PRICES AND ENDOGENOUS MARKET STRUCTURE IN OFFICE SUPPLY SUPERSTORES*

MARK D. MANUSZAK†
CHARLES C. MOUL‡

We consider the relationship between prices and market structure for office supply superstores in the U.S. which was central to the Federal Trade Commission’s opposition to the merger of Staples and Office Depot. Due to potential biases in a standard regression, we employ a two-stage approach in which a model of endogenous market structure provides correction terms for a second stage price regression. Using a cross-section of data on market structures and Staples’ prices, we find that excluding the correction term substantially distorts the importance of competitors as the two-stage model yields stronger negative relationships between prices and market structure variables.

I. INTRODUCTION

On September 4, 1996, two of the three major office supply superstore chains in the United States, Staples and Office Depot, proposed to merge. If consummated, the merger would have left OfficeMax as the only other major firm in that specific category. Due to concerns about potential anticompetitive effects, the Federal Trade Commission (FTC) challenged the merger. In mid-1997, the U.S. District Court issued a preliminary injunction in favor of the government’s position1 after which Staples and Office Depot abandoned the merger.

During the litigation, both sides introduced a variety of evidence including a substantial amount of econometric analysis about the potential effects of the merger (Ashenfelter, Ashmore, Baker, Gleason and Hosken...
In fact, analysts at the FTC conjecture that ‘(t)his case likely still holds the record for the most extensive commitment of resources to econometric analysis in government antitrust litigation, and by a wide margin’ (ABGH, pg. 2). Most of the FTC’s econometric evidence related to one apparent empirical regularity: prices at Staples tended to be lower in markets where an Office Depot was also present. This evidence led the FTC to conclude that prices would be higher, and consumers correspondingly harmed, if Office Depot were removed as a competitor for Staples.

The FTC documented the relationship between prices and market structure in a variety of ways. Perhaps most provocative was the simple observation that Office Depot’s price of file folders was $4.17 in Leesburg, Florida, where it was the only office superstore, whereas fifty miles away in Orlando, Florida, where Staples was also present, the price was $1.95. At the very least, small sample issues limit the validity of general conclusions from this comparison. More comprehensive evidence involved a variety of regressions in which Staples’ prices in a market were related to various variables, most importantly the presence of competing firms.

Beginning with the seminal work of Bain [1956], researchers in industrial organization have investigated the relationship between market structure and various market outcomes. One particular relationship of interest has been that of prices and market structure. Even absent the direct policy relevance that would apply in a merger case, interest in this relationship arises for a number of reasons (Weiss [1989]). First, prices are the crucial mechanism through which markets operate. Second, prices are not subject to measurement problems that plague many other variables such as economic profits. Finally, unlike the case with economic profits (Demsetz [1973]), multiple explanations do not clearly complicate the interpretation of consistent estimates of the relationship between prices and market structure.

However, empirical analysis of the relationship between prices and market structure involves substantial econometric problems. Market

2 We will refer to Ashenfelter, Ashmore, Baker, Gleason and Hosken [2004] as ABGH for purposes of brevity. In ABGH, participants in the case from the FTC provide an excellent summary of the econometric analysis and debates in the case. In addition, the paper provides a fascinating glimpse into the inner workings of merger litigation and the application of statistical analysis in legal settings. An abbreviated version of this paper, which omits discussion of some econometric details, can be found in Ashenfelter, Ashmore, Baker, Gleason and Hosken [2006]. See also Baker [1999].

3 Notably, the FTC analysts also bemoan the dearth of references to the econometric evidence in the court’s decision (ABGH, pg. 2). The reasons behind the court’s failure to cite the empirical evidence are unclear.

4 A preliminary divestiture agreement involving Office Depot stores was deemed to be an insufficient way to ensure competition in various markets.

5 A more fundamental criticism relates to the possibility that the actual pricing equilibrium might adjust following a merger in which case the proposed relationship between prices and market structure may change in response to a merger. Such a critique suggests the need for a structural approach to evaluate the potential impact of a merger. While a variety of studies have

© 2008 The Authors, Journal compilation © 2008 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.
structures are not randomly assigned, a necessary condition for correct inference about this relationship in a standard regression framework. Instead, firms base their entry decisions on demand and cost factors as well as the anticipated entry and pricing behavior of their competitors. Inclusion of additional covariates in a multivariate regression may alleviate this concern by accounting for some variables that impact both market structures and prices. As such, the regression approach is preferable to a simple comparison of prices across markets that vary in market structure regardless of how many observations are used in the latter approach. However, the presence of unobserved cost and demand factors that are correlated with both prices and market structure may prevent consistent estimation of the relationship between those two variables.6

The atheoretical nature of the typical price regression makes it difficult to infer the nature of the biases that may occur. However, a few intuitive possibilities might arise. Suppose, for the sake of argument, that prices are in fact negatively related to the number of competitors in a market. If costs in some markets are high in ways that the outside observer cannot easily measure, then prices may be unusually high where entry is unusually low. In such a case, a simple regression may overstate the negative relationship between the two variables. The analyst may infer that high prices are strongly associated with low numbers of firms, but this finding would be partially driven by the omitted unobserved costs that yield high prices and low entry. Conversely, unobserved positive demand shocks may support both higher prices and a higher number of firms. In this case, a negative relationship between the two variables may be understated since the data would suggest that increased competition yields relatively small decreases in prices. Again, this empirical result arises due to the common unobserved factor that induces both prices and entry to be higher than would otherwise occur.7

Analysts on both sides of the Staples case were well aware of this problem. To better support its claim that Office Depot constrained the pricing behavior of Staples, the FTC employed a variety of empirical approaches. Although the details of its analysis are unavailable,8 the FTC considered proposed such an approach, the corresponding modeling and data requirements are substantial. As a result, we do not consider this issue in the current paper and instead focus on the less ambitious goal of proposing a different empirical approach within the framework that the FTC considered.

