Do the Poor Pay More?
An Empirical Investigation of Price Dispersion in Food Retailing

Lashawn Richburg Hayes*
Department of Economics
Rutgers University

September 12, 2000

Abstract

On the question of whether prices are higher in poor, urban neighborhoods, the prior research is decidedly mixed. This paper revisits the question by analyzing unpublished price-level data collected by the Bureau of Labor Statistics for construction of the Consumer Price Index. Using this large, statistically representative sample of stores in poor and affluent neighborhoods, I first estimate if a price difference exists. I then empirically test the major arguments in support of disparate prices such as differences in quality, operating and consumer search costs. I also explore the relationship between pricing strategies and the racial and ethnic composition of poor neighborhoods. I find that market prices are up to 6 percent less in poor neighborhoods after controlling for a variety of covariates. In addition, I find that poor, predominantly white and Hispanic neighborhoods experience significant discounts, while market prices in poor, predominantly black neighborhoods are comparable to those in affluent white areas.

I would like to thank Cecilia Rouse, Orley Ashenfelter, Igal Hendel, Pinelopi Goldberg, and Alan Krueger for their time and advice while I worked on this paper, the participants of the Princeton University Labor Lunch for valuable comments, and the staff of the prices and living conditions section of the Bureau of Labor Statistics, especially William Cook, for their hospitality. This material is based upon work supported by Benheim-Thoman Center for Child Well-being, the Center for Domestic and Comparative Policy Studies, the Industrial Relations Section, all of Princeton University, and the National Science Foundation under Grant No. 9809232. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation or the Bureau of Labor Statistics.

*National Science Foundation Postdoctoral Fellow, Department of Economics, Rutgers University, New Brunswick, NJ 08903. Phone: (609)258-2363. Email: richburg@princeton.edu.

I. Introduction

Whether the poor pay more for basic food items has been a source of bitter empirical controversy for several decades. This paper contains the results of the first comprehensive empirical analysis of this issue based on a representative, national sample of data from the Bureau of Labor Statistics' (BLS) primary sampling frame for construction of the Consumer Price Index (CPI).

Whether the poor pay more for food, and if so, why, is an issue of great importance for recent discussions of public policies toward urban renewal, transfers to the urban poor, zoning requirements for "superstores," and for theories of market structure. The basic results indicate, contrary to some previous studies, that the poor pay less for food items.

The empirical literature on the price gap between poor and affluent neighborhoods focuses extensively on whether a price gap exists. The results from this literature are mixed. Groom [1966] and Alcaly and Klevorick [1971] find no relationship between market prices and the average income level of a neighborhood. In contrast, Kunreuther [1973] reports that low-income neighborhoods have higher prices, while Ambrose [1979] finds lower prices. More recent studies also produce a medley of results. Green [1991], Troutt [1993], and Frankel and Gould [1999] find that low-income neighborhoods have prices that are significantly higher than those in more affluent neighborhoods, while MacDonald and Nelson [1991] and Hayes [forthcoming] find no difference. There is much less research examining why prices may differ (e.g., cost or quality differences). Frankel and Gould [1999] find that higher prices are associated with greater income inequality. Finke *et al* [1997] find that in urban areas, low-income blacks pay significantly more than higher income blacks, and both low- and high-income whites.

Most of the prior research examines market prices through case studies. As a result, sample sizes are small. Further, the importance of the surveyed items in actual consumer baskets is often unknown. More importantly, a significant portion of the studies are based on improper survey methodology--focusing on stores known to engage in unfair pricing strategies, selecting stores on the basis of proximity to volunteer surveyors, or other non-random methods.[1]

The ideal survey to answer this question would provide prices for representative poor and affluent market baskets in both poor and affluent neighborhoods. Data from such a survey would allow researchers to study whether price differentials (if they exist) are due to where the poor shop or what the poor buy relative to more affluent consumers. Further, in such an ideal survey the sample of stores would be the primary shopping venue of the household, ensuring that these data match the demographic characteristics of the household to the survey outlet. In addition to accounting for consumer tastes in outlet and product selection, consumer costs would be collected as well. Most importantly, information pertaining to firm operating costs and quality (e.g., operating hours, the number of specialty departments) would also be obtained.

In this paper, I utilize price level data collected by the BLS to examine whether the poor pay more for food and why. In many ways, the price data mimic those from the ideal survey. The sample frame of stores is compiled from the universe of stores where consumers actually shop. In addition, the probability of selection of a particular product is proportional to its sales volume in the outlet. This allows both consumer behavior and supply factors to be accounted for simultaneously in the analysis of a price differential. I supplement the price level data with information on the service offerings, pricing strategies, and other proxies for operating costs faced by the firm. I find

[1] See e.g., Groom [1966], Alcaly and Klevorick [1971], Kunreuther [1973], Ambrose [1979], Green [1991], and Troutt [1993]. 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 of stores, the market basket is not identical

that the poor pay up to 6 percent less than their more affluent counterparts after accounting for operating costs, quality, market structure, consumer search, and geographic variation. Further, I find that poor, predominantly white (non-Hispanic) neighborhoods and poor predominately Hispanic black neighborhoods experience price discounts up to 9.6 and 18.3 percent, respectively, while poor predominantly black (non-Hispanic) neighborhoods face the going price in affluent white (non-Hispanic) neighborhoods.

A brief summary of the remainder of the paper is as follows: in sections 2 and 3 I present the analytical framework and empirical strategy. In section 4 I describe the data. In section 5 I report results on the existence of a price differential, evaluate the explanatory power of some economic theories offered to explain the differential, and address alternative explanations for the results. In section 6 I discuss some concerns with my interpretation of the results and I conclude in section 7.

2. Analytical Framework

Economic theory alone provides no clear answer as to whether the poor pay more for commodities. 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 Rothschild [1974]; Butters [1977]; Salop and Stiglitz [1977]; Varian [1980]; and Burdett and Judd [1983]). In addition, there are models of price discrimination and bargaining in which a gap in mean prices paid by different

---

across stores as a result of item unavailability. See Geithman and Marion [1993] for a critique of the ERS research design and methodology and Hayes [2000] for a detailed discussion of the prior literature.

groups may not exist, yet price discrimination occurs through differences in the variance of the price distribution [Goldberg 1996]. As the sign of the differential is theoretically indeterminate, this is an empirical question.

My empirical strategy is first to document whether prices differ between poor and affluent neighborhoods and then to control for a number of factors that may affect equilibrium price dispersion.[2] In this paper, price dispersion refers to the spread of prices attributable to differences in costs, quality, information acquisition, or market structure. This is a broad definition which encompasses spatial price dispersion (the distribution of prices for an identical item across space) and temporal dispersion (e.g., sales).

A. Operating Costs

The operating costs of food retailers include the marginal costs associated with a sale, quality-induced costs, and discretionary costs. According to the Food Marketing Institute (FMI) [1997], labor and rental property comprise 12 and 2 percent of sales, respectively. FMI [1997] estimates that payroll and employee benefits account for almost 54 percent of total operating cost for food retailers---more than the cost of supplies and insurance premiums combined. Labor costs are particularly high in seafood, bakery, and foodservice specialty departments [*Supermarket Business*, 1998]. In addition to providing a service, specialty departments are proxies for the quality of a supermarket. For example, only large, modern supermarkets contain in-store bakeries (where baking is done on the premises) and seafood counters. Discretionary costs include promotional activities, such as advertising, in-store demonstrations, and loyalty-card programs.

[2] The spread of prices for the same item is a well-documented phenomenon and economists have generated numerous models that predict dispersed prices in equilibrium [Salop and Stiglitz 1977; Pratt *et al* 1979; Burdett and Judd 1983].

Table 1 provides descriptive statistics on the average store characteristics in the U.S. and the average characteristics for stores in the BLS sample that is analyzed in this paper. A comparison of columns (4) and (5) of the table shows that BLS sample stores operating in low-income areas employ less labor and are less likely to be a branch of a chain. Since chain status proxies for economies of scale in purchasing, the lower proportion of chain stores may signal higher average inventory costs for inner-city retailers. Columns (4) and (5) of the table also show the average discretionary costs and offered services for stores operating in low-income and high-income areas. Stores operating in the inner-city are less likely to offer double-coupon, frequent shopper, and in-store discount activities. These activities encourage consumers to self-select into categories, which results in efficient promotional targeting. Not surprisingly, the greatest divergence between inner-city stores and stores operating in affluent neighborhoods occurs in the quality of the shopping experience. On average, inner-city stores have fewer specialty departments and offer fewer services.

B. Information Acquisition

While the theoretical literature on consumer search behavior is extensive [see e.g., Nelson 1970; Butters 1977; Salop and Stiglitz 1977; Burdett and Judd 1983; and Diamond 1987], empirical findings are limited. Alcaly [1976] finds that search activity by income group is positively related to the group's income elasticity of demand for the product. In an empirical test of joint search using food prices, Carlson and Gieseke [1983] find search behavior moderately increases for low-income consumers.

While theory predicts that increased search from an unchanging price distribution lowers the average transaction price, it is an empirical question as to whether the poor search more. As the wage rate rises, there are two competing forces influencing search. Higher income increases the opportunity cost of time and lowers search activity. However, increased income also allows the consumer to purchase better information from the market resulting in more search [Mincer 1963]. Further, direct costs (e.g., transportation costs) lower the marginal cost of search for the rich relative to the poor. So while the poor have a lower opportunity cost of time as measured by the wage rate, it is not clear whether the poor engage in greater search for the same product [Alcaly 1976].

In my empirical analysis, I proxy search costs with the proportion of households without a motor vehicle, the proportion of residents attaining a given educational level, and the number of stores in the neighborhood (per square mile) interacted with the neighborhood poverty rate.[3] Table 1 shows descriptive statistics for these variables. A comparison of columns (4) and (5) shows that households residing in low-income areas are three times less likely to own a vehicle and residents are almost twice as likely to drop out of high school.

C. Market Structure

Supermarkets comprise 24 percent of all food retailers but account for 77 percent of total food sales. Convenience stores account for 45 percent of all retailers and 6 percent of total sales, while wholesale clubs account for less than 1 percent of outlets but 5 percent of total sales.[4] The supermarket industry is regionally competitive and dynamic. Regionally, more than 85 metropolitan statistical areas (MSAs) have four-firm concentration ratios in excess of 80 percent

---

[3] Determining the variables and functional form to characterize search rules empirically is very difficult and the literature does not offer much guidance. I experimented with a number of different specifications of search such as accounting for the number of children in the household under age 5 (which should increase search costs) and the proportion of single parents. In my empirical work I assume search costs to be linear in income and transportation costs, and non-linear in education.

[4] *Progressive Grocer* [1999], p.10. Limited assortment stores, which sell less than 1,500 items, primarily dry grocery, with few (if any) perishables, account for the remainder of total sales.

[Geithman and Marion 1993].[5] Nationwide, the top four companies account for 20 percent of annual sales and almost 20 percent of individual supermarkets.[6] Despite the fact that four companies own a large share of individual stores, supermarkets are not very concentrated nationally; Albertson's acquisition of American Stores in 1998 made it the first supermarket with operations coast to coast.[7] Finally, the industry is constantly changing due to an average of 54 mergers/acquisitions each year [Kinsey 1998]. However, the number of stores remains fairly constant as entry and exit are nearly equal [USDA 1996].

Column 2 of Table 1 shows the means of the market structure variables I use in the empirical analysis. Surprisingly, columns (4) and (5) of the table show that there are significantly more stores (per square mile) in the inner-city than in more affluent neighborhoods, although these stores are smaller on average.

3. Empirical Framework

Measuring the net price differential

My empirical strategy is to try to account for variation in price levels due to costs and quality so as to isolate the relationship between the neighborhood income level and market prices. To begin, I first estimate price differentials by analyzing market prices in poor and rich neighborhoods for homogenous products. Because theory provides little guidance on the functional form, I estimate the following semi-log model:

[5] The four firm concentration ratio is the sum of the four largest market shares in the industry:

$$R_m \equiv \sum_{i=1}^{4} a_i ,$$

where a_i º q_i/Q is firm *i*'s market share and the firms are ordered such that $\alpha_1 \geq \ldots \geq \alpha_4 \geq \ldots \geq \alpha_n$.

