Unpublished Technical Appendix

Journal of Law and Economics

Volume 51, Number 2, May 2008, 209-250

WHEN LITTLE THINGS MEAN A LOT:
ON THE INEFFICIENCY OF ITEM PRICING LAWS

Mark Bergen
University of Minnesota

Daniel Levy*
Bar-Ilan University and Rimini Center for Economic Analysis

Sourav Ray
McMaster University

Paul H. Rubin
Emory University

Benjamin Zeliger
Cornell University

Last Revision: November 26, 2007

* Corresponding author: Daniel Levy, Department of Economics, Bar-Ilan University, Ramat-Gan 52900, ISRAEL; Email: Levyda@mail.biu.ac.il;Tel: +972-3-531-8331; Fax: +972-3-738-4034

TECHNICAL APPENDIX – EMPIRICAL ANALYSES

ANALYSIS 1

This analysis compares the stores that have the ESL exemption with stores that don't.

TABLE A1

COMPARISON OF DEMOGRAPHIC CHARACTERISTICS OF THE ESL AND NO-ESL MARKETS

Demographic Variable	type	N	Mean	t	Sig. (2-tailed)
(a) County Level Comparisons					
County_Dens (/km²)	IPL	12	1478.18	1.174	0.265
	ESL	3	545.00		
County_hhlds	IPL	12	514020.25	0.826	0.427
	ESL	3	324232.00		
County_Med Hhld Inc ($)	IPL	12	59869.50	-1.764	0.105
	ESL	3	65249.00		
County_Med Fam Inc ($)	IPL	12	73442.08	-1.067	0.309
	ESL	3	77690.00		
County_Per capita Inc ($)	IPL	12	33424.92	-2.413	0.034
	ESL	3	38350.00		
County_ Fam below pvrty (%)	IPL	12	5.92	1.478	0.167
	ESL	3	5.00		
County_ Pop below pvrty (%)	IPL	12	8.26	1.524	0.156
	ESL	3	6.90		
(b) City-Level Comparisons					
City_Dens (/km²)	IPL	12	2083.07	1.229	0.241
	ESL	3	1031.60		
City_hhlds	IPL	12	265108.58	0.923	0.376
	ESL	3	33780.00		
City_Med Hhld Inc ($)	IPL	12	81260.42	0.442	0.672
	ESL	3	73160.33		
City_Med Fam Inc ($)	IPL	12	95567.67	0.378	0.720
	ESL	3	86758.33		
City_Per capita Inc ($)	IPL	12	45064.00	-0.126	0.907
	ESL	3	47038.00		
City_ Fam below pvrty (%)	IPL	12	6.97	1.080	0.300
	ESL	3	4.30		
City_ Pop below pvrty (%)	IPL	12	9.51	1.106	0.289
	ESL	3	6.37		
(c) ZIP Level Comparisons					
Zip_hhlds	IPL	10	7861.20	-0.039	0.973
	ESL	2	7971.00		
Zip_Med Hhld Inc ($)	IPL	10	86387.60	0.079	0.948
	ESL	2	83401.00		
Zip_Med Fam Inc ($)	IPL	10	107643.60	0.248	0.841
	ESL	2	95917.50		
Zip_Per capita Inc ($)	IPL	10	51801.10	-0.249	0.843
	ESL	2	60947.50		
Zip_ Fam below pvrty (%)	IPL	10	3.13	-0.514	0.694
	ESL	2	4.80		
Zip_ Pop below pvrty (%)	IPL	10	5.37	-0.363	0.772
	ESL	2	6.80		

NOTE: Variables' definitions and other details of our new data collection efforts are discussed in the last two pages of this document.

Note that of the twenty comparisons, only two comparisons are significant (the median household income and per capita income, both at county levels). We did not have the density figures at the ZIP levels for comparison. Substantively, the results imply that the aggregate household and per capita incomes of the ESL adoptee stores may be different from those stores that do not adopt ESL. The important question is whether this can be construed as evidence that the adoption of ESL is indeed an endogenous choice based on the market characteristics of the stores. <u>Perhaps more importantly, is this evidence that the varying price levels observed between ESL and non-ESL (IPL) stores are in fact determined by market characteristics and not by the underlying pricing technology of the stores?</u>

There are many reasons to believe that the above hypothesis is not supported by the data above and that we cannot explain the differing price levels of ESL and non-ESL (IPL) stores with endogenous choice of prices depending on the market characteristics.

