Journal of Urban Economics **49,** 219–239 (2001)
doi:10.1006/juec.2000.2183, available online at http://www.idealibrary.com on IDEAL®

The Retail Price of Inequality*

David M. Frankel

Eitan Berglas School of Economics, Tel Aviv University, Ramat Aviv 69978, Israel

and

Eric D. Gould

Department of Economics, Hebrew University, Mount Scopus, Jerusalem 91905, Israel

Received November 23, 1999; revised March 13, 2000; published online December 13, 2000

This paper studies the relation between a city's income distribution and its retail price level using panel data. We find that an increase in the presence of lower-middle income households, relative to poor or upper income households, is associated with lower prices. Our findings suggest that greater income inequality raises the prices that poor households face, thus making it harder for them to invest in human capital.

Key Words: retail prices; income inequality; poverty; income distribution.

1. INTRODUCTION

The relation between retail prices and the local income distribution has long been controversial in the United States. Racial riots in the 1960s were sparked, in part, by complaints from the poor that they paid more for goods and services than their wealthier neighbors (see Alkaly and Klevorick [1]). Empirical studies have not supported this claim; poverty and prices are usually found to be more or less uncorrelated. We revisit this issue by considering the effects of the entire income distribution, not just the poverty rate as in prior studies. We also improve on previous research by using panel data with instrumental variables, which permits us to correct potential problems of endogeneity and unobserved

*We thank Avner Ahituv, Jan Brueckner, Saul Lach, John Pencavel, Yoram Weiss, two anonymous referees, and seminar audiences at Boston University, Hebrew University, Stanford University, and Tel Aviv University for helpful comments. We also thank Dan Feenberg for crucial data and Yiftah Gordoni and Ran Greenwald for research assistance. Frankel acknowledges financial support from the Israel Foundations Trustees. Gould thanks the Falk Institute for financial support.

0094-1190/00 $35.00

heterogeneity. In addition, we confirm the robustness of our results using a second source of price data.

Our results indicate that higher prices are a result not of greater poverty per se, but rather of an absence of lower-middle income households. We divide the population into three groups: poor (below the poverty line), lower-middle income (between one and two times the poverty line), and middle-upper income (above twice the poverty line).[1] If 1% of the population moves from the lower-middle income group to the poor group, prices rise by about 0.7% using ordinary least squares (OLS) and 1.1% using instrumental variables (IV). If the same 1% moves to the middle-upper income group, price rise by about 0.6% under OLS and 0.7% using IV.[2] A higher concentration of people in either the top or the bottom income group, relative to the middle group, leads to higher retail prices.

These effects are economically significant. The standard deviation of the poverty rate across cities in 1990 was 7%; so a one-half standard deviation increase in the poverty rate, coming at the expense of the lower-middle income group, would lead to a 2.5% increase in prices. This is large compared to the 1990 standard deviation of the retail price index across cities, which was 4.6%.

These findings suggest that the recent trend toward greater income inequality may have hurt the poor by raising their cost of living. In 1980, the average poor central-city resident lived in a city in which 18.1% of the population were poor and 21.3% were between one and two times poverty.[3] By 1990, the first figure had risen to 20.3% while the second had fallen to 19.9%. The average poor person has become increasingly isolated from the lower-middle income group, whose presence holds prices down. In addition to lowering current consumption, higher prices leave the poor with fewer resources to invest in human capital, thus potentially exacerbating the ‘‘poverty trap.’’[4] This suggests a new way in which current inequality may lead to greater inequality in the future. The growth in the return to education is likely to have strengthened this effect in recent years.

We present results from both OLS and IV estimation. The OLS results answer the question of whether ‘‘the poor pay more’’: they do, but only to the

[1] We also verify the robustness of these results using a different schema that is based on income quantiles.

[2] These estimates are for fixed-effects regressions on a panel dataset; similar estimates are obtained in cross section.

[3] Each percentage is a weighted average across central cities, where the weights are the numbers of poor persons in each central city. A ‘‘central city’’ is defined by municipal boundaries; e.g., the central city of Boston does not include the adjacent municipalities of Cambridge or Brookline. There were 367 central cities in 1980 and 1990. Data are from the U.S. Census Bureau [19].

[4] This assumes that the poor are credit constrained. For a theoretical treatment, see, e.g., Galor and Zeira [10].

extent that they are isolated from lower-middle income households. The IV results suggest that changes in the income distribution *cause* the changes in prices that we observe. This is important for two reasons. First, it rules out reverse-causality as an explanation: the findings are not due to the locational choices of the near-poor being more sensitive to cost-of-living differences. Second, some urban policies have the potential to affect the distribution of income (e.g., welfare policy, income taxation, choices of public good bundles). Our IV results suggest that such policies may have large effects on retail prices.

We test several explanations for our findings. One is based on search behavior. There is empirical evidence that search intensity is greatest among buyers with moderate incomes (see Carlson and Gieseke [6] and Goldman and Johansson [11]). This may be because search costs are U-shaped in income; in particular, the poor lack cars while the rich have a high cost of time.[5] Poor families also tend to have limited storage space and small budgets (see Kunreuther [13]). This leaves them captive to their local grocery store, while the nonpoor are freer to search for bargains outside their immediate area.

Two proxies for the costs and benefits of search are used in the analysis. One reason poor families tend to have higher search costs is that they are less likely to own cars. Hence, to capture the role of search costs, we use the percentage of households who do not own cars. To capture the *benefits* of search, we use average household size: search should pay off more for bigger households, which buy larger quantities. Hence, this variable should be negatively correlated with prices.

We also consider several competing hypotheses. If crime raises retail costs and poverty causes crime, the poverty rate could simply be picking up the effect of crime on prices. To test this, we include the property crime rate in the regression. Another hypothesis is that an increase in the middle-high income group leads real estate prices to increase. We test this using an index of housing costs. Finally, higher prices in poorer cities may be due to a prevalence of small convenience stores or ‘‘mom-and-pop’’ grocery stores, which have higher average costs. To test this, we include the percentage of retail stores that have 10 or more employees; this variable should be negatively related to the price level.

