Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart

Jerry Hausman and Ephraim Leibtag[1]
MIT and Economic Research Service, U.S. Department of Agriculture
Draft, November 29, 2004[2]

I. Introduction

Consumers often benefit from increased competition in differentiated product settings. In previous research Hausman (1997a, 1997b, 1999, 2002) has estimated the increased consumer welfare from the introduction of new brand, e.g. Apple Cinnamon Cheerios, and new products, e.g. mobile telephones. In this paper we consider consumer benefits from increased competition in a differentiated product setting: the spread of supercenters. Supercenters are extremely large stores that sell a large variety of products. They are differentiated from more traditional shopping outlets that often specialize in a specific category: supermarkets specialize in selling food, although they carry other products, while a supercenter will sell food, clothing, prescription drugs, clothing, home office supplies, and electronic equipment.

These supercenters are new outlets within a given geographic market defined by household shopping patterns. They originally began in the southern and southwestern areas of the US.[3] Over the past few years they have spread to the central U.S. and they are now attempting to move into the northeast and west coast areas of the U.S. However, they have encountered problems in entering markets in these areas, with the restrictions often created by zoning ordinances.

By far the largest and most controversial of these supercenter outlets is Wal-Mart. Wal-Mart typically encounters significant opposition from competing outlets and from labor unions, who often represent workers in these competing outlets.[4] Wal-Mart charges significantly lower prices than traditional outlets. The traditional outlets typically respond to Wal-Mart entry by decreasing their prices and attempting to reduce wages and

[1] We thank Jie Yang and Ketan Patel for outstanding research assistance. Author contact: jhausman@mit.edu
[2] First draft: do not quote or cite without permission. No views in this paper reflect the U.S. Department of Agriculture's position on these issues.
[3] Somewhat similar outlets, called hypermarkets in France, have existed for a number of years.
[4] Wal-Mart has no unions in the US. It has recently permitted unions in China. One Wal-Mart store in Quebec,

benefits to their unionized workforce. Thus, neither competitive outlets nor their unionized workforces favor Wal-Mart entry. Indeed, from late 2003 to early 2004, a protracted work stoppage took place in Los Angeles area supermarkets prior to expected Wal-Mart entry into the market.

In this paper we estimate consumer benefits from supercenter entry and expansion into markets for food. We estimate a discrete choice model for household shopping choice of supercenters and traditional outlets for food. We have panel data for households so we can follow their shopping patterns over time and allow for a fixed effect in their shopping behavior. Most households shop at both supercenters and traditional outlets during the period. Given a model of shopping behavior we estimate the compensating variation of household from the presence of supercenters. We find the benefits to be substantial. Thus, while we do not estimate the costs to workers who receive lower wages and benefits, we find the effects of supercenter entry and expansion to be sufficiently large so that overall we find it to be extremely unlikely that the expansion of supercenters does not confer a significant overall benefit to consumers.

II. Market Description

Over the past decade, "non-traditional" shopping formats have captured significant share from "traditional grocery." P. Little (2004) describes the two categories of alternative retail outlets as "high-spend" outlets, which are low price, one-stop shopping destinations, and "low and medium-spend" stores which are mostly convenience stores that serve a "fill-in" role in between trips to the "high-spend" outlets He includes supercenters (Wal-Mart, Kmart, Meijer, etc.), warehouse clubs (Sam's Club, Costco and BJ's), and mass merchants (Wal-Mart, Kmart, Target, etc.) as the primary outlets for these "high-spend" expenditures.[5] Using 2003 data, he estimates that these outlets have 24.8% of food expenditures, with supercenters having 45.6% of the category. Over the past few years Wal-Mart has become the largest supermarket chain in the U.S. Wal-Mart, excluding its Sam's Club, now has supermarket- related revenues approximately 51% larger than the runner-up Kroger, and larger than Albertsons and Safeway, the third and

Canada has been involved in a controversy over whether its workers will form a union.

[5] Sam's Club is owned by Wal-Mart.

fourth largest supermarket chains, combined. Nationwide Wal-Mart has a 14% market share (in 2003), despite not being in a number of regional markets, and an 18% share when Sam's Clubs are included. Within the "medium-low spend" category, Little estimates convenience stores that also sell gasoline as the fastest growing store type with 85.5% of the 12.4% total share for the category. Little calculates that total traditional grocery outlets, including conventional supermarkets and superstores (a larger version of the conventional supermarket), have decreased to a 56.3% dollar share in 2003. He also forecasts that in 5 years, the "high-spend category" will grow from 24.8% to 31%, with supercenters comprising 54.8% of the total while traditional grocery outlets decrease from 56.3% to 48.3%. Thus, he expects Wal-Mart to become increasingly important over the next few years, continuing the trend of change over the past decade.