1. Of the twenty comparisons, *only two* are significant. If indeed these were driving characteristics, one would expect other measures of spending power and market size to be significantly different as well.
2. There is no significance at the city level comparisons. Arguably, cities are a better approximation of grocery stores' trade areas than counties (which are larger) or ZIP code areas (which are smaller).
3. Even for the significant comparisons, the sign of the differences seem to be counterintuitive. For example, in our analysis in the main paper, we found ESL prices to be systematically LOWER than IPL prices. However, the comparison above shows that the income levels on ESL markets are systematically HIGHER than in non-ESL (IPL) markets. It is not immediately clear why the stores would systematically charge lower prices to those customers who can afford more. There may indeed be a reason that we can ascribe to this, but in conjunction with the other evidence, we feel very strongly that these demographic variables may not explain the differing price levels (between ESL and non-ESL IPL stores) we observe.
4. While the validity of the t-statistic itself could be suspect given the small samples involved, we note the relatively small differences between the economic variables. The only comparisons that show a big difference are the population figures (number of households). However, this is largely driven by the fact that we included the county and city of New York which is substantially larger than the others.

To conclude, our detailed comparison of the market level characteristics of the stores, finds no compelling evidence that the variation of price levels between stores that adopt ESL and those that do not, can be explained as an endogenous choice of stores driven by their market characteristics. We therefore feel confident that the observed variation in prices is driven more by the differences in the pricing technology than market characteristics.

We revisit this question further in Analysis 4 in the framework of regression analysis as discussed below on pages 6–8. The key question we ask there is whether or not these differences can explain the price gaps we find between the ESL and Non-ESL stores.

ANALYSIS 2

This analysis compares the market characteristics of stores that are subject to the IPL regime with those that are not.

TABLE A2

COMPARISON OF MARKET CHARACTERISTICS OF IPL AND NO-IPL REGIMES

Demographic Variable	type	N	Mean	t	Sig. (2-tailed)
(a) County Level Comparisons					
County_Dens (/km²)	IPL	15	1291.55	-0.268	0.792
	NOIPL	5	1476.20		
County_hhlds	IPL	15	476062.60	1.182	0.255
	NOIPL	5	254616.40		
County_Med Hhld Inc ($)	IPL	15	60945.40	1.234	0.258
	NOIPL	5	54769.20		
County_Med Fam Inc ($)	IPL	15	74291.67	1.522	0.172
	NOIPL	5	64616.80		
County_Per capita Inc ($)	IPL	15	34409.93	2.270	0.056
	NOIPL	5	26991.80		
County_ Fam below pvrty (%)	IPL	15	5.73	-1.014	0.361
	NOIPL	5	7.68		
County_ Pop below pvrty (%)	IPL	15	7.99	-0.907	0.406
	NOIPL	5	10.04		
(b) City-Level Comparisons					
City_Dens (/km²)	IPL	15	1872.77	0.050	0.961
	NOIPL	5	1835.36		
City_hhlds	IPL	15	218842.87	1.039	0.317
	NOIPL	5	10756.60		
City_Med Hhld Inc ($)	IPL	15	79640.40	0.835	0.415
	NOIPL	5	69002.00		
City_Med Fam Inc ($)	IPL	15	93805.80	0.817	0.425
	NOIPL	5	81940.80		
City_Per capita Inc ($)	IPL	15	45458.80	1.516	0.149
	NOIPL	5	33162.40		
City_ Fam below pvrty (%)	IPL	15	6.43	2.109	0.050
	NOIPL	5	2.20		
City_ Pop below pvrty (%)	IPL	15	8.88	2.220	0.040
	NOIPL	5	3.78		
(c) ZIP Level Comparisons					
Zip_hhlds	IPL	12	7879.50	1.772	0.119
	NOIPL	5	4614.60		
Zip_Med Hhld Inc ($)	IPL	12	85889.83	1.581	0.135
	NOIPL	5	67578.80		
Zip_Med Fam Inc ($)	IPL	12	105689.25	1.835	0.087
	NOIPL	5	82391.60		
Zip_Per capita Inc ($)	IPL	12	53325.50	2.138	0.049
	NOIPL	5	36649.60		
Zip_ Fam below pvrty (%)	IPL	12	3.41	1.075	0.308
	NOIPL	5	2.36		
Zip_ Pop below pvrty (%)	IPL	12	5.61	0.824	0.427
	NOIPL	5	4.40		

These comparisons intend to show the differences between the geographic markets that are subjected to the IPL regime and those that are not. In general, the significance of the figures should be treated with caution given the small sample. The population figures (number of households) are different, probably driven by the size of New York City. The poverty figures (% below poverty) are different at the city level. Interestingly, the cities with IPL regime seem to have greater poverty levels than the cities with No-IPL regimes. In keeping with the notion of IPL regimes being an endogenous choice, this could of course be a tentative suggestion that policy intervention was well-intentioned. (At the same time, this point to a double whammy, where the poorer markets were being subjected to higher prices due to the IPL regimes.)