We find only weak evidence for any of these hypotheses: although the variables usually have the predicted signs, they are never significant in a statistical sense except for the housing cost index. Even after controlling for these additional variables, the income distribution coefficients are still statisti-

[5]Goldman and Johansson [11, p. 178] also suggest that the poor may be less able to ‘‘identify, assess, and exploit marketplace opportunities.’’ This can be thought of as a higher cognitive cost of search. For survey evidence indicating that the rich have a higher cost of time, see Marmorstein *et al*. [16].

cally significant, which suggests either that we are not testing these hypotheses with accurate measures or that some other explanation is driving the results. We leave it to future research to examine this further.

The paper is organized as follows. The next section discusses the relevant literature. Section 3 discusses the data and Section 4 presents the basic OLS results. IV estimation and results are presented in Section 5 and 6. Explanations are tested in Section 7 and we conclude in Section 8.

2. LITERATURE REVIEW

The first to study the relation between income and retail prices was Caplovitz [5] in 1963. He found that the poor paid considerably more than the nonpoor for major durables such as televisions and washing machines. However, Caplovitz's reliance on retrospective data was seen as problematic. Subsequent studies of food prices based on store surveys tended instead to find that prices are uncorrelated with or slightly increasing with neighborhood income (Alcaly and Klevorick [1], Donaldson and Strangways [8], Goodman [12], MacDonald and Nelson [14], and Marcus [15]). Most recently, MacDonald and Nelson [14] find that a 1% increase in median income is associated with a (statistically insignificant) 0.013 to 0.019% food price increase.

All of the cited studies limit themselves to cross-sectional OLS regressions; thus, MacDonald and Nelson's [14] results may be contaminated by unobserved heterogeneity or endogeneity. In contrast, we confirm our results using both cross-sectional and panel data, thus reducing the likelihood that unobserved heterogeneity across cities is driving the results. Unlike prior studies, we also use instrumental variables to address problems of endogeneity. In addition, we base our price index on a broader consumption basket than either consumer durables (as in the case of Caplovitz [5]) or food (as in the other studies). We also examine the effects of the income distribution on different components of the overall price index.

Our study also differs from previous work by looking at price differences across cities, while prior work examined differences across neighborhoods. We believe that our approach is better suited to measuring the causal effects of income distribution on prices for two reasons. The first is that our results are insensitive to the problem of cross-neighborhood shopping, which appears to be widespread (see, e.g., Mitchell [17]). Since a person's neighborhood of residence is not perfectly correlated with where she shops, studies of neighborhoods in a single city are less likely to detect a relationship between the income distribution and retail prices. The second advantage of conducting the analysis on the city level is that it avoids problems of intracity sorting. Residents of a city sort themselves according to neighborhood characteristics (including perhaps the cost of living), and therefore it may be that prices determine the income distribution of a neighborhood rather than vice versa. Sorting across cities is likely to be less of a problem since intercity mobility is more limited.[6] (This does not entirely satisfy our concerns about endogeneity, so we also use instrumental variables.)

3. THE DATA

Data were collected from 184 cities in 1979/80 and 1989/90.[7] Summary statistics are shown in Table 1. For each city and biannual period, a retail price index was constructed using data from the American Chamber of Commerce Research Association (ACCRA) [2].[8] To confirm our findings, we also examine a panel of 15 metropolitan areas using price data from the Bureau of Labor Statistics (BLS). This second dataset is described in Section 4.

The ACCRA index is an expenditure-weighted average of price indices for four categories: groceries, transportation, health, and miscellaneous goods and services.[9] ACCRA also presents price indices for housing and utilities, which were excluded since we wish to focus on private sector retail markets.[10] ACCRA attempts to control for quality, either by careful specification or by restricting to given brand names. The following is a partial list of components of the ACCRA price index. The lists are complete in the case of transportation and health. For complete descriptions of the other two indices, see [2].

- **Groceries:** One pound of 100% pork, Jimmy Dean brand sausage; a 6.125-oz can of Starkist chunk light tuna; a half-gallon carton of whole milk; a 5-lb package of sugar; one dozen grade A eggs.

[6] Reasons include job market frictions, the costs of creating new social networks, and the relative difficulty of finding a house or apartment in a new city.

[7] Hence, for each city we have two biannual price indices. The 187 cities are those for which sufficient data existed to construct indices for both time periods.

[8] ACCRA data are available only for selected cities and quarters. Each city's index in each quarter is reported by ACCRA as an index relative to the average of cities that reported that quarter. This means that the base of the index depends on the set of reporting cities, which changes from quarter to quarter. To eliminate this dependence, we recomputed each quarterly index so that prices are measured relative to average price level of the 49 cities that reported data in all 16 quarters in our 4-year sample (1979, 1980, 1989, and 1990). These recomputed indices were then used to compute biannual averages. One remaining problem is that the number of quarters for which prices are available varies by city and biannual period. To eliminate the resulting heteroskedasticity, we weighted our regressions by the number of quarters used to compute each biannual index. We also weighted by city size to make the results more representative of the urban population.

[9] The results we report use the expenditure weights that ACCRA reports for each year. These weights differ between 1979/80 and 1989/90 because of changes in expenditure patterns. We also computed price indices using, for each biannual period, the average of the weights in the years 1979, 1980, 1989, and 1990. The regression results (available on request) were virtually identical. We also present regressions for the separate indices, which are independent of the weighting scheme.

[10] The transportation and health indices also include some components that are not sold in private sector retail markets. We address this below by analyzing each of the four categories separately.

