration. Further, we find strong empirical evidence suggesting that retail concentration is endogenous to retail price determination on the account of the correlation between concentration and unobserved market and product-level idiosyncrasies. Results from our fixed-effects IV regressions that account for this type of endogeneity show that fixed-effects estimator of market concentration effects has a significant downward bias. Importantly, we find that our instruments satisfy the relevance tests at the standard significance-levels.

In general, are results are in agreement with findings from other similar studies (e.g., Evans, Froeb, and Werden, 1993; Cotterill, 1999; Aalto-Setala, 2002; Manuszak and Moul, 2008) both in terms of the direction and the magnitude of the bias in the concentration coefficient that results from the use of the standard econometric techniques (e.g., OLS and fixed-effects panel data model that ignore the endogeneity of concentration).

References

Aalto-Setälä, V., 2002, The Effect of Concentration and Market Power on Food Prices: Evidence from Finland,' *Journal of Retailing*, 78, pp. 207-216.

Angrist, J. and Pischke, J. S., 2009, *Mostly Harmless Econometrics* (Princeton. NJ: Princeton University Press).

Balagtas, J. V., 2010, 'Changing Structure and Competition in Food and Agricultural Markets,' *Choices*, 25.

Barros, P. P.; Brito, D. and de Lucena, D., 2006, 'Mergers in the Food Retailing Sector: An Empirical Investigation,' *European Economic Review* 50, pp. 447-468.

Benjamin, J. and Chinloy, P., 2004, 'The Structure of a Retail Lease,' *Journal of Real Estate Research*, 26(2), pp. 223-236.

Berry, S. T., 1992, ‘Estimation of a Model of Entry in the Airline Industry,’ *Econometrica: Journal of the Econometric Society*, pp. 889-917.

Biscourp, P.; X. Boutin, X. and Vergé, T., 2013, ‘The Effects of Retail Regulations on Prices: Evidence from the Loi Galland,’ *The Economic Journal*, 123, pp. 1279-1312.

Clark, T. S. and Linzer, D. A., 2015, ‘Should I Use Fixed or Random Effects?’ *Political Science Research and Methods*, 3(02), pp. 399-408.

Correia, S., 2014, ‘REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing any Number of Digh-dimensional Fixed Effects. *Statistical Software Components.*

Cotterill, R. W., 1999, ‘Market Power and the Demsetz Quality Critique: An Evaluation for Food Retailing,’ *Agribusiness*, 15, pp. 101-118.

Dafny, L.; Duggan, M. and Ramanarayanan, S., 2009, ‘Paying a Premium On Your Premium? Consolidation in the US Health Insurance Industry,’ *National Bureau of Economic Research* (No. w15434).

Demsetz, H., 1973, ‘Industry Structure, Market Rivalry, and Public Policy,’ *Journal of Law and Economics*, pp. 1-9.

Ellickson, P., 2015, *Market Structure and Performance*, in James D. Wright (Ed.), International Encyclopedia of the Social and Behavioral Sciences, 2nd Edition, Vol. 14, pp. 9211-9216 (Oxford, Elsevier).

Evans, W. N.; Froeb, L. M. and Werden, G. J., 1993, ‘Endogeneity in the Concentration--Price Relationship: Causes, Consequences, and Cures,’ *The Journal of Industrial Economics*, pp. 431-438.

Froeb, L. M. and Werden, G. J., 1991, 'Endogeneity in the Concentration-Price Relationship: Causes and Consequences,' US Department of Justice, Antitrust Division, Economic Analysis Group Discussion Paper, July 1.

Gyourko, J. and Voith, R., 1993, 'Leasing as a Lottery: Implications for Rational Building Surges and Increasing Vacancies,' *Real Estate Economics*, 21, pp. 83-106.

Hovhannisyan, V.; Stiegert, K.W. and Bozic, M., 2014, 'On the Endogeneity of Retail Markups in an Equilibrium Analysis: A Control-Function Approach,' *Journal of Agricultural and Resource Economics* 39(2), pp. 188-200.

Lin, H. and Wang, I. Y., 2015, "Competition and Price Discrimination: Evidence from the Parking Garage Industry,' *The Journal of Industrial Economics*, 63, pp. 522-542.

Manuszak, M. D. and Moul, C. C., 2008, 'Prices and Endogenous Market Structure in Office Supply Superstores,' *The Journal of Industrial Economics*, 56, pp. 94-112.