[6] *Food Industry Review* [1998]. In 1998 the top four food retailers were Albertsons, Inc., The Kroger Co., Wal-Mart Stores, Inc., and Safeway, Inc.

$$\ln p_{ijzst} = g + b_1 Poor_{zs} + X_{jzs}\Pi + Z_{ijst}\Phi + w_s + u_{ijzst}, \qquad (1)$$

where p_{ijzst} is the unit price plus applicable sale taxes of item i in store j located in neighborhood z, in local area s, at time t; g is the item fixed effect; $Poor_{zs}$ is an indicator variable which equals 1 if the neighborhood is low-income; X_{jzs} represents a set of covariates (operating and discretionary costs, consumer search, quality, market structure, and neighborhood demographics); Z_{ijst} is a set of other covariates to be discussed below; w_s is the local area fixed effect; and u_{ijzst} is a random error term that can be decomposed into:

$$u_{ijzst} = h_j + e_{ijzst}, \qquad (2)$$

such that,

$$E(u_{ijzst}, u_{ikzst}) = s_h^2 = \left[\frac{s_h^2}{s_h^2 + s_e^2}\right] s^2 = r s^2, \forall j \neq k, \qquad (3)$$

This error structure allows for the correlation of item prices within a store j. The approach most used in the literature omits covariates X_{jzs} and Z_{ijst} from specification (1).

I also analyze the effect of the interaction between race/ethnicity and residence in a low-income neighborhood on neighborhood price levels by estimating the following model:

$$\ln p_{ijzst} = g + b_1 Poor_{zs} + R_{zs}\Gamma + (R_{zs} * Poor_{zs})\Psi + X_{jzs}\Pi + Z_{ijst}\Phi + w_s + u_{ijzst}, \qquad (4)$$

where R_{zs} represents race and ethnicity indicator variables and Y is the vector of coefficients on the race/ethnicity and poor interaction terms.

In models (1) and (4), income plays a multiple role making the interpretation of b_1 less straightforward. Income at once represents an identifiable characteristic, which may be used by retailers to segment consumers for price discrimination (i.e., in the sense of Pigou [1920]), in

[7] There is no consensus as to whether concentration increases price in this industry. The most comprehensive studies find polar results. See Marion *et al* [1979] and Kaufman and Handy [1989].

addition to representing gross demand and the cost of search.[8] Despite embodying these multiple effects, one would only expect the coefficient b_1 to be positive for a limited number of reasons. For example, assuming the good is normal, the first possibility is for the search effect to dominate the income effect in such a way that the rich search more *à la* Mincer [1963] and pay lower prices. However, Rothschild [1974] and others provide strong arguments against this possibility. The second, and more plausible, possibility is price discrimination, in which the seller captures some consumer surplus from the captive market.

Measuring item dispersion

Equation (1) estimates the *average* net price differential across poor neighborhood types. While the average net price difference is informative, it has two potential drawbacks for this analysis. First, the average price differential constrains the relationship between price and the income status of a neighborhood to be the same for each item. Since anecdotal evidence suggests that item prices vary widely, this constraint may be unreasonable. Second, the weighting scheme employed in the estimation of the average price differential does not account for the relative importance of items in the market basket of poor consumers. Therefore, to analyze item price dispersion across neighborhood types, I also estimate a modified version of specification (1) by item for the poor and rich separately:

$$\ln p_{ijzst} = a_0 + b_i' I_{zs} + X_{jzs}\Pi' + Z_{ijst}\Phi' + w_s' + u_{ijzst}', \tag{5}$$

[8] Price discrimination exists if (and only if) the same variety of a commodity is sold to two different buyers at different net prices [Phlips 1983]. In order for price discrimination to be a viable strategy, a retailer must have market power, a means to segment the market, and the ability to prevent resale. The first condition is no longer interpreted in the strong sense of absolute monopoly power (see e.g., Greenhut and Greenhut [1975] and Norman [1981]). Duopolists, oligopolists, and small competitors in differentiated markets can practice price discrimination. Price discrimination by inner-city food retailers may be facilitated by the possession of some monopoly power through the lack of effective competition, the ability to sort consumers according to their intensity of demand (known from food stamp or other

where I_{zs} is the median household income of the neighborhood z. While I can compare the differences in b_i between the poor and the rich for item i, such a difference does not measure the concept of differential prices discussed in the popular press. This is because high prices for certain goods may be offset by low prices for other goods across neighborhood types. The usual concept of a price gap seeks to capture a situation in which the net result of all pricing strategies is higher prices in poor neighborhoods. To address this, I derive an aggregate measure that weights the price differences across items by the expenditure shares of the poor. The net price differential in poor neighborhoods is then:

$$\bar{b}_{poor} = \sum_{i=1}^{I} s_i \hat{b}_i', \tag{6}$$

where s_i is the relevant food category expenditure share for low-income consumers taken from the Consumer Expenditure Survey (CEX), and $\hat{b}_i'$ refers to the estimated price difference in specification (5).

In summary, I first examine mean gross price differences and relative dispersion. My regression analysis begins by estimating the relationship between the market price and income level of the neighborhood using prices for homogenous items priced in almost every sampling area. I initially impose restrictions on the price relationship (the price differential is assumed to be the same across sampling areas, as are operating costs and other covariates), then relax the assumptions. In my most general specification, I derive an overall measure of the relationship between price and income that is simply the weighted-average of the individual item differences, where the weights are item budget-shares. This final measure, $\bar{\beta}$, can be interpreted as the price

welfare redemption, information from shopper club plans, etc.), and the practice of quantity restrictions on discounted items. A corollary to the second condition is that the intensity of demand must actually differ across segmented groups.

differential which accounts for the relative importance of items to the low-income consumer. A comparison of $\overline{\beta}_{poor}$ to $\overline{\beta}_{rich}$ yields another measure of the price differential.

4. Data

I. BLS Data

The BLS collects the price data analyzed in this paper to compile the monthly CPI. The BLS data are uniquely suited for the analysis of income price differentials for several reasons. First, the prices are national in scope. The BLS selects 87 urban regions in the U.S. to survey based upon various demographic factors.[9] Map 1 shows the distribution of BLS survey areas. Although the survey is limited to urban areas (encompassing 86 percent of the population), the map shows sampling areas to be geographically diverse. Second, the BLS prices are representative of where consumers shop and what they actually purchase. This is because the sampling strategy derives the frame of survey outlets from the Telephone Point of Purchase Survey (TPOPS), a separate unpublished consumer survey. Once the outlet is selected based upon total expenditures, the unique item to price is selected using probability proportional to sales.[10] These selection techniques imbed consumer behavior in the collected price, which is essentially the market price. Another benefit of the BLS sample is that prices are collected from a variety of neighborhood types, allowing the study of the average poor consumer. Finally, the BLS weights price indices to be representative of the nation. Therefore using their sample, properly weighted, results in representative national price differentials.

[9] The primary sampling units are chosen to be representative of the current U.S. demography [Williams 1996] and are selected based upon the following factors the BLS finds to be highly correlated with price change: region, population size, mean interest and dividend income per housing unit, mean wage and salary income per housing unit, percent of housing units heated by electricity, percent of housing units heated by fuel oil, percent black, and percent retired. See *BLS Handbook of Methods*, chapter 17.

[10] See the Data Appendix for details on the BLS sampling strategy.

One potential drawback of the BLS data is that the same item is not priced everywhere, raising concerns of comparability. I avoid comparing "apples and oranges" by limiting my analysis to five homogenous items surveyed in the majority of sampling areas. Another concern is related to the unique item selection process, which does not account for product availability. For example, it is not fair to compare Brand 1 flour at store 1 to prices at other stores when store 1 only stocks Brand 1. This procedure is appropriate for a price index reflecting inflation, but is a concern for comparing price differentials because the lack of choice can be regarded as a price premium. If product choice is positively correlated with the neighborhood income level, then my estimate of the net price differential will be upward biased. This is a real possibility as stores in low-income neighborhoods tend to be substantially smaller than average.

I limit my analysis to prices from the food and beverages major group of the CPI. I further limit my sample to the following products: whole chicken, eggs, milk, bananas, oranges, and lettuce.[11] The analysis sample consists of 10,170 prices from 2,181 stores in 43 states over the 12 months of 1998.[12] Columns 1 and 2 of Table 1 compare the means for selected demographics of my analysis sample to all neighborhoods in the U.S. As expected, the analysis sample differs significantly from the average neighborhood in the U.S. in dimensions that reflect the CPI's goal of tracking inflation in urban areas.[13]

[11] This group of products is priced in almost all survey areas. This sample is derived from a larger, more heterogenous sample of products including flour, white bread, ground beef, pork chops, bananas, potatoes, non-carbonated juice, and salad. This larger sample consists of 63,557 prices from 2,181 stores. The analysis of this sample yields qualitatively similar results. See the Data Appendix for details.

[12] The BLS does not survey Montana, North Dakota, and Wyoming. Further, my analysis sample does not contain stores in Iowa, Maine, New Mexico, Rhode Island, and West Virginia because the homogenous items in the sample are not surveyed in these states.

[13] There are no significant differences when the BLS sample is compared to all urban neighborhoods in the U.S.

II. Outlet information

Since the price data contain only address information and store type, I supplement them with more extensive data from SPECTRA, Inc., a private marketing firm.[14] The SPECTRA sample consists of 19,836 observations on supermarkets, grocery stores, and large-scale discounters and provides information on service offerings, outlet size, and costs. Columns 2 and 3 of Table 1 report the means of selected covariates for the BLS sample and the subsample for which I have SPECTRA data. The subsample generally resembles the entire BLS sample.[15] I use additional data from InfoUsa, Inc., a private marketing firm, to derive measures of competition from the number of stores in a zip code. I use the 1990 Census, Summary Tape File 3B, to obtain demographic data by zip code.

III. Defining the poor

Defining the poor is not a straight-forward task, as there are a number of ways to classify a household as being poor, such as being cash poor, food and clothing poor, or being subject to detrimental social conditions as in Wilson [1987]. In the empirical work below, I analyze different definitions of the poor based on various percentiles of the state poverty rate. Poverty percentiles based on each state's distribution of poverty are relative measures, which is important to note because the composition of the poor changes with the distribution rate used. Alternate distributions include the U.S. population and the distribution of poverty in the BLS sample. However, the results are qualitatively the same regardless of the poverty rate distribution used.[16] Therefore, for

[14] The match rate is approximately 78 percent. In adherence to the confidentially of the data, I do not reveal outlet names and product brands in this analysis.
[15] See Appendix Table 1 for a comparison of the full SPECTRA sample to the BLS outlet sample. The table reveals that the SPECTRA sample consists of stores of various sizes located in a diverse set of neighborhoods.
[16] At issue is a choice about absolute versus relative poverty. For example, percentiles based upon the distribution of poverty across the U.S. population are an absolute measure of poverty.

comparative purposes, most of the empirical work below defines the poor as households residing in neighborhoods in which more than 20 percent of the residents are below the state poverty level.

III. Racial and ethnic composition

I measure the racial and ethnic composition of a neighborhood using dummy variables based upon segregation indices calculated at the county level. I utilize the dissimilarity [Duncan and Duncan 1955] and isolation [Bell 1954] indices.[17] The dissimilarity index proxies whether two groups are evenly distributed throughout an area. The index varies from 0 to 1 and is minimized when all parts of a county have the same relative number of minority and majority members as the county as a whole. At a value of 1, minority members live in completely different areas of a county than majority members. The index of isolation measures the extent to which minority members interact with only minority members. It is intended to capture the characteristic of exposure. The index of isolation is minimized when the minority group is a relatively small proportion of the county, so that the minority group will have to interact in some capacity with the majority.[18] This index also varies from 0 to 1, with higher values indicating a greater isolation from the majority population. I define a predominantly black county as one in which the index of isolation exceeds 0.3.[19] Similarly, I define a predominantly Hispanic county as one in which the Hispanic index of isolation exceeds 0.3. In the empirical work below, I refer to "black neighborhoods" and "Hispanic

[17] Segregation indices are used extensively in the sociology literature. The index definitions I use are based upon Massey and Denton [1988]. See the Data Appendix for the mathematical definitions.
[18] While I utilize the conventional dissimilarity and isolation indices only, Massey and Denton [1988] propose three other indices intended to capture the concentration, centralization, and clustering aspects of segregation. As these indices require detailed geographic information, I cannot use these indices in the analysis.
[19] I have experimented with a number of alternative definitions, including defining a predominantly black neighborhood as one in which the index of dissimilarity exceeds 0.6 and the index of isolation exceeds 0.3---the definition used in Cutler, Glaeser, and Vigdor [1997]. See Table 2 for comparisons. The proportion of counties satisfying this condition differs from that in Cutler, Glaeser, and Vidgor [1997] because of my use of zip codes as my neighborhood proxy and the calculation of segregation at the county level. See the Data Appendix for details.

neighborhoods" as indicating neighborhoods located in predominantly black counties and Hispanic counties, respectively.

5. Estimation results

I. Summary statistics

As noted earlier, Table 1 presents the means of the covariates I use to represent the factors that may influence price levels. I group the covariates according to the primary mechanism through which they may affect price. The top rows of Table 1 show the means of the various racial and ethnic neighborhood classifications I use. The top-most five rows show the proportion of predominantly black or Hispanic counties in the BLS sample as a whole in column (2) and by poverty status in columns (4) and (5). Neighborhoods located in predominantly black counties are significantly more likely to be poor than affluent, using either measure of racial composition. This is also true for predominantly Hispanic neighborhoods.