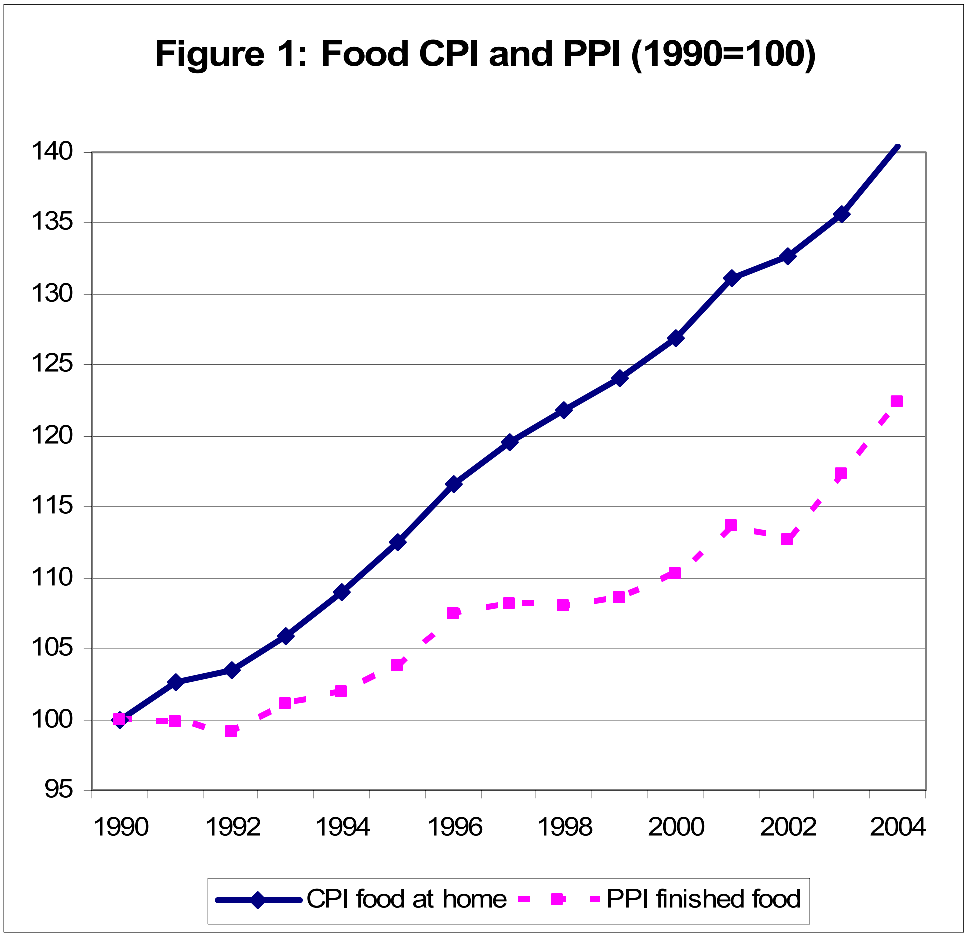
Wal-Mart began selling food in 1988 and in 2002 became the largest grocery chain in the U.S, now larger than Kroger, Albertsons, and Safeway, which are the next largest supermarket chains. Significant consolidation has occurred in the supermarket industry, but Wal-Mart continues to grow at a significantly faster rate than these supermarket chains. The majority of Wal-Mart's grocery sales arise from its over 1400 (as of April 2004) supercenters which average 180,000 square feet per store and contain both discount and grocery store items, although it also has some "Neighborhood Market" stores that are about the size (40,000 sq. feet) of an average supermarket.[4] While most of the stores are in the South and Southwest, Wal-Mart is increasingly moving into urban centers with openings expected in Los Angeles and Chicago, along with other urban areas.[6]

Over the 10-year period from 1991-2001 margins increased in supermarkets as the price of food sold at supermarkets grew at approximately twice the rate of the PPI for food. Over this period the PPI for finished food increased by 13.9% while the CPI for food at home increased by 27.7% as demonstrated in Figure 1. Despite significant recent

[4] Wal-Mart management has given guidance that it expects to open between 230-240 new supercenters in 2005 for an increase of about 16%. See Dow Jones, "Factiva," April 19, 2004. Morgan Stanley reports that Wal-Mart is seeking 16%-17% growth in supermarket sales compared with 3% industry growth. See M. Wiltamuth and R. Fariborz, "Food Retail," June 2004. Wal-Mart has grown at a 16% rate over the past three years.

[6] Wal-Mart has sometimes had difficulty in receiving planning approval for its stores. Currently, Wal-Mart has either no presence or an extremely limited presence in New England, the New York metro area, California, and the Pacific Northwest. However, its expansion into new areas has proceeded over the past few years.

research on competition among supermarkets, the importance of this factor does not seem to have been recognized.



Source: BLS

The CPI for food at home is the portion of the consumer price index that measures price change for food items primarily consumed at home most of which is purchased at supermarkets.[7] The PPI for finished food products is the portion of the producer price index which approximates wholesale prices for the majority of supermarket food sales. While other cost components, e.g. labor costs, enter the cost function of supermarkets, wage and other inflation was sufficiently lower during the 1991-2001 period that it is very

[7] Because of an incorrect "linking procedure" used by the BLS in computing the CPI, the prices from Wal-Mart and other supercenters do not affect the CPI when these stores enter new market or expand in existing markets. For a

likely that gross margins (price minus average variable cost) increased significantly during the 1990s. This increased gross margin presented supercenters, and especially Wal-Mart, an invitation to enter markets and to expand with their lower than supermarket prices.

Various studies have demonstrated that food items at Wal-Mart are 8%-27% lower priced that at the large supermarket chains, even after discounts for loyalty card and other special are taken into account.[8] After entry by Wal-Mart conventional supermarkets typically decrease their prices (or do not increase them as much as in non-Wal-Mart markets) because of the increased competition. Basker (2004) explored the decreased prices in a market with Wal-Mart present for drug stores, convenience stores, and clothing stores, but did not discuss the dynamics of entry and expansion nor the consumer benefits from these market outcomes.

Some previous studies and the U.S. Bureau of Labor Statistics have posited a "compensating service effect" explanation where consumers do not receive benefits because of inferior service from supercenters. However, this explanation is inconsistent with the "indirect price effect" that we estimate subsequently, where we find that as expenditure at supercenters increases in a given market, the prices at traditional supermarkets decrease. For example, after two Wal-Mart supercenters opened in Houston, a nearby Kroger's sales dropped 10%, the Kroger store reduced worker hours by 30%-40%, and it decreased its prices.[9] Presumably this price decrease is caused by greater competition. We would expect this indirect price effect in a model of differentiated shopping outlet competition. For a related problem see Hausman and Leonard (2002) which finds an indirect price effect and test for its consistency with a Nash-Bertrand model of competition among competing differentiated bath tissue products.

---

discussion see Hausman (2003), and Hausman and Leibtag (2004).

[8] A recent December 2003 study by UBS Investment Research found a price gap of 17.3% to 26.2%, "Price Gap Tightens, Competition Looks Hot Hot Hot." The previous year UBS found a price gap of 20.8% to 39.1%. For example for a specified identical market basket UBS finds Wal-Mart supercenters to have an average price 19.1% less expensive in Tampa and 22.8% less expensive in Las Vegas. In 2002, Salomon Smith Barney estimated the price gap to be between 5% and 25%. See L. Cartwright, "Empty Baskets, September 12, 2002.

[9] P. Callahan and A. Zimmerman (2003) report on these effects. The regional head of Kroger's stated, "Wal-Mart made us look at ourselves and reinvent ourselves."

Thus, consumers demonstrate with their expenditure choices that they prefer lower priced outlets, and the higher priced supermarket must respond in a competitive manner. Consumers benefit both from the "direct price effect" of lower prices at supercenters and the "indirect price effect" of lower prices at competing outlets due to the entry and expansion of supercenters. We measure the consumer welfare effects of both the direct price effect and indirect price effect in our empirical results.