TABLE 1

Summary Statistics for 184 Cities in Sample, 1990 Values

Variable	Mean	Std. Dev	Minimum	Maximum
Retail price index	100.00	4.60	90.10	116.73
Grocery prices	99.81	4.16	90.51	112.93
Health prices	99.17	13.27	73.28	157.83
Transportation prices	99.39	6.44	85.85	126.06
Miscellaneous prices	100.56	4.77	89.64	118.43
Population	186003	327758	10034	3485398
Percent under poverty	0.188	0.070	0.033	0.580
Pct between 1X and 2X poverty	0.201	0.035	0.062	0.306
Pct over 2X poverty	0.610	0.091	0.203	0.905
Pct in 1st income quintile	0.253	0.077	0.042	0.681
Pct in 2nd income quintile	0.219	0.035	0.083	0.322
Pct in 3rd income quintile	0.201	0.031	0.083	0.280
Pct in 4th income quintile	0.174	0.034	0.045	0.277
Pct in 5th income quintile	0.152	0.070	0.016	0.525
Population density	1046.29	610.93	202.67	3891.38
10-Year population growth rate	0.076	0.198	−0.317	1.398
Pct stores with > 9 employees	0.341	0.039	0.204	0.469
Total sales tax rate	0.0603	0.0137	0	0.090
Property crime rate	0.0413	0.024	0.008	0.217
Pct households with no cars	0.135	0.072	0.021	0.453
Mean household size	2.586	0.220	2.139	3.662
Pct aged 65+	0.131	0.036	0.036	0.257
Pct hispanic	0.083	0.147	0.002	0.769
Pct female	0.524	0.015	0.481	0.554
Pct nonfamily households	0.368	0.069	0.190	0.554

Population density equals number of persons per square mile. Ten-year population growth rate is the log change in the population from 1980 to 1990. Total sales tax is the sum of state, city, and county statutory sales tax rates. Property crime rate equals property crimes per capita. The following are shown in absolute form, but logs are used in regressions: all price indices, population density, the property crime rate, and mean household size.

- **Transportation:** One-way commuting fare to central business district, up to 10 miles; average price to spin-balance one front wheel; one gallon of regular unleaded gasoline, cash price at self-service pump, including all taxes.
- **Health:** Average cost per day of semiprivate hospital room; fee for routine doctor's office visit; fee for routine adult teeth cleaning and oral examination; price of 100-tablet bottle of Bayer 325-mg aspirin.
- **Miscellaneous:** McDonald's Quarter-Pounder with cheese; 12-in. Pizza Hut thin crust cheese pizza; dry cleaning of man's two-piece suit; ticket to

first-run movie at indoor cinema, evening, no discount; Miller Lite or Budweiser beer, 6-pack, 12-oz containers, excluding deposit.

Prices do not include sales taxes except for the gasoline component of the transportation price index.

Data on the income distribution come from the decennial U.S. Census.[11] We used two income schemas. The first is based on the ratio of a person's pretax family income to the poverty line for a family of the given size and composition.[12] We use the percentages of persons in three groups: below the poverty line, between one and two times the poverty line, and over twice the poverty line. (These were the natural delineations because no breakdown of the third group was available.) The second classification is based on after-tax family income. It consists of the share of families that fall into each national after-tax income quintile.[13] (For example, if 25% of a city's population falls into the poorest national quintile, then the city is poorer than the nation as a whole.) This schema has the advantage of taking federal and state income taxes into account. Moreover, a finer breakdown of the middle-upper segment is available, so we can test the validity of treating it as a single group. The drawback of this schema is that no family size adjustment is made, unlike the poverty classification.

Prices can also depend on sales taxes. We control for this using the combined city, state, and county sales tax rate from the *National Sales Tax Directory* [24]. A small number of our cities (10 in 1982 and 11 in 1992) also have city or county income taxes. We tried entering the average combined city and county income tax rate.[14] Its coefficient was always insignificant, so we removed it from the specifications.

[11] We used Census places as our geographic unit, since they correspond to the places used by ACCRA. A Census place is defined by metropolitan political boundaries, like a central city. In fact, a central city is just a large Census place that is located in a Census metropolitan area.

[12] The data come from Summary Tape File 3C of the decennial *Census of Population and Housing* [19]. Pretax income includes market income as well as cash transfer payments (AFDC, unemployment compensation, Social Security, etc.). The poverty line is adjusted for inflation over time but does not vary by region.

[13] The Census Bureau gives distribution information for 17 and 25 family income categories in 1980 and 1990, respectively. For each city−year, we converted this to an after-tax income distribution by reducing the category boundaries by the combined state and federal taxes paid by a typical family with income at the boundary. We then aggregated the categories, interpolating where necessary, to estimate the proportion of families in each of the national after-tax income quintiles. Tax rates were computed using the internet TAXSIM program of the National Bureau of Economic Research (Feenberg and Coutts [9]); we thank Dan Feenberg for working to make this program available and helping us use it. In computing tax rates, we assumed a married couple filing jointly with two children and no other deductions.

[14] We used the rates in 1982 and 1992 for observations in 1979/80 and 1989/90, respectively. Rates were computed as total tax receipts (from the *Census of Governments* [18]), divided by aggregate personal income in the city (from the *Census of Population and Housing* [19]).

All regressions using this dataset also control for a fixed set of other demographic variables.[15] These are the percentage of persons aged 65 and over, percentage Hispanic, percentage female, percentage of households that are not family households, the log of population density, the growth rate of the population in the last 10 years, and dummy variables for four regions (Northeast, South, North Central, West). We began with a larger set of demographic controls but removed a subset whose effects were always insignificant.[16] All regressions are also weighted by city size to make the results more representative of the urban population. A city's weight is the average of its 1980 and 1990 populations.

4. OLS RESULTS

We first estimate the relationship between the income distribution and the price level using simple OLS regressions. In order to reduce problems of unobserved heterogeneity across cities, we exploit the panel structure of the dataset by regressing the 1979/80 to 1989/90 differences in the dependent variable on the differences in the independent variables. (Later in this section we present pooled, cross-sectional results, which are broadly similar.)

Results for the overall retail index are presented in Table 2.[17] In the first column, the only income measure included is the percentage of households below the poverty line.[18] The coefficient estimate is not statistically significant. This is consistent with the existing literature, reviewed above, which has found a small or insignificant relation between poverty and prices.