The remainder of Table 1 shows the means of the other cost, consumer search, quality, and market structure covariates. Given that most of the variables that proxy search costs are significantly different for the poor, it is likely that search may explain a portion of the price patterns. In terms of the operating cost variables, stores operating in non-poor neighborhoods provide quicker checkout as measured by the number of available cashiers, greater availability of service or replenishment of products as represented by the number of full- and part-time workers, and on average, more opportunities to enjoy discount prices. Although crime is significantly higher in poor areas, the table indicates that it is more costly to do business in non-poor areas.[20]

[20] Insurance costs for supermarkets comprise less than 0.5 percent of sales [FMI 1997]. Even if insurance costs were 50 percent higher for stores operating in low-income areas, such costs would continue to be smaller in magnitude and importance than labor, utilities, and supplies.

II. Mean gross price differences

I limit my analysis to five homogenous items--milk, whole chicken, eggs, navel oranges, and Iceberg lettuce--that are priced in most sampling areas. Table 2 shows the mean (gross) price differences for these items by alternative definitions of poor neighborhoods. The first panel of the table defines the poor as households residing in neighborhoods in which more than 20 percent of the residents are beneath the poverty line. Using this definition, the poor pay less for chicken (10 cents less) and eggs (7 cents less). The table shows that there is a significant discount for chicken (6 cents lower) and eggs (12 cents lower) when poor neighborhoods are defined as being in the 80th percentile of the state poverty distribution. The significance of the discount for eggs disappears when the definition of poor changes to the 90th percentile of the state poverty distribution. Under this definition, the only significant difference is for poultry, where the price is 9 cents lower per pound in poor neighborhoods. When prices are allowed to vary nonlinearly across neighborhood types, there is a statistically significant difference in orange and lettuce prices of 16 and 10 cents per pound, respectively, between the first and fourth quintiles (where the first quintile consists of the most affluent neighborhoods). This suggests that the poor pay less for these goods. Similarly, concentrated poverty neighborhoods (as represented by the fifth quintile) have prices different from those in the first quintile only for eggs, with eggs being 14 cents cheaper in poor neighborhoods. All other differences are not statistically significant.

The last column shows a weighted-sum of the average item prices, where the weights are the category expenditure shares for low-income consumers taken from the CEX. Although the individual item averages fluctuate, the overall basket cost is negative and significant for three of the five poverty definitions.

Two patterns are evident in this table. First, all of the price differences are either zero or negative, indicating that market prices are likely to be lower in poor neighborhoods. Second, the small significant differences that do appear are not constant across definitions, indicating that the gross price discounts are not robust to alternative definitions of the poor. The last panel of the table shows essentially no price differences across the poverty distribution for milk, chicken, and eggs. In fact, the last quintile (which is equivalent to 80th percentile of the state poverty distribution) is sufficient to model the relationship between price and neighborhood income status based on F-tests (by item) constraining the dummy variables for the other quintiles to be zero. However, this is not true for oranges and lettuce, which show differences in other quintiles.

In short, a clear pattern of differences in mean gross prices between poor and affluent neighborhoods is not evident. However, this result may not be true throughout the price distribution and/or the variation in prices may differ across neighborhoods. For example, Goldberg [1996] finds no mean difference between the prices paid by white and black males for automobiles, but the variance in prices paid suggest the presence of price discrimination. To investigate relative dispersion across neighborhood types, I calculate the difference in the cumulative density functions of the neighborhood gross prices.[21] Figure 1 shows the dispersion in gross prices for poor and affluent neighborhoods for milk, chicken, and lettuce. The price distribution in poor neighborhoods appears to lie to the left of both the mean and the price distribution in affluent neighborhoods. The poor-affluent difference in the cumulative density functions (CDFs) is positive and significant for all items based on a F-test, indicating a greater likelihood of lower prices in poor neighborhoods.[22]

In summary, gross prices do not appear to be higher in poor neighborhoods. If anything, it appears that gross market prices may be lower on average in low-income neighborhoods. In the

[21] I calculate the kernel density estimates of the gross prices using the Epanechnikov kernel.
[22] The CDFs are measured by summing the item price density functions.

next section I attempt to explain these patterns in gross prices by differences in operating costs, quality, and consumer search. As I begin to add covariates in subsequent work below, I follow the literature and define poor neighborhoods as those in which more than 20 percent of the households are below the poverty level (i.e., the first 3 rows of Table 2).

III. Explaining the (non-)existence of a price gap

Table 3 isolates the relationship between price and poverty composition from other factors such as costs, quality, and consumer search by sequentially adding other covariates. The top panel of the table shows the average poor price differential. All regressions include a constant, month dummies, item and local area fixed effects. In addition, with the exception of column (1), all regressions include product size and a dummy variable for whether the price is a sale price. These latter variables comprise the Z_{ijst} in equation (1).

As indicated by comparing mean prices in Table 2, the poor pay approximately 6 percent less for these goods. Controlling for neighborhood demographics and crime increases the discount, while the other factors decrease the discount. Each change in the discount from the prior column occurs in the expected direction for all of the factors. For example, when discretionary costs are unaccounted for in column (5), the price differential is larger than when such costs are included as in column (6). The final column shows that after accounting for many factors to represent price dispersion, the net price gap is about the same size as the gross price gap but the difference is now insignificant.[23]

Overall, the results suggest that the price differential can be explained by differences in costs, as the statistical significance of the average poor price gap disappears after controlling for a

[23] I obtain similar results when I conduct the analysis with a continuous measure of neighborhood income.

number of cost factors. However, the robust negative price difference provides some evidence that third-degree (classical) price discrimination [Pigou 1920] may play a role in the explanation as well. This is because the price discount may be a result of firms reacting strategically to the higher demand elasticities of poor consumers with lower prices.[24] In addition, the results appear to be consistent with imperfect information models that relate relative prices paid to search costs (see e.g., Salop and Stiglitz, 1977).

IV. Item price dispersion

The analysis performed above estimated empirical specification (1) outlined in section 3, in which the item characteristics are constrained to be the same across neighborhood types. This procedure is most similar to what has been done in the literature and is the conventional approach for this type of survey data (see e.g., Primont and Kokoski 1990, 1991). Two main conclusions can be drawn from the results presented. First, the poor coefficients are generally negative or zero. Second, the covariates contribute significantly to explaining the variation in prices and the point estimates are fairly robust to their inclusion. However, as the number of variables in the models presented are large (e.g., there are 123 model degrees of freedom in the model estimated in column (9), Table 3), the pattern of interactions among the product characteristics is unknown. In order to allow the price differential and product characteristics to vary freely, I follow an alternative weighting strategy which allows each item price to fluctuate within each neighborhood type. I do this by estimating empirical specification (5) by item for each neighborhood type and deriving a weighted-sum of poor-price differentials from each model using the poor expenditure shares taken

[24] Hoch *et al* [1995] estimate food demand elasticities and find low-income consumers to be more sensitive to price changes.

from the CEX as weights.[25] Because I evaluate the price gap for poor and affluent neighborhoods separately (using the definition of poor noted above), I substitute a continuous measure of neighborhood income (e.g., median household income) for the poor indicator variables used in the above analysis.[26]

I show the averages with and without covariates for the homogeneous items in Table 4. The coefficient on median household income varies a great deal across items and by neighborhood type. For example, in column (2) milk prices increase significantly (25 percent) as income increases one unit, or 10,000 dollars, for the poor but column (12) shows no relationship between income and price for milk in affluent areas. In general, the variation across items is enlarged by the inclusion of the full set of covariates, showing that the importance of the covariates differs across items.

The bottom panel of the table calculates the weighted average of the income coefficients ($\overline{\beta}$), with and without covariates. The weighted-average of the income coefficients without covariates for the poor neighborhoods is essentially zero. The result of no relationship between the median household income of a neighborhood and market prices has been reported by a number of researchers (e.g., Groom [1966] and Alcaly and Klevorick [1971]). However, the weighted-average of the income coefficients with covariates for the poor neighborhoods is large and significantly different from zero, indicating that prices increase 42.7 percent as median household income increases by 10,000 dollars. This is in contrast to the weighted-coefficient in affluent neighborhoods, which is negative and insignificant. I interpret these results as suggesting that prices are more sensitive to income in poor areas, conditional on covariates. The difference

[25] *Consumer Expenditure Survey* [1997], Table 2.

[26] I use this measure in lieu of the continuous poverty rate because of collinearity problems in the poor neighborhood sample. Multicollinearity is encountered in this procedure because of the smaller sample sizes and relatively large set of covariates.

$\left(\bar{b}_{rich} - \bar{b}_{poor}\right)$ indicates that, given a 10,000 increase in the income level of a neighborhood, prices are more likely to increase.[27]

The weighted-coefficients $\bar{\beta}$ verify the earlier results. In more concentrated poverty areas, the poor do not face higher prices. Although the pricing strategies in relation to income differ across products, these strategies do not result in higher prices in the inner-city. One interpretation of this is that price discrimination may be a local phenomenon: stores may exploit poor consumers for some products but on average this practice does not result in the entire market basket price being higher than in affluent areas.

The results from this approach are not qualitatively different from those reported above, suggesting that the constraints are not substantively binding in specification (1). Therefore, I return to specification (1) to evaluate the influence of other factors unaccounted for above.

V. Other explanations

The results from the comparison of means and the above regression analysis indicate that the net price discount in poor neighborhoods ranges between zero and 6.1 percent. While informative, this is a large range so it is helpful to examine other possible explanations that may affect the net price differential. One such explanation is the format of the store. While the above analysis accounted for the possible economies of scale experienced by chain stores, I did not explicitly account for differences in prices across store formats, which may be substantial. For example, while both Wal-Mart and Kroger's are chain stores they practice very different pricing strategies.[28]

[27] I have also conducted this analysis on a larger, more heterogeneous sample of items. The results are qualitatively the same.

[28] Wal-Mart employs a everyday low price strategy, while Kroger focuses more extensively on a hi-low pricing strategy. This example should not be construed to imply that prices from either Wal-Mart or Kroger's are represented in the data.

Table 5 examines whether prices differ by the type of store. A superstore has at least 30,000 square feet in retail space and annual sales in excess of 12 million dollars. Superstores offer a variety of specialty departments and services. A conventional supermarket is any full-line, self-service grocery store with annual sales of 2 million or more.[29] The table reveals significant differences across store types for the homogenous items, with discounts in poor neighborhoods at superstores (23 percent). There are negligible differences at other store types. The evidence suggests that superstores located in low-income neighborhoods offer larger savings than conventional supermarkets.

Another explanation that may affect the net price differential is the location of a store in a central city. Since poor households (especially poor, black households) tend to live in central cities and higher prices in central cities is reported in prior work [McDonald and Nelson 1991], the 6.1 percent discount reported in the last column of Table 3 may be upward biased. Columns (4) and (5) of Table 5 show estimates of the net price differential for central city and suburban neighborhoods.[30] The differentials are insignificant in both the central city and suburbs, and not statistically different from one another. In contrast to MacDonald and Nelson [1991], I find no difference in the price differential between poor central city and poor suburban neighborhoods.

Though the regressions presented above allow for sampling area fixed effects, regional variation may also affect differentials, as in certain regions (e.g., the south) the cost of living is low and the poverty rate is high, possibly inducing a downward bias in the estimate of the poor price differential. I investigate regional variation in columns (6)-(9) of Table 5. Generally, prices are

[29] Store format information is based on data from Spectra Marketing, Inc. While I do not know the actual store format from the Spectra data for the stores in column (3) of the table, the operating names indicate that this column contains bakeries, delicatessens, vegetable stores, independent supermarkets, etc. Estimates for warehouses are omitted due to collinearity between the poor indicator variable and other covariates.

[30] I define an area as being located in a central city if the zip code of the store is located in a MSA declared a central city by the Office of Management and Budget in June 1996.

lower for the poor when the differences are allowed to vary by region and the largest discounts for the poor occur in the Midwest (26 percent) and the south (10 percent).

Another potentially important factor affecting the price differential may be the racial/ethnic composition of the neighborhood. Race/ethnicity may matter if retailers charge higher prices in certain neighborhoods because of the *perceived* higher cost of conducting business. This may result in Becker [1957] discrimination, in which retailers act as if $c_i(1+d_k)$ were the true marginal cost of providing item i, where d_k is the discrimination coefficient against group k.[31] In columns (10) -(13) I analyze the effect of racial and ethnic neighborhood composition on the poor price differential. Column (10) shows that neighborhoods primarily comprised of black residents have market prices similar to those in neighborhoods comprised mostly of affluent, white residents. Insignificant discounts in market prices are evident in neighborhoods where the predominant group is either white (non-Hispanic) or of Hispanic ethnicity.[32] In contrast, there is a large, significant discount in neighborhoods that are jointly black and Hispanic. This discount remains after accounting for the region, central city status, and store composition in the neighborhood. One possible explanation of this pattern is "reverse" Becker-type discrimination.[33] As stores in poor Hispanic, black neighborhoods are likely to be owned by Hispanics [Bates 1985], the price discount could reflect a preference to serve one's own community.