III. A Utility-Consistent Economic Model of Shopping Destination

Households choose among differentiating shopping outlets by trading off prices and other shopping costs with quality and convenience. While the goods purchased are often physically the same, e.g. a 12 ounce box of Cheerios, the prices are typically lower at supercenters but the service quality may be lower and the checkout lines may be longer.[10] However, the assumption of an totally offsetting compensating service differential is inconsistent with real world market behavior that finds when Wal-Mart opens a store in a new geographic market, it rapidly gains share while conventional supermarkets lose share.[11] Some consumers find the choice to be superior while others continue to shop at conventional supermarkets.[12] Thus, the arrival of Wal-Mart in a given geographic market is similar to the introduction of a new differentiated good into the geographic market. The effect on consumers is similar to previous research by Hausman (1997a, 1997b, 1999) and Hausman and Leonard (2002) since consumers now have increased choice in their shopping trip destinations.

For our economic model we consider the conditional choice of consumers to shop at either a conventional supermarket or at a lower price, and perhaps lower service quality, supercenter. For ease of exposition, we use a two-stage choice model in which at

[10] Personal experiences are often a dangerous guide to market outcomes. Nevertheless, in my limited experience in supermarkets and supercenters I have not found the service better or the checkout lines shorter in supermarkets on average.

[11] Supermarket chains sometimes exit a geographic market after Wal-Mart enters. Albertsons exited the Houston market after Wal-Mart entry. However, in our model we assume that consumers continue to have access to traditional supermarkets, even if a given chain exits the market.

[12] As we discussed above, these conventional supermarkets typically decrease price because of the increased competition from Wal-Mart. If the BLS consistently applied its "quality adjustment" procedure it would ignore these price decreases at conventional supermarkets because presumably they arise from reduced service quality. However, the BLS fully incorporates these price decreases, demonstrating that its approach is based on no correct economic assumptions.

the lower stage the consumer considers his or her shopping behavior conditional on type of store. The consumer calculates a price index for shopping at either type of store, takes account of service and other quality differences, and then at the upper stage decides which type of store to shop at. Because of the stochastic term in the choice decision, many consumers can be expected to shop at both types of stores during a period. We use the two-stage approach of Hausman (1985) and Hausman, Leonard and McFadden (1995), although neither of the models was designed precisely for the situation of shopping destination choice.

We allow for consumers choice of shopping at either a conventional supermarket, j=1, or at a supercenter, j=2. Conditional on choosing to shop at one of these two types of stores the consumer has a *conditional expenditure function*

$$y = e(p_{0,} p_1{}^j, p_2{}^j, ..., p_n{}^j; \bar{u}) = e(p, \bar{u}) \text{ solves } \min \sum_i p_i x_i \textit{ such that } u(x) = \bar{u} \quad (3.1)$$

where p_0 is a vector of prices of all non-food items assumed the same for destination choice, $p^j = \{ p_1{}^j, p_2{}^j, ...p_n{}^j \}$ are the prices of the n goods in the two types of outlets denoted by the superscript j, and $\bar{u}$ is the utility level of the consumer.[13] The conditional demand for each type of product, depending on the type of outlet j chosen is:

$$x_i{}^j = \frac{\partial e(p_0, p^j, \bar{u})}{\partial p_i{}^j} = \frac{-\dfrac{\partial v(p_0, p^j, y)}{\partial p_i{}^j}}{\dfrac{\partial v(p_0, p^j, y)}{\partial y}} \quad i = 1..., n \qquad (3.2)$$

where the indirect utility function v(p,y) is derived from the duality relationship with the expenditure function. Using duality corresponding to any level of utility in equation (3.1) and any vector of prices, a price index exists that corresponds to the minimum expenditure required to achieve a given level of utility $\bar{u}$. Indeed, the utility consistent price index is the level of expenditure needed to achieve the utility level:

[13] As written, equation (3.1) assumes that both types of stores carry all goods. To the extent that supermarkets carry a wider variety of products that supercenters, the prices for supercenters can be entered as virtual prices that set

$$\Pi(p^j,\bar{u}) = e(p^j,\bar{u}) = y^j(p^j,\bar{u}) = y^j = \sum_i p_i{}^j x_i{}^j \qquad (3.3)$$

An "average price" $\bar{p}^j$ can then be calculated by dividing y^j by a quantity index $\bar{x}^j$ so that $y^j = \bar{p}^j\bar{x}^j$.[14]

We now move to the top level where the consumer decides whether to shop at the conventional supermarket or at the supercenter outlet. We expect $y^1 > y^2$ because most prices in supermarkets exceed the prices in supercenters. Consider the use of a binomial choice model for choice between traditional supermarkets and supercenters.[15] We specify the model with a household fixed effect that controls for household characteristics. The probability of household i in choosing the traditional supermarket is:

$$pr(j=1) = pr(U_{i1} > U_{i2}) = \alpha_i + (X_{i1} - X_{i2})\beta + \varepsilon_{i1} - \varepsilon_{i2} \qquad (3.4)$$

where α_i is the household fixed effect, X_{ij} are the attributes of choice j for household i, and ε_{ij} are the stochastic disturbances. Letting the stochastic disturbances be extreme value, we derive the binomial logit model with fixed effects used by Andersen (1973) and Cox (1978):

$$pr(j=1) = \frac{1}{1+\exp(\alpha_i + \beta_1(\bar{p}^1 - \bar{p}^2))} \qquad (3.5)$$

where a log price index or other type of price index (e.g. a Stone price index) can also be used depending on the precise form of the underlying expenditure (utility) and demand functions in equation (3.1) and (3.2).[16]

---

demand to zero. See Hausman (1997) for an explanation of virtual prices.

[14] Instead of the average price we can also divide expenditure by utility to get a "cost of utils" index.

[15] Because of only two choices, the independence of irrelevant alternative assumption does not create a problem here. With more than two choices a nested logit or multinomial probit model could be used. See Hausman et. al. (1995) for a derivation with the nested logit model.

[16] An exact aggregation approach when using a Gorman generalized polar form appears in Hausman et. al. (1995).

IV. Data Description

This study uses a customized subset of the ACNielsen Homescan scanner panel data for the four years 1998-2001. The ACNielsen Homescan data is a consumer panel consisting of approximately 61,500 randomly selected households across the U.S. and includes purchase as well as demographic information for all households in the sample. Homescan households are randomly recruited to join the panel using sampling techniques to ensure household representation for demographic variables such as household income, family composition, education, and household location. Each household is equipped with an electronic home-scanning unit, and household members record every UPC-coded food purchase they make by scanning in the UPC of the food products that they buy from all retail outlets that sell food for home consumption.