In order to capture more features of the income distribution, column two of Table 2 includes the percentage of households over two times the poverty line. The suppressed variable is the size of the intermediate group: the percentage of households between the poverty line and twice the poverty line. Both income coefficients are now positive and statistically significant. Relative to the intermediate group, a one percentage point increase in the poor group or the higher income group is associated with an increase of 0.68 and 0.64%, respectively, in retail prices. The adjusted R-squared increases from 0.36 to 0.43 by adding the second income variable.

[15] Below, we also report regressions on a smaller dataset of 15 metropolitan areas; these regressions do not include any control variables.

[16] These include the percentage aged 17 and under, percentage black, percentage of households that are married-couple households, percentage unemployed, male and female labor force participation rates, and the percentage of persons living in urban areas.

[17] Each specification in Table 2 controls for the sales tax and the demographic covariates described in Section 3. For parsimony, we present only the coefficient estimates for the income variables.

[18] All income variables are numbers between 0 and 1, measuring the proportion of persons or families in a given group.

TABLE 2

OLS Analysis for the Retail Index

Independent variables	Dependent variable: Δ Log of retail price index			
Δ Percent under poverty	−0.14 (−1.35)	0.68 (3.37)		
Δ Pct over 2x poverty		0.64 (4.65)		
Δ Pct in 1st income quintile			0.37 (1.58)	0.46 (2.28)
Δ Pct in 3rd income quintile			0.45 (1.29)	
Δ Pct in 4th income quintile			0.19 (0.79)	
Δ Pct in 5th income quintile			0.62 (3.44)	
Δ Pct in top 3 income quintiles				0.54 (3.67)
Δ Other control variables	Yes	Yes	Yes	Yes
Adjusted R^2	0.36	0.43	0.42	0.41
N	184	184	184	184

Other control variables are four regional dummy variables, as well as changes in the following: the total state and local sales tax, percentage of the population over 65 years old, percentage Hispanic, percentage female, percentage of households that are not family households, log of the population density, and the 10-year population growth rate. All regressions include an intercept. *t*-statistics are in parentheses.

Column three in Table 2 uses the income quintiles. The suppressed group is the set of households falling in the second national after-tax quintile. (The ‘‘first quintile’’ refers to the poorest group, with the others following in order.) This is the group that most closely matches the suppressed group using the poverty schema. The coefficient estimates for the other four quintiles in the third column are all positive, which is consistent with the effects found in column two using the poverty classification.

An F-test for the equality of the coefficients of the top three income groups yields a p-value of 0.19. Therefore, we cannot reject that these categories have the same effects on prices. In the last column of Table 2, the top three income quintiles are aggregated together. With this specification, the coefficient estimates are now statistically significant.[19] Relative to the second quintile, a one percentage point increase in the poorest quintile (respectively, the upper three

[19]Combining the quintiles also reduces problems of multicollinearity. The standard errors for quintiles 3, 4, and 5 in column 3 are 0.35, 0.24, and 0.18, respectively, while the standard error for the combined category in column 4 is just 0.15.

quintiles) is associated with a price increase of 0.46% (0.54%). In subsequent regressions, we continue to use this more parsimonious specification.

Up to now, our dependent variable has been the (logged) overall retail price index. This index is an expenditure-weighted average of four individual price indices: groceries, miscellaneous, health, and transportation. We now examine each individual index in turn.

Table 3 presents OLS estimates for each price component using the poverty classification; Table 4 uses the quantile categories. The results are similar to what was found for the overall index: an increase in either the low or middle-high income group is usually associated with higher prices. However, the transportation index does not follow the pattern of results for the other indices. Most notably, the estimates for the quantile categories are negative, although statistically insignificant. For the poverty categories, the estimations are positive but still far from significant. Hence, transportation goods appear to be different. This difference could be explained by two factors. One is that public transportation is one of three components of the transportation index, and these prices are set by local governments rather than private bodies. Another reason is that gasoline (the second component of the index) is subject to high state excise taxes. In fact, gasoline is the only component of the overall retail price index that includes taxes. For these reasons, we might not expect transportation to follow a pattern similar to that of the other indices.

The results for the health index in Tables 3 and 4 mirror the overall results, which may seem puzzling since the items in the health index (aside from aspirin) are not typically thought of as search goods. The reason may be due in

TABLE 3

OLS Analysis for the Change in Each Index Using Poverty Categories

	Dependent variable				
Independent variables	Δ All	ΔGroc	ΔMisc	ΔHealth	ΔTrans
Δ Pct under poverty	0.68	0.38	1.20	1.16	0.05
	(3.37)	(1.44)	(3.22)	(2.08)	(0.10)
Δ Pct over 2x poverty	0.64	0.35	1.11	0.68	0.27
	(4.65)	(1.94)	(4.38)	(1.79)	(0.83)
Δ Other control variables	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.43	0.39	0.31	0.17	0.30
N	184	184	184	184	184

Other control variables are four regional dummy variables, as well as changes in the following: the total state and local sales tax, percentage of the population over 65 years old, percentage Hispanic, percentage female, percentage of households that are not family households, log of the population density, and the 10-year population growth rate. All regressions include an intercept. *t*-statistics are in parentheses.