---

[31] Becker-style discrimination differs from classical price discrimination in that it results in differential prices for non-economic reasons; price differences are not driven by differences in the costs of service or differences in the intensity of demand of the discriminated group. This is clearly seen in the mark-up. The Becker-type discriminator prices according to:

$$p_{ijk} = \frac{c_i(1+d_k)}{1-(a_j/e)},$$

where a_j is firm j's market share and e is the elasticity of demand (which is the same for all consumers). Only d_k differentiates him from his competition and d_k is unrelated to marginal costs or demand intensities.

[32] Hispanic ethnicity can be of any race. I define a predominately Hispanic neighborhood in the table as one in which the race is predominately white.

[33] While classical price discrimination by race/ethnicity is possible, I assume that retailers cannot segment their market on the basis of race/ethnicity (i.e., race/ethnicity is not identifiable through sales receipts) and demand intensities do not vary by race. These two assumptions render discrimination by race/ethnicity less profitable than a nondiscriminatory

6. Caveats

There are several caveats that should be kept in mind when interpreting the above results. First, zip codes may not be a good proxy for the neighborhood since zip codes are delineated by the U.S. Postal Service to ensure efficient delivery of the mail. The boundaries are not intended to reflect the amenities and characteristics that may distinguish one neighborhood from another. A related potential problem is the use of county level segregation measures. In general, aerial units used to measure segregation are arbitrary and indices calculated from different units will differ in their correlation and magnitude--a problem known as aggregation bias. Since zip codes are larger than census tracts (another proxy for neighborhood) in urban areas, my indices of segregation may be smaller because of less racial and ethnic concentration and homogeneity [Massey and Denton 1988]. As a result, my estimates of the net price differential across racial and ethnic neighborhood types may be a lower bound of the true range.

The second source of bias arises from unmodeled mobility. I identify the price effects using intra-area variation, but I do not account for geographic mobility by households. If less mobile blacks choose to live in concentrated counties and mobility is positively correlated with price (which may occur because more mobile people may be younger, wealthier, or may have a higher value for amenities such as specialty coffees), then my estimate of the net price differential may be downward biased. In this case, poor black neighborhoods may have higher prices than affluent white neighborhoods. Cross-neighborhood shopping may induce a related source of potential bias. Since consumers can theoretically shop anywhere, the relationship between neighborhood income and price may be weakened. One possible fix for this problem of measurement error in the poverty status variables is to use the average income of the store's patronizing consumers as the

---

pricing strategy. In contrast, classical price discrimination is at least as profitable as a nondiscriminatory policy due to the extraction of consumer surplus [Phlips 1983].

income measure. While the BLS sample frame represents where consumers shop, the information linking the income of the correspondent to the store patronized is not currently available.

The third source of bias arises from unmodeled product availability. As the BLS survey does not explicitly account for product breadth in its multi-level probability of selection technique, I cannot control for this source of variation in item prices. This is a potentially important omission because the lack of variety can be modeled as a price premium. Further, since smaller stores necessarily lack extensive variety and smaller stores are more likely to be located in poor and predominantly black and Hispanic neighborhoods, my net price differential may be downward biased. However, I believe that the full effect of this source of bias is mitigated by my inclusion of store quality characteristics since larger stores offering a mix of auxiliary services are more likely to have large varieties of every product.

Finally, the analysis sample is comprised largely of perishable items for which quality differences may exist despite the characteristic controls. However, the sample composition does not appear to drive the main results as I have conducted the analysis in Table 4 across a mixture of items.[34] The analysis shows that the relationship between the prices of fresh fruit and meat and income is not systematically different from that of the non-perishable items.

7. Conclusion

This analysis looks at price differences and attempts to estimate whether there is a price gap between poor and affluent neighborhoods both before and after controlling for factors such as costs, quality, and consumer search. Although there are many ways to define the poor, I find that independent of classification, the most deprived neighborhoods in the U.S. do not face higher

[34] I conduct the analysis on a larger, more heterogenous sample of items that include non-carbonated juice, coffee, and flour.

market prices for goods. In fact, I find that the poor face discounted net prices that can be as much as 6 percent lower than those faced by the more affluent. In addition, I find that store format is an important determinant of price differences, with large discounts available for the poor at superstores. Although I analyze only 5 items, the results likely extend to a variety of foods that are prepared at home by virtue of the BLS survey strategy.

The price gap appears to be most consistent with price dispersion generated by various costs to both the consumer and the firm. Quality differences and consumer search, in particular, go far in explaining price differences between stores both within and between different neighborhood types. While price dispersion accounts for much of the observed variation in prices, it does not completely explain the price gap as significant discounts remain for some poor subgroups. One possible explanation for the robust significant discounts is classical price discrimination. The results presented are consistent with food stores offering discounts to those consumers who may have greater price elasticities of demand.

Finally, while poor (non-Hispanic) white and predominantly Hispanic, black neighborhoods have market prices considerably lower than those in affluent (non-Hispanic) white neighborhoods for some items, net prices in poor, predominantly black neighborhoods do not significantly differ. This result is robust to the definition of poor and to the item. The lack of a difference by race may reflect omitted variables (e.g., the race of the store owners, consumer mobility, etc.) or suggest there is no difference.

8. References

Alcaly, R. (1976). "Information and Prices," *Bell Journal of Economics*, 7, 658-667.

Alcaly, R., Klevorick, A. (1971). "Food Prices in Relation to Income Levels in New York City," *Journal of Business* 44, 380-397.

Ambrose, D. (1979). "Retail Grocery Pricing: Inner-City, Suburban, and Rural Comparisons", *Journal of Business* 52, 95-102.

Bates, T. (1985). "An Analysis of Minority Entrepreneurship: Utilizing the Census of public Use Samples." (Burlington, VT: *Fourth Progress Report on MBDA Contract*).

Becker, G. (1957). *The Economics of Discrimination*. (Chicago, IL: University of Chicago Press).

Bell, W. (1954). "A probability Model for the Measurement of Ecological Segregation," *Social Forces*, 32, 357-364.

Blau, P. (1977). *Inequality and Heterogeneity*: A Primitive Theory of Social Structure. (Free Press).

Blodgett, J. and CIESIN (1998). *MABLE/Geocorr V2.5 Geographic Correspondence Engine* [Computer File].

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.

Carlson, J. and Gieseke, R. (1983). "Price Search in a Product Market," *Journal of Consumer Research*, 9, 357-65.

Cutler, D., Glaeser, E., and Vigdor, J. (1997). "The Rise and Decline of the American Ghetto," NBER Working Paper 5881.

Diamond, P. (1987). "Consumer Differences and Prices in a Search Model," *Quarterly Journal of Economics*, 102(2), 429-436.

Duncan, O. and Duncan, B. (1955). "A Methodological Analysis of Segregation Indices," *American Sociological Review*, 20, 210-217.

Finke, M., Chern, W., and Fox, J. (1997). "Do the Urban Poor Pay More for Food? Issues in Measurement," *Advancing the Consumer Interest*, 9, 13-17.

Food Industry Review (1998). (New Jersey: The Food Institute).

Food Marketing Institute (1997). "Key Industry Facts," http://www.fmi.org/keyfacts/ index.html.

Frankel, D. and Gould, E. (1999). "The Retail Price of Inequality," Working Paper, Tel Aviv University.

Geithman, F. and Marion, B. (1993). "Testing for Market Power in Supermarket Prices: A Review of the Kaufman-Handy/ERS Study," in Ronald Cotterill (ed.), *Competitive Strategy Analysis in the Food System*, (Boulder, CO: Westview Press, Inc.), pp. 253-291.

Goldberg, P. (1996). "Dealer Price Discrimination in New car Purchases: Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 104(3), 622-654.

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.

Hayes, L. (forthcoming). "Do the Poor Pay More in New York City?" *Journal of Consumer Policy*.

_____ (2000). "Estimates of the Food Price Differential Between Poor and Affluent Neighborhoods," Princeton University mimeo.

Hoch, S., Kim, B., Montgomery, A. and Rossi, P. (1995). "Determinants of Store-Level Price Elasticity," *Journal of Marketing Research*, Vol. XXXII, 17-29.

Kaufman, P. and Handy, C. (1989). Supermarket Prices and Price Differences: City, Firm, and Store-level Determinants. Economic Research Service, United State Department of Agriculture, Technical Bulletin 1776. (Washington, D.C.: Government Printing Office).

Kinsey, J. (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.

Lane, W. (1996). "Changing the Item Structure of the Consumer Price Index," *Monthly Labor Review*, December 1996, 18-25.

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).

Massey, D. and Denton, N., (1988). "The Dimensions of Residential Segregation," *Social Forces*, 67(2), 281-315.

Massey, D. and Denton, N., (1993). *American Apartheid: Segregation and the Making of the Underclass*. (Cambridge, MA: Harvard University Press).

Mincer, J. (1963). "Market Prices, Opportunity Costs, and Income Effects," *in Measurement in Economics*. (Stanford, CA: Stanford University Press).

Nelson, J. (1970). "Information and Consumer Behavior," *Journal of Political Economy*, 78(2), 311-329.

Phlips, L. (1983). *The Economics of Price Discrimination*. (Cambridge, UK: Cambridge University Press).

Pigou, A. (1920). *The Economics of Welfare*. (London, UK: Macmillan).

Pratt, J., Wise, D., and Zeckhauser, R. (1979). "Price Differences in Almost Competitive Markets," *Quarterly Journal of Economics*, 93(2), 189-211.

Primont, D. and Kokoski, M. (1990). "Comparing Prices Across Cities: A Hedonic Approach," BLS Working Paper No. 204 (Washington, DC:U.S. Bureau of Labor Statistics).

Primont, D. and Kokoski, M. (1991). "Differences in Food Prices across U.S. Cities: Evidence From CPI Data," BLS Working Paper No. 209 (Washington, DC:U.S. Bureau of Labor Statistics).

Progressive Grocer (1999). "66th Annual Report of the Grocery Industry," (New York: Trade Division, Butterwick, Co.).

Rothschild, M. (1974). "Searching for the Lowest Price When the Distribution of Prices Is Unknown," *Journal of Political Economy*, 82(4), 689-711.

Salop, S. and Stiglitz, J. (1977). "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," *Review of Economic Studies*, 44, 493-510.

Stigler, G. (1961). "The Economics of Information," *Journal of Political Economy*, 69, 213-225.

Supermarket Business (1998). (New York: Gralla Publications).

Troutt, D., (1993). *The Thin Red Line: How the Poor Still Pay More*. (San Francisco, CA: West Coast Regional Office, Consumers Union of the US, Inc.).

United States Department of Agriculture (1996). *Food Marketing Review 1994-1995*. Economic Research Service, Agricultural Economics Report No. 743. (Washington, D.C.: Government Printing Office).

United States Department of Justice, Federal Bureau of Investigation (1997). *Uniform Crime Reporting Program Data [United States: Agency-Level Offenses Known and Clearances by Arrest, 1997]* [Computer File]. ICPSR 9028 ed. (Ann Arbor, MI: Inter-university Consortium for Political and Social Research) [producer and distributor].

United States Department of Labor, Bureau of Labor Statistics (1998). *Consumer Price Index C&S Initiation Data Collection Manual*. (Washington, DC: Government Printing Office).

United States Department of Labor, Bureau of Labor Statistics (1997). *The Consumer Expenditure Survey*. (Washington, DC: Government Printing Office).

United States Department of Labor, Bureau of Labor Statistics (1997). *BLS Handbook of Methods*, Chapter 17.

Varian, H. (1980). "A Model of Sales," *American Economic Review*, 70, 651-659.

Williams, J. (1996) " The Redesign of the CPI Geographic Sample," *Monthly Labor Review*, December 1996, 10-17.

Wilson, W. (1987). *The Truly Disadvantaged: The Inner-city, the Underclass, and Public Policy*. (Chicago, IL: University of Chicago Press).

9. Data Appendix

Commodities and Services Database (C&S)

The Commodities and Services Database (C&S) contains the prices collected by the Bureau of Labor Statistics (BLS) for use in the compilation of the Consumer Price Index (CPI). The CPI is designed to measure the (monthly) change in prices of goods and services purchased by typical urban American consumers. It is calculated by comparing the cost of a fixed set of goods and services at current prices with the cost of an identical market basket at prices prevailing during a reference period. Price data are collected from a survey of stores without regard to coupon use or special discounts (i.e. senior citizen and quantity discounts). However, the survey does collect sale prices when they are available to every consumer.

