WORKING PAPER #423
INDUSTRIAL RELATIONS SECTION
PRINCETON UNIVERSITY
SEPTEMBER, 1999
REVISED FEBRUARY, 2000
http://www.irs.princeton.edu/pubs/working_papers.html

Are Prices Higher For the Poor in New York City?

Lashawn Richburg Hayes
Department of Economics
Princeton University
Princeton, New Jersey 08544
Tel: 609-258-2363
Email: richburg@princeton.edu

Are Prices Higher For the Poor in New York City?

Lashawn Richburg Hayes
Department of Economics
Princeton University
Princeton, New Jersey 08544
Tel: 609-258-2363
Email: richburg@princeton.edu

ABSTRACT

Despite earlier evidence to the contrary, recent inquiries appear to reach a consensus that the poor pay more for food. However, these studies utilize samples drawn on the basis of prior knowledge of unfair pricing strategies, proximity of volunteer surveyors, or other non-random methods. This paper revisits the issue of price discrimination by analyzing price data collected using a stratified, random sample design to answer the question of whether prices are higher in poor, urban neighborhoods. Contrary to the recent literature, I find that market prices in poor neighborhoods are not higher than those in more affluent areas.

KEYWORDS: Imperfect competition; Price differential; Supermarket; Urban poor.

JEL classification: D4; J7; L1

A current public policy debate in cities such as New York, Los Angeles, and Detroit is whether the poor pay more for food.¹ Although this debate has smoldered for years, it has not received as much attention in the academic literature.² One reason for this lack of research may be the comparative magnitudes: although food expenditures exhaust 13 percent of the average income, individual transaction expenditures are small, unlike the magnitudes involved in purchasing “big ticket” items such as shelter or an automobile.³ Nevertheless, it is important for policy makers to know both the level and spread of food prices charged to different consumer types because the equity of income transfers such as food stamps, Temporary Aid to Needy Families (TANF), and other governmental transfer programs may be affected. For example, if prices are higher in low-income neighborhoods then the allocation of certain transfers may not be optimal and the real value of such transfers may be spatially eroded.⁴

Market prices for food may be different in low-income neighborhoods as compared to more affluent neighborhoods for a number of reasons. First, marginal costs of operation may vary with neighborhood composition. For example, delivery costs may be higher in more congested neighborhoods or greater inventory pilferage may require increased insurance coverage. Second, fixed costs of operation (e.g., commercial rental rates) may vary.

¹ The discrimination literature addresses related issues in commodities such as housing [Yinger 1986], lending [Munnell, *et al* 1996], and automobiles [Ayres 1991; Ayres and Siegelman 1995; Goldberg 1996]. See the *Journal of Economic Perspectives* [1998] symposium on discrimination for a review. Pratt, *et al* [1979] provides evidence of price dispersion across different products and Sheppard [1996] provides evidence of price discrimination among gasoline retailers.

² See Kinsey [1998] for a review of the concentration and price relationship in the food industry. One exception is Graddy [1997], which analyzes dispersion in fast food prices. Graddy [1997] finds that prices are higher in predominantly minority neighborhoods.

³ Consumer Expenditure Survey, Standard table 1982-1997.

⁴ Glaeser [1998] posits that some indexing of transfers to local price levels may be welfare improving when the transfer population is immobile or highly risk adverse. Kaplow [1995] suggests efficiency increases when transfers are designed in such a way as to induce the poor to reside in low-cost areas.

Alternatively, stores operating in low-income neighborhoods may enjoy monopoly rents.⁵ In addition, there may be a different mix of outlets available in low-income neighborhoods. Outlets may limit offerings to small sizes or lower quality, or limit payment options that result in unit or quality-adjusted prices being higher [Alwitt and Donley 1996]. Finally, stores located in poor neighborhoods may employ a different mix of price discriminatory devices. For example, inner-city stores may limit advertising and couponing expenditures.⁶ These can be thought of as ways of increasing consumer knowledge of the price distribution.

The empirical literature concerning price dispersion in food retailing reports mixed results.⁷ Groom [1966] and Alcay and Klevorick [1971] find little difference between the prices charged in low and upper-income neighborhoods. In contrast, Kunreuther [1973] reports that residents in low-income neighborhoods pay more, while Ambrose [1979] reports that the poor pay less.⁸ More recent studies tend to conclude that residents in low-income neighborhoods face prices that are significantly higher than those faced by residents of upper-income neighborhoods [see e.g., Green 1991 and Finke, *et al* 1997].⁹

⁵ Darden and Lumpkin [1984] find that “out-shoppers,” consumers who shop outside of their neighborhood, are typically more affluent.

⁶ See e.g., Salop and Stiglitz [1977] and Butters [1977] for models relating prices to advertising.

⁷ Economic theory alone provides no clear answer as to whether the poor pay more for goods. The neoclassical model under perfect competition and the Bertrand duopoly model predict that firm prices should be the same in equilibrium (and equal to the marginal cost of the homogenous good) regardless of geographical location. However, the Bertrand model with product differentiation and Hotelling’s spatial-monopoly model can generate higher prices in poor neighborhoods. (For other models that predict price differences see Burdett and Judd [1983]; Rothschild [1974]; Salop and Stiglitz [1977]; and Varian [1980].) In addition, there are models of price discrimination and bargaining in which a difference in mean prices paid by different groups may not exist, yet price discrimination occurs through differences in the variance of the price distribution [Goldberg 1996]. Therefore, this is an empirical question.

⁸ See Sexton [1971] for a review and critique of the early literature on price dispersion in food retailing.

⁹ MacDonald and Nelson [1991] find that the poor pay more when they compare mean basket prices across central and non-central cities without controlling for other covariates. They find no difference in prices when they control for variation in supplier costs and quality. Troutt [1993] finds an insignificant difference, but concludes that the poor pay more because of the lack of grocery stores in poor neighborhoods.

Although the recent empirical literature appears to reach a consensus that prices are higher in low-income neighborhoods, the studies--with the exception of MacDonald and Nelson [1991]--are based on data collected from stores that are either known to engage in unfair pricing strategies, are close to volunteer surveyors, or other non-random methods.¹⁰ The effect of these informal selection criteria on the results is unknown.

In this paper I investigate whether the poor pay more for food. Ideally to answer this question I would analyze market prices charged by (initially identical) stores randomly distributed across neighborhood types, taking consumer behavior as given.¹¹ Data from such an experiment would ensure that market prices are mapped to the neighborhood.¹² Unfortunately, such data do not exist. As a result I analyze data collected from a stratified, random sample of stores in New York City (NYC), which consists of New York, Bronx, Kings, Queens, Richmond, and Westchester counties. I use the term “inner-city” throughout the paper to refer to poor, urban neighborhoods. In contrast to the recent literature, I find no evidence of a systematic price differential between food retailers located in inner-city and upper-income neighborhoods.

A brief outline of the paper is as follows: section 1 describes the data, section 2 reports summary statistics of the market basket survey, details the empirical specification, and presents results. Section 3 investigates possible explanations for the apparent difference in results from the literature. Section 4 contains concluding remarks.

¹⁰ See e.g., Alcay and Klevorick [1971]; Ambrose [1979]; Groom [1966]; Green [1991]; Kunreuther [1973]; and Troutt [1992]. MacDonald and Nelson [1991] utilize data collected by the Economic Research Service (ERS) of the U.S. Department of Agriculture. While the data are from a random sample, the market basket is not identical across stores. See Geithman and Marion [1993] for a critique of the ERS research design and methodology.