TABLE 4

OLS Analysis for the Change in Each Index Using Quantile Categories

Independent variables	Dependent variable				
	Δ All	ΔGroc	Δ Misc	Δ Health	ΔTrans
Δ Pct in 1st income quintile	0.46	0.33	0.99	1.15	−0.68
	(2.28)	(1.29)	(2.67)	(2.11)	(1.46)
Δ Pct in top 3 income quintiles	0.54	0.40	1.01	0.82	−0.30
	(3.67)	(2.14)	(3.77)	(2.08)	(0.88)
Δ Other control variables	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.41	0.39	0.29	0.17	0.31
N	184	184	184	184	184

Other control variables are four regional dummy variables, as well as changes in the following: the total state and local sales tax, percentage of the population over 65 years old, percentage Hispanic, percentage female, percentage of households that are not family households, log of the population density, and the 10-year population growth rate. All regressions include an intercept. *t*-statistics are in parentheses.

part to differences in health insurance coverage. Middle-upper income people are much more likely to have health insurance than either the poor or the lower-middle income group. These two groups are about equally likely to have insurance, but the near-poor typically have private insurance while the poor usually have Medicaid.[20] Medicaid reimbursement rates are generally lower than rates for private insurance, so the prevalence of poor Medicaid patients may lead providers to raise their ‘‘list prices’’ in order to cover their losses. It may also be that the uninsured poor *behave differently* from uninsured lower-middle income households. Poor uninsured patients, who cannot afford to pay for routine doctors’ visits, may tend to wait until their problems become urgent and then seek health care in emergency rooms. Since they cannot pay for emergency care either, hospitals must raise their list prices to cross-subsidize. On the other hand, uninsured lower-middle income households *can* pay doctors’ fees if they are not too high. To the extent that bargaining is rare (so that

[20] During a 32-month period in 1991−1993, 49.2% of the poor and 51.0% of the lower-middle income group had continuous health insurance coverage, compared to 83.3% of persons over twice the poverty line. Of the poor, 36.2% were covered by Medicaid for the 32-month period while only 5.5% were covered by private health insurance; 34.5% of the lower-middle group were covered by private health insurance for the full period while only 6.2% were covered by Medicaid. (These percentages do not sum to the total percentage covered for the full 32 months because some persons with continuous coverage had a combination of Medicaid and private insurance.) The figures are the authors’ computations from the Census Bureau [21].

TABLE 5
Pooled Cross-Sectional OLS Analysis

Independent variables	Dependent variable: Log of retail price index	
Percent under poverty	0.62	
	(4.70)	
Pct Over 2x poverty	0.48	
	(5.46)	
Pct in 1st income quintile		0.62
		(5.43)
Pct in top 3 income quintiles		0.59
		(6.91)
Other control variables	Yes	Yes
Adjusted R^2	0.60	0.62
N	368	368

Log of retail price index (not differenced) regressed on income distribution variables, a year dummy for 1980−1990, and the other core control variables: four regional dummy variables, the total state and local sales tax, percentage of the population over 65 years old, percentage Hispanic, percentage female, percentage of households that are not family households, log of the population density, and the 10-year population growth rate. All regressions include an intercept. *t*-statistics are in parentheses.

list prices matter), the presence of these households might lead doctors and dentists to *lower* their list prices.

The preceding results come from 10-year difference regressions. Table 5 presents cross-sectional regressions. The two periods (1979/80 and 1989/90) are pooled, with the addition of a dummy year to capture changes in the average price. Results for both income schemas are consistent with those found in the difference regressions in the preceding tables. This strengthens the evidence for a positive relation between inequality and prices.

Alternative Dataset: 15 *Metropolitan Areas*

To test whether our results depend on the source of the price data, we also studied the 15 metropolitan areas for which the U.S. Bureau of Labor Statistics collected price data during the 1980s. The price index consists of all items except shelter. The 15 metropolitan areas are listed in Table 6. (Metropolitan areas tend to be considerably larger than places, the census unit studied in our main dataset.)

The results for this dataset are shown in Table 7. Similar to Tables 2, 3, and 4, the 10-year change in the dependent variable is regressed on the 10-year

TABLE 6

Metropolitan Areas for Which BLS Price Data Are Available

Metropolitan area name	1990 Population
Boston−Lawrence−Salem, MA−NH (CMSA)	4,171,643
Chicago−Gary−Lake County, IL−IN−WI (CMSA)	8,065,633
Cleveland−Akron−Lorain, OH (CMSA)	2,759,823
Dallas−Fort Worth, TX (CMSA)	3,885,415
Detroit−Ann Arbor, MI (CMSA)	4,665,236
Houston−Galveston−Brazoria, TX (CMSA)	3,711,043
Los Angeles−Anaheim−Riverside, CA (CMSA)	14,531,529
Miami−Fort Lauderdale, FL (CMSA)	3,192,582
New York−Northern NJ−Long Island, NY−NJ−CT (CMSA)	18,087,251
Philadelphia−Wilmington−Trenton, PA−NJ−DE−MD (CMSA)	5,899,345
Pittsburgh−Beaver Valley, PA (CMSA)	2,242,798
San Francisco−Oakland−San Jose, CA (CMSA)	6,253,311
Baltimore, MD (MSA)	2,382,172
St. Louis, MO−IL (MSA)	2,444,099
Washington, DC−MD−VA (MSA)	3,923,574

TABLE 7

OLS Regression Results for 15 Metropolitan Areas Using BLS Price Data

Independent variables	Dependent variable: Δ Log of BLS price index	
ΔPercent under poverty	1.406	
	(1.48)	
ΔPct between 1x and 2x poverty		−0.491
		(−2.67)
ΔPct over 2x poverty	0.880	
	(2.01)	
Adjusted R^2	0.30	0.31
N	15	15

The 1980−1990 changes in the dependent and independent variables are used in the regressions. Price index is all items excluding shelter. Regressions are weighted by mean population in 1980 and 1990. All regressions include an intercept. *t*-statistics are in parentheses.

change in the independent variables.[21] Despite the very small sample size, the coefficient estimates are broadly similar to those for our main dataset and mostly significant. The data used in the second column of Table 7 are plotted in Fig. 1. (Names of metropolitan areas are abbreviated.) Note the pronounced

[21] To maintain consistency with our primary dataset, the 1980 (1990) price index computed here is actually the average of the years 1979−1980 (1989−1990). Using only 1980 and 1990 data gives essentially the same results.