The BLS maintains price information at several levels of aggregation. See Appendix Figure 1 for a schematic diagram. Prices are available for eight major groups: food and beverages; housing; apparel; transportation; medical care; recreation; education and communication; and other goods and services. These groups represent the highest level of price aggregation. Within each major group items are arranged according to expenditure classes (e.g., cereal and cereal products, bakery products, etc.) in order to group like products and to allow imputation of price change for the CPI when actual prices are unavailable. Within each expenditure class are item strata (e.g., flour and prepared flour mixes, cereal, etc.), which are generally a group of products that are expected to have similar price movements (Lane (1996), p. 19). The lowest level of aggregation is entry-level items--the products surveyed in stores (e.g., flour, cereal, rice, etc.). The C&S Database contains price information for the 69 expenditure categories that comprise the 8 major groups, which in turn are divided into 207 item strata and 364 ELIs (*BLS Handbook of Methods* (1997), p. 178). Price data are collected monthly for the food and beverages major group and bimonthly for all other major groups.

The price survey samples 87 geographic areas referred to as primary sampling units (PSUs) which comprise most of the contiguous states, as well as Alaska and Hawaii *(BLS Handbook of Methods* (1997), p.177). Montana, North Dakota, and Wyoming are not sampled. The frame of outlet respondents is obtained from the Telephone Point-of-Purchase Survey (TPOPS), an unpublished supplemental survey conducted by the Bureau of the Census under contract with the BLS. The TPOPS questions qualified households about where they typically shop for a number of commodities and the amount they expend. Eligible households for the TPOPS include all civilian, non-institutional persons, including persons residing in boarding houses, housing facilities for students and workers, mobile home parks, permanent-type living quarters in hotels and motels, and staff residing in institutions (*BLS Handbook of Methods* (1997), p. 179). The probability of outlet selection for the price survey is proportional to consumer expenditures derived from the TPOPS. The unique item (i.e., brand, size, etc.) is chosen through disaggregation--a multistage probability sampling procedure in which all goods within an entry-level item category are given a probability for selection in proportion to their dollar sales in the store (U.S. Department of Labor, Consumer Price Index C&S Initiation Data Collection Manual, January 1998, Chapter 6, p. 1).

As a result of this sampling strategy, each item strata is surveyed in every PSU, but different unique items are selected in each store. Thus, the market basket differs across sampling units precluding the comparison of identical market basket prices in this study. I use the itemized specification list completed after disaggregation to create indicator variables for the hedonic regressions I compute.

Analysis Sample

The data used in this study are derived from the C&S Database for the food and beverages major group. The sample includes 19 expenditure categories, 65 item strata, and 92 entry-level items. Data were extracted for the period covering January 1998 to December 1998 for the food at home item strata. All prices are analyzed at the entry-level item level in this paper. The sample averages approximately 42,500 monthly price observations from 4,790 stores. The sample contains only the last nine monthly observations of round steak due to a coding error and subsequent archiving of the data. These omissions exclude 4,740 observations---less than 1 percent of the extracted sample.

I use several other criteria to limit the sample. First, I select only observations that are available for use in the final compilation of the CPI. This eliminates prices collected for evaluation purposes such as experimental indices, reducing the sample by 104,295 observations or 22 percent. Second, I eliminate price quotes that are flagged as "awaiting central office clearance," "temporarily unavailable to be priced in outlet," "out-of-season," "outlet status unknown," or "deletion of price quote pending" to insure price quotes are obtained from established outlets. This affects 12 observations. I further limit the sample to outlets for which food may be purchased for home preparation and consumption. This criterion excludes food service establishments such as restaurants, cafeterias, and food vending machines, effecting 30,327 observations. Overall, these combined criteria reduce the initial sample size by 134,544 observations or approximately 30 percent. In this paper I focus on the following items only: flour, white bread, ground beef, pork chops, whole chicken, eggs, milk, bananas, oranges, potatoes, lettuce, salad, and non-carbonated juice. My usable sample from the BLS consists of 63,557 observations from 2,181 outlets across 1,813 zip codes in 43 states. Survey data are collected from outlets and respondents on a voluntary basis and are confidential. In adherence to this confidentially, I do not reveal outlet names and product brands in this analysis.

Marketing Data

Because the C&S Database contains only cursory data on outlets, specifically name, address, and phone number, I supplement it with more extensive data from SPECTRA, Inc., a private marketing firm in Illinois. The SPECTRA sample consists of 19,836 observations on supermarkets, grocery stores, and large-scale discounters from every state, except Alaska, Delaware, Kansas, New Mexico, Oklahoma, Rhode Island, and South Dakota. These data are matched to the C&S Database sample using a statistical matching technique provided by AUTOMATCH software. AUTOMATCH matches using a probabilistic algorithm. The file match utilized in this study is accomplished in four passes through the data, with the first pass matching on zip code and the second matching on the soundex (a four-digit alphanumeric code that represents the phonetic pronunciation) of the parsed outlet name. The third pass matches on the soundex of the parsed street name of the outlet, while the final pass matches on a combination of the dwelling number and state of the outlet. In this manner 1721 or 79 percent of BLS outlets were matched. A dummy variable equal to 1 if the BLS outlet lacks characteristic data is included in all regressions using the SPECTRA data. Missing data are imputed utilizing a hot-deck procedure.

I derive measures of competition from the number of outlets in a zip code using another database obtained to InfoUsa, Inc., a marketing company in Nebraska. An outlet is defined as operating in a "competitive" environment if more than 5 outlets operate in its zip code. It is defined as being an oligopolist if it does not operate in a monopoly, duopoly, or competitive environment.

Shopper density is defined as the total population per zip code divided by the number of stores in a zip code using the InfoUsa sample.

Segregation indices

Demographic data for this project were collected from a variety of sources. Demographic data by zip code were obtained from the 1990 Census utilizing Summary Tape File 3B. Segregation indices were calculated using zip code data aggregated to the county level. The dissimilarity and isolation indices are based on formulations outlined in Massey and Denton [1988]. The dissimilarity index is calculated as:

$$D = \sum_{z=1}^{Z} \frac{t_z |p_z - P|}{2TP(1-P)}, \tag{A1}$$

where t_z and p_z are the total population and subgroup proportion in zip code z, and T and P are the total population and subgroup proportion in the county. The index varies between 0 and 1 and measures the proportion of subgroup members that would have to change their area of residence to ensure an even distribution of groups in the county. Index values above 0.6 are considered large [Massey and Denton 1993].

The isolation index measures the likelihood that subgroup members come into contact only with other subgroup members. The index is calculated as:

$$_xP_x^* = \sum_{z=1}^{Z} \frac{(x_z/X)}{(x_z/t_z)}, \tag{A2}$$

where x_z is the count of subgroup members in zip code z and X is the total number of subgroup members county-wide. This index also varies from 0 to 1, with higher index values indicating greater isolation from the majority population.

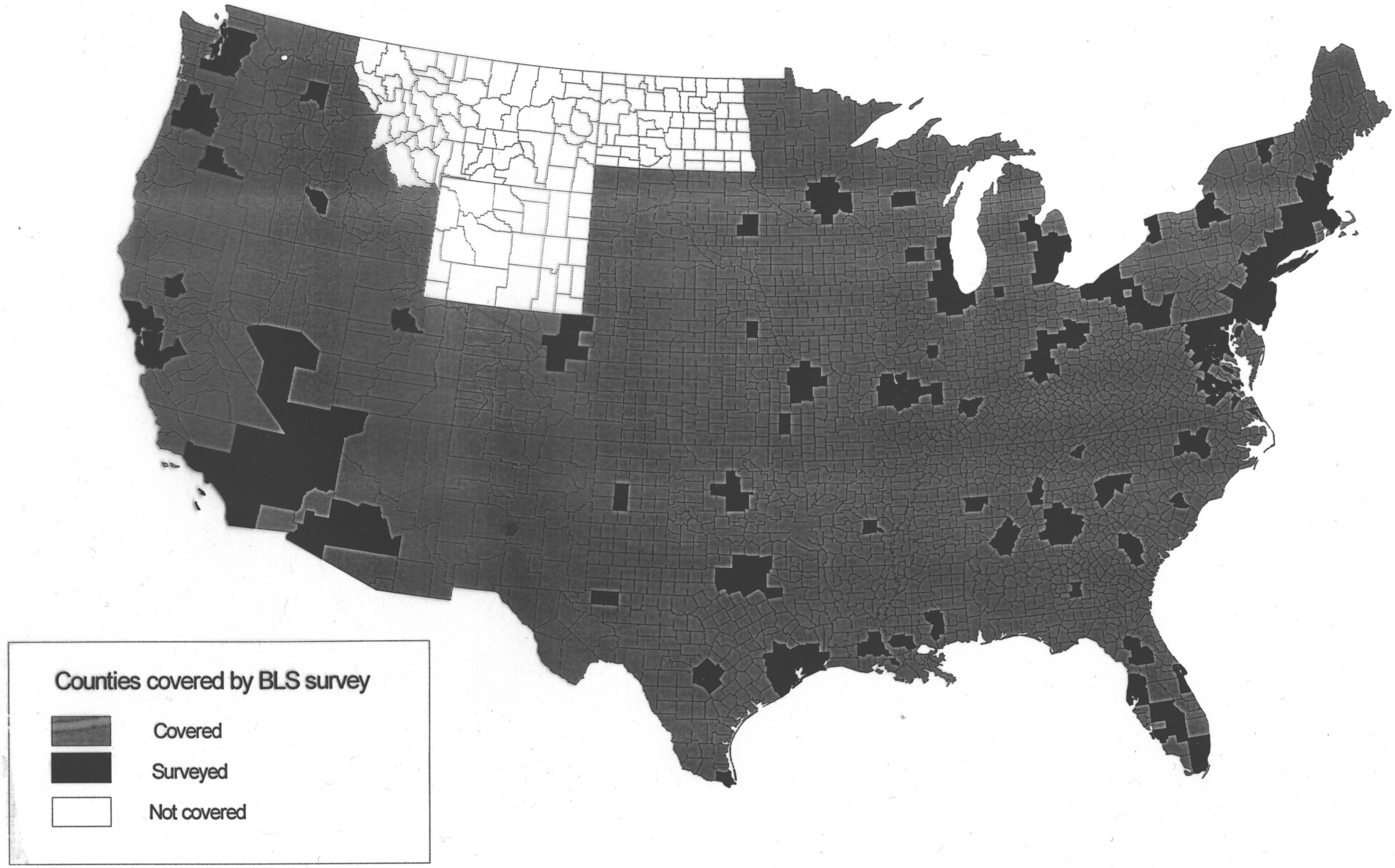
While the indices appear very similar they are conceptually distinct because it is possible to simultaneously have a low index of dissimilarity and a high index of isolation. This would occur if the subgroup members were a relatively large proportion of the zip code, but experience very little contact with majority members (Blau, 1977).

The indices used in this paper may be fairly low due to my choice of the zip code as the aerial unit. The spatial unit of observation largely determines the magnitude of the segregation index. This is because smaller aerial units (e.g., census tracts) may be more homogenous, which generally yields higher indices of segregation.