Marcus & Millichap Real Estate Reports, 2008-2012. Available at: http://www.marcusmillichap.com/research/researchreports

Mourouzi-Sivitanidou, R., 2002, 'Office Rent Processes: The Case of US Metropolitan Markets,' *Real Estate Economics*, 30, pp. 317-344.

Newman, A. J. and Peter Cullen, 2002, *Retailing: Environment and Operations*, (Cengage Learning EMEA).

Newmark, C. M., 2004, *Price-concentration Studies: There You Go Again*, Available at SSRN 503522.

Parente, P. M. and Silva, J. S., 2012, 'A Cautionary Note on Tests of Overidentifying Restrictions,' *Economics Letters*, 115, pp. 314-317.

Schmalensee, R. 1989, *Inter-industry Studies of Structure and Performance*, Handbook of Industrial Organization, Vol. II, Chapter 16 (North-Holland, Amsterdam).

Sivitanidou, R. and Sivitanides, P., 2000, 'Does the Theory of Irreversible Investments Help Explain Movements in Office–Commercial Construction?' *Real Estate Economics,* 28, pp. 623-661.

Stock, J. H. and Yogo, M., 2005, *Testing for Weak Instruments in Linear IV Regression*, In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, ed. Andrews, D. W. K. and Stock, J.H., pp. 80-108 (Cambridge and New York: Cambridge University Press).

U.S. Census Bureau, 2012, Annual Retail Trade Survey, Available at: https://www.census.gov/retail/arts/how_surveys_are_collected.html

U.S. Department of Commerce, Bureau of Economic Analysis. 2015. Population and Per Capita Income by Metropolitan Areas. Available at: http://bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=5#reqid=70&step=1&isuri=1

U.S. Department of Justice, 2011, *Agriculture and Antitrust Enforcement Issues in Our 21st Century*.

Ward, M. B.; Shimshack, J. P.; Perloff, J. M. and Harris, J. M., 2002, 'Effects of the Private-Label Invasion in Food industries,' *American Journal of Agricultural Economics*, 84, pp. 961-973.

Weiss, L. W., 1989, *Concentration and Price* (MIT Press, Cambridge, Mass.).

Zhu, T.; Singh, V. and Manuszak, M. D., 2009, 'Market Structure and Competition in the Retail Discount Industry,' *Journal of Marketing Research*, 46, pp. 453-466.

Table 1. Descriptive Statistics for the Market-Level HHI Estimates, 2008-2012

Market	State	Mean	SD	Min	Max	CV (%)
Charlotte	NC	624	36	574	670	5.77
Chicago	IL	532	19	509	562	3.57
Cincinnati	OH	1053	30	1003	1078	2.85
Columbus	OH	631	20	606	655	3.17
Dallas	TX	525	38	486	574	7.24
Houston	TX	518	32	469	559	6.18
Indianapolis	IN	698	27	661	736	3.87
Jacksonville	FL	759	56	665	801	7.38
Louisville	KY	1194	74	1079	1267	6.20
Milwaukee	WI	721	32	666	744	4.44
Minneapolis	MN	579	13	557	590	2.25
New York	NY	567	31	523	604	5.47
Phoenix	AZ	685	21	649	701	3.07
Sacramento	CA	503	19	478	522	3.78
San Antonio	TX	1242	58	1170	1328	4.67
San Diego	CA	628	65	567	705	10.35

Source: Author Calculations based on Nielsen TDLinx Data, 2008-2012.

Table 2. Descriptive Statistics for Retail Newly Completed Area by Market, 2008-2012

Market	State	New space ($footage^2$)			New space/ Total space (%)		
		Mean	SD	CV (%)	Mean	SD	CV (%)
Charlotte	NC	827	592	71.58	12.26	7.90	64.46
Chicago	IL	2102	1934	92.01	15.98	13.27	83.05
Cincinnati	OH	793	823	103.78	10.88	9.90	91.03
Columbus	OH	448	257	57.37	6.51	3.40	52.29
Dallas	TX	3300	2923	88.58	39.41	31.83	80.78
Houston	TX	2103	1926	91.58	9.92	8.33	83.95
Indianapolis	IN	630	661	104.92	7.94	7.74	97.47
Jacksonville	FL	615	758	123.25	8.08	8.97	110.93
Louisville	KY	158	98	62.03	2.45	1.35	54.92
Milwaukee	WI	660	273	41.36	10.82	3.78	34.94
Minneapolis	MN	596	321	53.86	7.23	3.74	51.80
New York	NY	1079	450	41.71	15.34	5.95	38.82
Phoenix	AZ	2349	2907	123.75	24.88	27.29	109.69
Sacramento	CA	715	467	65.31	12.92	7.74	59.90
San Antonio	TX	1375	1458	106.04	13.66	13.44	98.38
San Diego	CA	301	114	37.87	3.72	1.24	33.42