The panel is recruited on a permanent basis, subject to turnover from normal attrition or adjustments to demographic targets necessitated by Census revisions.[17] The Homescan panel is considered by many in the food industry as the most reliable household based panel data due to its long-standing reputation in the marketplace and its utilization of hand-held technology that minimizes the recording burden for participants. The ACNielsen Homescan consumer panel collects consumer shopping and purchase data from all outlet channels, including grocery, drug, mass and convenience stores. The panel is geographically dispersed and is demographically balanced so the sample profile matches the US population as closely as possible. The panel data is also projected to census estimates that are updated regularly to reflect population changes.

Household panel data allows for observation of the ongoing purchase habits and practices of household and demographic groups. Tracking and analyzing this information over time can reveal the dynamics of consumer behavior such as who is buying what products, what different products are purchased during a given shopping trip, and how often a product is purchased. Panel data quantifies the composition of category or brand volume which can be used to measure the impact of store choice on the purchase level of product quantities and prices. Data are collected after each panelist shopping trip. Members of the panel record their purchases, capturing not only what is

[17] Households lost through attrition are replaced with others having similar key characteristics.

purchased, but also where the purchase was made, and whether the purchase was a promotional, sale, or coupon item.

These data are useful in price analysis since we are able to observe actual purchase choices by consumers. However, in terms of food purchase behavior, the key missing information is consumer purchases of food away from home (primarily restaurant meals) so one needs to assume that the unknown levels of food away from home purchases do not somehow bias the average prices paid by an individual household for their food at home purchases. Once this assumption is made these data are useful for analysis of the impact of store choice on average prices paid for food at home items. Consumer panel information can be used to measure the average prices paid by a representative group of households over time. This measurement of average price paid can be aggregated across households and/or across time to measure price change for different categories of products.

Along with the description of each product, the price and quantity that was purchased is recorded on a daily basis. National and regional level aggregates can be calculated using transaction data from households located in 50 local U.S. markets as well as households in non-metro/rural areas that are included in this data set. For 21 of these 50 markets, a large enough number of panelists are included to enable comparisons across markets for all UPC-coded products. [18]

The Economic Research Service (ERS) of the USDA purchased a sub-sample of transaction level data from the Fresh Foods Homescan Panel[19] comprised of households that not only recorded their UPC-coded transactions, but also recorded their random-weight (non-UPC coded) food purchases over the year(s) that they participated in the panel. This sub-sample was used for this study in order to be able measure the entire market basket of household purchases of food for at-home consumption[20]. Of this group

[18] Albany, Atlanta, Baltimore, Birmingham, Boston, Buffalo-Rochester, Charlotte, Chicago, Cincinnati, Cleveland, Columbus, Dallas, Denver, Des Moines, Detroit, Grand Rapids, Hartford-New Haven, Houston, Indianapolis, Jacksonville, Kansas City, Little Rock, Los Angeles, Louisville, Memphis, Miami, Milwaukee, Minneapolis, Nashville, New Orleans-Mobile, New York, Oklahoma City-Tulsa, Omaha, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland, Raleigh-Durham, Richmond, Sacramento, Salt Lake City, San Antonio, San Diego, San Francisco, Seattle, St. Louis, Syracuse, Tampa, Washington, D.C.

[19] The Fresh Foods Homescan Panel contained 12,000 households in 1998 and 1999 and was expanded to 15,000 households in 2000 and 2001.

[20] If only UPC-coded products were used to measure food-at-home expenditures, many fruit, vegetable, meat, and poultry purchases would not be recorded in the data and food-at-home expenditure shares by store type would not

of 15,000 households per year, the sample was restricted to households that participated in the panel for at least 10 out of 12 months per year[21].

Standard demographic information is collected on an annual basis from each household and each household's home market/city and census region is identified for stratification purposes (see below). Each household is then assigned a projection factor (weight) based on its demographics in order to aggregate the data to be representative at the market, regional, and national level.[22]

These data were constructed based on a stratified random sample with households as the primary sampling unit. A stratified random sample is used to ensure that the sample of households matches Census-based demographic and geographic targets. One function of the design is to allow description of 8 major markets for cross-market comparisons.[23]

The strata for 1998 and 1999 are based on six cities (ACNielsen major markets) Atlanta, Baltimore/Washington, Chicago, Los Angeles, New York, and San Antonio. All other households fall into one of four census regions: East, Central, South, and West.

1998-1999

Stratum	Description
1	Atlanta
2	Baltimore-Washington
3	Chicago
4	Los Angeles

accurately measure true household and market expenditure shares. This is especially true in this situation when alternative channel stores sell less random weight items than conventional retailers. Leaving out random weight items would then tend to overstate the shares of food expenditures of alternative retail outlets.

[21] In total, there were 9,501 unique households in the data with some subset participating each year creating a total of 28,996 household by year observations. In 1998 there were 7,624 households, 7,124 households in 1999, 7,523 households in 2000, and 8,216 households in 2001. Some households participated in the panel for more than one year. Of the 9,501 households in the data, 5,247 households participated for all four years, 1,877 households participated for three years, and 2,377 households were one year participants.

[22] Age, gender, education, occupation, of head(s) of household, number of household members, household income, household composition, race, and ethnicity.

[23] Atlanta, Baltimore/Washington, Chicago, Los Angeles, New York, Philadelphia, San Antonio, San Francisco.

5	New York
6	San Antonio

For all other households- Census Regions are used as strata:

7	East
8	Central
9	South
10	West

ACNielsen augmented their stratification scheme in 2000, selecting 2 additional major markets.

2000-2001

Stratum	Description
1	Atlanta
2	Baltimore-Washington
3	Chicago
4	Los Angeles
5	New York City
6	Philadelphia
7	San Antonio
8	San Francisco

For all other households- Census Regions are used as strata:

9	East
10	Central
11	South
12	West

There was no known or intentional clustering in the sample construction. The projection factor (weight) reflects the sample design and demographic distribution within the strata.

The information that is captured on a transaction level basis includes: date of purchase, store name and channel type identifier[24], store department identifier[25], item

[24] Grocery, Drug, Mass Merchandiser, Supercenter, Club, Convenience, Other (including dollar stores, bakeries,

description, brand name, number of units purchased, price paid, promotions/sales/coupons used (if any). For retail stores that ACNielsen tracks with their store-level scanner data[26], prices are verified through store-level price and promotion checks.