Demographic Data

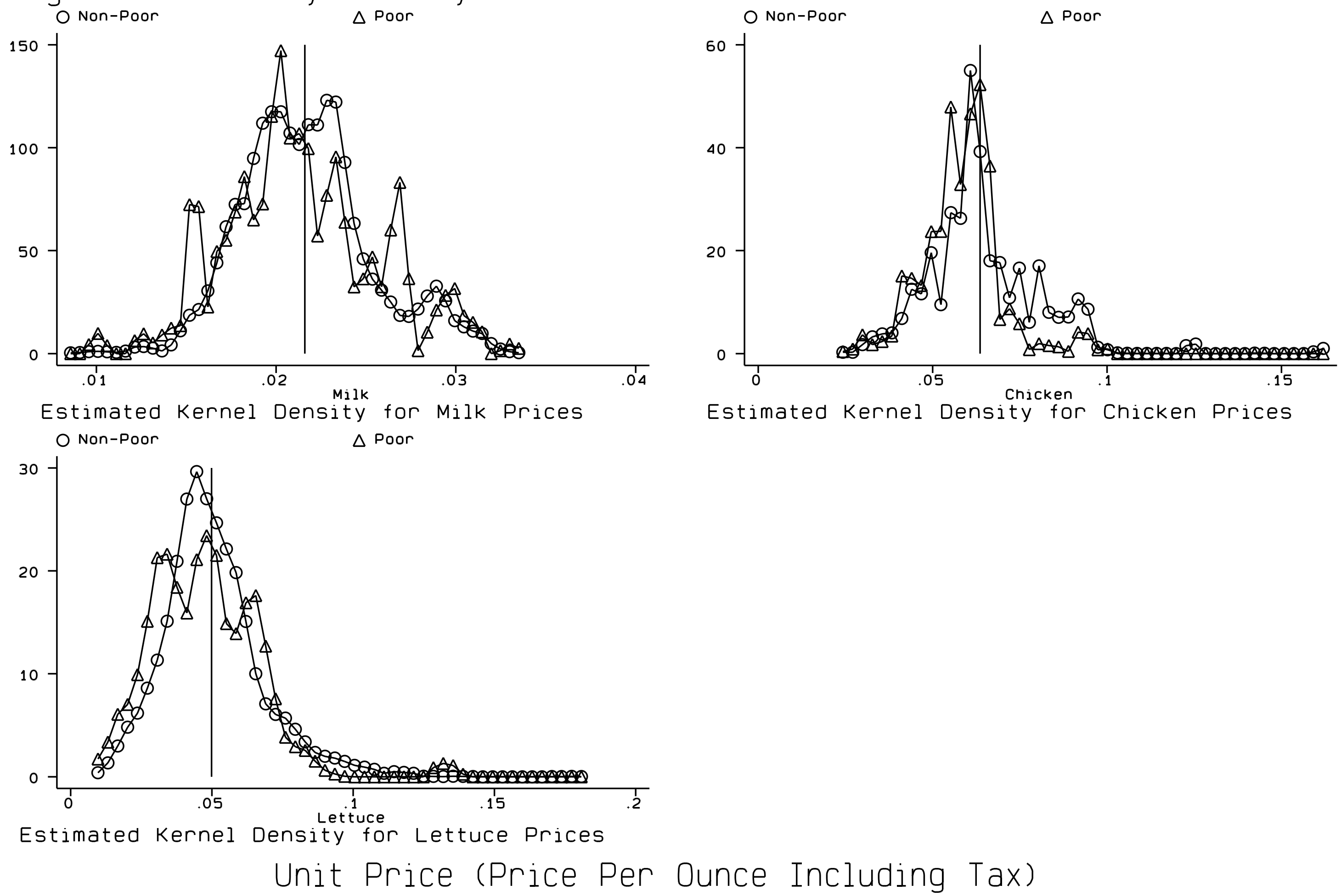
Expenditure shares are taken from Table 2 of the *Consumer Expenditure Survey* (1997). Information on central city status was obtained from U.S. Census Bureau (http:// www.census.gov/ geo). Data on land area by zip code were obtained from the MABLE/Geocorr V2.5 geographic correspondence engine (http://www.census.gov/plue/geocorr). Agency-level crime information by zip code was compiled from the Uniform Crime Reports (ICPSR Study No. 9028) provided by the U.S. Department of Justice, Federal Bureau of Investigation and data files provided from Marianne Bertrand (University of Chicago) and Brian Doyle (Federal Reserve Board of Governors). Crime counts are assigned to neighborhoods by zip code. All demographic variables are matched to the price sample by zip code.

BLS Primary Sampling Units (PSUs) in 1998



Source: Monthly Labor Review, Appendix 2, December 1996.

Figure 1 - Distribution of Unit Prices of Selected Homogenous Items by Poverty Class



Distribution Comparisons by Poverty Class.

Figure 2 - Distribution of Unit Prices of Selected Homogenous Items by Racial Composition

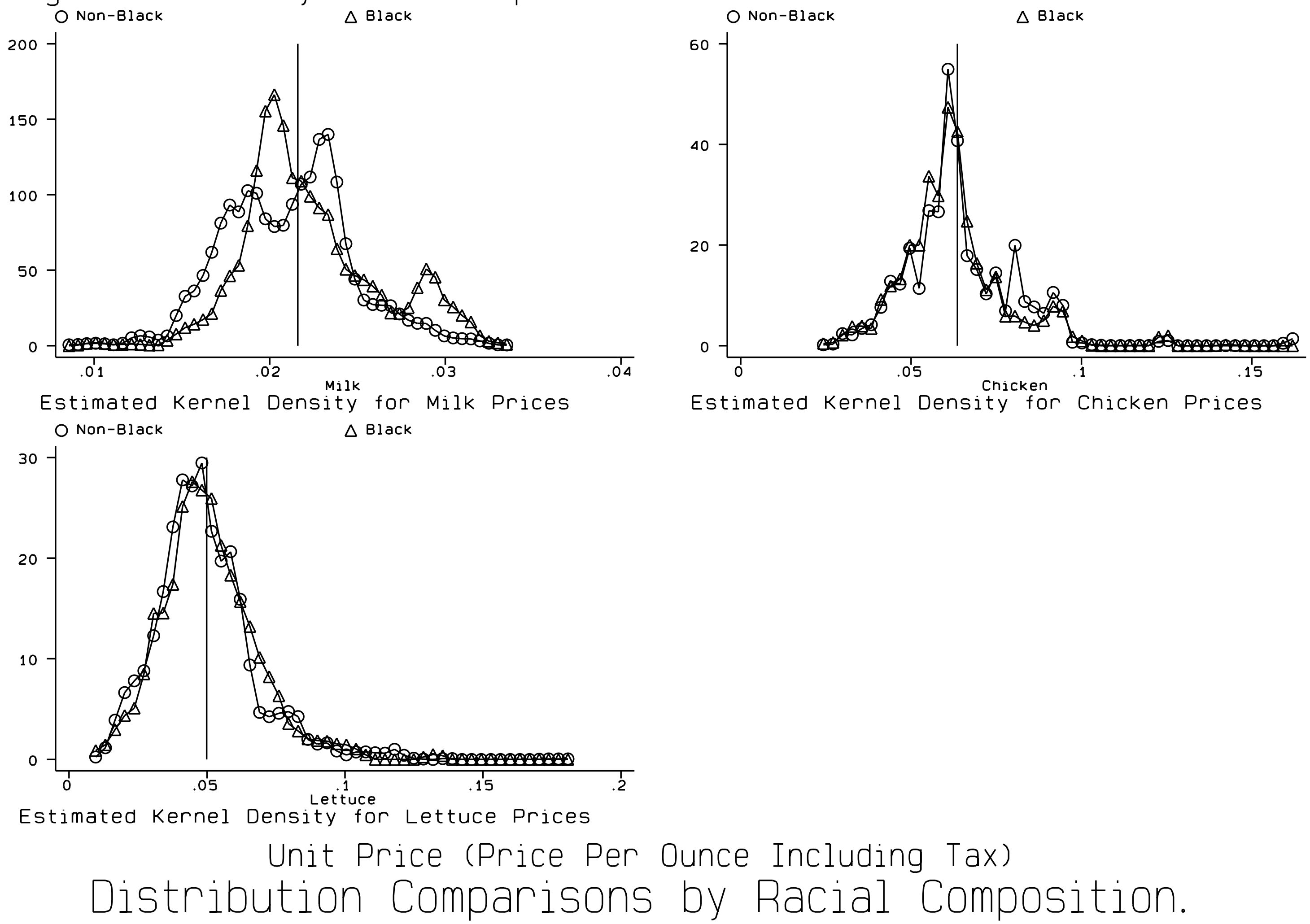
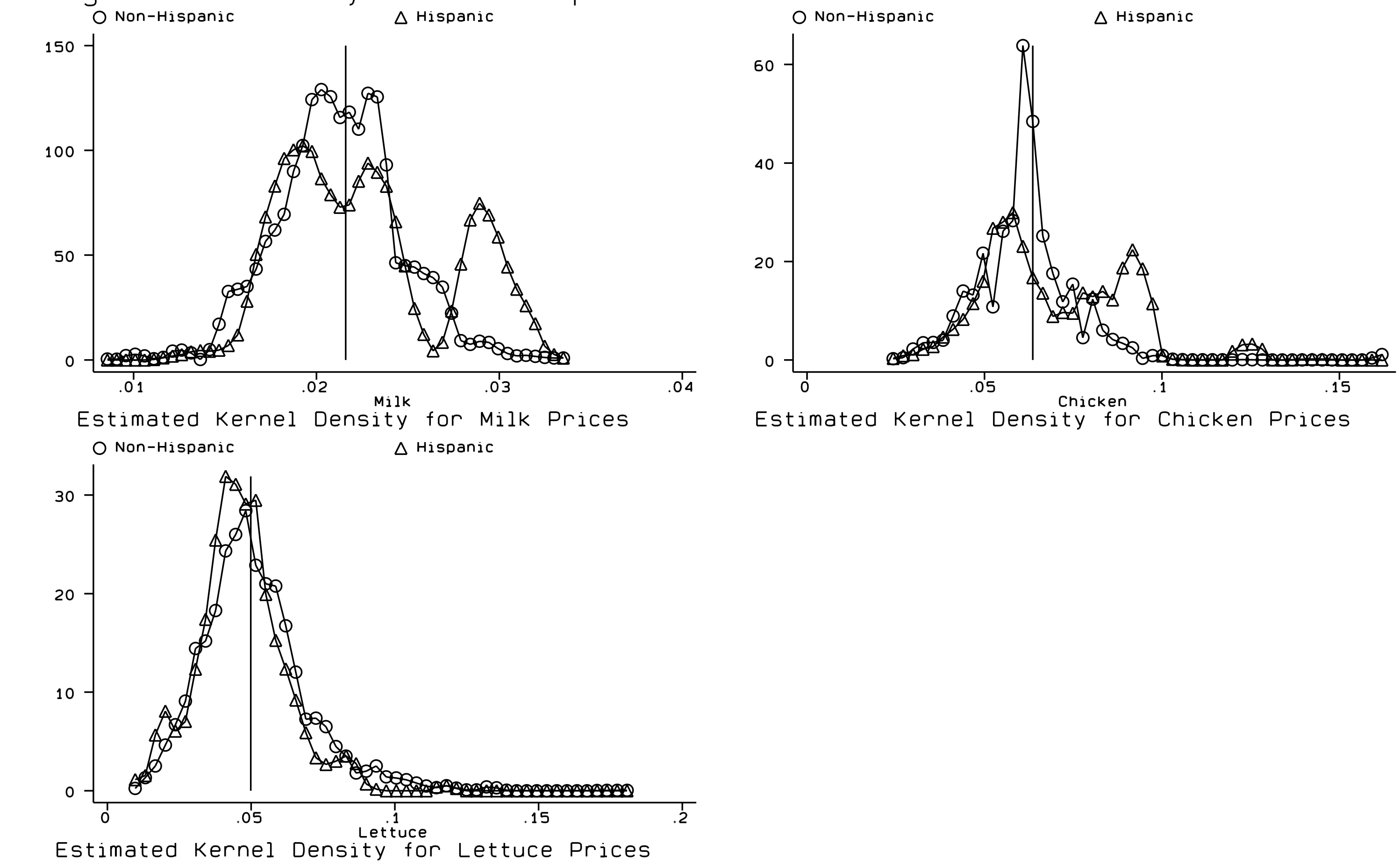


Figure 3 - Distribution of Unit Prices of Selected Homogenous Items by Ethnic Composition



Unit Price (Price Per Ounce Including Tax)

Distribution Comparisons by Ethnic Composition.

Appendix Figure 1: Example of CPI Item Structure

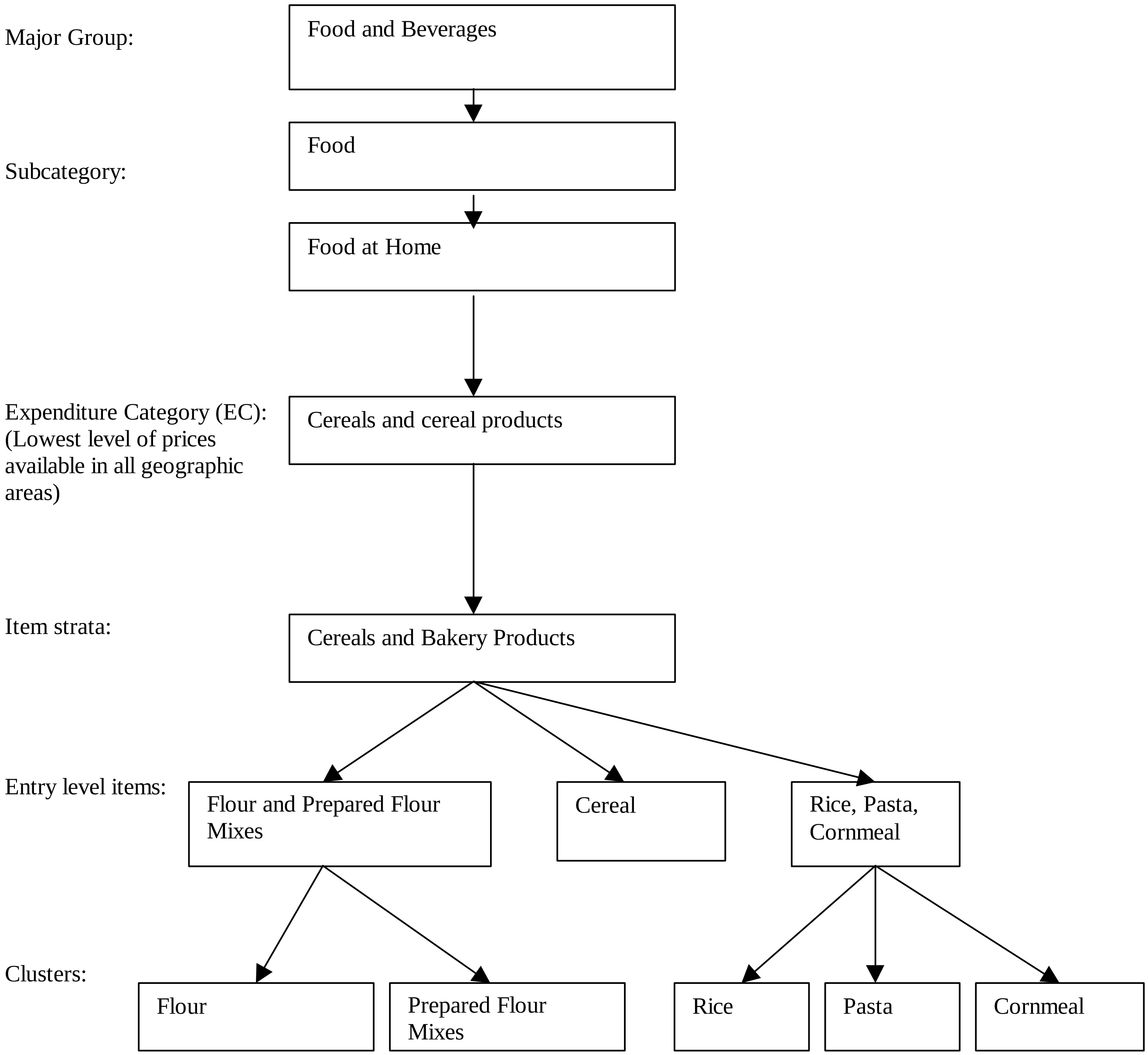


Table 1. Descriptive statistics of independent variables: Comparison of BLS sample means of search costs, neighborhood demographics, store costs, and market structure to all neighborhoods in the U.S., Spectra-available data subsample, and BLS sample by poverty status