Note also that, while the income figures are quite comparable across the regimes at the county and city levels, they differ at the ZIP levels. At the ZIP level, the income figures for IPL regimes were higher than for non-IPL regimes. This of course could be tentative suggestion that IPL stores' higher prices are driven more by market conditions than the IPL regime.

However, it is not clear, if this analysis is strong evidence against our results. Many of the same arguments as in Analysis 1 apply. Notice that of the twenty comparisons, *only five* are significant. While ZIP level comparisons of income are significant, the City level comparisons are not. Even within the ZIP level comparisons, the household level income is not significantly different. Yet, the results are tentative enough to suggest even further analysis.

We revisit this question further in Analysis 3 (which follows below) in the framework of regression analysis. The key question we ask there is whether or not these differences can explain the price gaps we find between the IPL and non-IPL stores and between the ESL and Non-ESL stores.

ANALYSIS 3

In the following, we attempt to deal with the issue of unobserved factors and endogeneity in a more direct fashion by incorporating these additional variables in our regression framework. Note that we restrict our analyses to data set II because data set I only has four stores in three states which are not enough to capture the regime level variation.

We start with the model in regression (3) on page 11 of the manuscript. Recall that this model was:

$$P = \alpha + \theta_k CHAIN_k + \gamma_j CATEGORY_j + \beta_1 IPL + \beta_2 \left(IPL \times ESL \right) + v_k + \varepsilon$$

We now modify this first to explicitly take into account the city level variables, population density (population per square km) and median family income. We consider the city level variables first (rather than county or zip level ones) for the unobserved store level characteristics because they may be better proxies for the trade area characteristics of the store.

For completeness however, we follow up with another regression by including ZIP level variables. However, since we do not have density figures at that level, we include number of households in the regression in lieu thereof. The two models, therefore, are:

(4)
$$P = \alpha + \theta_k CHAIN_k + \gamma_j CATEGORY_j + \beta_1 IPL + \beta_2 \left(IPL \times ESL \right) + \beta_4 City\_Density + \beta_5 City\_MedFamInc + v_k + \varepsilon$$

(5)
$$P = \alpha + \theta_k CHAIN_k + \gamma_j CATEGORY_j + \beta_1 IPL + \beta_2 \left(IPL \times ESL \right) + \beta_4 Zip\_Hhlds + \beta_5 Zip\_MedFamInc + v_k + \varepsilon$$

We ran the models with and without heteroskedastic error correction (corresponding to Regressions 3 and 3-Modified in Table 8 of the manuscript, respectively). The following two tables have the City and Zip level regression results respectively.

	Regression (4)			Regression (4, Modified)		
Variable	Estimate	T	Sig.	Estimate	T	Sig.
Intercept	2.887	17.266	0.000	2.887	17.266	0.000
[chain2=Stop & Shop]	-0.171	-1.225	0.221	-0.171	-1.225	0.221
[category=Condiments]	-0.892	-10.468	0.000	-0.892	-10.468	0.000
IPL	0.211	1.984	0.048	0.211	1.984	0.048
IPLxESL	0.014	0.076	0.939	0.014	0.076	0.939
City_Density	0.000	2.778	0.006	0.000	2.778	0.006
City_MedFamInc	0.000	0.119	0.905	0.000	0.119	0.905

Dependent Variable: price
NOTE: The “Chain=Other” and the Household category dummy variables are excluded from the regression because of their redundancy.

	Regression (5)			Regression (5, Modified)		
Variable	Estimate	T	Sig.	Estimate	t	Sig.
Intercept	3.135	12.178	0.000	3.138	11.892	0.000
[chain2=Stop & Shop]	-0.205	-1.083	0.307	-0.207	-1.283	0.205
[category=Condiments]	-0.890	-9.691	0.000	-0.890	-9.698	0.000
IPL	0.281	1.783	0.108	0.277	1.762	0.104
IPLxESL	-0.005	-0.015	0.988	-0.004	-0.014	0.989
Zip_hhlds	0.000	-0.438	0.672	0.000	-0.441	0.669
Zip_MedFamInc$	0.000	-0.485	0.639	0.000	-0.462	0.655

Dependent Variable: price
NOTE: The “Chain=Other” and the Household category dummy variables are excluded from the regression because of their redundancy.