6 Evans, Froeb and Werden [1993] note some potential biases in regressions of price on market structure variables including those that may arise due to measurement error in covariates. They propose panel data and instrumental variable techniques to address the problem.

7 Singh and Zhu [2006] look to formally derive these biases in their study of the pricing of airport rental cars.

8 As with many merger cases, one unfortunate feature of the Staples case is the inability of the parties to report many of the results from their analysis, notwithstanding the description in
instrumental variables techniques to estimate this relationship. However, variables that are correlated with market structure, but uncorrelated with unobserved factors that impact price are notoriously difficult to find. Such problems underlie the long-standing criticisms of the Structure-Conduct-Performance (SCP) approach to empirical analysis of industry performance (Schmalensee [1989]).

An alternative approach involved exploiting the panel structure of the data to estimate market fixed effects. Such fixed effects account for various market-specific factors that are time invariant and may impact both prices and market structure. Ideally, a fixed effects approach will yield consistent estimates of the relationship between prices and market structure through analysis of the relationship between changes in market structure and changes in prices over time. The FTC relied heavily on this approach to contest the merger (ABGH, pg. 6).

Despite its appeal, this approach still has problems. Precise estimation of a relationship based on variation over time requires actual variation in the explanatory variable, in this case market structure, over time. In other words, non-negligible entry and exit are required to accurately estimate the importance of market structure for prices. As the FTC’s analysts note, such variation was limited in this case (ABGH, pg. 8). In light of the data that we will subsequently employ, this issue is not the focus of our paper.

A more fundamental difficulty with this approach, which is related to the general focus of this paper, is the possibility that unobserved factors may still impact both prices and market structure. In the panel data context, the concern is the presence of unobservable changes in some demand or cost variable that impacts changes in both prices and market structure. Hopefully, such unobservable changes are sufficiently short-run that they do not substantially impact entry or exit decisions. The FTC relied on such a rationale in its use of a fixed effects model (ABGH, pg. 7).

The validity of this assumption is not easily testable. Taken to its logical extreme, the arguments underlying the panel data approach suggest that the use of panel data can alleviate many of the concerns associated with the SCP empirical approach. In both the current setting and other SCP studies, however, one could still face the general difficulty that changes in the error

ABGH. For example, both parties were unable to enter many of their empirical results into the record. This complicates the ability of observers to completely evaluate the empirical analysis. Moreover, as we will subsequently discuss, these restrictions also hinder research intended to revisit the analysis in this case.

9 The FTC’s analysts particularly note the possible aggravation of ‘errors in variables’ problems in a panel data setting (ABGH, pg. 8).

10 Interestingly, the FTC’s analysts note that the panel and cross-sectional results largely coincided. It is difficult to discern whether the decision to oppose the merger was primarily based on cross-sectional or panel analysis despite the reliance on the latter due to econometric concerns.
term may be correlated with changes in market structure. In effect, the panel
data approach shifts the concern about correlation between unobservables
and market structure to differences in those two items over time.\textsuperscript{11}

We employ an alternative approach to analyze the relationship between
prices and market structure that the FTC considered in the \textit{Staples} case.
Beginning with Bresnahan and Reiss [1990], researchers have considered
ways to analyze empirically the determinants of market structure (Berry
[1992], Mazzeo [2002b]). Formally, these models consider observed market
structure as the outcome of a strategic game between potential entrants.
Following the suggestion of Bresnahan and Reiss [1991], these models can be
used to account for the non-random determination of market structure in a
regression of an outcome such as price on market structure variables as in
Mazzeo [2002a]. This approach involves a correction term that is derived
from the model of endogenous market structure. As in the standard selection
model (Heckman [1979]), this correction is used to account for possible
correlation between market structure and the error term in the price
equation.

In this paper, we note that such an approach provides a fairly
straightforward alternative to the one that the FTC used in the Staples-
Office Depot merger. We begin by constructing a model that characterizes
the realized market structures across a cross-section of heterogeneous
markets. As in Mazzeo [2002a], we then use this model to yield correction
terms in a cross-sectional price regression analogous to the one that the FTC
considered in the \textit{Staples} case.

This approach is not without its weaknesses.\textsuperscript{12} First, we require
substantial distributional assumptions to derive the appropriate correction
term in our second-stage regression. Second, the validity of the correction
depends on correct specification of the first-stage model that characterizes
the market structure. Despite these caveats, we believe that this approach
provides an additional avenue for empirical investigation in the event that
parties in a merger case appeal to such evidence.

In our application, we consider data related to office supply superstores in
a way that parallels as closely as possible the approach of the FTC in the
\textit{Staples} case. We gather data on the presence of the three major office supply
superstores across markets which, following the precedent of the \textit{Staples}
case, we define as metro- and micropolitan statistical areas (MSAs).\textsuperscript{13}

\textsuperscript{11} A tension exists in the reasoning behind the panel regression since the need to observe non-
trivial entry and exit may require a time window long enough to raise doubts about the validity
of the assumed lack of correlation between changes in the unobservable and changes in market
structure.