Warehouse shipment data are used to supplement scanner-generated data collected from households or provided to ACNielsen through their store-level scanner data. Warehouse shipment data is used to estimate the balance of sales moving through other food retailers. This information is Census data (i.e., non-projected, actual shipment data) supplied to ACNielsen by wholesale co-operators.

Some question the quality of household panel data when they try to reconcile it with store-level scanner data. There is the perception that the volumetric data from each source should be the same. However, panel data and store data are not always equal because measurement methodologies differ. Store-level data records millions of shopping transactions while panel data records a specific group of shoppers. In addition, panel data only represents household-based purchases, so there are no small businesses or other institutional purchases included in the panel.

Both types of information have their uses, and by combining the two, one can quantify the composition of volume, understand the reasons behind consumer behavior changes, and measure the impact of store choice on average prices. Store-level scanning data may show that sales were down in a particular store for some group of products in a given time period. Panel data provide insight into whether the lost volume is due to fewer buyers or if the existing buyers purchased less at the given store or chain of stores. Panel data also provide information on which competitors gained the lost expenditures of the store in question.

V. Effects on Prices

Our empirical approach first investigates the effect of supercenters, mass merchandisers, and club stores, (hereafter SMC) on prices paid by households. Two

---

military stores, online purchases, health food stores, and vending machines)

[25] Dry Grocery, Dairy, Frozen-Produce-Meat, Random Weight.

[26] The ACNielsen store-level sample is updated through both replacement of canceled or closed stores and *Continuous Sample Improvement Program* -- when the sample is changed intentionally to ensure that changes in the

effects are present. The direct effect is that as more of these superstores operate in a given geographic market, the average prices paid by households will decrease. Prices for food categories in superstores are typically 5%-48% less than prices for the same product in supermarkets and other conventional retail outlets. Thus, as a high proportion of households buy their food at non-traditional retail outlets, the average price paid in a market will decrease.

A. Price Difference between Supermarkets and Superstores

In Table 5.1 we calculate the ratios of average prices across different types of outlets for 20 food categories. Column 2 compares the prices for the food categories in traditional supermarkets compared to prices for these same categories in SMCs (non-traditional stores).

Table 5.1: Ratio of Supermarket and Other Outlet Prices to Superstore Prices

Product	Supermarkets/SMC	All Other/SMC
Apples	1.546	1.531
Apple Juice	1.585	1.596
Bananas	1.384	1.368
Bread	1.108	1.098
Butter/Margarine	1.096	1.096
Cereal	1.172	1.166
Chicken Breast	1.408	1.411
Coffee	1.373	1.383
Cookies	1.223	1.214
Eggs	1.312	1.305
Ground Beef	1.372	1.367
Ham	1.967	1.984
Ice Cream	1.320	1.331
Lettuce	2.117	2.107
Milk	1.207	1.199
Potatoes	1.412	1.402
Soda	0.891	0.974
Tomatoes	1.358	1.321
Bottled Water	1.058	1.165
Yogurt	1.413	1.411
Average	1.300	1.306

universe are reflected in the sample.

The largest difference in average price was for lettuce where SMC prices were about 50% lower than traditional supermarkets over the 48 month period. Bottled water was the lowest price difference with SMC prices about 5% less expensive. Soda was the only item with a lower price in traditional supermarkets than in SMCs. When we take an average across all of the food categories we find that SMCs have prices that are 27% lower than traditional supermarkets. We find this difference to be quite large.[27]

In column 3 of Table 5.1 we compare the price in all non-SMC outlets, including traditional supermarkets, to the price of these food categories in superstores. We find the results to be quite similar with the main differences occurring in soda and bottled water. We find the same overall results that SMC stores offer significantly lower prices than other retail outlets.

We do not find any indication that SMC stores change (increase) their prices at a greater or lower rate than traditional supermarkets and other retail outlets. However, we cannot do the comparison of price changes in equilibrium because as the presence of SMC stores increases, traditional retail outlets, and most importantly traditional supermarkets, decrease their prices as a competitive response.

B. Direct and Indirect Effects on Prices from Superstores

Another important effect exists from the expansion of SMC stores. Their increasing presence also increases competition among traditional food retailers. These supermarkets must decrease prices to remain competitive. The well-publicized strike in the Los Angeles area in late 2003 through early 2004 when traditional supermarkets wanted to decrease health benefit for their employees demonstrates the effect that potential entry of supercenters can have on competition. We call this SMC effect on traditional supermarkets the indirect price effect. The indirect price effect is consistent with a Nash-Bertrand model of differentiated shopping outlet with increased competition arising from entry or expansion of SMCs. Both the direct and indirect price effects lead to lower average prices for households.

[27] The estimated difference is in line with stock analyst reports who have previously sampled the difference in prices over a very few markets.

To investigate both the direct and indirect effects on average prices, we do an econometric analysis using the ACNielsen Homescan data. These data are particularly useful since they provide household data and allow for a stratified random sample of all households. Importantly they provide both price and quantity data across all stores. Since Wal-Mart and some other large superstores no longer participate in the IRI or ACNielsen store level data collection, household data collection provide a source of price and quantity data that are not available elsewhere.[28]

We analyze data at the market level using a fixed effects specification with 48 monthly observations for each market during the period 1998-2001:

$$p_{it} = \alpha_i + \delta_t + \beta e_{it} + \varepsilon_{it} \qquad i = 1, 34 \quad t = 1, 48 \tag{5.1}$$

where p_{it} is the average log price paid for a given product, α_i is a fixed effect for a market, δ_t is a monthly fixed effect, e_{it} is percentage expenditure for a given product in superstores, and β is thee elasticity coefficient that wee estimate. We use market fixed effects rather than random effects because expenditure in SMC stores is unlikely to be uncorrelated with the stochastic disturbance, e.g. Hausman (1978). In this situation a fixed effects estimator yields the efficient estimator. However, we make two further econometric adjustments. First, expenditure in superstores on a given product may well not be econometrically pre-determined. Thus, we use instrumental variable estimation (2SLS) where as the instrument we use the overall proportion of food expenditure in SMC stores in a given market as the instrumental variable. Also, we use an autoregressive model for the stochastic disturbance (AR1) to capture the time series aspect of the data and to achieve more efficient estimates. However, least squares with robust standard errors leads to quite similar results.