¹¹ Of course, consumer behavior affects prices actually paid. Differences in household characteristics (e.g., size and composition) and purchasing habits (e.g., buying items in bulk) may result in different market baskets. I do not consider consumer behavior issues in this paper because I am interested in market prices. Future analyses should also consider consumer behavior.

¹² Further, changes in outlet characteristics in response to the random location could be observed.

1. DESCRIPTION OF DATA

1.1 Overview

Because there are no applicable data sets readily available to study the question of price differentials in NYC I conducted my own survey.¹³ I utilize the Food and Safety Inspection Report (FSI Report) compiled by the New York State Department of Health as the principal sampling frame for stores in NYC.¹⁴ The FSI Report provides data on the address, ownership, store description, and size of firms operating in New York State that sell or manufacture food. I also use many other directories and computer databases to collect information as detailed in the Appendix.

The compiled database contains stores with at least 4000 square feet of retail space. I truncate at this size in order to isolate supermarkets from convenience stores because price differences between supermarkets are economically relevant: supermarkets account for over 77 percent of the food sales volume in the United States, despite comprising only 24 percent of all food stores.¹⁵ Second, the truncation provides a lower bound of the differential if market prices are higher in low-income areas; the inclusion of smaller stores in the analysis would likely increase the differential.¹⁶ However, the truncation at 4000 square feet is nontrivial as it assumes consumers incur larger transportation costs to manage their moderate to heavy shopping

¹³ Many researchers use data from the Consumer Expenditure Survey (CEX) to study consumption patterns among groups. However, these data are of limited use for my purposes because I am interested in whether stores set prices in response to different demographic groups; the CEX provides expenditures by group, but does not provide information on quantities, prices, or where expenditures are made. Another data source, the Nationwide Food Consumption Survey (NFCS), analyzed by Finke, *et al* [1997], attempts to link the consumer unit to the transaction prices it faces. However, the survey does not provide outlet information.

¹⁴ An alternative sampling strategy would be to use the 1990 Census as the sampling frame to draw a random sample of people or neighborhoods by population-weighted zip codes. This would oversample neighborhoods with monopoly supermarkets. Such a sample would be suitable for testing the monopoly pricing hypothesis discussed in the introduction.

¹⁵ *Progressive Grocer* [1999], p. 10.

requirements and it restricts market competition to occur solely between “moderately-sized” supermarkets. There are 1408 documented stores in NYC satisfying the size requirement.¹⁷

I divided the universe of stores into three strata based upon median household income by zip code. Although zip codes are somewhat arbitrary geographic units, they are the most readily available neighborhood proxies.¹⁸ The poor group consists of zip codes with median household income less than 18,500 dollars, while the upper group consists of zip codes with median income exceeding 31,000 dollars. Stratification ensures an adequate sample of stores operating in low-income areas for analysis. The poor, middle, and upper strata contain 190, 407, and 811 firms, respectively. Appendix Table I tabulates the composition of the universe of stores. See the Appendix for a more extensive analysis of store representation.

1.2 Market Basket Composition and Implementation of Survey

The items that comprise the market basket of goods were selected to represent every class of the Consumer Price Index’s defined “food at home” categories. These categories are (1) cereals and bakery products, (2) meats, poultry, and eggs, (3) dairy products, (4) fruits and vegetables, and (5) other foods including categories of fats and oils, sugars and sweets. I selected items that were representative of a typical consumption basket. For example, I selected tuna and red delicious apples because they are more common than dried codfish and Crispin apples, respectively. This

¹⁶ The inclusion of smaller stores in the analysis would likely increase the differential because of the higher costs of conducting business that smaller stores can reasonably be believed to incur.

¹⁷ Based upon *Progressive Grocer’s 1997 Marketing Guidebook* (1997, p.187) which reports convenience stores and supermarkets combined in NYC total 906, *Supermarket Census* (1996, p. 8) which reports 1,634 supermarkets in NYC, and the numerous sources utilized to compose the database, I believe I have compiled an essentially exhaustive listing.

¹⁸ The supermarket industry typically regards the metropolitan statistical area (MSA) as the relevant market area. Chevalier [1995] and Marion, *et al* [1979] use the MSA as the relevant market area. MSAs are obviously too coarse

selection principle was intended to assure that most firms would carry the market basket. The brand names and sizes of the items were largely chosen to coincide with those used in Green [1991] for purposes of comparison. Brand name is also used to control for the quality of the product.

The survey compared the “every-day” prices of identical items at different stores in accordance to the schedule discussed in the Appendix, where “every-day” refers to the regular sticker or shelf price of the item. If an item was unavailable at a site, no substitutions were made and the regular price of the particular item was recorded as missing. This ensured that the resultant prices of the market basket were not influenced by haphazard substitutions so that they could be compared with confidence. The sale price, if applicable, was recorded for every item as well as the price of the most identical generic item, if available. Analysis of these prices is necessary to capture the effect of different pricing strategies on the existence of a price differential between poor and upper-income residents. A sample of the survey sheet listing the items is provided in Appendix Table II.

I did not survey 13 stores (14 percent) in the sample because they were no longer in business. I was unable to survey another 20 stores (22 percent) within the survey period.¹⁹ The final analysis sample consists of 57 stores located in all parts of NYC, except Staten Island, in 28 zip codes. Figure 1 presents a map of NYC and the survey sites. The boundary lines indicate zip codes, while the shaded region represents NYC.²⁰

Although I was unable to survey a sizeable portion of the original stratified, random sample, the exclusion of these stores should not effect the results I obtain if the omitted sites do not differ

for this analysis. While MacDonald and Nelson [1991] examine price differences in several MSAs, they also use the zip code of the store to proxy the neighborhood.

¹⁹ See Appendix section A. 3.

²⁰ The clustering of stores in the analysis sample is the result of stratifying the stores by the median household income of its zip code. As there are multiple stores in each zip code, the sample tends to be weighted toward those zip codes with more stores.

systematically from the surveyed sites. To determine this I analyze the demography of the neighborhood surrounding the store by regressing the dependent demographic variable on a constant and a dummy variable indicating whether I surveyed the store. Table 1 presents the coefficient and standard error of the indicator variable from this regression. Columns (1) – (4) of the table compares the 33 excluded stores to the 57 sampled stores for the labeled dependent demographic variable. As a group, the excluded stores differ from the sampled stores. In columns (5) – (10), I omit the 13 stores that were not in business.²¹ The omitted 20 stores do not differ significantly from the 57 stores in the analysis sample.²² As a result I conclude that the surveyed stores are representative of all stores.

All subsequent analysis is weighted in order to account for the sample design in the calculation of standard errors. The weights are the inverse of the ex-ante probability for the strata multiplied by the inverse of the survey response rate for the strata.

1.3 Demographic Characteristics

Table 2 presents descriptive information on the characteristics of consumers in NYC and those served by the firms in the analysis sample. As expected, the population in the inner-city exceeds that in upper-income neighborhoods significantly. There are approximately 50 and 25 percent more people and households, respectively, residing in poor neighborhoods than in upper-income neighborhoods. Although inner-city neighborhoods are roughly 20 percent more dense, population density does not differ significantly between the two groups. The median age of the head

²¹ This exclusion is justified on the basis that the closed stores are ex-ante different from the remaining sample because some factor(s) not widely present induced the stores to go out of business.

²² In addition, stores were surveyed randomly each day of the survey period. This additional random component should mitigate the effects of unobserved differences between the sampled and un-sampled stores.

of the household differs significantly between the poor and upper groups, with the household head being 9 years older on average in the upper group.