\textsuperscript{12} These concerns have led researchers in other fields, such as labor economics, to consider
models that are less structural and parametric to deal with selection problems.

\textsuperscript{13} Dalkir and Warren-Boulton [2003] discuss the market definition problem that faced the
FTC in more detail including the treatment of other firms such as discount retailers.
Demographic and other market characteristics for MSAs are readily available from the Census. These data allow us to estimate the initial model related to strategic entry by the three chains. We then use the results of this model to account for possible correlation between market structure and the unobservable in a regression that considers Staples’ prices of various products across markets.

Ideally, we would implement our approach using data from the original case. However, confidentiality concerns prevent access to those data. As a result, we cannot hope to replicate the data or the exact results of either side in the case. Most notably, the price data available to the FTC were much more comprehensive than those available to us. We must consider a much smaller menu of prices than the parties in the initial lawsuit used to construct their price indices, although we still focus on prices for products that were particularly relevant in the case. In addition, we consider price and market structure data ten years after the initial case. Clearly, circumstances may have changed over the last decade implying that our results are not directly comparable to those from the original case. Nevertheless, the same three firms continue to dominate the industry, leading us to conclude that the basic concerns are still relevant. Most importantly, to the extent that future cases in this or other industries involve similar analysis, we feel that this approach is relevant.

A final notable difference between our data and those of the initial case is the fact that we have a single cross-section of prices and market structures rather than a panel. While we could potentially obtain a panel dataset of market structures, we cannot realistically compile an analogous dataset of prices. To the extent that our approach provides a solution to econometric problems in cross-sectional data, however, the panel data analysis used by the FTC is extraneous. In addition, future merger cases may involve settings in which panel data are unavailable or, as noted earlier, insufficient intertemporal variation exists in the relevant explanatory variables.

II. A PRICE REGRESSION AND AN ENDOGENOUS MARKET STRUCTURE MODEL

In this section, we present a very simple regression model relating price to a market structure variable. We then introduce a model that describes observed market structures. As in Mazzeo [2002a], we use this latter model to generate correction terms for the price regression to account for potential correlation between the error term and the market structure regressor.

Admittedly, the model that we propose in this section ignores many potential complications. However, the nature of the relationship that we consider is generally relevant in settings such as the Staples case. In addition, as we discuss in the conclusion, both the price regression and the market structure variables are inherently cross-sectional.
structure model can be amended substantially to deal with more complicated situations.\textsuperscript{14}

We envision a situation in which undifferentiated firms enter and compete in a number of different markets. The issue of interest is the pricing behavior of one of those firms given the presence of its competitors. A reduced form regression\textsuperscript{15} intended to examine this issue is

\begin{equation}
    p_m = X_m \beta + g(N_m; \gamma) + u_m
\end{equation}

where $p_m$ is the (log) price of the primary firm in market $m$, $X_m$ are demand and cost characteristics in market $m$, $N_m$ characterizes the market structure in market $m$, and $u_m$ reflects unobserved factors in market $m$ that impact prices.

The fundamental concern with this simple regression is potential correlation between $u_m$ and $N_m$ which would yield inconsistent estimates of $\beta$ and, more importantly, $\gamma$. If available, an instrument that is correlated with $N_m$, but uncorrelated with $u_m$ could be used to obtain consistent estimates of the regression’s parameters. As noted in the introduction, such instruments are difficult to find in such an application. Moreover, even if an instrument is available, instrumental variables estimation will yield less efficient estimates than the approach that we use in this paper if our model is correctly specified.

An alternative approach involves a model that directly describes the determination of $N_m$. Following Bresnahan and Reiss [1991], we propose a latent payoff function for the homogeneous firms in market $m$ when $N_m$ firms are present as

\begin{equation}
    \Pi(Z_m, N_m, e_m; \theta) = \pi(Z_m, N_m; \theta) + e_m
\end{equation}

where $Z_m$ are market-specific factors that impact profitability in market $m$, $e_m$ are unobserved factors in that market, and $\theta$ are unknown parameters of the payoff function. The payoff function $\Pi(Z_m, N_m, e_m; \theta)$ should be interpreted as the reduced form of the expected present discounted value of profits that result from post-entry competition between firms and that all firms observe. We assume that profits are decreasing in $N$ so that $\Pi(Z_m, N, e_m; \theta) > \Pi(Z_m, N + 1, e_m; \theta)$. In principle, we allow for variables that impact profits, $Z_m$, to potentially differ from those that impact prices. In practice, those variables could coincide.\textsuperscript{16}

\textsuperscript{14}Extensions that consider, for example, firm identities would require relatively lengthy modifications of the current model without augmenting the basic point of our analysis.

\textsuperscript{15}As Whinston [2006] discusses in more detail, (1) is not actually the reduced form of a formal structural model. This fact is the primary source of the difficulties with the price regression.