The coefficient estimates of the ‘Modified’ regressions were almost identical to their corresponding model without correction for heteroskedastic errors. In both cases, the model fit was similar based on the likelihood ratio tests. Both the magnitude and the significance of the coefficient for IPL provide encouraging signs that our original conclusions are valid.

In the City level regressions (4 and 4-Modified), the coefficient for the IPL variable is 0.211 for both regressions and significant at $p < 0.05$, suggesting that IPL stores are 21.1 cents more expensive than non-IPL stores on average. In the ZIP level regressions (5 and 5-Modified), the coefficient for the IPL variable is 0.281 and 0.277 respectively (significant at approximately p = 0.10), suggesting that IPL stores are 28.8 cents and 27.7 cents more expensive than non-IPL stores on average. All these figures are in the ballpark range of 0.230 that was obtained earlier (see Table 8 of the manuscript).

On the other hand, only one of the two city level variables (density) and none of the ZIP level coefficients were significant. However, the significant coefficient (density) was very small (5.94×10^{-5}) relative to that of the IPL. The coefficients of the ZIP level variables were even smaller (less than 0.1×10^{-5}). Therefore, it is not clear how big a role these variables play in determining the variation in price levels.

The coefficients for the interaction term of IPL and ESL (IPLxESL) are statistically insignificant ($p > 0.9$ in all regressions), as they are in the regression estimation results we have reported in the manuscript for data set II (see Table 8 in the manuscript). In the City level regression the coefficient on the interaction term is 0.014. In the ZIP level regressions we observed the expected negative sign but the magnitudes was lower, –0.004. However, neither estimate is statistically significant. We also note that the validity of the results in the City level analysis is uncertain because the final Hessian matrix based on our estimates was not positive definite as it should be. For the ZIP level regressions, Regression 5-Modified also did not result in a positive definite Hessian. We therefore move to more detailed consideration of the price variation between ESL and no-ESL regimes.

<u>ANALYSIS 4</u>

This analysis investigates the possibility that the price differences observed between ESL and non-ESL IPL stores may be driven by unobserved variables, possibly an endogenous outcome driven by market conditions faced by the stores. For this we further restrict our sample to consider only the stores in Connecticut. Recall that Connecticut is subject to the IPL regime and allows ESL exemption. Therefore in our sample we have both stores that have adopted the ESL exemption and stores that have not.

We first run a simple regression model of price against the ESL variable. While not significant, the coefficient obtained in the regression is -0.166, suggesting that ESL stores are 16.6 cents cheaper in the aggregate than non-ESL stores. Note that this essentially mimics the data in Table 6 (column 6). 16.6 cents is the difference between the IPL and ESL stores in that column ($2.744 – $2.578 = $0.166).

Variable	Estimate	t	Sig.
(Constant)	2.744	21.034	0.000
ESL	-0.166	-0.984	0.327

Dependent Variable: price

Now we undertake more refined analyses consistent with the ones reported in the manuscript, specifically the regression model (3) in page 11.

We need to make some adjustments to this model however, in order to run it on the Connecticut sample. First, note that in conducting the regression analyses in data set II, we kept the Shop Rite stores out of the regression sample (see footnote 39, page 10). In keeping with this, we keep out the Shop Rite stores from the sample used in the above regression. Second, in keeping out the Shop Rite stores, we limit our ability to control for the random effect of chains because the ESL variable is now perfectly correlated with the chain variable (the only ESL stores in the reduced sample are all Stop & Shop). Third, the IPL and IPLxESL variables of regression (3) are no more relevant in this sample because there is no variation in the IPL variable. Therefore, we are left with the following model:

$$P = \alpha + \gamma_j CATEGORY_j + \beta_3 ESL + \varepsilon$$

Variable	Estimate	t	Sig.
Intercept	3.142	21.545	0.000
[category=Condiments]	-0.795	-4.724	0.000
ESL	-0.185	-1.097	0.275

Dependent Variable: price
NOTE: The variable Household is excluded from the regression because of its redundancy.

The estimate of the coefficient of the ESL variable is -0.185, suggesting that ESL stores are 18.5 cents cheaper than non-ESL IPL stores on the average. This figure is slightly different from the 16.6 cents obtained with the larger sample (with Shop Rite) but is still in the ballpark range. Note however, that the coefficient is not significant, limiting our ability at making definitive conclusions.