The incidence of poverty, according to the 1990 poverty cutoffs, is approximately 4 times higher for the poor group as compared to the upper group. This provides evidence that the poor are suitably portrayed in the analysis sample.²³ The upper group's median income is approximately 10 percent higher than the average median income in New York State of 33,788 dollars in 1995.²⁴ On this basis, the analysis sample firmly captures the middle class in the upper group variable rather than the "rich."

In terms of race and ethnicity, the inner-city neighborhoods of the analysis sample are primarily composed of minorities. African-Americans make up over 40 percent of residents in the inner-city, whereas they are only a tenth of all residents in upper-income areas--a difference that is statistically significant at any level. Asians are more likely to reside in upper-income neighborhoods, as are whites.

The competition variables included in the table are based upon the 237 unique zip codes for NYC and the analysis sample, respectively, that satisfy the 4000+ square footage requirement. On average, there are 9 stores per square mile in a zip code meeting the size requirement in NYC, but slightly more than 19 stores per square mile per zip code in the analysis sample. A joint test of the store concentration variables rejects the hypothesis that the poor and upper groups differ in these dimensions.

²³ The official 1996 poverty cutoff for this family type is 15,911 dollars [U.S. Bureau of the Census 1996]. Citro and Michael [1995] propose a poverty cut-off range of 13,700-15,900 (in 1992) dollars for a two-adult, two-child family.

²⁴ *Sourcebook of Zip Code Demographics 1996*, p. 212-C.

Overall, close to 70 percent of stores in the analysis sample are independently operated and almost half of the stores are equipped with scanners and are capable of processing debit or credit payments for purchases. The average square footage of retail space in the survey sample is slightly more than 16,000. In the poor group, over 80 percent of stores are independently operated as compared to roughly 60 percent of stores operating in upper-income neighborhoods. The average retail space in an inner-city supermarket in the survey sample is slightly greater than 10,000 square feet, with the largest store having less than 13,000 square feet. In contrast, the average retail space of stores operating in upper-income neighborhoods is approximately 18,000, with the smallest store having over 10,000 square feet of retail space. The difference in retail space between the poor and upper groups is significant at the 5 percent level. All other differences in store investment are not significant at conventional levels.

2. RESULTS OF MARKET BASKET SURVEY

2.1 Mean Differences

By regressing the total basket cost (the sum of item prices weighted by the expenditure shares of low-income consumers) on a constant, group indicator variables, and indicator variables signaling missing item prices, I find that the mean cost is 52.20 and 53.04 dollars in poor and upper-income neighborhoods, respectively.²⁵ Paying approximately 2 percent more in affluent neighborhoods costs an average family of four close to 58 dollars more per year in 1996, unadjusted for inflation, using the 1997 family food expenditure of 2,880 dollars. This figure represents roughly 0.2 percent of the average New York state median income. The difference in average costs by inner-city and suburban consumers is not statistically significant at conventional levels, nor economically meaningful. The magnitude and insignificance of this differential is similar to that

²⁵ Low-income expenditure shares are taken from the Consumer Expenditure Survey [1993], Table 43.

reported by MacDonald and Nelson [1991], although the direction of the differential differs. In contrast, Green [1991] finds that inner-city residents pay a significant 5.1 percent more in supermarkets. Possible reasons for the difference in results are Green's [1991] lack of random sampling to select the stores he surveys and little price variability within neighborhood types.²⁶

Figure 2 presents a histogram of the market basket cost showing the location of the low-income stores in the overall distribution. The total market basket cost depicted is the predicted values from a regression of the total basket cost on a constant and indicator variables signaling missing item prices. From the figure, it is readily apparent that there is considerable variation in prices across stores in NYC and that this variation is fairly similar for stores operating in low-income areas and other stores in NYC. This provides evidence that the lower cost attributed to poor-area stores is not the result of the poor stores having a different price distribution than the remaining stores in the analysis sample. To formally test the hypothesis that the low-income and the other stores are from populations with the same distribution, I conduct a Mann-Whitney U-test, which yields a p-value of 0.560. This indicates that the medians of the distributions do not significantly differ and I cannot reject the null hypothesis of a random distribution of low-income stores. I also conduct Mann-Whitney U-statistics by individual item. There is considerable variation from store to store in individual item prices, but the variation is largely unrelated to the location of low-income stores in the distribution.

In an attempt to address the large variation evident in figure 2, I conducted a small-scale reliability test on February 7, 1998 to evaluate whether any of the variation in prices I encountered was due to measurement error.²⁷ The test consisted of the following: 2 volunteers conducted the

²⁶ The low variability in prices may be due to Green's [1991] handling of missing values, which is unclear.

²⁷ For example, the collection of incorrect prices for various goods.

same price survey as described in this paper in 5 randomly selected supermarkets in the morning and I surveyed the same establishments in the evening of the same day, for a total of 10 stores surveyed and 200 prices collected. A regression of the log of prices collected by the volunteers on the log of prices collected by me revealed a coefficient of 0.995, which does not significantly differ from 1. This is evidence the variation is not solely attributable to measurement error.

Table 3a displays the average individual item prices charged in the poor and upper-income neighborhoods of the analysis sample. Columns (1) through (3) of the table report the means by poor, middle, and upper income status, respectively, while column (4) computes the poor and upper-income difference. Column (4) reproduces the above result of a 2 percent difference and shows that the only significant price differences between the groups occur in cereal, orange juice, apples, and bananas. There is no difference in prices for necessities such as milk, bread, and cheese. It is interesting to note that inner-city consumers do not pay more for 80 percent of the goods in the market basket. That said, it is possible the poor pay more for lower quality. I find that the greatest divergence in prices between the poor and upper groups occur in items where quality is difficult to control (e.g., fresh fruits).

Table 3b reports the coefficient on the poor indicator variable from a model which controls for the investment in scanner equipment, the number of checkouts, the ownership structure of the store, the square feet of retail space, and the number of stores per square mile. A joint test of the significance of the store investment variables rejects the hypothesis that the variables are simultaneously equal to zero. The regression-adjusted poor-upper difference in market basket cost is 1.1 percent.²⁸

²⁸ I replicate the model in Table 3b using per capita income (in thousands) in place of the income group variables. As per capita income increases by 1000, the total (regression-adjusted) market basket price increases 0.4 percent. Results available upon request.

2.2 Sales and Generic Prices

Table 4 tabulates the effect of sales and the availability of generic brands on the average cost of the market basket by group. Sale and generic brand pricing can be thought of as strategies that the firm uses to classify its clientele and to compete with its rivals. The average market basket cost under each strategy is derived by regressing the dependent variable on a constant, group indicator variables, and indicator variables for missing items in a similar fashion as in Table 3b. After controlling for pricing strategies, Table 4 shows that consumers shopping in inner-city areas do not pay more. Given that a consumer will search for a sale for every item purchased, that consumer spends approximately 4 percent less shopping in inner-city supermarkets. A consumer who will buy generic brands whenever possible pays roughly 3 percent less in the inner-city for a market basket comprised largely of generic or generic sale products.