Variables	All neighborhoods in the U.S.[1] (1)	All BLS sample outlets (2)	BLS sample outlets for which store characteristics are available (3)	BLS Sample outlets located in neighborhoods where the proportion in poverty exceeds 20 percent (poor) (4)	BLS Sample outlets located in neighborhoods where the proportion in poverty is less than 20 percent (non-poor) (5)
Proportion of population below the poverty level	0.132 [0.099]	0.123 [0.099]	0.125 [0.100]	0.312 [0.086]	0.089 [0.050]
Proportion of counties with Black index of dissimilarity>0.6 and Index of Isolation>0.3[2]	0.019 [0.136]	0.320 [0.466]	0.361 [0.480]	0.400 [0.490]	0.305 [0.461]
Proportion of counties with Black index of Isolation>0.3	0.164 [0.370]	0.507 [0.500]	0.563 [0.496]	0.634 [0.482]	0.485 [0.500]
Proportion of counties with Hispanic index of dissimilarity>0.6 and Index of Isolation>0.3	0.009 [0.093]	0.210 [0.407]	0.223 [0.416]	0.312 [0.463]	0.192 [0.394]
Proportion of counties with Hispanic index of Isolation>0.3	0.052 [0.223]	0.308 [0.462]	0.323 [0.468]	0.530 [0.499]	0.268 [0.443]
<u>Search</u>					
Proportion of households without a vehicle	0.113 [0.126]	0.135 [0.156]	0.145 [0.166]	0.300 [0.216]	0.105 [0.121]
Number of stores in neighborhood (per square mile)*Poverty rate	n/a[3]	1.074 [1.298]	1.150 [1.362]	2.877 [1.923]	0.748 [0.793]
Proportion of population completing up to grade 12, no diploma	0.250 [0.127]	0.238 [0.128]	0.243 [0.130]	0.402 [0.151]	0.209 [0.097]
Proportion of population obtaining high school diploma, including GED	0.302 [0.082]	0.279 [0.074]	0.279 [0.074]	0.249 [0.071]	0.284 [0.073]
Proportion of population completing some college, no degree	0.206 [0.064]	0.215 [0.060]	0.212 [0.057]	0.190 [0.094]	0.220 [0.050]
Proportion of population with associate, bachelor, or graduate/professional degree	0.241 [0.127]	0.267 [0.129]	0.266 [0.130]	0.159 [0.096]	0.287 [0.124]
<u>Neighborhood demographics</u>					
Proportion of unoccupied housing units	0.087 [0.076]	0.066 [0.046]	0.065 [0.044]	0.100 [0.051]	0.061 [0.042]
Population density (per square mile, per zip code, ÷ 1,000)	3.938 [9.436]	7.135 [14.642]	8.063 [15.976]	13.156 [19.652]	6.048 [13.253]
Proportion located in a central city	0.381 [0.486]	0.504 [0.500]	0.500 [0.500]	0.753 [0.431]	0.460 [0.498]
Proportion non-Hispanic White	0.756 [0.262]	0.697 [0.279]	0.683 [0.290]	0.348 [0.293]	0.761 [0.225]
Proportion non-hispanic Black	0.119 [0.198]	0.125 [0.199]	0.134 [0.209]	0.317 [0.308]	0.090 [0.146]
Proportion of Hispanic origin	0.089 [0.161]	0.124 [0.192]	0.129 [0.197]	0.296 [0.315]	0.093 [0.139]
<u>Crime</u>					
Total property crime per capita	n/a	0.072 [0.144]	0.071 [0.138]	0.126 [0.286]	0.062 [0.096]
Total crime per capita	n/a	0.079 [0.166]	0.078 [0.161]	0.140 [0.331]	0.068 [0.109]
<u>Operating Costs</u>					
Number of checkouts	8.700 ----	9.822 [4.531]	9.770 [5.011]	8.739 [4.398]	10.017 [4.527]
Number of full-time employees	n/a	39.862 [21.935]	38.814 [24.142]	34.916 [23.365]	40.755 [21.546]
Number of part-time employees	n/a	56.279 [28.758]	53.113 [30.942]	46.625 [32.390]	58.022 [27.697]

-- (continued) --

(Table 1, continued)

Variables	All neighborhoods in the U.S.[1] (1)	All BLS sample outlets (2)	BLS sample outlets for which store characteristics are available (3)	BLS Sample outlets located in neighborhoods where the proportion in poverty exceeds 20 percent (poor) (4)	BLS Sample outlets located in neighborhoods where the proportion in poverty is less than 20 percent (non-poor) (5)
Proportion chain stores	0.621	0.722	0.743	0.593	0.746
	----	[0.448]	[0.437]	[0.491]	[0.436]
Proportion with scanning equipment	n/a	0.777	0.774	0.737	0.784
		[0.378]	[0.418]	[0.411]	[0.371]
Discretionary costs					
Proportion using circulars	0.899	0.855	0.857	0.863	0.854
	----	[0.295]	[0.327]	[0.288]	[0.297]
Proportion using in-store demonstrations	0.664	0.624	0.617	0.601	0.628
	----	[0.416]	[0.460]	[0.425]	[0.414]
Proportion doubling coupons	0.437	0.367	0.367	0.273	0.383
	----	[0.412]	[0.456]	[0.374]	[0.417]
Proportion with frequent shopper program	0.215	0.213	0.213	0.173	0.220
	----	[0.353]	[0.390]	[0.321]	[0.358]
Proportion using in-store coupons	0.829	0.717	0.716	0.689	0.722
	----	[0.383]	[0.424]	[0.401]	[0.380]
Quality					
Proportion with from-scratch bakery	0.466	0.692	0.688	0.564	0.715
	----	[0.394]	[0.436]	[0.436]	[0.381]
Proportion with delicatessen	0.772	0.851	0.848	0.740	0.872
	----	[0.302]	[0.334]	[0.398]	[0.277]
Proportion with butcher department	0.604	0.719	0.721	0.706	0.722
	----	[0.380]	[0.421]	[0.391]	[0.378]
Proportion with seafood department	0.433	0.662	0.662	0.522	0.687
	----	[0.403]	[0.445]	[0.437]	[0.391]
Proportion with pharmacy	0.264	0.379	0.374	0.279	0.397
	----	[0.420]	[0.465]	[0.385]	[0.424]
Proportion with full-service bank	0.221	0.226	0.225	0.197	0.231
	----	[0.263]	[0.291]	[0.190]	[0.274]
Proportion with automatic teller machine	0.618	0.701	0.704	0.589	0.722
	----	[0.387]	[0.428]	[0.430]	[0.375]
Proportion offering check cashing services	0.614	0.585	0.588	0.547	0.591
	----	[0.421]	[0.466]	[0.431]	[0.419]
Proportion with warehouse aisles	0.160	0.122	0.125	0.085	0.129
	----	[0.266]	[0.294]	[0.218]	[0.273]
Market structure					
Number of stores per zip code (per square mile)	n/a	1.769	2.029	2.935	1.559
		[6.029]	[6.604]	[6.635]	[5.888]
Population per store (÷ 1,000)	n/a	6.753	6.566	6.304	6.834
		[6.077]	[5.985]	[4.622]	[6.300]
Grocery selling area (in square feet, ÷ 1,000)	27.341	31.173	30.882	26.592	32.000
	----	[15.013]	[16.595]	[16.896]	[14.494]
Yearly sales volume (in dollars, ÷ 1,000,000)	$11.328	$11.659	$11.600	$10.630	$11.846
	----	[10.423]	[11.530]	[9.507]	[10.570]
Market share (share of yearly sales volume)	n/a	0.328	0.316	0.301	0.333
		[0.225]	[0.247]	[0.221]	[0.225]
Number of observations	28619	2181	1728	308	1873

Note: Standard deviations are in brackets. Neighborhood refers to the postal zip code. Means are weighted by the total population in the zip code. The BLS does not survey Montana, North Dakota, and Wyoming, therefore these states are omitted from the above table. In addition, I do not have cost and quality data for stores in Arkansas, Delaware, Kansas, and Rhode Island. The number of observations in the BLS sample is the number of unique outlets. There are multiple outlets in a zip code.

[1]Average store characteristic data for the U.S. are taken from the *Supermarket Census* [Trade Dimensions 1998].

[2]Segregation indices are calculated at the county level. The figure reported for the U.S. is the average of segregation indices in 3,005 counties. Racial indices refer to non-Hispanic Blacks.

[3]Data not available.

Source: Author's calculations using the 1990 Census STF3B, Bureau of Labor Statistics survey data, the *Supermarket Census* [Trade Dimensions 1998], and data obtained from Spectra Marketing, Inc.

Table 2. Mean price differences of homogenous items by alternative definitions of poor

Group status definition:		One Gallon of Vitamin D fortified, whole milk (1)	Nonkosher, broiler/fryer whole chicken, per pound (2)	One-dozen, large, Grade A white eggs (3)	Loose Navel oranges, per pound (4)	Individually packaged Iceberg lettuce, per pound (5)	Total market basket price (weighted) (6)
Proportion in which 20 percent or	nonpoor:	2.775	1.036	1.090	0.832	0.805	1.321
more of residents are below the		(0.042)	(0.018)	(0.025)	(0.022)	(0.017)	(0.012)
poverty level	poor:	2.732	0.933	1.020	0.680	0.758	1.235
		(0.090)	(0.023)	(0.048)	(0.094)	(0.045)	(0.034)
	poor-nonpoor:	-0.043	-0.103	-0.070	-0.152	-0.047	-0.086
		[.666]	[.000]	[.024]	[.114]	[.334]	[.016]
Proportion in 80th percentile of state	nonpoor:	2.768	1.029	1.107	0.814	0.795	1.312
poverty distribution		(0.044)	(0.018)	(0.026)	(0.023)	(0.017)	(0.013)
	poor:	2.779	0.963	0.982	0.766	0.811	1.283
		(0.062)	(0.018)	(0.039)	(0.075)	(0.043)	(0.027)
	poor-nonpoor:	0.011	-0.066	-0.125	-0.048	0.016	-0.029
		[.886]	[.010]	[.009]	[.536]	[.725]	[.326]
Proportion in 90th percentile of state	nonpoor:	2.777	1.027	1.088	0.824	0.798	1.316
poverty distribution		(0.040)	(0.017)	(0.024)	(0.022)	(0.017)	(0.012)
	poor:	2.624	0.934	1.013	0.679	0.798	1.220
		(0.101)	(0.019)	(0.069)	(0.124)	(0.047)	(0.042)
	poor-nonpoor:	-0.153	-0.093	-0.075	-0.145	0.000	-0.096
		[.166]	[.000]	[.310]	[.252]	[.999]	[.028]
First quintile of poverty rate	nonpoor:	2.788	1.016	1.123	0.847	0.852	1.336
		(0.109)	(0.035)	(0.039)	(0.039)	(0.023)	(0.028)
Second quintile of poverty rate	nonpoor:	2.789	1.048	1.104	0.813	0.787	1.319
		(0.074)	(0.042)	(0.037)	(0.042)	(0.034)	(0.023)
Third quintile of poverty rate	nonpoor:	2.749	1.036	1.087	0.862	0.747	1.309
		(0.071)	(0.033)	(0.072)	(0.053)	(0.034)	(0.024)
Fourth quintile of poverty rate	poor:	2.734	1.022	1.077	0.685	0.749	1.260
		(0.089)	(0.035)	(0.108)	(0.044)	(0.022)	(0.025)
Fifth quintile of poverty rate	poor:	2.779	0.963	0.982	0.766	0.811	1.283
		(0.062)	(0.018)	(0.039)	(0.075)	(0.043)	(0.027)
	poor(fourth quintile)-nonpoor(first quintile):	-0.054	0.006	-0.046	-0.162	-0.103	-0.076
		[.705]	[.901]	[.689]	[.007]	[.001]	[.044]
	poor(fifth quintile)-nonpoor(first quintile):	-0.009	-0.053	-0.141	-0.081	-0.041	-0.053
		[.942]	[.176]	[.012]	[.344]	[.403]	[.172]
Overall mean price		2.765	1.018	1.081	0.803	0.798	1.306
Standard deviation of overall mean price		1.783	0.821	0.909	1.211	0.532	0.012
Total number of observations		2176	2825	1634	2418	1117	10170

Note: The unit of observation is the price including tax. Standard errors (robust to correlation of residuals within stores) are in parentheses. P-values are in brackets. The last column is the weighted-sum of the average item prices, where the weights are the low-income expenditure shares taken from the Consumer Expenditure Survey. Prices are weighted by the base-period quantity weight (see the *BLS Handbook of Methods* , chapter 17).