\textsuperscript{16}In this case, identification relies on the non-linearity of the model, although Cohen and Mazzeo [2004] argue that such a model necessarily involves non-linearities.
In a Nash equilibrium of this model, firms enter the market until no additional firms would be profitable. Specifically, if $N_{\text{max}}$ is the maximum number of potential firms in a market, then the equilibrium number of firms in market $m$ is characterized by the following restrictions on latent profits:

$$
\begin{align*}
N_m &= 0 &\iff& \pi(Z_m, N = 1; \theta) + \epsilon_m < 0 \\
N_m &= N^* < N_{\text{max}} &\iff& \pi(Z_m, N = N^*; \theta) + \epsilon_m > 0, \pi(Z_m, N = N^* + 1; \theta) + \epsilon_m < 0 \\
N_m &= N_{\text{max}} &\iff& \pi(Z_m, N = N_{\text{max}}; \theta) + \epsilon_m > 0
\end{align*}
$$

The first inequality states that a market will have no firms when a monopolist would be unprofitable. The second set of inequalities states that $N^*/C^3$ firms will operate when each of the $N^*/C^3$ firms would be profitable, but additional firms cannot profitably enter. As in Bresnahan and Reiss [1991], we can estimate the parameters of the latent payoff functions, $\theta$, by maximum likelihood given distributional assumptions on the error term $\epsilon_m$. This approach, which mirrors the estimation of latent utilities in single agent decision problems, obtains estimates of $\theta$ through the restrictions on payoffs implied by observed market structures across markets that vary in $Z$.

We end our model by imposing distributional restrictions on the two error terms. Specifically, we assume that, conditional on $X_m$ and $Z_m$,

$$
\begin{pmatrix}
\mu_m \\
\epsilon_m
\end{pmatrix}
\sim N
\begin{pmatrix}
0 \\
0 \\
\sigma_u^2 \\
\sigma_{ue}
\end{pmatrix}
$$

independent across $m$. Since the scale of a discrete outcome model is not identified, we normalize the variance of the latent payoff function, $\sigma_{e}^2$, to one.

Given the distributional and structural assumptions, it is straightforward to show that

$$
E[p_m|X_m, Z_m, N_m] = X_m\beta + g(N_m; \gamma) + \sigma_{ue}h(N_m, Z_m; \theta)
$$

where

$$
h(N, Z_m; \theta) = \begin{cases} 
\phi[\pi(N, Z_m; \theta)] - \phi[\pi(N+1, Z_m; \theta)] & \text{for } 0 < N < N_{\text{max}} \\
\Phi[\pi(N, Z_m; \theta)] - \Phi[\pi(N+1, Z_m; \theta)] & \text{for } N = N_{\text{max}} \\
\Phi[\pi(N_{\text{max}}, Z_m; \theta)] - \Phi[\pi(N_{\text{max}}, Z_m; \theta)] & \text{for } N = N_{\text{max}}
\end{cases}
$$

with $\phi(.)$ and $\Phi(.)$ representing the pdf and cdf of the standard normal distribution. We can then consider the regression model

$$
p_m = X_m\beta + g(N_m; \gamma) + \sigma_{ue}h(N_m, Z_m; \theta) + \epsilon_m
$$

where $\epsilon_m \equiv \mu_m - \sigma_{ue}h(N_m, Z_m; \theta)$ is mean zero conditional on $X, Z$ and $N$ by construction. The term $\sigma_{ue}h(N_m, Z_m; \theta)$ reflects the possibility that $E[\mu_m|N_m]$ is not equal to zero, so that least squares estimates of the initial regression model may be inconsistent. This term plays a similar role to the correction term for selection bias in the standard model proposed by Heckman [1979]. As Mazzeo [2002a] notes, this similarity suggests a two-step estimation...
procedure in which the market structure model is first estimated using an ordered probit maximum likelihood routine. The resulting estimates of $\theta$ can be used to compute correction terms $h(N, Z_m; \hat{\theta})$ which are then included in a second-stage price regression. Derivation of the asymptotic distribution of this two-stage estimator involves straightforward but tedious calculations analogous to those in Heckman [1979] and elsewhere. We omit these calculations for the sake of brevity.

III. DATA AND ESTIMATION RESULTS FOR OFFICE SUPPLY SUPERSTORES

Estimation of the model in the previous section requires information on market structures, market characteristics, and prices for a sample of $M$ markets. We gathered data to mimic those that the FTC employed in the Staples case, although various constraints prevented us from creating a dataset that is anything more than a rough approximation of its data.

Following the FTC, we define the relevant markets to be metro- and micropolitan statistical areas (MSAs). Determined by the Office of Management and Budget, MSAs are combinations of counties that have a high degree of social and economic integration as measured by commuting patterns. For MSAs, we have information about total population (Pop), per cent of population in urban areas (% Urban), per cent of population below the poverty level (% Poor), average household income (Avg Inc), average value of owner occupied housing (Avg Home Val), per cent of white collar workers in the employed adult civilian population (% White Collar), and the number of business establishments (Establishments.) We gathered the first six variables from the county-level data in the 2000 Census while we obtained the last variable from the 2004 County Business Patterns survey. In all cases, we aggregated the county-level data up to the level of MSAs as necessary. We also measured the distance from each MSA to the corporate headquarters of the three chains using interior latitude-longitude points from the Census as the point of reference for each MSA. As we subsequently discuss, we employ the minimum of these three distances in our empirical analysis.

We compiled a comprehensive list of Staples, Office Depot and OfficeMax stores by searching the online store directories for the three chains. Each chain assigns unique identification numbers to their stores. We iterated over possible store identification numbers on the websites of the three chains.

---

17 The corporate headquarters of Staples and Office Depot are respectively Delray Beach, Florida, and Framingham, Massachusetts. OfficeMax is currently headquartered in Naperville, Illinois, although for reasons that should become clear, we use its original headquarters in Shaker Heights, Ohio, for this calculation.