2.3 Differences weighted by expenditure-share

The fluctuation in item prices evident in Table 3a may camouflage true price differences if high prices for certain goods are offset by low prices for other items. To address this concern, I derive an aggregate measure that weights the price differences across items by their expenditure shares as determined by the CEX. Using generalized least squares, I estimate the following model:

$$\ln p_{ijz} = \alpha_0 + \sum_{i=2}^{20} \alpha_i D_i + \beta_0 PC_{jz} + \sum_{i=2}^{20} \beta_i D_i PC_{jz} + X_{jz} \Phi' + u_{ijz} \quad (1)$$

where p_{ijz} is the price of item $i=1,...,20$ at firm $j=1,...,57$ located in zip code z ; D_i is an indicator variable for item i , PC_{jz} is the per capita income of the zip code of store j , X_{jz} is a vector of store cost variables, and u_{ijz} is an error term that can be decomposed to $u_{ijz}=v_{jz}+e_{ijz}$, with the usual properties. I

use the inverse of the ex-ante probabilities of selection multiplied by the inverse of the strata response rate as weights in all estimation procedures that pool data across strata to account for the survey design and implementation in the calculation of standard errors. This specification amounts to estimating individual item prices across firms and allows estimation of the variance-covariance structure across items. The difference between poor and rich neighborhoods is:

$$\bar{\beta} = s_1\beta_0 + \sum_{i=2}^{20} s_i(\beta_i + \beta_0) = \beta_0 + \sum_{i=2}^{20} s_i\beta_i \quad (2)$$

where s_i is the relevant food category expenditure share from taken from the CEX and β_0 and β_i refer to the coefficients on the per capita income variable in specification (1).²⁹

Table 5 presents the GLS estimates of specification (1) and reports the weighted-average differences in prices. In all models, the omitted item is cereal and the standard error of the weighted-average is computed utilizing the variance-covariance matrix of the β_i coefficients. The first column controls only for the full interaction of per capita income with the item indicator variables. The weighted-average coefficient of the price differences is not significant. Column (2) controls for the investment in scanning equipment, which may represent a cost-saving investment that reduces cashier and customer theft, efficiently reports inventory, and minimizes data entry errors. The inclusion of the scanner variable does not help explain any of the remaining variation in the logarithm of price. However, addition of the chain variable in column (3) does help; the coefficient on the chain variable is positive and significant at the 1 percent level. This contradicts the pass-through hypothesis of economies of scale attributed to chains [e.g., Alcally and Klevorick 1971]; I find chain stores charge significantly more than independents. The inclusion of this variable decreases the magnitude of β -bar and increases its standard error negligibly. Column (4) looks at the

effect of holding constant scanning equipment, chain status, and square feet of retail space. In this specification, only the coefficient on the chain variable is significant indicating the mode of operation may be an important determinant of a price differential. In the last column I control for the number of checkouts in the store which may proxy for the speed of consumer exit. In general, the price difference is stable across models. The sign verifies the mean results reported in Tables 3a and 3b; the poor do not pay more for food.

3. DISCUSSION OF RESULTS

There are a few potential problems with my sample that may have biased the results. First, the results may stem from the various temporal pricing strategies in place at the time the survey was conducted. If I conducted the survey at another point in time, say a week earlier or later, it is possible I would not have observed the same pattern in either the item prices or the total market basket cost. If this is the case, then the results are unrepresentative.

The best way to evaluate whether the sample is idiosyncratic would be to re-conduct the entire survey. However, that would be time-intensive and expensive. Instead, as discussed above, I surveyed a subsample of stores on February 7, 1998. These data were collected from 10 randomly chosen stores from the original stratified, random ex-ante sample of 90 stores. I can evaluate the robustness of the results in Section 2 by testing for whether the prices in the subsample are significantly different from those in the original sample.³⁰ Such an analysis reveals that the

²⁹ Consumer Expenditure Survey [1993], Table 43.

³⁰ More precisely I estimate the following model:

$$\ln p_{jz} = \alpha_0 + \beta_1 PC_{jz} + \beta_2 R_{dum} + \beta_3 R_{dum} * PC_{jz} + \varepsilon_{jz} \quad (3)$$

where p_{jz} is the cost of the market basket at firm $j=1, \dots, 57$ in zip code z ; PC_{jz} is the per capita income of store j in zip code z ; R_{dum} is an indicator variable that equals 1 if the market basket cost of store j is from the reliability study; and the third term is an interaction of the indicator variable with the per capita income of the neighborhood of store j . The test consists of evaluating the sign and significance of $\beta_1 - \beta_3$, where β_3 is the coefficient on the

difference of the coefficients is not significant at conventional levels.³¹ This provides evidence that the results reported in Section 2 are not likely idiosyncratic.

A second problem may be that the results are due to my small sample size. The power of the survey sample may be too low to detect a significant difference between the mean market basket costs in poor and upper-income neighborhoods. Unfortunately, I would need to conduct a census of the approximately 1408 stores in NYC, which would require more resources than I had available. A useful extension of this paper would be to use a much larger sample.

Finally, section 2 revealed that the upper income group largely represented the middle class and not the most affluent consumers in NYC. To assess the effect of this on my results, I examined a subset of the poorest and richest neighborhoods. My results appear robust to this potential problem as the price distributions for the poor and upper-income groups continue to overlap a great degree.³²

While my estimate of 1.6 percent lower prices in poor neighborhoods differs from Green [1991] and Finke, *et al* [1997], who find prices to be 5.1 and 2.4 percent higher, respectively, in poor neighborhoods, it is not different from the literature as a whole. Groom [1966] and Ambrose [1979] report that the poor pay 2.3 and 2.4 percent less, respectively, while MacDonald and Nelson [1991] and Troutt [1993] report poor neighborhoods have prices that are an insignificant 1.8 and 2.1 percent higher, respectively.

4. CONCLUSION

I estimate that a statistically insignificant 1.6 percent differential exists between the market prices in inner-city neighborhoods and that of their more affluent counterparts, with inner-city

interaction term. The idea of this test is that the level of prices may vary over time, so that β_2 may differ positively or negatively, but the relationship of price and per capita income is relatively stable.

³¹ The results are available from the author on request.

market prices not being higher. However, it is worth noting that the differentials reported in this paper do not take into account the purchasing pattern of consumers nor do they adjust for product quality.³³

There are several interesting extensions to this question. The first would be to expand this analysis from a case study to one encompassing many large MSAs. Such a large, randomized study could provide the answer to the question of whether the poor pay more generally, and how different infrastructures (e.g., availability of transportation, store ownership, operating costs, etc.) may influence the results. Other extensions are the effect on prices faced by inner-city consumers given convenience store competition with supermarkets and the effect of "copy-catting" behavior on the part of firms in terms of investments and pricing strategies (i.e., sales and generics) that could be captured in a spatial autocorrelation context. Still another extension is to analyze those goods purchased by consumers whose income are supplemented by Women, Infants, and Children (WIC) program and food stamps to evaluate the presence of price discrimination among these goods and the corresponding welfare impact, if any.

³² As evidenced by applying the estimation model presented in Table 3b.

³³ During the survey, I observed that the cheapest meat and poultry products were often undesirable to purchase. In addition, the freezers in most inner-city neighborhoods did not appear to be set at an appropriate temperature and some meat and poultry products did not appear fresh. Among items for which quality may not be such an issue (e.g., cereal, oil, pampers, etc.) almost half showed prices were higher in poor neighborhoods as reported in Table 3a. To further evaluate the difference between perishable and nonperishable items, I estimated the model presented in Table 4 separately for nonperishable items. The results are robust to this restriction.