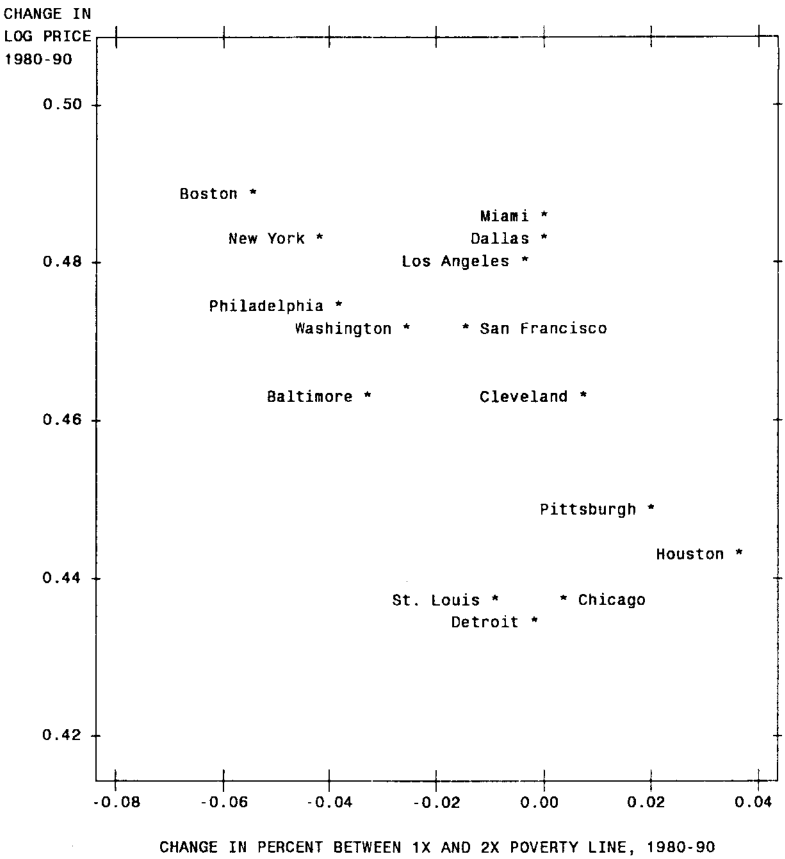

FIG. 1. Using BLS price data from 15 metropolitan areas.

negative relation between the change in the price level and the change in the proportion of lower-middle income persons.

Discussion of OLS Results

Despite differences in data and techniques, our first regression replicates the findings of the existing literature: by itself, the poverty rate is essentially uncorrelated with retail prices (see column 1 of Table 2). However, the results change sharply when a more complete income schema is used. Growth in the poor or middle-upper income group, relative to the lower-middle income group, is associated with higher prices. This illustrates the shortcomings of using a dichotomous poor/nonpoor income schema. These results appear in both cross-section and panel regressions using our main dataset and can also be detected in panel regressions using a different source of price data. Furthermore, the magnitude of these effects is economically important. The standard deviation of the poverty rate across cities in 1990 was 7%, so a one-half standard deviation increase in the poverty rate, coming at the expense of the lower-middle income group, would lead to a 2.5% increase in prices. This is large compared to the 1990 standard deviation of the retail price index across cities, which was 4.6%.

5. IV ESTIMATION

The OLS results show that prices are not higher in high poverty cities per se, but rather in cities with high inequality: those with relatively few lower-middle income households. One potential explanation for this is reverse causality: in deciding where to live, lower-middle income households may be especially sensitive to high retail prices, as they have tighter budgets than the rich while still having sufficient resources to move (unlike many poor households). If so, they may be more likely than the rich or the poor to choose to live in low-cost cities. In this section we test this explanation via instrumental variables.

Our strategy consists of using information about a city's industrial composition in 1980 to predict the change in the income distribution within that city from 1980 to 1990.[22] The idea is that nationwide factors such as technological change and international trade cause changes in the industrial composition that have different effects on different cities. For example, the national decline in the manufacturing sector from 1980 to 1990 reduced demand for manufacturing jobs more in areas that were manufacturing-intensive in 1980. This in turn affected the income distribution by lowering wages for some manual workers and leading others to leave the area. Thus, the size of a city's manufacturing sector in 1980 can be used to predict exogenous changes in the income distribution from 1980 to 1990.

The 1980 city-level employment shares of 15 industries were used as instruments.[23] After controlling for changes in our set of demographic variables, the 15 industrial shares explain 39% of the remaining variation of the changes in the poverty rates within cities, and 50% of the remaining variation of the changes in the percentage of households above twice the poverty line. Thus, the instruments capture a substantial portion of the variation in the changes in the income distributions within cities.

[22] Bartik [3] and Blanchard and Katz [4] employ a similar strategy.

[23] The industries are: agriculture, forestry, fishing, and mining; construction; manufacturing nondurable; manufacturing durable; transportation, communications and other public utilities; wholesale trade; retail trade; finance, insurance, and real estate; business and repair services; personal, entertainment, and recreation services; health services; educational services; other professional and related services; and public administration.

TABLE 8

IV Estimation for Δ Log of Each Index Using Poverty Categories

Independent variables	Dependent variable				
	ΔAll	ΔGroc	ΔMisc	ΔHealth	ΔTrans
Δ Pct under poverty	1.08	0.85	2.12	1.63	−0.12
	(2.90)	(1.75)	(3.10)	(1.63)	(0.14)
Δ Pct over 2x poverty	0.74	0.52	1.46	0.84	0.10
	(3.34)	(1.80)	(3.57)	(1.41)	(0.19)
Δ Demographic controls	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.39	0.38	0.28	0.16	0.30
N	184	184	184	184	184

Other control variables are four regional dummy variables, as well as changes in the following: the total state and local sales tax, percentage of the population over 65 years old, percentage Hispanic, percentage female, percentage of households that are not family households, log of the population density, and the 10-year population growth rate. All regressions include an intercept. Income distribution variables are instrumented with 15 variables defined as the percentage of employed workers 16 years and older within 15 industrial sectors. *t*-statistics are in parentheses.

6. IV RESULTS

Table 8 presents two-stage least-squares results using the poverty classification. For the overall price index, the coefficient estimates for the poorest group and the middle-upper income group are positive and significant. The IV estimates for both coefficients are slightly larger in magnitude than the OLS estimates (Table 3). For the overall index, the IV coefficient on the percentage poor is 1.08 compared to 0.68 with OLS. The IV coefficient for the upper income group is 0.74 compared to 0.64 with OLS.