18 The Census reports interior points for counties, but not MSAs. In cases where multiple counties comprise an MSA, we used the average distance from the relevant counties for this calculation.
When a search yielded a store with a particular number, we obtained latitude and longitude information for that store from the resulting maps on the website. These coordinates allowed us to place each store in a MSA. In total, we found 1,010 Office Depot stores, 913 OfficeMax stores and 1,240 Staples stores in MSAs across the United States, numbers that closely correspond to the store counts in the annual reports of the three companies.

In its empirical analysis, the FTC used comprehensive price data on basic office supplies at Staples stores to create a price index for each store. The focus on basic office supplies excludes products such as computers or furniture that presumably involve important market participants beyond the office supply superstores. The FTC categorized the underlying products into a number of classes including ‘price-sensitive’ items that consumers purchase frequently such as copy paper and ‘leadership’ items such as pens and pencils that sell at low margins to increase traffic in the stores. Both sides in the case combined prices across these classes into a final price index that was used in the estimation (ABGH, pg. 9–10).

We cannot hope to replicate the price index that the FTC employed in the Staples case. Instead, we hired a team of Washington University undergraduates to gather data on a small set of goods by calling Staples stores in each MSA on Saturday, January 28, 2006. We selected products that were similar to the types of goods about which the FTC was concerned and that would almost certainly be sold in every Staples. The four products that we selected were: a case (10 reams with 500 pages per ream) of basic 20 lb. Staples copy paper; a box of 500 #10 standard gummed envelopes; a box of 12 Sanford Uniball pens; and a 12 pack of 12 lb. 50-sheet writing pads. Unfortunately, a temporary sale of Uniball pens in some stores prevented us from obtaining useful data on that product. The data collection process was time consuming which prevented us from gathering data on more products or from the other chains.

The distribution of population for MSAs is highly skewed. In large MSAs, all three chains are almost always present. In addition, the proliferation of outlets is much greater in large MSAs. Because competitive concerns associated with the number of stores is likely more important than chain presence in large MSAs, we focus our attention on MSAs with population

---

19 For example, from the ‘Find a store’ section of Staples’ website (http://stores.staples-locator.com/staples/), store #665 is located in Garden City, Kansas. An examination of the source code for the search result yields a latitude of 37.981392 and a longitude of — 100.838799. These coordinates reveal that this is the only Staples outlet, and the only office supply superstore, in the Garden City MSA.

20 We found a few additional stores that are not located in MSAs or are located in U.S. territories or Canada. We exclude these stores from our estimation.

21 Our callers had particular difficulties obtaining price information from stores in large metropolitan areas such as New York City. In such cases, they spent a substantial amount of time listening to the greatest hits of the 1980’s while on hold.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.
less than the 90th percentile. In these smaller MSAs, multiple outlets of a given chain are relatively rare, allowing us to examine the more straightforward issue of chain presence. Of the 842 MSAs in our final data, 97 have more than one store of any chain, although 70 of these markets involve multiple stores for only one of the three chains. Table I describes the demographic characteristics and market structures that we observe in these markets.

Table II provides summary statistics of the Staples price data by market structure. For most of these products, this table appears to indicate a decreasing relationship between the prices that Staples charges and the degree of competitor presence. Our subsequent analysis will investigate whether this relationship is robust to inclusion of additional covariates and the correction provided by our market structure model.
We begin by estimating the market structure model by maximum likelihood using the sample of all 842 smaller MSAs. As discussed in the previous section, this model is an ordered probit where the outcome variable describes the number of firms present in a market. The maximum number of firms is conveniently capped by the existence of three major firms. We use a simple functional form for the latent payoff function given by

\[
\Pi(Z_m, N_m, e_m; \theta) = Z_m \delta + \Delta_1 \times 1(N_m > 1) + \Delta_2 \times 1(N_m > 2) + e_m
\]

where \(1(.)\) is an indicator equal to one if the expression in parentheses is true. Hence, \(\Delta_1\) measures the difference between latent monopoly and duopoly payoffs while \(\Delta_2\) reflects the incremental impact of adding a third firm to the market. We then use the market structure model to compute correction terms that we include in the price regressions for the subsample of 324 markets in which Staples is present.\textsuperscript{22,23} The functional form that we employ for the price regression is

\[
p_m = X_m \beta + \gamma_1 \times 1(N_m > 1) + \gamma_2 \times 1(N_m > 2) + u_m
\]

so that \(\gamma_1\) measures the difference between monopoly and duopoly prices and \(\gamma_2\) measures the additional impact of a third firm. The actual estimated regression then includes \(\sigma_{ \text{uc}\theta}(N_m, Z_m; \hat{\theta})\) computed using the first-stage maximum likelihood estimates of \(\theta = \{\delta, \Delta_1, \Delta_2\}\).

We estimate this model separately for each of the three products using the logarithm of the prices. To illustrate the importance of the market structure model, we estimate each price regression with and without the correction term. To account for some of the cost and demand factors that may impact prices in this industry, the variables \(X\) include the demographic variables that we described earlier, namely Pop, % Urban, Avg Inc, % Poor, Avg Home Val, % White Collar, and Establishments. The variables \(Z\) include those variables along with the distance to the nearest corporate headquarters. We include this variable in the market structure model due to the possible contemporary effect of historical expansion patterns of these chains around their respective corporate headquarters.\textsuperscript{24} Such expansion may

\textsuperscript{22} One observation for envelopes and two observations for writing pads have prices that are clearly incorrect. We drop these problem observations in both cases.