ACKNOWLEDGEMENTS

I would like to thank Cecilia Rouse and David Card for their time and advice while I worked on this paper, and Orley Ashenfelter, Angus Deaton, Pinelopi Goldberg, and Christina Paxson for independent discussions. I would also like to thank Lenience Akins, Linda Dickson, and Nelson Hayes for excellent research assistance and the participants of the Princeton University Labor Lunch for valuable comments. This material is based upon work supported under a National Science Foundation Graduate Fellowship.

APPENDIX: DATA DESCRIPTION

A.1 Data Sources

I utilize the New York State Food and Safety Inspection Report (FSI Report) as the principal source of information for my database. The FSI Report is not exhaustive as store closings and openings are recorded at unpredictable lags and the quality of the data largely depend on the number of inspectors deployed to inspect the stores. I supplemented the FSI Report with *the Directory of Single Unit Supermarket Operators*, the *Directory Supermarket, Grocery and Convenience Store Chains*, the *1996 American Big Business Directory*, the *Million Dollar Directory*, and *SN Retailers and Wholesalers* to obtain information on the presence of private and generic labels, the availability of automatic teller machines, and whether the store is a chain or independent grocer. Private label denotes whether the company carries merchandise with its trade name. Generic label denotes whether the company carries non-trademark merchandise. Demographic data is obtained from the *CensusCD 2.0*.

A.2 Developing the Random Sample

I drew a random sample from the stratified universe of stores by applying a random number generator and sorting observations in each of the strata based on the random component. I selected 90 observations such that each record had a one-third probability of being in the poor, middle, or upper stratum. The size of this ex-ante sample was determined by applying the formula for a two-sample test of the equality of means (assuming normality) to the average market basket prices and standard deviations for poor and affluent neighborhoods in NYC noted in Green [1991, app. A]. Using a t-statistic of 2.0, I calculated a required sample size of 17 stores in each of the poor and

upper neighborhoods (34 total stores) at 0.90 power. I just about doubled this suggested sample size in anticipation of slightly larger standard deviations due to my implementation of random sampling. The resultant self-weighting sample of 90 stores is the basis for the stores contained in the survey.

I randomized this sample of 90 survey targets, then labeled A, B, C, D, successively, and sorted by label to determine the survey order. The objective of this last step was to ensure that all stores labeled “A” were surveyed on the same day, regardless of location and stratum. This introduces another random component that should eliminate any effect of surveying neighboring stores within the labeled group. As an additional precaution, I recorded the time and day all stores were surveyed to determine if these variables influenced the market basket price. Regressing total market basket price on day and time indicator variables resulted in coefficients, with a p-value of the regression equal to 0.774. The stores were surveyed over a span of four weekdays, September 19, 1996 - September 22, 1996 (Monday to Thursday), to avoid the distortion caused by crossing sale weeks and to control for pricing strategies practiced by the firm.

A.3 Store Representation

Appendix Table 1 summarizes the composition of the universe of stores. Firms operating in inner-city areas comprise close to 14 percent of the universe of stores, whereas firms operating in upper-income areas comprise almost 58 percent of the universe. The latter figure was much larger than expected. The cause is the very crude division of strata based on income level.

Appendix Table 1 presents the ex-ante sample’s representation of the universe of stores. The table shows that the universe of stores contains slightly more than 13 and 57 percent of firms located in poor and upper-income neighborhoods, respectively, whereas the analysis sample contains 32 and

34 percent, respectively, of such stores. However, the percentage of stores in each stratum is fairly similar for the ex-ante survey sample and the analysis sample. Although I was unable to survey all 90 stores, the composition of the stores surveyed does not differ from the ex-ante sample based on a joint test of the column differences.

Appendix Table I also exhibits the survey sample's representation of the universe of stores. The table confirms the expectation that the poor and upper groups are over- and under-represented, respectively, as they were in the ex-ante sample. A firm operating in a poor neighborhood is almost three times as likely to appear in the survey sample as it is in the universe of stores.

A.4 Missing Values

Of the 1140 prices (57 stores multiplied by 20 items) I attempted to collect from the analysis sample, 220 prices (19 percent) were not in stock and were recorded as missing. I use several methods to analyze the effect of the missing values on my results. First, I analyze the complete data case in accordance with Little and Rubin (1987), whereby I create a subsample that contains all price data (no missing prices) and evaluate the difference in the values of available data from the complete subsample item by item. With the exception of apples, there are no significant differences.

Second, I evaluate the impact of replacing the missing values with the overall item mean, which should decrease the differential between groups, if one exists. I then compare this unconditionally imputed subsample to the complete case subsample. The variances of the prices with this method are smaller due to the imputations thereby making these tests less conservative. Again with the exception of apples, there are no significant differences. I also experimented with

regression-adjusted imputation utilizing the fixed investment variables as regressors with no change in results.

Finally, I evaluate the effect of imputing values with the mean of the item by income strata. I then compare this conditionally imputed sample to the complete data subsample. There were no significant differences. On the basis of these results, I impute missing values with the mean of the item by income strata. I believe my results are robust to the imputation method employed.

A.5 Sale and Generic Prices

In this study, the cost of the market basket under the sale pricing strategy is calculated by substituting item sale prices for their respective regular prices. Sale prices include shelf-posted and store advertised sales, and quantity discounts. The sale price does not include unadvertised sales, e.g. sales that only take effect at the checkout. The generic sale price was calculated in a similar manner in which the lowest price of the most similar generic item was substituted for the regular price whenever applicable, but the sale price was used if no generic brand was available. The generic sale price was calculated using the sales price of the most similar generic brand in lieu of the generic brand price when applicable. In this way, the generic sale price is the absolute lowest price a consumer can pay for the market basket of goods without further consumer specific discounts such as senior citizen discounts.

REFERENCES

- Alcaly, R., Klevorick, A., (1971). "Food Prices in Relation to Income Levels in New York City," *Journal of Business* 44, 380-397.
- Alwitt, L. and Donley, T. (1996). *The Low-Income Consumer*. (Thousand Oaks, CA: Sage Publications, Inc.).
- Ambrose, D., (1979). "Retail Grocery Pricing: Inner-City, Suburban, and Rural Comparisons", *Journal of Business* 52, 95-102.
- Ayres, I. (1991). "Fair Driving: Gender and Race Discrimination in New Car Negotiations," *Harvard Law Review* 104, 817-872.
- Ayres, I. and Siegelman, P. (1995). "Race and Gender Discrimination in Bargaining," *American Economic Review* 85, 304-321.
- Burdett, K. and Judd, K. (1983). "Equilibrium Price Dispersion," *Econometrica* 51(4), 955-969.
- Butters, G. R. (1977). "A Survey of Advertising and Market Structure," *American Economic Review* 66(2), 392-397.
- CensusCD 2.0, (1995). (New Jersey: Geolytics, Inc.).
- Chevalier, J. (1995). "Capital Structure and Product-Market Competition: Empirical Evidence from the Supermarket Industry," *American Economic Review* 85(3), 415-435.
- Citro, C. F. and Michael, R. T., eds. (1995). *Measuring Poverty: A New Approach*. (Washington, D.C.: National Academy Press).
- Darden, W. and Lumpkin, J. (1984). "Psychographic and Demographic Profile of Convenience Food Store Users: Why People Convenience Shop," *Review of Business and Economic Research* 19(2), 68-80.
- Demographics USA, Zip Edition 1996*, (1996). (New York: Market Statistics).
- Directory of Single Unit Supermarket Operators*, (1996). (Florida: CSG Information Services).
- Directory of Supermarket, Grocery and Convenience Store Chains*, 1996. (Florida: CSG Information Services).
- Finke, M., Chern, W., Fox, J., (1997). "Do the Urban Poor Pay More for Food? Issues in Measurement," *Advancing the Consumer Interest* 9, 13-17.
- Geithman, F. and Marion, B. (1993). "Testing for Market Power in Supermarket Prices: A Review of the Kaufman-Handy/ERS Study," in *Competitive Strategy Analysis in the Food System*, Ronald Cotterill, ed. (Boulder, CO: Westview Press, Inc.).
- Glaeser, E. (1998). "Should Transfer Payments be Indexed to Local Price Levels?" *Regional Science and Urban Economics* 28(1), 1-20.
- Goldberg, P. (1996). "Dealer Price Discrimination in New car Purchases: Evidence from the Consumer Expenditure Survey," *Journal of Political Economy* 104(3), 622-654.
- Graddy, K. (1997). "Do Fast-Food Chain Price Discriminate on the Race and Income Characteristics of an Area?" *Journal of Business and Economic Statistics*, 15(4), 391-401.
- Green, M., (1991). "The Poor Pay More for Less. Part I: Grocery Shopping," City of New York Department of Consumer Affairs.
- Groom, P., 1966. "Prices in poor Neighborhoods," *Monthly Labor Review* 89, 1085-1090.
- Kaplow, L. (1995). "Regional Cost of Living Adjustments in Tax Transfer Schemes," NBER Working Paper #5008.