For our econometric investigation of 20 food products we use 34 markets, each with over 12,000 food transactions per year. The 34 markets are listed in Table 5.1:

28 Since the SMC stores continue to collect the data for internal use, we believe that if the BLS were to begin to use scanner data these large stores would likely allow use of their data under confidentiality restrictions. However, the household data provide an adequate source of information for most uses.

Table 5.1: Markets Used in Econometric Analysis

BOSTON	DENVER	HARTFORD-NEW HAVEN
CHICAGO	DETROIT	PHOENIX
HOUSTON	MIAMI	SALT LAKE CITY
INDIANAPOLIS	MILWAUKEE	COLUMBUS
KANSAS CITY	MINNEAPOLIS	CHARLOTTE
LOS ANGELES	PHILADELPHIA	DES MOINES
NEW YORK	PITTSBURGH	GRAND RAPIDS
SAN FRANCISCO	PORTLAND, OR	OMAHA
SEATTLE	ST. LOUIS	SAN ANTONIO
ATLANTA	TAMPA	SYRACUSE
CINCINNATI	BALTIMORE	
CLEVELAND	BUFFALO-ROCHESTER	

For each of these markets we standardized purchases on a physical unit measure and estimated the effect of increasing purchases in SMC stores. Since we have fixed effects for each market, persistent cost and price differences should be take account of as well as seasonal effects given the presence of monthly fixed effects. We give the econometric estimates for these 20 food categories across the 34 markets in Table 5.2:

Table 5.2: Average Price for Food Products across 34 Markets

National Results

AR(1) IV Results

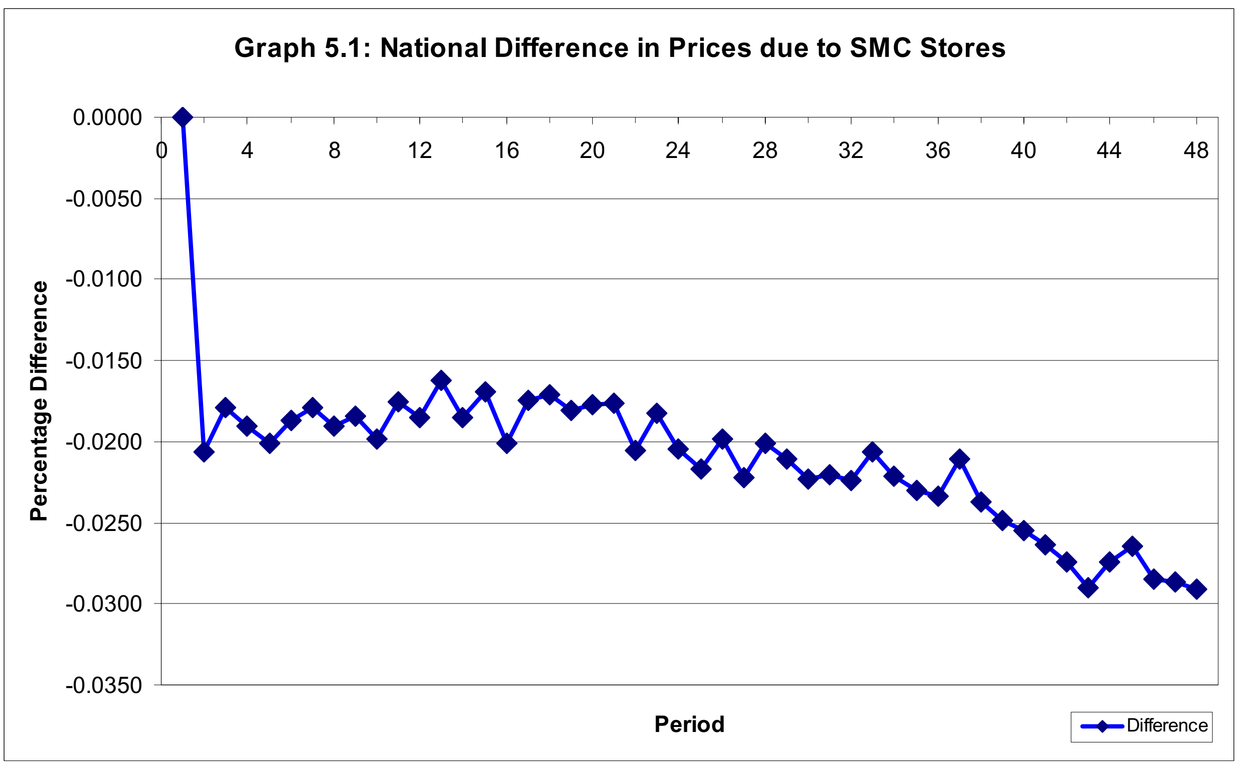
(Asymptotic Standard Errors)

Product	**All Stores**
Apples	-0.1036
	(0.2298)
Apple Juice	-0.2769
	(0.3799)
Bananas	-0.1545
	(0.1747)
Bread	-0.0642
	(0.0898)
Butter/Margarine	-0.8192
	(0.2445)
Cereal	-0.1079

	(0.1275)
Chicken Breast	-0.5597
	(0.4402)
Coffee	-0.6548
	(0.4774)
Cookies	-0.4850
	(0.1294)
Eggs	-0.4324
	(0.0995)
Ground Beef	-0.0679
	(0.1637)
Ham	-1.3032
	(0.7580)
Ice Cream	-0.3516
	(0.3053)
Lettuce	-1.6194
	(1.0106)
Milk	-0.2411
	(0.0748)
Potatoes	-0.6406
	(0.2346)
Soda	-0.3756
	(0.1489)
Tomatoes	-0.8157
	(0.4942)
Bottled Water	-0.7231
	(0.9446)
Yogurt	-0.1832
	(0.1635)

All of the estimated elasticity coefficients are negative as expected. Thus as households spend increasing amounts of expenditure at SMCs, the average prices paid for food items decrease. While the effects are estimated with varying amount of precision, overall the results are highly significantly different from zero. No obvious pattern of coefficient size seems to exist: we find the largest effects for ham, lettuce, butter/margarine, tomatoes, potatoes, and coffee, which are a mix of branded and unbranded products. Yet, we find relatively small effects for ground beef, apples and bananas, which are typically unbranded products, but we also find relatively small effects for cereal and yogurt, which typically are branded products. Overall, we find a statistically negative effect on average prices as shopping in superstores increases. Thus, we find the "direct effect" operates as household shift their expenditure from traditional supermarkets to lower priced superstore outlets.