\textsuperscript{23} The requirement that each market must contain a Staples in our subsample for the price regression is not problematic in light of our assumption about the homogeneity of the competitors. If heterogeneity were an issue of interest, as could be the case in a more complex analysis, we would have to amend the model as we discuss in the conclusion.

\textsuperscript{24} Similar expansion patterns are well documented in a wide variety of retail industries such as discount retailers (Basker [2005], Zhu, Singh and Manuszak [2005]). While anecdotal evidence suggests that these patterns also occurred for office supply superstores, we cannot directly verify those patterns, nor are we aware of any formal evidence of them. An exception is Staples which provides store opening dates in the source code on its website. Examination of opening dates and proximity of MSAs to Staples’ headquarters for stores of that chain corroborates this relationship.
impact the market structures that we currently observe at different distances from those headquarters. Implicitly, such patterns involve correlation in payoffs across markets arising from dynamic factors. Rather than explicitly modeling spatial relationships in payoffs, we intend inclusion of this variable to crudely reflect them, although we notably do not account for chain identity in these effects. We allow the coefficient on this variable to differ across market structures so that the prevalence of monopoly, duopoly and triopoly may differ depending on proximity to the nearest corporate headquarters.25 While state dependence may have persistent effects on the market structures that we observe due to the relatively long-run nature of entry decisions, we assume that any cost effects that may have been associated with proximity to corporate headquarters are no longer important.26 Hence, we omit this variable from the price regression. This feature of our estimation implies that identification of the model is not purely due to functional form. Finally, both the price regression and latent payoff function include constants.27

Table III presents the maximum likelihood estimates of the market structure model. Due to the scale normalization for $\sigma_c^2$ and the reduced form

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.3611</td>
<td>0.2958</td>
</tr>
<tr>
<td>Population</td>
<td>1.1654</td>
<td>0.1494</td>
</tr>
<tr>
<td>% Urban</td>
<td>0.0154</td>
<td>0.0039</td>
</tr>
<tr>
<td>Avg Inc</td>
<td>-0.0451</td>
<td>0.0923</td>
</tr>
<tr>
<td>% Poor</td>
<td>-0.0229</td>
<td>0.0132</td>
</tr>
<tr>
<td>Avg Home Val</td>
<td>0.1498</td>
<td>0.0896</td>
</tr>
<tr>
<td>% White Collar</td>
<td>0.017</td>
<td>0.007</td>
</tr>
<tr>
<td>Establishments</td>
<td>0.4429</td>
<td>0.1546</td>
</tr>
<tr>
<td>Min HQ Dist (1(N = 1))</td>
<td>-0.1785</td>
<td>0.0717</td>
</tr>
<tr>
<td>Min HQ Dist (1(N = 2))</td>
<td>-0.0525</td>
<td>0.0668</td>
</tr>
<tr>
<td>Min HQ Dist (1(N = 3))</td>
<td>0.0867</td>
<td>0.0786</td>
</tr>
<tr>
<td>(1(N &gt; 1))</td>
<td>-1.8201</td>
<td>0.0981</td>
</tr>
<tr>
<td>(1(N &gt; 2))</td>
<td>-3.4498</td>
<td>0.1696</td>
</tr>
<tr>
<td>LogLik</td>
<td>-600.9439</td>
<td>0.0981</td>
</tr>
<tr>
<td>N</td>
<td>842</td>
<td></td>
</tr>
</tbody>
</table>

Note: Population, Avg Inc, Avg Home Val, Establishments, and Min HQ Dist are standardized.

25 In principle, coefficients for the distance variable that differ across market structures could yield payoffs that are not ordered for all values of that variable. Given the estimated parameter values, however, payoffs do exhibit an ordered structure for the range of distances that we observe.
26 Proximity to a distribution center of a given chain is likely an important cost factor for stores of that chain. Unfortunately, we have no information about the location of distribution centers for these chains.
27 To facilitate estimation of the first stage model, we standardize the variables Pop, Avg Inc, Avg Home Val, Establishments, and Min HQ Dist using the means and standard deviations in the complete sample of 842 markets.
nature of the payoff function, it is difficult to concretely interpret these estimates beyond noting their signs and relative magnitudes. Nevertheless, based on these estimates, market size reflected in both population and the number of establishments is clearly an important determinant of market structure, a fact that is evident from the sample statistics in Table I. Markets that are more urban also support more firms as do those that have more white collar workers while markets with a high proportion of residents below the poverty line have fewer firms. The estimates related to the distance variable indicate that an otherwise unattractive market can support a firm when that market is closer to a headquarters. For a monopoly, a decrease of 100 miles in the distance variable is equivalent to approximately 2,300 consumers all else equal. Conversely, some weak evidence exists that such a market is less likely to support a triopoly, although the point estimate for that market structure is far from statistically significant. The estimates of $\Delta_1$ and $\Delta_2$ provide evidence about the extent of competition in this industry. When all other variables are evaluated at their sample means, our estimates imply that the approximate breakeven populations for a monopoly, duopoly, and triopoly are 47,000, 155,000 and 360,000 respectively.

In Table IV, we present regression estimates for the logarithm of price for the three products sold by Staples. For each product, the first column presents the uncorrected regression estimates while the second column presents estimates and asymptotically consistent standard errors when we use the market structure model to generate correction terms.