- Kinsey, J. D. (1998). "Concentration of Ownership in Food Retailing: A Review of the Evidence about Consumer Impact," Working Paper 98-04, The Retail Food Industry Center, University of Minnesota.
- Kunreuther, H., (1973). "Why the Poor Pay More for Food: Theoretical and Empirical Evidence", *Journal of Business* 46, 368-383.
- Little, R., Rubin, D., (1987). *Statistical Analysis with Missing Data*, (New York: John Wiley & Sons, Inc.).
- MacDonald, J., Nelson, P., (1991). "Do the Poor Still Pay More? Food Price Variations in Large Metropolitan Areas", *Journal of Urban Economics* 30, 344-359.
- Marion, B., Mueller, W., Cotterill, R., Geithman, F., and Schmelzer, J. (1979). *The Food Retailing Industry*. (New York, NY: Praeger Publishers).
- Million Dollar Directory*, (1996). Dun & Bradstreet, Inc., Pennsylvania.
- Munnell, A., Tootell, G., Browne, L., and McEneaney, (1996). "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review* 86(1), 25-53.
- 1996 American Big Business Directory*, (1996). American Big Business Directories, Inc., Illinois.
- Progressive Grocer* (1999). "66th Annual Report of the Grocery Industry," (New York: Trade Division, Butterwick Co.).
- Pratt, J., Wise, D., and Zeckhauser, R. (1979). "Price Differences in Almost Competitive Markets," *Quarterly Journal of Economics* 93(2), 189-211.
- Rothschild, M. (1974). "Searching for the Lowest Price When the Distribution of Prices Is Unknown," *Journal of Political Economy* 82(4), 689-711.
- The Sourcebook of Zip Code Demographics*, (1996). (Virginia: CACI Marketing Systems).
- Salop, S. and Stiglitz, J. (1977). "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," *Review of Economic Studies* 44, 493-510.
- Sexton, D. E. (1971). "Comparing the Cost of Food to Blacks and to Whites—A Survey," *Journal of Marketing* 35, 40-46.
- Sheppard, A. (1991). "Price Discrimination and Retail Configuration," *Journal of Political Economy* 99(1), 30-53.
- Supermarket Census*, (1997). (Connecticut: Trade Dimensions).
- Supermarket News, 1996. *SN Retailers and Wholesalers*. (New York: Fairchild Books).
- Troutt, D., (1993). "The Thin Red Line: How the Poor Still Pay More," West Coast Regional Office, Consumers Union of the US, Inc.
- Yinger, J. (1986). "Measuring Discrimination with Fair Housing Audits: Caught in the Act," *American Economic Review* 76, 881-893.
- United States Bureau of the Census (1997). *Poverty in the United States: 1996*. (Washington, DC: Government Printing Office).
- United States Department of Labor Statistics, (1993). *The Consumer Expenditure Survey*, Bulletin 2462. (Washington, DC: Government Printing Office).
- Varian, H. (1980). "A Model of Sales," *American Economic Review* 70, 651-659.

Figure 1: Map of NYC and Analysis Sample

Note: The shaded region represents the survey counties in New York City. The data points represent the zip codes of the analysis sample. The symbols labeled “1,” “2,” and “3” represent the poor, middle, and upper, zip codes respectively.

Westchester County, NY

Bronx County, NY

New York County, NY

Queens County, NY

Kings County, NY

Richmond County, NY

Figure 2: Histogram of Market Basket Price

Note: Market basket cost depicted is derived from a regression of total basket cost (the sum of item prices) on a constant and indicator variables signalling missing item prices. The histogram shows the predicted values from this regression.

Table 1: Analysis of the difference between Surveyed and Non-surveyed Stores

	Dependent Variable			
	Number of households (in thousands) (1)	Population (in thousands) (2)	Median household income (in thousands) (3)	Per capita income (in thousands) (4)
Store surveyed	-2.195 (1.586)	-6.200 (5.275)	0.241 (2.035)	1.720 (1.312)
N	90 (5)	90 (6)	90 (7)	90 (8)
Store surveyed	-1.820 (1.745)	-6.867 (5.903)	0.188 (2.193)	2.160 (1.454)
N	77	77	77	77

Note: Data from *CensusCD 2.0*. The regression coefficient in columns (1) - (4) are from the model which compares the 33 excluded stores to the 57 sampled stores. The coefficients in columns (5) - (10) are from the model that omits the 13 stores that were not in business at the time of the survey. The p-value is 0.048 of the test that the excluded stores (columns (1) - (4)) are identical to the sampled stores. The equivalent test for the group of column (5) - (10) yields a p-value of 0.063. All regressions include an intercept. All regressions account for stratification and sampling weights in the calculation of the standard errors, which are listed in parentheses. The weights are the inverse of the ex-ante probability of selection.

Table 2: Mean Characteristics of Neighborhoods and Supermarkets in New York City by Zip Code