In Graph 5.1 we depict the difference in average prices paid by households due to the spread of SMC stores over the period. During the sample period from January 1998 to December 2001 the expenditure share of SMC stores increased from 10.9% to 16.9%, a 55.3% increase over the 48 months or 11.6% per year. We take the econometric estimates from Table 5.2 and use them to estimate the decrease in average price for each food category. We then average across food categories and plot the results in Graph 5.1, which demonstrates the increasing effect on average food prices as SMCs become more available and households increase their expenditures at these retail outlets. We find that food prices are 3.0% lower than otherwise, or an effect of about 0.75% year.



We now repeat the econometrics to test for the "indirect effect" of lower conventional supermarket prices because of increased competition from superstores. In equation (5.1) we replace the left-hand variable p_{it}, which is the average log price paid for a given product, with $\tilde{p}_{it}$, which is the average price paid in supermarkets. We give the results in Table 5.3:

Table 5.3: Average Price for Food Products in Supermarkets across 34 Markets

National Results for Supermarkets

AR(1) IV Results

(Asymptotic Standard Errors)

Product	Supermarkets
Apples	-0.2307
	(0.2233)
Apple Juice	-0.5385
	(0.5104)
Bananas	-0.0437
	(0.1447)
Bread	0.0066
	(0.0890)
Butter/Margarine	-0.6853
	(0.2089)
Cereal	0.0832
	(0.1538)
Chicken Breast	-0.5812
	(0.5352)
Coffee	-0.4763
	(0.6005)
Cookies	-0.4366
	(0.1966)
Eggs	-0.1915
	(0.0922)
Ground Beef	-0.0303
	(0.1538)
Ham	-2.1172
	(1.2448)
Ice Cream	-0.3985
	(0.2895)
Lettuce	-2.4217
	(1.5517)
Milk	-0.1247
	(0.0887)

Potatoes	-0.5092
	(0.2244)
Soda	-0.2728
	(0.1513)
Tomatoes	-0.6956
	(0.4791)
Bottled Water	-0.5950
	(0.8155)
Yogurt	-0.0759
	(0.1833)

We estimate 18 of the 20 coefficients to be negative, with the only exceptions being bread and cereal, neither of which is statistically significant.[29] As would be expected from economic theory, the effects of increased SMC expenditures are smaller for most of the products. Thus, the "direct effect" on average prices paid by household arising from substitution to lower priced SMCs typically exceeds the "indirect effect" of decreased prices in supermarkets. Nevertheless, we do find some quite large indirect effects as in lettuce, butter/margarine, coffee, ice cream, potatoes, tomatoes, and bottled water. The spread of supercenters leads to lower prices both for households that shift their food shopping from supermarket to SMC stores but also for households who continue to shop at supermarkets because of lower prices caused by the increased competition from expanding food offerings at SMCs.

In terms of one of the questions we posed at the beginning of the paper, the spread of supercenters does significantly affect prices paid by households. Holding prices fixed as households shift their expenditures to non-traditional retail outlets, we find the average prices they pay decrease. However, prices also change because as households shift their purchasing behavior, the increased competition forces supermarkets to lower their prices. Both of these effects, the direct effect and indirect effect, lead to lower average prices paid by households for food items.

VI. Effects on Consumer Welfare

[29] We find very similar results if we group the remaining Nielsen categories with supermarket: drug stores, convenience, and "other". These other outlet categories have relatively low expenditure levels compared to traditional supermarkets.

We now use the binomial choice model of Section III to estimate the changes in consumer welfare that arise from the entry and expansion of supercenters. We use the "virtual price approach" of Hausman (1997) and Hausman and Leonard (2002). The total effect on consumers of the introduction of a new shopping outlet, i.e., the compensating variation, can be written as the difference in the consumers' expenditure function before and after the introduction, holding utility constant at the post-introduction level:

$$CV = e(p_1, p_N, r, u_1) - e(p_0, p_N^*(p_0), r, u_1) \tag{6.1}$$

where p_1 is the price index of post-introduction prices of the competing shopping outlets, p_N is the post-introduction price of the new supercenter, r is a vector of prices of products outside the market (which are assumed to be unaffected by the introduction), and u_1 is the post-introduction utility level. The pre-introduction utility level could also be used which would yield an equivalent variation measure. The function $p_N^*(p_0)$ defines the "virtual" price index for the new outlet, i.e., the reservation price at which demand for the new outlet would be zero given the prices of the other products.

This total benefit to consumers can be broken into two parts,

$$\begin{aligned} CV = & \left[e(p_1, p_N, r, u_1) - e(p_1, p_N^*(p_1), r, u_1) \right] + \\ & \left[e(p_1, p_N^*(p_1), r, u_1) - e(p_0, p_N^*(p_1), r, u_1) \right] \end{aligned} \tag{6.2}$$

and written as CV = VE + IPE. The first term ("VE"—the "variety effect") represents the increase in consumer welfare due to the availability of the new outlet, holding the prices of the existing brands constant at their post-introduction level.

The second term ("IPE"--the "indirect price effect") represents the change in consumer welfare due to the change in the prices of existing outlets after the introduction. By changing the competitive structure of the industry, the new outlet introduction will lead to a decrease in the prices at existing outlets. The more closely the new outlet competes with the existing outlets, the greater the downward effect on prices. Thus, in addition to providing additional variety, the introduction of a new outlet can change consumer welfare through an effect on the prices of existing outlets.

We now estimate the fixed effects binomial logit model of equation (3.5). Given that we have multiple monthly observations across households, we can estimate a fixed effects model. We find the fixed effects to be quite important in explaining household shopping behavior. We also find that fixed effects are required or otherwise a random effects (or regular) binomial choice model fails a Hausman (1978) specification test because unobserved household components are correlated with shopping behavior. We estimate the coefficient of the log of the price index ratio in equation (3.5) to be -.040 with an asymptotic standard error of (0.014). We also include the market ratio of supercenter expenditure as another variable (which may be jointly endogenous), but we find our estimate of β to remain nearly the same with an estimate of -.035 (0.014), which is not statistically different from our first estimate. We find a coefficient of log income to be estimated at 0.024 (0.010).[30] Thus, we find both the price index and income to significantly effect outlet shopping choice.