The uncorrected regression results exhibit a number of common features across the products. First, the estimated coefficients for number of establishments, extent of urban population, and percentage of white collar workers are negative for each product while the population coefficient is negative for all products except paper. Second and more notably, the estimated coefficients for a second firm are negative. The estimates range from a price decrease of 2% for envelopes to 12.6% for writing pads. Conversely, the estimates suggest that a third firm has a lesser effect. The regressions for envelopes and paper indicate that the incremental effect of a third firm is also negative, but these estimates are statistically and economically insignificant. For writing pads, the point estimate implies that a triopoly involves higher prices than a duopoly, although this estimate is also statistically insignificant.

Since we really have two highly positively correlated measures of market size, namely population and establishments, we perform these calculations for breakeven population with the explanatory variable mean establishments scaled by mean population.

Bresnahan and Reiss [1991] draw inferences about the competitive effects of entry based on the estimated market sizes required to support different numbers of firms, albeit using a different functional form than we use.
Taken together, the estimates would seem to suggest that most of the effect on prices arises with the second firm while the third firm is immaterial. If robust, these results would indicate that, when relying on this type of evidence for this industry, the FTC should be concerned with ensuring the presence of at least two firms in a market, while additional competitors are relatively unimportant.

The corrected regressions, however, suggest that these findings are not, in fact, robust despite the general decrease in precision that results from the two-stage estimation approach. Compared to the initial estimates, the price regressions for each of the three products exhibit the same basic patterns when we include the correction term. First, the relationship between prices and the second firm becomes more negative, ranging from a decrease of 4.85% for envelopes to 20.1% for writing pads. For all of the products, the change in these coefficient estimates is striking with an approximate doubling of the values for paper and envelopes. Second, unlike in the simple least squares case, the corrected regressions indicate that the third firm also has a substantial impact. This relationship ranges from a decrease of 4.76% for the price of envelopes to 10.56% for a case of paper. Except for writing pads, these estimates notably suggest that the effects of a second and third firm are roughly the same.

### Table IV

Uncorrected and Corrected Regressions of Log(Price) for Various Staples Products

<table>
<thead>
<tr>
<th>Variable</th>
<th>Case of paper</th>
<th>Writing pads</th>
<th>Envelopes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uncorrected</td>
<td>Corrected</td>
<td>Uncorrected</td>
</tr>
<tr>
<td>Constant</td>
<td>3.2713</td>
<td>3.2338</td>
<td>1.8175</td>
</tr>
<tr>
<td></td>
<td>(0.0195)</td>
<td>(0.0243)</td>
<td>(0.0326)</td>
</tr>
<tr>
<td>Pop</td>
<td>0.0038</td>
<td>0.0386</td>
<td>−0.0024</td>
</tr>
<tr>
<td></td>
<td>(0.0095)</td>
<td>(0.0174)</td>
<td>(0.0161)</td>
</tr>
<tr>
<td>% Urban</td>
<td>−0.00035</td>
<td>0.00043</td>
<td>−0.00017</td>
</tr>
<tr>
<td></td>
<td>(0.00023)</td>
<td>(0.0003)</td>
<td>(0.00038)</td>
</tr>
<tr>
<td>Avg Inc</td>
<td>0.01296</td>
<td>0.01329</td>
<td>−0.0115</td>
</tr>
<tr>
<td></td>
<td>(0.0065)</td>
<td>(0.007)</td>
<td>(0.0109)</td>
</tr>
<tr>
<td>% Poor</td>
<td>−0.0024</td>
<td>0.0014</td>
<td>−0.00011</td>
</tr>
<tr>
<td></td>
<td>(0.00104)</td>
<td>(0.00103)</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>Avg Home Val</td>
<td>−0.0031</td>
<td>−0.00073</td>
<td>0.0064</td>
</tr>
<tr>
<td></td>
<td>(0.0045)</td>
<td>(0.0045)</td>
<td>(0.00075)</td>
</tr>
<tr>
<td>% White Coll</td>
<td>−0.00037</td>
<td>0.000298</td>
<td>−0.0019</td>
</tr>
<tr>
<td></td>
<td>(0.00047)</td>
<td>(0.00056)</td>
<td>(0.00079)</td>
</tr>
<tr>
<td>Establishments</td>
<td>−0.0051</td>
<td>0.0024</td>
<td>−0.0048</td>
</tr>
<tr>
<td></td>
<td>(0.0094)</td>
<td>(0.0091)</td>
<td>(0.0158)</td>
</tr>
<tr>
<td>1(N &gt; 1)</td>
<td>−0.0423</td>
<td>−0.1058</td>
<td>−0.1264</td>
</tr>
<tr>
<td></td>
<td>(0.0708)</td>
<td>(0.0233)</td>
<td>(0.0131)</td>
</tr>
<tr>
<td>1(N &gt; 2)</td>
<td>−0.0074</td>
<td>−0.1056</td>
<td>0.0216</td>
</tr>
<tr>
<td></td>
<td>(0.0119)</td>
<td>(0.0368)</td>
<td>(0.0199)</td>
</tr>
<tr>
<td>(\sigma_{\text{uc}})</td>
<td>0.0441</td>
<td>0.05199</td>
<td>0.05199</td>
</tr>
<tr>
<td></td>
<td>(0.0168)</td>
<td>(0.0255)</td>
<td>(0.0255)</td>
</tr>
</tbody>
</table>

**Note:** Sample sizes equal 324 for case of paper, 323 for envelopes and 322 for writing pads. Standard errors in parentheses. Pop, Avg Inc, Avg Home Val, and Establishments have been standardized.
In each case, the estimate of the (scaled) covariance between the error terms from the two models is positive and statistically significant.\textsuperscript{30} The positive value of the coefficient estimates of $\sigma_{ue}$ in each case indicates that, conditional on other market characteristics, unusually high prices are associated with markets that are unusually likely to support additional firms. Although the reduced form nature of the price regression prevents clear interpretation of this finding, it is consistent with the possibility that unusually attractive demand conditions encourage entry while also supporting prices that are higher than would otherwise obtain. Alternatively, prices and presence of office supply superstores could both be unusually high in markets that lack other types of competitors such as discount retailers, a variable about which we have no information in this paper.