	New York City	Analysis Sample (Weighted)	Analysis sample sub-groups			Poor - Upper Difference
			Poor	Middle	Upper	
<u>Population</u>						
1990 Census (in thousands)	48.930 (0.655)	52.768 (3.106)	60.873 (3.945)	71.655 (2.526)	40.454 (2.593)	20.419 (4.721)
Density (in thousands)	41.645 (0.821)	41.847 (3.336)	51.439 (3.193)	32.714 (2.463)	44.763 (5.604)	6.675 (6.450)
Households (in thousands)	19.651 (0.305)	18.800 (0.851)	19.909 (1.215)	23.494 (0.921)	15.971 (0.947)	3.938 (1.540)
Median age ¹	---	34.617 (0.411)	28.005 (0.500)	32.799 (0.556)	37.298 (0.338)	-9.293 (0.603)
<u>Household Income</u>						
Median income (in thousands)	\$33.942 (0.392)	\$30.606 (1.418)	\$14.498 (0.424)	\$26.339 (0.668)	\$37.048 (2.124)	-\$22.550 (2.166)
Per capita income (in thousands)	\$19.663 (0.363)	\$14.974 (1.040)	\$7.354 (0.269)	\$10.736 (0.395)	\$19.225 (1.427)	-\$11.871 (1.452)
<u>Proportion with income:</u>						
Less than \$15,000	0.258 (0.004)	0.263 (0.009)	0.507 (0.008)	0.284 (0.011)	0.190 (0.011)	0.317 (0.014)
\$15,000 to \$24,999	0.146 (0.001)	0.165 (0.005)	0.181 (0.003)	0.190 (0.002)	0.147 (0.006)	0.034 (0.006)
\$25,000 to \$49,999	0.295 (0.001)	0.315 (0.007)	0.222 (0.004)	0.342 (0.006)	0.325 (0.011)	-0.102 (0.011)
\$50,000 to \$99,999	0.221 (0.002)	0.206 (0.009)	0.080 (0.003)	0.164 (0.006)	0.260 (0.010)	-0.180 (0.011)
\$100,000 to \$149,000	0.047 (0.001)	0.034 (0.005)	0.006 (0.001)	0.016 (0.001)	0.051 (0.007)	-0.045 (0.007)
\$150,000 or more	0.034 (0.001)	0.017 (0.005)	0.004 (0.001)	0.004 (0.001)	0.028 (0.009)	-0.024 (0.009)
Proportion with public assistance income	0.121 (0.003)	0.124 (0.008)	0.342 (0.013)	0.147 (0.012)	0.057 (0.006)	0.284 (0.014)
Proportion in poverty	0.168 (0.330)	0.182 (0.923)	0.427 (0.008)	0.201 (1.209)	0.109 (1.021)	0.318 (1.022)
<u>Race/Ethnicity</u>						
Proportion white	0.581 (0.008)	0.515 (0.035)	0.267 (0.044)	0.345 (0.029)	0.671 (0.040)	-0.404 (0.059)
Proportion African-American	0.237 (0.007)	0.208 (0.026)	0.432 (0.049)	0.297 (0.030)	0.103 (0.039)	0.329 (0.062)

- Continued -

Table 2: Continued

	New York City	Analysis Sample (Weighted)	Analysis sample sub-groups			Poor - Upper Difference
			Poor	Middle	Upper	
Proportion Asian	0.075 (0.002)	0.121 (0.010)	0.016 (0.002)	0.117 (0.018)	0.150 (0.016)	-0.134 (0.016)
Proportion other	0.104 (0.003)	0.153 (0.017)	0.281 (0.021)	0.239 (0.023)	0.074 (0.011)	0.207 (0.024)
<u>Competition²</u>						
Number of stores per square mile in zip code	9.338 (0.526)	19.397 (1.090)	16.637 (1.707)	22.580 (1.150)	18.262 (1.775)	-1.625 (2.462)
Population per store (in thousands)	17.359 (1.057)	8.389 (1.731)	7.664 (1.311)	5.080 (0.993)	10.309 (2.824)	-2.645 (3.113)
<u>Supermarket Investment</u>						
Proportion chain stores	---	0.320 (0.072)	0.190 (0.088)	0.205 (0.094)	0.416 (0.109)	-0.226 (0.140)
Square feet of retail space (in thousands)	---	16.031 (2.904)	10.230 (1.253)	15.143 (6.462)	18.000 (3.581)	-7.770 (3.794)
Proportion with scanning equipment	---	0.464 (0.076)	0.476 (0.111)	0.289 (0.106)	0.555 (0.109)	-0.079 (0.156)
Number of checkouts	---	5.274 (0.620)	4.714 (0.576)	3.721 (0.655)	6.259 (0.968)	-1.545 (1.126)
Proportion with credit/debit equipment	---	0.493 (0.078)	0.429 (0.110)	0.424 (0.151)	0.544 (0.110)	-0.116 (0.156)
Proportion offering private label goods	---	0.439 (0.068)	0.667 (0.105)	0.354 (0.145)	0.424 (0.098)	0.243 (0.143)
Proportion offering generic label goods	---	0.668 (0.070)	0.810 (0.088)	0.523 (0.152)	0.704 (0.093)	0.105 (0.128)
Number of observations	1408	57	21	14	22	

Note: Data from *CensusCD 2.0*, unless otherwise noted. Data are weighted by the inverse of the ex-post probabilities of inclusion in the analysis sample. Standard errors are in parentheses.

¹Data from the *Sourcebook of Zip Code Demographics* and *Demographics USA, Zip Edition 1996*.

²There are 237 and 28 unique zip codes in the NY-MSA and the analysis sample, respectively. There are 10, 7, and 11 unique zip codes in the poor, middle, and upper strata, respectively, of the analysis sample.

Table 3a: Mean Prices of Market Basket Items by Group

Item Price (in logarithms)	Subgroup			Poor-Upper Difference
	Poor (1)	Middle (2)	Upper (3)	
Total market basket (weighted by expenditure-shares)	3.955 (0.008)	3.968 (0.013)	3.971 (0.014)	-0.016 (0.016)
Total market basket (unweighted)	3.925 (0.008)	3.938 (0.014)	3.942 (0.015)	-0.016 (0.017)
Cereal	0.927 (0.005)	0.926 (0.016)	0.876 (0.009)	0.052 (0.010)
Long-Grain Rice	1.104 (0.017)	1.116 (0.013)	1.088 (0.018)	0.016 (0.025)
Spaghetti	-0.025 (0.044)	0.071 (0.039)	0.002 (0.032)	-0.027 (0.054)
Flour	0.648 (0.046)	0.700 (0.024)	0.676 (0.026)	-0.029 (0.052)
Sugar	0.957 (0.008)	0.895 (0.025)	0.990 (0.019)	-0.033 (0.020)
Tuna	-0.059 (0.026)	0.034 (0.058)	-0.051 (0.031)	-0.008 (0.040)
Eggs	0.192 (0.020)	0.222 (0.035)	0.265 (0.039)	-0.074 (0.044)
Butter	1.106 (0.009)	1.060 (0.030)	1.068 (0.027)	0.037 (0.028)
Milk	0.932 (0.008)	0.885 (0.029)	0.948 (0.020)	-0.015 (0.022)

- Continued -

Table 3a: Continued

Item Price (in logarithms)	Subgroup			Poor-Upper Difference (4)
	Poor (1)	Middle (2)	Upper (3)	
Bread	0.468 (0.023)	0.538 (0.012)	0.504 (0.014)	-0.036 (0.027)
Cheese	1.180 (0.021)	1.166 (0.046)	1.152 (0.027)	0.028 (0.034)
Orange Juice	0.919 (0.026)	0.915 (0.031)	1.011 (0.022)	-0.092 (0.034)
Apples	-0.120 (0.032)	0.058 (0.032)	0.082 (0.028)	-0.202 (0.043)
Bananas	-0.799 (0.038)	-0.854 (0.067)	-0.636 (0.049)	-0.163 (0.062)
Oil	0.533 (0.005)	0.570 (0.017)	0.542 (0.009)	-0.009 (0.011)
Ground Beef	1.444 (0.042)	1.490 (0.077)	1.530 (0.044)	-0.086 (0.061)
Chicken	0.788 (0.044)	0.755 (0.069)	0.862 (0.052)	-0.074 (0.068)
Pampers	2.168 (0.021)	2.168 (0.020)	2.103 (0.041)	0.065 (0.046)
Soap	0.369 (0.010)	0.414 (0.014)	0.410 (0.017)	-0.042 (0.020)
Detergent	1.534 (0.013)	1.567 (0.021)	1.557 (0.017)	-0.023 (0.022)
N	21	14	22	

Note: Prices are in logarithms. Standard errors are in parentheses. The total market basket price is the sum of the item prices. Expenditure shares for low-income consumers are taken from the Consumer Expenditure Survey [1993], Table 43.