Table 9 presents the IV results using the quantile classifications. Again, the effects are larger under IV than OLS. For the lowest quintile, the IV coefficient for the overall index is 1.13 versus 0.46 with OLS in Table 4. For the upper income group, the IV estimate is 0.90 in contrast to 0.54 using OLS.

These results show that reverse causality does not underlie our findings.[24] They also indicate that OLS induces a *downward* bias in the coefficients. This may indicate that the lower-middle income group is *less* sensitive to a city's average retail price level in deciding where to live than the other two groups.

[24] The instrument set does not pass the Basmann test for overidentifying restrictions. However, the Basmann test is well known to reject too often in small samples. We have tried many different IV strategies, some of which consistently passed the Basmann test; all give essentially the same results as those presented here. Other instruments we have tried include: (1) the 1980 income distribution, (2) changes in the state-level income distribution, (3) extreme weather conditions in 1980, and (4) changes in extreme weather conditions from 1980 to 1990. Results are available on request.

TABLE 9
IV Estimation for Δ Log of Each Index Using Quantile Categories

Independent variables	Dependent variable: Δ All	ΔGroc	Δ Misc	Δ Health	ΔTrans
Δ Pct in 1st income quintile	1.13	0.64	2.31	1.16	−0.33
	(2.10)	(0.95)	(2.34)	(0.82)	(0.27)
Δ Pct in top 3 income quintiles	0.90	0.52	1.75	0.89	−0.10
	(2.58)	(1.19)	(2.72)	(0.97)	(0.13)
Δ Demographic control	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.38	0.38	0.25	0.16	0.30
N	184	184	184	184	184

For definitions of demographic controls, see note to Table 8. Income distribution variables are instrumented with the 15 variables defined as the percentage of employed workers 16 years and older within 15 industrial sectors. *t*-statistics are in parentheses.

The reason may be search costs. Upper and lower income households do not search intensively; thus, they tend to pay close to the average retail price in a city. Middle income households do search intensively; they care less about the average price than about the presence of low-price stores. There are more low-price stores if the intracity price variance across stores is high. Thus, if cities with high average prices also tend to have high variances, they may repell low and high income households more than middle income households.

7. EXPLANATIONS

One explanation for our findings is that lower-middle income households have lower search costs than either of the other two groups. This is supported by the finding of Carlson and Gieseke [6] and Goldman and Johansson [11] that search intensity is greatest among buyers with moderate incomes. In this section we attempt to test the search hypothesis. Our test is necessarily indirect since we do not have observations of the search costs of different households. Instead, we examine the effects of adding proxies for the costs and benefits of search in the regression. To proxy for the benefits of search, we use the log of average household size: search should pay off more for bigger households, which buy larger quantities. Hence, this variable should be negatively correlated with prices. Our proxy for the cost of search is the percentage of households that do not own a car. Since not having a car raises search costs, an increase in the percentage of households without cars should discourage search and lead to higher prices.

We also consider several competing hypotheses. If higher crime rates lead to higher retail costs, and if crime is concentrated in poorer areas, our measure for the poor could be picking up the effect of higher crime rates in poorer areas.

Consequently, we include the log of reported property crimes per capita in the regression.[25]

Another hypothesis could be that an increase in the middle-high income group leads real estate prices to increase. This drives up retailers' operating costs, forcing them to raise prices. We test this using ACCRA's housing cost index, which is an average of the cost of renting an unfurnished two-bedroom apartment and the monthly mortgage payment on a new 1800 square foot home.

Finally, prices may move in response to unmeasured changes in the quality of service. Smaller stores provide a service to poor residents by locating closer to their homes and selling smaller quantities. If smaller stores have higher unit costs than larger stores, it could be that the poor pay for this service through higher prices.[26] To test this, we include the percentage of retail stores that have 10 or more employees using data from the *Zip Code Business Patterns* [20]. If prices in smaller stores are higher because their costs are higher, this variable should be negatively related to the price level.[27]

The OLS and IV estimates including these alternative variables are presented in Table 10. The first two columns show results for the poverty categories while the last two columns use the quantiles. The dependent variable is the change in the overall retail index (in log form). The specifications include all of the variables used previously, in addition to the new hypothesis variables. As predicted by the theory, household size is negatively related to the price level. However, this variable is statistically insignificant in all specifications. As predicted, the coefficient for the percentage of households without cars is positive, but also is not statistically significant.

Even with all the hypothesis variables in the equation, the estimated coefficients of the income groups are still large and significant. This may indicate that other, undiscovered explanations play a role. However, it may simply be that our proxies are too indirect for our purposes. Further work is needed to give a more definitive explanation for our findings.

[25] These data are from a CD-ROM of the U.S. Census Bureau [22]; the original source is the Federal Bureau of Investigation. We did not have city-level data, so we used the county crime rate. In a few cases, a city overlapped several counties; here, we used the overall crime rate for the overlapping counties taken together. Property crimes comprise burglaries, robberies, and larcenies. We also tried replacing the overall property crime rate with each of these components in turn; the results were essentially the same.

[26] Kunreuther [13] presents evidence that unit prices are indeed higher in small stores.

[27] Another service-related hypothesis could be that retail firms in richer areas invest more in the upkeep of their stores or hire better workers. We have no way to test this theory. However, this would offer at most a partial explanation for our findings, since it does not explain why growth in the poor group pushes prices up.