The correction also yields interesting results for some of the other coefficients. The coefficients on the market size variables population and establishments as well as those for per cent of population urban and white collar change signs in many cases, although the exact signs and statistical significance varies across products. These findings illustrate the commonly known, but sometimes overlooked, fact that correlation between a regressor and the error term in a multivariate regression may contaminate all of the estimates, not just that associated with the problem regressor.

Overall, these results indicate that analysis based on a basic cross-sectional regression of prices on market structure variables may yield misleading results.\textsuperscript{31} Simple least squares estimates appear to understate the importance of the first competitor while completely misstating the impact of the second competitor. Was the FTC correct to oppose the Staples-Office Depot merger? Not least due to the time gap between the actual case and our analysis, the answer to that question is unclear.

Were such a merger proposed today in this industry, however, standard regression analysis of the relationship between prices and market structure would suggest that the presence of two firms is the crucial concern for prices whereas a third firm is largely irrelevant.\textsuperscript{32} If the FTC were to employ such evidence when evaluating a proposed merger, these findings would indicate that a merger could be relatively innocuous in those markets where all three firms operate with concerns arising only for markets where the merging firms are currently a duopoly. This evidence could encourage the FTC to generally

\textsuperscript{30} Note that the normalization of the variance of $e_{it}$ does not impact the sign of $\sigma_{ue}$.

\textsuperscript{31} We should emphasize that any results from such a model apply only to the specific setting being examined.

\textsuperscript{32} More comprehensive evaluation could employ price data for all three chains, although data for a single chain could still be useful in a merger case depending on the identities of the merging firms, as appears to have been the case in the FTC’s analysis for the Staples case (ABGH, pp. 9–10). Having noted this point, we ignore the fact that we only have price data for Staples in our subsequent discussion.
approve a merger subject to conditions related to the problematic duopoly markets.

The corrected regressions paint a different picture of this industry with substantially different policy implications. In contrast to the implications of the uncorrected estimates, the corrected estimates predict that a merger could have negative ramifications for triopoly markets. Competition from a third firm, rather than being inconsequential, has a large effect on prices and is perhaps as important as that of a second firm. Moreover, from a more long-run point of view, detrimental effects could arise by removing a competitor who could potentially enter duopoly markets that would otherwise be unaffected by the merger. In sum, our findings when we explicitly model the determination of market structure would suggest a position that is almost completely opposite to that which would follow from the uncorrected estimates: the FTC should generally view the merger in a negative light.33

IV. CONCLUSION

The general relationship between price and market structure that we examine in this paper is still relevant for policy concerns. In 2005, Blockbuster Video and Hollywood Video abandoned a proposed merger following the threat of opposition from the FTC. Anecdotal evidence indicates that the FTC relied on analysis similar to that in the Staples case when considering whether to oppose the merger (Wilke and Flint [2005]). In light of the continued use of this type of evidence in merger cases and elsewhere, we feel that this paper’s approach, in which we explicitly model the determination of market structure to correct a basic price regression, can supplement analysis based on panel data or other approaches.

As we previously noted, the model that we consider in this paper ignores many complications. As may be necessary in an actual merger case, it could be extended in a variety of directions subject, in part, to the availability of suitable data. To begin with, the price regression could be modified to examine more subtle issues. For example, our price regression assumed that different market structures yield a shifted outcome where the identity of the competitors does not matter. In principle, one could allow the impact of covariates to differ across market structures. Moreover, one could allow the impact of competitors to depend on the identity of competing firms or their characteristics such as proximity to the firm of interest.

In the latter case, the market structure model would also have to be adjusted to reflect the importance of heterogeneity across firms. Various

33 Presumably, any predicted price effects of a merger based on this analysis could be short-run in nature since new entry could occur over a longer time horizon. The current analysis is not informative about when in the future such entry might take place.
models have been proposed to deal with these types of situations. Among many others, such models consider discrete heterogeneity (Mazzeo [2002b]), firm location (Seim [2006]), idiosyncratic heterogeneity in fixed costs (Berry [1992]), or differences across specific firms such as the retail discount chains examined in Zhu, Singh and Manuszak [2005]. The last example is particularly relevant for the current application where heterogeneity across specific firms may be of special interest. However, as heterogeneity is added to the market structure model, difficulties may arise. Multiple equilibria are endemic in more complex models of endogenous market structure. Estimation of the market structure model and the correction term for the price regression become correspondingly more complex and computationally intensive. Various assumptions and alternative estimation approaches such as those employed by Mazzeo [2002b] or Zhu, Singh and Manuszak [2005] can deal with some of the complications in these types of model. In general, the nature of the market structure model and any required assumptions will depend on the particular application and available data. Nonetheless, the current paper suggests that this approach can provide a tractable, and significant, improvement over simple cross-sectional regression techniques.

REFERENCES


© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.