Table 3b: OLS Estimates of the Effect of Group Status,
Supermarket characteristics, and Competition on Item

Dependent Variable (in logarithms)	Coefficients on Subgroup variables		
	Intercept (1)	Poor (2)	R ² (3)
Total market basket (weighted by expenditure-shares)	3.998 (0.031)	-0.011 (0.018)	0.320
Total market basket (unweighted)	3.970 (0.032)	-0.011 (0.019)	0.320
Cereal	0.915 (0.020)	0.040 (0.010)	0.414
Long-Grain Rice	1.059 (0.043)	0.026 (0.028)	0.255
Spaghetti	0.021 (0.060)	-0.042 (0.056)	0.259
Flour	0.664 (0.057)	-0.040 (0.052)	0.176
Sugar	1.035 (0.032)	-0.028 (0.017)	0.521
Tuna	0.028 (0.046)	-0.004 (0.038)	0.339
Eggs	0.268 (0.065)	-0.045 (0.034)	0.413
Butter	1.033 (0.063)	0.049 (0.025)	0.340
Milk	0.951 (0.051)	0.006 (0.025)	0.432

- Continued -

Table 3b: Continued

Dependent Variable (in logarithms)	Coefficients on Subgroup variables		
	Intercept (1)	Poor (2)	R ² (3)
Bread	0.531 (0.020)	-0.045 (0.024)	0.281
Cheese	1.173 (0.048)	0.030 (0.034)	0.166
Orange Juice	1.077 (0.042)	-0.088 (0.035)	0.504
Apples	0.207 (0.071)	-0.185 (0.045)	0.517
Bananas	-0.484 (0.116)	-0.159 (0.073)	0.452
Oil	0.584 (0.019)	-0.007 (0.010)	0.558
Ground Beef	1.609 (0.138)	-0.072 (0.075)	0.180
Chicken	0.747 (0.110)	-0.047 (0.079)	0.235
Pampers	2.104 (0.084)	0.071 (0.055)	0.362
Soap	0.486 (0.027)	-0.040 (0.015)	0.434
Detergent	1.639 (0.037)	-0.016 (0.022)	0.437

Note: Each row represents a separate regression. The dependent variables are in logarithms. Standard errors are in parentheses. There are 57 observations in each regression. All regressions control for scanning equipment, number of checkouts, square feet of retail space, the number of stores per square mile in the zip code, and chain status. In addition, each regression includes a dummy variable indicating whether the price was imputed. See text for details on imputation of missing values. Expenditure shares for low-income consumers are taken from the Consumer Expenditure Survey [1993], Table 43.

Table 4: OLS Estimates of the Effect of Income Group on Regular, Sale, and Generic Shelf Prices

	Dependent Variable			
	Total cost of basket using <i>regular</i> shelf prices (1)	Total cost of basket using <i>sale</i> shelf prices (2)	Total cost of basket using <i>generic item</i> shelf prices (3)	Total cost of basket using <i>generic sale</i> <i>item</i> shelf prices (4)
Intercept	3.998 (0.031)	3.673 (0.095)	3.670 (0.091)	3.651 (0.092)
Poor	-0.011 (0.018)	-0.043 (0.083)	-0.034 (0.081)	-0.027 (0.082)
Middle	-0.022 (0.020)	-0.153 (0.122)	-0.159 (0.119)	-0.160 (0.120)
R ²	0.320	0.323	0.308	0.293
N	57	57	57	57

Note: The dependent variables are in logarithms. Standard errors are in parentheses. See the Data Appendix for a description of the construction of "sale," "generic," and "generic sale" prices. All prices are expenditure-share weighted sums of individual item prices, where the expenditure shares for low-income consumers are taken from the Consumer Expenditure Survey [1993], Table 43. All regressions control for scanning equipment, number of checkouts, square feet of retail space, chain status, and the number of stores per square mile in the zip code. In addition, each regression includes a dummy variable indicating each item whose price is imputed.

Table 5: Generalized Least Squares Regression Results

	Dependent Variable: $\text{Log}(P_{ijz})$				
	(1)	(2)	(3)	(4)	(5)
Constant	0.9544 (0.0141)	0.9555 (0.0151)	0.9601 (0.0146)	0.9606 (0.0148)	0.9683 (0.0182)
B-bar (in thousands)	0.0027 (0.0017)	0.0028 (0.0017)	0.0021 (0.0018)	0.0018 (0.0019)	0.0031 (0.0014)
Scanners		-0.0055 (0.0169)	-0.0247 (0.0167)	-0.0265 (0.0172)	-0.0030 (0.0141)
Chain			0.0444 (0.0162)	0.0450 (0.0162)	0.0642 (0.0152)
Square feet (in thousands)				0.0003 (0.0003)	0.0006 (0.0002)
Number of checkouts					-0.0091 (0.0021)
R ²	0.9562	0.9562	0.9568	0.9569	0.9582
N	1140	1140	1140	1140	1140

Note: All regressions include item indicator variables and per capita income*item interactions. Regressions are weighted by the inverse of the ex-ante probability of inclusion multiplied by the inverse of the response rate for the strata. Standard errors are listed in parentheses.

Appendix Table 1: Sample Representation of the Universe by Group

Strata	Universe of Stores		Ex-Ante Survey Sample		Analysis Sample	
	Number of Stores	Proportion	Number of Stores	Proportion	Number of Stores	Proportion
Poor	190	0.135	30	0.333	21	0.368 (0.576)
Middle	407	0.289	30	0.333	14	0.246 (0.205)
Upper	811	0.576	30	0.333	22	0.386 (0.402)
Totals*	1408	1.000	90	0.064	57	0.040

Note: Proportion refers to the fraction of stores in the column that are in each stratum. P-values from the test of the hypothesis that the analysis sample proportions equal the ex-ante survey sample proportions are listed in parentheses.

*Proportion of the universe represented by the universe of stores, the ex-ante sample, and the analysis sample, respectively.

Appendix Table II. Sample of price survey sheet.

COMPARATIVE GROCERY PRICE SURVEY, FALL 1996

Name of Surveyor: _____ Date: _____ Time: _____ Debit/Credit? _____
NAME OF STORE: _____ Phone: _____ #Chkots _____
STORE ADDRESS: _____ ZIP _____ Day of Week: _____ Scanners? _____

ITEM	Amount	REG. PRICE	Sale Price	Generic Brand Name	Amount	Reg. Price	Sale Price
Kellogs Corn Flakes	12oz						
Carolina Long Grain Rice	5 lbs						
Ronzoni Spaghetti #8	1 lb						
Gold Medal All-Purpose flour	5 lbs						
Domino Sugar	5 lbs						
Bumble bee Tuna, chunk light in oil	6 oz						
Dozen eggs	large						
Land O'Lakes Butter, lightly salted	1 lb						
Gallon Whole Milk							
Wonder Bread, white	medium						
Kraft American Cheese, yellow	8 oz						
Tropicana Pure Premium OJ, original, container	12 oz						
Red Delicious apples	half gallon						
	1 lb						
Bananas	1 lb						
Wesson Oil	16 oz						
Ground Beef, not lean	per pound						
Chicken, whole leg	per pound						
Huggies Ultrathin pampers, girls, 16-28 lbs	medium, 32 ct.						
Ivory soap, personal	4 bars, 3.5 oz						
Tide Ultra 2, regular	92 oz						
	39 oz						