Source: Marcus & Millichap Real Estate Data, 2008-2012.

Note: Retail real estate data vary annually over the period 2008-2012.

Table 3. Parameter Estimates and Standard Errors from OLS and IV-OLS Regressions

	OLS					IV-OLS				
	2008	2009	2010	2011	2012	2008	2009	2010	2011	2012
Convenience	Ref.	0.153***	0.326***	0.475***	0.560***	Ref.	0.155***	0.366***	-0.451***	0.566***
		0.004	*0.004*	*0.004*	*0.004*		*0.009*	*0.010*	*0.022*	*0.009*
Dollar	Ref.	0.096***	0.210***	0.316***	0.279***	Ref.	0.096***	0.185***	0.806***	0.291***
		0.003	*0.003*	*0.003*	*0.003*		*0.009*	*0.009*	*0.013*	*0.009*
Drug	Ref.	0.140***	0.239***	0.439***	0.536***	Ref.	0.145***	0.273***	-0.158***	0.561***
		0.003	*0.003*	*0.003*	*0.003*		*0.008*	*0.008*	*0.015*	*0.008*
Grocery	Ref.	0.049***	0.152***	0.332***	0.358***	Ref.	0.037***	0.072***	0.179***	0.355***
		0.003	*0.003*	*0.003*	*0.003*		*0.009*	*0.009*	*0.032*	*0.008*
Mass Mer.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HHI	0.058***	0.051***	0.054***	0.067***	0.054***	0.160***	0.101***	-0.107***	0.440***	0.091***
	0.001	*0.001*	*0.001*	*0.001*	*0.001*	*0.005*	*0.003*	*0.008*	*0.090*	*0.004*
Population	0.061***	0.079***	0.083***	0.080***	0.059***	0.078***	0.093***	0.047***	0.881***	0.064***
	0.001	*0.001*	*0.001*	*0.001*	*0.001*	0.002	*0.002*	*0.003*	*0.017*	*0.002*
Income	0.036***	-0.004**	0.010***	0.007***	0.002*	0.112***	0.022***	-0.062***	0.192***	0.024***
	0.001	*0.002*	*0.001*	*0.001*	*0.001*	0.004	*0.003*	*0.004*	*0.040*	*0.003*
Ad. controls	(Year x month, product x year, product x channel)					(Year x month, product x year, product x channel)				
No. obs.	1,457,861	1,525,662	1,542,870	1,603,345	1,625,428	1,457,861	1,525,662	1,542,870	1,603,345	1,625,428

Notes: ***P<0.01, **P<0.5, *P<0.1. Store-level cluster-robust standard errors are italicized. The final sample includes a total of 7,755,166 observations.

Table 4. Parameter Estimates and Standard Errors from Fixed-Effects and Fixed-Effects IV Regressions