We now use these parameter estimates to estimate the gain in consumer welfare from the entry and expansion of Wal-Mart and other supercenters. Because food expenditure is a significant amount of total expenditure, about 12%, and we find that income plays a role in destination choice we estimate the exact competing variation of equations (6.1) and (6.2) using the approach and numerical methods developed in Hausman (1981), Hausman and Newey (1995), and Small and Rosen (1981). These methods integrate the compensated choice function of equation (3.5), thus evaluating the expenditure function, from the current price to the "virtual" price at which demand for shopping at supercenters would be zero.

We estimate the compensating variation for each household for each period and each month and aggregate over each market. Average food expenditure across markets and periods is approximately $151 per month. The average variety affect across market and period is substantial at an estimated exact compensating variation 20.2% of food expenditure. Thus consumers benefit from the availability of supercenters. The effect

[30] Note that this coefficient is estimated from changes in household income over the sample period. It may suffer from an errors in variables problem because it is self-reported. For a further discussion of errors in variables in panel

varies significantly across markets and periods as supercenters enter and expand during our sample period: the minimum estimated compensating variation from the variety effect is estimated to be 9.6% of household expenditure on food and the maximum compensating variation is estimated to be 32.7% of household expenditure on food.

We now estimate the exact compensating variation from the indirect price effect. Note that this estimate will be less than the amount that traditional outlets decrease their prices when supercenters enter or expand in a market, because only savings on actual expenditures are estimated. We estimate the average exact compensating variation from the indirect price effect to be 4.8% of food expenditure across markets and time periods. Again we find significant variation across markets and across time with the range of our estimates 1.3% to 7.3%.

We now add the two sets of estimates together using equations (6.1) and (6.2) and the fact that our approach uses integrable estimates of compensating variation since we have used the compensated demand functions. We find the overall average increase in the ratio of exact compensating variation to overall food expenditure to be 25.0%. Considering the minimum and maximum estimates over markets and time periods we find an estimated range of 14.9% to 36.6% of household expenditure. We give the results in Table 6.1:

Table 6.1: Exact Percentage Compensating Variation Estimates

	Average	Minimum	Maximum
Variety Effect	20.2%	9.6%	32.7%
Indirect Price Effect	4.8%	1.3%	7.3%
Total Effect	25.0%	14.9%	36.6%

Thus, we find a substantial and significant effect of increased compensating variation for households who have the choice to shop at supercenters. A direct effect arises from the lower prices at supercenters and an indirect effect arises from the competitive effect on

data see Griliches and Hausman (1986).

traditional food outlets. Our estimates take account of price and service differentials across stores as well as household attributes.

VII. Conclusion

Over the past 15 years the largest development in food retailing has been the introduction of Wal-Mart supercenters that compete most closely with traditional supermarkets. Wal-Mart has expanded greatly, mostly in the South and Southwest, and become the largest supermarket chain in the U.S. Wal-Mart is now expanding into additional geographic markets in California and the upper Midwest, so its effects will become even more important.[31] Wal-Mart offers many identical food items at an average price about 15%-25% lower than traditional supermarkets. Wal-Mart's entry into a new geographic market creates a direct price effect by offering a lower price option to consumers and an indirect price effect by causing traditional supermarkets to lower their prices because of the increased competition. This paper estimates the effect on consumer welfare of the entry and expansion of Wal-Mart and other supercenters into geographic markets.

We find that an appropriate approach to the analysis is to let the choice to shop at Wal-Mart be considered as a "new good" to consumers when Wal-Mart enters a geographic market. Some consumers continue to shop at traditional supermarkets while other consumers choose to shop at Wal-Mart. Many consumers shop at both types of stores. Thus, we specify a utility consistent two level model of choice among types of shopping destinations. We then estimate a fixed effects binomial logit choice model to estimate the parameters of the utility model that differs across households. We use the estimated parameters to calculate the exact compensating variation that arises from the direct variety effect of the entry and expansion of supercenters and find the average estimate to be 20.2% of average food expenditure. We similarly estimate the exact compensating variation from the indirect price effect that arises from the increased competition that supercenters create. We find this average effect to be 4.8%. Thus, we estimate the average effect of the total the compensating variation to be 25% of food expenditure, a sizeable estimate.

[31] Wal-Mart has announced plans to open 40 supercenters in California in the next 3-5 years, Wiltamuth op. cit.

Since we find that lower income households tend to shop more at these low priced outlets, a significant decrease in consumer surplus arises from zoning regulations and pressure group tactics that restrict the entry and expansion of supercenters into particular geographic markets.

References

Basker, E., “Selling a Cheaper Mousetrap: Entry and Competition in the Retail Sector,” mimeo, 2004

Cage, R., “New Methodology for Selecting CPI Outlet Samples,” Monthly Labor Review, December 1996, p. 49

Callahan, P. and A. Zimmerman, “Grocery Chains Fighting Wal-Mart for Market Share,” Wall Street Journal, May 31, 2003.

Griliches Z. and J. Hausman, "Errors in Variables in Panel Data," Journal of Econometrics, 1986

Hausman, J., "Specification Tests in Econometrics," Econometrica, 46, 1978

Hausman, J., “Exact Consumers Surplus and Deadweight Loss,” American Economic Review, 71, 1981

Hausman, J., "The Econometrics of Nonlinear Budget Sets," Econometrica, 53, 1985

Hausman, J., "Valuation of New Goods Under Perfect and Imperfect Competition," ed. T. Bresnahan and R. Gordon, The Economics of New Goods, University of Chicago Press, 1997a.

Hausman, J., "Valuation and the Effect of Regulation on New Services in Telecommunications," Brookings Papers on Economic Activity: Microeconomics, 1997b.

Hausman, J., "Cellular Telephone, New Products and the CPI," Journal of Business and Economics Statistics, 1999.

Hausman J., “Sources of Bias and Solutions to Bias in the CPI”, Journal of Economic Perspectives, 2003

Hausman J. and E. Leibtag, CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?, NBER discussion paper, September 2004.

Hausman, J. and G. Leonard, “The Competitive Effects of a New Product Introduction: A Case Study,” with G. Leonard, Journal of Industrial Economics, 50, 2002

Hausman, J., G. Leonard, and D. McFadden, "A Utility-Consistent Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses Due to Natural Resource Damage," Journal of Public Economics, 56, 1995.

Hausman J. and W. Newey, "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," Econometrica, 63, 1995

Little, P., “Channel Blurring Redefines the Grocery Market,” Competitive Edge, June 2004.

Small, K and H. Rosen, “Applied Welfare Economics with Discrete Choice Models.” Econometrica 49, 1981.