Source: Author's calculations using the Bureau of Labor Statistics price level data and 1990 Census STF 3B.

Table 3. FE estimates of the effect of sequentially controlling for factors related to price dispersion on the unit price (price per ounce plus tax) differential for the poor

	Overall								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Poor	-0.067	-0.074	-0.092	-0.093	-0.090	-0.079	-0.081	-0.064	-0.061
	(0.024)	(0.029)	(0.029)	(0.029)	(0.028)	(0.028)	(0.028)	(0.035)	(0.034)
Other covariates included?	no	yes	yes	yes	yes	yes	yes	yes	yes
P-values of joint-test of the explanatory contribution of covariate groups:									
Neighborhood demographics	----	----	0.028	0.029	0.033	0.027	0.013	0.022	0.038
Crime variables	----	----	----	0.890	0.694	0.354	0.379	0.856	0.644
Operating costs	----	----	----	----	0.082	0.652	0.911	0.918	0.817
Discretionary costs	----	----	----	----	----	0.023	0.047	0.049	0.068
Market structure variables	----	----	----	----	----	----	0.052	0.041	0.085
Search variables	----	----	----	----	----	----	----	0.006	0.018
Quality variables	----	----	----	----	----	----	----	----	0.004
Adjusted R^2	0.967	0.977	0.978	0.978	0.978	0.978	0.978	0.979	0.979

Note: The dependent variable is the logarithm of the unit price plus tax (price per ounce). Sample size is 10,170. All regressions include an intercept, item dummies, and local area fixed-effects. The other covariates are product size, a dummy variable indicating a sale price, and month dummies. Prices are weighted by the base-period quantity weight (see the *BLS Handbook of Methods*, chapter 17). Standard errors (robust to the correlation of residuals within stores) are in parentheses. The specific variables included in each category are listed in Table 1. The neighborhood demographics category excludes the race/ethnicity variables.

[1]The coefficients reported are the sum of the poor and race/ethnicity main effects and interaction.

Source: Author's calculations using the 1990 Census STF3B, Spectra, Inc. data on store characteristics, InfoUSA, Inc. data on market structure, and Bureau of Labor Statistics price level data.

Table 4. Estimates accounting for between-item dispersion on the calculation of the price differential between poor and affluent neighborhoods (neighborhood income measured by median household income)

	One Gallon of Vitamin D fortified, whole milk		Nonkosher, broiler/fryer whole chicken		One-dozen, large, Grade A white eggs		Loose Navel oranges		Individually packaged Iceberg lettuce	
	Neighborhoods where the proportion in poverty exceeds 20 percent (poor)									
Neighborhood income measure:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Median household income	-0.115	0.251	-0.040	-0.332	-0.052	0.234	-0.014	0.604	0.151	1.239
(÷ 10,000)	(0.071)	(0.001)	(0.054)	(0.014)	(0.080)	(0.002)	(0.317)	(0.028)	(0.243)	(0.024)
Other covariates included?	no	yes	no	yes	no	yes	no	yes	no	yes
Number of observations	393	393	492	492	224	224	404	404	167	167
Adjusted R^2	0.108	0.778	0.290	0.821	0.051	0.865	0.235	0.809	0.344	0.610
	Neighborhoods where the proportion in poverty is less than 20 percent (affluent)									
Neighborhood income measure:	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Median household income	0.016	0.008	0.021	-0.067	0.086	0.092	0.020	-0.004	0.059	0.006
(÷ 10,000)	(0.014)	(0.030)	(0.014)	(0.024)	(0.014)	(0.046)	(0.016)	(0.038)	(0.020)	(0.005)
Other covariates included?	no	yes	no	yes	no	yes	no	yes	no	yes
Number of observations	1783	1783	2333	2333	1410	1410	2014	2014	950	950
Adjusted R^2	0.172	0.796	0.255	0.714	0.310	0.785	0.454	0.624	0.368	0.709
Budget share weighted-average of price differentials (β-bar):										
	Without covariates:			With covariates:						
β-bar(affluent):	0.033	(0.007)		-0.008	(0.013)					
β-bar(poor):	-0.006	(0.090)		0.427	(0.009)					
β-bar(affluent) - β-bar(poor):	0.039	(0.090)		-0.435	(0.015)					

Note: The dependent variable is the logarithm of the unit price plus tax (price per ounce). All regressions include an intercept, a dummy variable indicating a sale price, and month dummies. The other covariates include the neighborhood, crime, operating and discretionary cost, market structure, search, and quality covariates indicated in Table 1 (excluding the race variables from the neighborhood demographics category), dummies for store format, regional dummies, and local-area fixed effects. The top panel shows the coefficients for the poor neighborhoods in the sample, while the middle panel shows the coefficients for the affluent neighborhoods. The bottom panel shows the expenditure-share weighted average for the column coefficients. Each weighted average is calculated using the expenditure share for the poor. Prices are weighted by the base-period quantity weight (see the *BLS Handbook of Methods*, chapter 17). Standard errors (robust to the correlation of residuals within stores) are in parentheses.

Source: Author's calculations using the 1990 Census STF3B, Spectra, Inc. data on store characteristics, and Bureau of Labor Statistics price level data.

Table 5. Investigating the effect of store type, central city status, and regional variation on FE estimates of the poor price gap

	Store type[1]			Central city status		Region				Race/Ethnicity			
	Superstore	Conventional supermarket	Store format unknown	Central city	Suburb	Northeast	Midwest	South	West	Predominately black only	Predominately Hispanic only	Predominately black and Hispanic	Predominately white, non-Hispanic
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Poor neighborhoods	-0.262	-0.054	-0.116	-0.054	-0.151	-0.107	-0.305	-0.103	-0.103	0.042	-0.114	-0.183	-0.096
	(0.055)	(0.041)	(0.061)	(0.045)	(0.087)	(0.061)	(0.091)	(0.041)	(0.101)	(0.069)	(0.097)	(0.071)	(0.051)
Number of observations	3794	2997	2985	4967	5203	2152	2083	3582	2353	3109	789	1288	4984
Adjusted R^2	0.986	0.989	0.973	0.981	0.982	0.988	0.990	0.987	0.856	0.988	0.940	0.980	0.983

Note: The dependent variable is the logarithm of the unit price plus tax (price per ounce). All regressions include product size, a dummy variable indicating a sale price, and month dummies, as well as the neighborhood, crime, operating and discretionary cost, market structure, search, and quality covariates indicated in Table 1 (excluding the race variables from the neighborhood demographics category), and local area fixed-effects. Prices are weighted by the base-period quantity weight (see the *BLS Handbook of Methods* , chapter 17). Standard errors (robust to the correlation of residuals within stores) are in parentheses.

[1]A superstore is a supermarket with at least 30,000 square feet and annual sales in excess of $12 million, offering specialty departments and extensive services. Store type is based on data obtained from Spectra Marketing, Inc.

[2]The coefficients reported are the sum of the poor and race/ethnicity main effects and interaction.

Source: Author's calculations using the 1990 Census STF3B, Spectra, Inc. data on store characteristics, and Bureau of Labor Statistics price level data.

Appendix Table 1. Descriptive Statistics: Comparison of BLS sample means of search costs, neighborhood demographics, store costs, and market structure to the Spectra sample of stores

Variables	Spectra sample	BLS sample outlets
Search		
Proportion of households without a vehicle	0.208	0.135
	[0.220]	[0.156]
Number of stores in neighborhood (per square mile)*Poverty rate	1.930	1.074
	[2.545]	[1.298]
Proportion of population completing up to grade 12, no diploma	0.272	0.238
	[0.144]	[0.128]
Proportion of population obtaining high school diploma, including GED	0.273	0.279
	[0.074]	[0.074]
Proportion of population completing some college, no degree	0.201	0.215
	[0.057]	[0.060]
Proportion of population with associate, bachelor, or graduate/professional degree	0.254	0.267
	[0.141]	[0.129]
Neighborhood demographics		
Proportion of unoccupied housing units	0.071	0.066
	[0.046]	[0.046]
Population density (per square mile, per zip code, ÷ 1,000)	12.934	7.135
	[22.222]	[14.642]
Proportion located in a central city	0.513	0.504
	[0.500]	[0.500]
Proportion non-Hispanic White	0.613	0.697
	[0.318]	[0.279]
Proportion non-Hispanic Black	0.179	0.125
	[0.250]	[0.199]
Proportion of Hispanic origin	0.157	0.124
	[0.210]	[0.192]
Crime		
Total property crime per capita	0.139	0.072
	[0.433]	[0.144]
Total crime per capita	0.159	0.079
	[0.508]	[0.166]
Operating costs		
Number of checkouts	6.985	9.822
	[4.784]	[4.531]
Number of full-time employees	24.804	39.862
	[22.117]	[21.935]
Number of part-time employees	32.077	56.279
	[29.725]	[28.758]
Proportion chain stores	0.652	0.722
	[0.269]	[0.448]
Proportion with scanning equipment	0.484	0.777
	[0.500]	[0.378]

-- (continued) --

(Appendix Table 1, continued)

Variables	Spectra sample	BLS sample outlets
Discretionary costs		
Proportion using circulars	0.812	0.855
	[0.309]	[0.295]
Proportion using in-store demonstrations	0.555	0.624
	[0.394]	[0.416]
Proportion doubling coupons	0.370	0.367
	[0.380]	[0.412]
Proportion with frequent shopper program	0.179	0.213
	[0.304]	[0.353]
Proportion using in-store coupons	0.640	0.717
	[0.380]	[0.383]
Quality		
Proportion with from-scratch bakery	0.573	0.692
	[0.392]	[0.394]
Proportion with delicatessen	0.783	0.851
	[0.330]	[0.302]
Proportion with butcher department	0.632	0.719
	[0.381]	[0.380]
Proportion with seafood department	0.557	0.662
	[0.393]	[0.403]
Proportion with pharmacy	0.297	0.379
	[0.382]	[0.420]
Proportion with full-service bank	0.173	0.226
	[0.208]	[0.263]
Proportion with automatic teller machine	0.596	0.701
	[0.388]	[0.387]
Proportion offering check cashing services	0.531	0.585
	[0.394]	[0.421]
Proportion with warehouse aisles	0.121	0.122
	[0.248]	[0.266]
Market structure		
Number of stores per zip code (per square mile)	4.700	1.769
	[11.748]	[6.029]
Population per store (÷ 1,000)	5.192	6.753
	[4.651]	[6.077]
Grocery selling area (square feet in thousands)	21.202	34.029
	[16.413]	[21.676]
Yearly sales volume (in thousands)	$6.452	$11.659
	[8.462]	[10.423]
Market share (share of yearly sales volume)	0.221	0.328
	[0.226]	[0.225]
Number of observations	19836	2181

Note: Standard deviations are in brackets. Means are weighted by the total zip code population. The unit of observation is a postal zip code. The BLS does not survey Montana, North Dakota, and Wyoming, therefore these states are omitted from the above table. In addition, I do not have cost and quality data for stores in Arkansas, Delaware, Kansas, and Rhode Island. The number of observations in the BLS sample is the number of unique outlets. There are multiple outlets in a zip code.

Source: Author's calculations using the 1990 Census STF3B, Bureau of Labor Statistics survey data, and data obtained from Spectra Marketing, Inc.