	Fixed-Effects					IV Fixed-Effects				
	2008	2009	2010	2011	2012	2008	2009	2010	2011	2012
Convenience	Ref.	0.098***	0.151***	0.196***	0.187***	Ref.	0.109***	0.178***	0.207***	0.198***
		0.009	*0.012*	*0.015*	*0.016*		*0.010*	*0.016*	*0.016*	*0.017*
Dollar	Ref.	0.138***	0.177***	0.217***	0.263***	Ref.	0.139***	0.172***	0.217***	0.264***
		0.008	*0.011*	*0.013*	*0.015*		*0.008*	*0.014*	*0.014*	*0.015*
Drug	Ref.	0.074***	0.164***	0.267***	0.299***	Ref.	0.075***	0.175***	0.277***	0.303***
		0.007	*0.011*	*0.013*	*0.015*		*0.008*	*0.014*	*0.014*	*0.015*
Grocery	Ref.	0.026***	0.090***	0.192***	0.194***	Ref.	0.013	0.057***	0.180***	0.191***
		0.008	*0.011*	*0.014*	*0.015*		*0.009*	*0.014*	*0.015*	*0.015*
Mass Mer.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HHI	0.013	0.013	0.030***	0.037***	0.033***	0.351***	0.303***	0.236***	0.287***	0.321***
	0.009	*0.009*	*0.009*	*0.009*	*0.009*	*0.044*	*0.042*	*0.040*	*0.038*	*0.044*
Population	0.078***	0.081***	0.100***	0.095***	0.078***	0.307***	0.296***	0.298***	0.287***	0.247***
	0.021	*0.021*	*0.021*	*0.021*	*0.020*	*0.050*	*0.050*	*0.050*	*0.047*	*0.043*
Income	*0.009*	*0.007*	*0.025****	*0.036****	*0.033****	0.234***	0.175***	0.154***	0.180***	0.201***
	0.008	*0.008*	*0.007*	*0.007*	*0.007*	*0.026*	*0.020*	*0.018*	*0.017*	*0.020*
Ad. controls	(Year x month, product x year, store fixed effects)					(Year x month, product x year, store fixed effects)				
No. obs.	1,457,861	1,525,662	1,542,870	1,603,345	1,625,428	1,457,861	1,525,662	1,542,870	1,603,345	1,625,428

Notes: ***P<0.01, **P<0.5, *P<0.1. Store-level cluster-robust standard errors are italicized. The final sample includes a total of 7,755,166 observations.

APPENDIX

Table A.1. Parameter Estimates and Standard Errors from Fixed-Effects and Fixed-Effects IV Regressions that Include Product x Retail Format Interactions

	Fixed-Effects					IV Fixed-Effects				
	2008	2009	2010	2011	2012	2008	2009	2010	2011	2012
Convenience	Ref.	0.098***	0.151***	0.196***	0.187***	Ref.	0.109***	0.178***	0.207***	0.198***
		0.009	*0.012*	*0.015*	*0.016*		*0.010*	*0.016*	*0.016*	*0.017*
Dollar	Ref.	0.138***	0.177***	0.217***	0.263***	Ref.	0.139***	0.172***	0.217***	0.264***
		0.008	*0.011*	*0.013*	*0.015*		*0.008*	*0.014*	*0.014*	*0.015*
Drug	Ref.	0.074***	0.164***	0.267***	0.299***	Ref.	0.075***	0.175***	0.277***	0.303***
		0.007	*0.011*	*0.013*	*0.015*		*0.008*	*0.014*	*0.014*	*0.015*
Grocery	Ref.	0.026***	0.090***	0.192***	0.194***	Ref.	0.013***	0.057***	0.180***	0.191***
		0.008	*0.011*	*0.014*	*0.015*		*0.009*	*0.014*	*0.015*	*0.015*
Mass Mer.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HHI	0.013	0.013	0.030***	0.037***	0.033***	0.351***	0.303***	0.236***	0.287***	0.321***
	0.009	*0.009*	*0.009*	*0.009*	*0.009*	*0.044*	*0.042*	*0.040*	*0.038*	*0.044*
Population	0.078***	0.081***	0.100***	0.095***	0.078***	0.307***	0.296***	0.298***	0.287***	0.247***
	0.021	*0.021*	*0.021*	*0.021*	*0.020*	*0.050*	*0.050*	*0.050*	*0.047*	*0.043*
Income	*0.009*	*0.007*	*0.025****	*0.036****	*0.033****	0.234***	0.175***	0.154***	0.180***	0.201***
	0.008	*0.008*	*0.007*	*0.007*	*0.007*	*0.026*	*0.020*	*0.018*	*0.017*	*0.020*
Ad. controls	(Year x month, product x year, product x retail format, store fixed effects)					(Year x month, product x year, product x retail format, store fixed effects)				
No. obs.	1,457,861	1,525,662	1,542,870	1,603,345	1,625,428	1,457,861	1,525,662	1,542,870	1,603,345	1,625,428

Notes: ***P<0.01, **P<0.5, *P<0.1. Store-level cluster-robust standard errors are italicized. The final sample includes a total of 7,755,166 observations. Additional controls include dummy variables representing interactions between products and retail formats. This allows for the possibility of different products being priced differently across the retail formats.