TABLE 10

OLS and IV Estimation for Retail Index with Alternative Hypothesis Variables

	Dependent variable: ΔAll			
Independent variables	OLS	IV	OLS	IV
Δ Pct under poverty	0.64	1.01		
	(3.04)	(2.67)		
Δ Pct over 2x poverty	0.53	0.45		
	(3.31)	(1.66)		
Δ Pct in 1st income quintile			0.43	1.16
			(2.10)	(2.14)
Δ Pct in top 3 income quintiles			0.41	0.73
			(2.55)	(1.98)
Δ Log household size	−0.04	−0.27	−0.08	−0.19
	(0.28)	(1.41)	(0.53)	(1.14)
Δ Pct households with no cars	0.03	0.05	0.06	0.06
	(0.65)	(0.86)	(1.21)	(1.03)
Δ Log property crime	0.003	0.003	0.003	0.003
	(0.48)	(0.44)	(0.55)	(0.45)
Δ Pct stores with $>$ 9 employees	−0.11	−0.06	−0.08	−0.05
	(1.17)	(0.53)	(0.87)	(0.48)
Δ Log of housing price index	0.02	0.04	0.02	0.03
	(2.02)	(2.79)	(1.80)	(2.17)
Δ Demographic controls	Yes	Yes	Yes	Yes
Adjusted R^2	0.43	0.40	0.42	0.39
N	184	184	184	184

See Table 9 for definitions of instruments and other controls. t-statistics are in parentheses.

8. DISCUSSION

We find that prices increase when lower-middle income households in a community are replaced by either poor or middle-higher income residents. One implication is that greater isolation of the poor from lower-middle income households will raise the prices paid by poor residents. In addition to lowering the current consumption of the poor, this is likely also to make it more difficult for them to invest in human capital, which may intensify the ‘‘poverty trap.’’

Our results are stronger and more reliable than those of prior studies for several reasons. Using different data and methodology, we reproduce the insignificant effects found in prior studies, but show that this is a result of using a dichotomous poor/nonpoor income schema. Significant and economically important effects emerge when we turn to a finer schema that captures the income distribution in greater detail. We also address problems of simultaneity and unobserved heterogeneity by using city-level panel data with instrumental variables. Our findings are shown to be robust to OLS (cross-section and panel analysis), instrumental variables, the inclusion of several alternative hypotheses variables, and an alternative source of price data.

We also considered several competing hypotheses that can explain these results. However, empirical attempts to test these hypotheses were inconclusive, if not negative. Further work is needed to explain the mechanism behind our findings. But one clear conclusion from our study is that changes in the income distribution can have a large causal effect on retail prices, which suggests that cities should consider these effects in their assessment of policies that may affect the local distribution of income.

REFERENCES

1. R. E. Alcaly and A. K. Klevorick, Food Prices in Relation to Income Levels in New York City, *Journal of Business*, **44**, 380–397 (1971).
2. American Chamber of Commerce Researchers Association, ''Inter-City Cost of Living Index'' (various years).
3. T. Bartik, ''Who Benefits from State and Local Economic Development Policies?,'' W. E. Upjohn Institute for Employment Research, Kalamazoo, MI (1991).
4. O. J. Blanchard and L. F. Katz, Regional evolutions, *Brookings Papers on Economic Activity*, **0**:1, 1–69 (1992).
5. D. Caplovitz, ''The Poor Pay More,'' The Free Press of Glenco, New York (1963).
6. J. Carlson and R. Gieseke, Price search in a product market, *Journal of Consumer Research*, **9**, 357–365 (1983).
7. D. F. Dixon and D. J. McLaughlin, Jr., Low income consumers and the issue of exploitation: A study of chain supermarkets, *Social Science Quarterly*, **51**, 320–328 (1970).
8. L. Donaldson and R. S. Strangways, Can ghetto groceries price competitively and make a profit?, *Journal of Business*, **36**, 61–65 (1973).
9. D. Feenberg and E. Coutts, An introduction to the TAXSIM Model, *Journal of Policy Analysis and Management*, **12**, 189–194 (Winter 1993).
10. O. Galor and J. Zeira, Income distribution and macroeconomics, *Review of Economic Studies*, **60**, 35–52 (1993).
11. A. Goldman and J. K. Johansson, Determinants of search for lower prices: An empirical assessment of the economics of information theory, *Journal of Consumer Research*, **5**, 176–186 (1978).
12. C. S. Goodman, Do the poor pay more?, *Journal of Marketing*, **32**, 18–24 (1968).
13. H. Kunreuther, Why the poor pay more for food: Theoretical and empirical evidence, *Journal of Business*, **46**, 368–383 (1973).
14. J. M. MacDonald and P. E. Nelson, Jr., Do the poor still pay more? Food price variations in large metropolitan areas, *Journal of Urban Economics*, **30**, 344–359 (1991).
15. B. H. Marcus, Similarity of ghetto and nonghetto food costs, *Journal of Marketing Research*, **6**, 365–368 (1969).
16. H. Marmorstein, D. Grewal, and R. P. H. Fishe, The value of time spent in price-comparison shopping: Survey and empirical evidence, *Journal of Consumer Research*, **19**, 52–61 (1992).
17. A. Mitchell, Where markets are never super: Some urban neighborhoods fight absence of major chains, *The New York Times*, late edition, Section 1, p. 25 (June 6, 1992).

18. U.S. Department of Commerce, Bureau of the Census, ‘‘Census of Governments,’’ Washington, D.C. (1982, 1992).
19. U.S. Department of Commerce, Bureau of the Census, ‘‘Census of Population and Housing, Summary Tape File 3C,’’ Washington, D.C. (1980, 1990).
20. U.S. Department of Commerce, Bureau of the Census, ‘‘Zip Code Business Patterns,’’ Washington, D.C. (various years).
21. U.S. Department of Commerce, Bureau of the Census, ‘‘Dynamics of Economic Well-Being: Health Insurance Statistics, 1991 to 1993,’’ Washington, D.C. (1995).
22. U.S. Department of Commerce, Bureau of the Census, ‘‘USA Counties 1996 (CD-ROM),’’ Washington, D.C. (1996).
23. U.S. Department of Labor, Bureau of Labor Statistics, ‘‘Consumer Price Index,’’ Washington, D.C. (various years).
24. Vertex Systems Inc., ‘‘National Sales Tax Rate Directory,’’ Berwyn, PA, (various years).