Now, to investigate the possible role of unobserved store level factors, we expand the above model to include city level variables as proxies for the stores' trade area characteristics. We first run a model including the city level variables: density (population per square km) and median family income.

$$P = \alpha + \gamma_j CATEGORY_j + \beta_3 ESL + \beta_4 City\_Density + \beta_5 City\_MedFamInc + \varepsilon$$

Variable	Estimate	t	Sig.
Intercept	3.806	3.835	0.000
[category=Condiments]	-0.815	-5.361	0.000
ESL	-0.272	-1.132	0.260
City_Density	0.000	-0.459	0.647
City_MedFamInc	0.000	-0.695	0.488

Dependent Variable: price
NOTE: The variable Household is excluded from the regression because of its redundancy.

Note that the coefficients for the two city level variables are both insignificant and very small. The coefficient of ESL is -0.272, suggesting that after controlling for city level factors, the ESL stores are on average about 27.2 cents cheaper than non-ESL IPL stores. However, interpretation of this result requires some caution because the coefficient is not significantly different from zero.

In a slight modification of the above model, we now include city level fixed effects rather than the city level variables used above. That is we now run the following model with η_l being the fixed effect corresponding to $City_l$:

$$P = \alpha + \gamma_j CATEGORY_j + \eta_l City_l + \beta_3 ESL + \varepsilon$$

Variable	Estimate	t	Sig.
Intercept	3.225	10.576	0.000
[category=Condiments]	-0.815	-5.347	0.000
[category=Household]	0.000	.	.
[City=Canaan]	-0.173	-0.508	0.612
[City=Greenwich]	0.027	0.111	0.912
[City=Norwalk]	0.070	0.290	0.772
[City=Stamford]	0.000	.	.
ESL	-0.271	-1.126	0.262

Dependent Variable: price
NOTE: The variables Household and Stamford are excluded from the regression for their redundancy

Note that as one would expect, the results here are very similar to the earlier regression. The coefficients for the city effects as well as for ESL are not significant. The coefficient of -0.271 for the ESL variable is also consistent with the earlier result.

In a further refinement of the earlier analyses we now attempt to use ZIP level characteristics as proxies for unobserved store effects. However, our ability to do this is severely limited, because we have data only for four of the five ZIP code areas that make up the sample. With the Shop Rite store being excluded, we have only three usable ZIP code areas. This creates some problems because we are now left with only one ESL store (S18), the other two being non-ESL IPL stores. So, not only would the explanatory power of the model be limited but the ESL variable itself would become perfectly correlated with any ZIP level fixed effects. So, while we attempted to run several specifications with ZIP level variables we ran into identification problems, given the small sample of stores.
Yet, we run the following model for completeness and offer it as an illustrative check for the role of ZIP level characteristics.

$$P = \alpha + \gamma_j CATEGORY_j + \beta_3 ESL + \beta_4 Zip\_MedFamInc + \varepsilon$$

Variable	Estimate	t	Sig.
Intercept	3.538	6.817	0.000
[category=Condiments]	-0.782	-3.906	0.000
ESL	-0.167	-0.787	0.433
Zip_MedFamInc	0.000	-0.815	0.417

Dependent Variable: price
NOTE: The variable Household is excluded from the regression because of its redundancy.

Again, within the constraints of the sample, the only real information from the regression above is the lack of significance of either the ESL or the ZIP level variable. The coefficient of the ESL variable is -0.167 suggesting ESL stores are 16.7 cents cheaper than non-ESL IPL stores on average. This is very similar to the 16.6 cents figure in the data of Table 6 (column 6) in the manuscript. The coefficient of the ZIP level variable is both insignificant and small (-2.8×10^{-6}).

To conclude, our analyses have not been able to find any conclusive evidence that the price variation between ESL and non-ESL stores can be attributed to store level variation in market conditions. On the other hand, the results of the new analyses seem to indicate greater confidence in not only the gradient but also the magnitude of our results.

<u>ANALYSIS 5</u>

Exploring county level variations as an explanation for our results.

In the earlier analyses, when we considered the role of trade area differences as an explanation for our results, we mainly looked at city and zip level variations. County level variation was included in the descriptive analyses more for completeness because we had the county level data. Yet, we did not include county level factors in our regressions framework. This was driven by our belief that the store trade area variations are more appropriately indexed by city and zip level variations.

Nevertheless, county level contribution to the observed price variation may be a reasonable argument that merits exploration. We therefore run an additional analysis in the same manner as equations (4) and (5) in the earlier analyses (pages 4-7, Analysis 3). In this new analysis, we replace the city variables with the corresponding county variables. For convenience we reproduce all the equations below, equation (6) being the new specification.[1]

$$\begin{aligned} P &= \alpha + \theta_k CHAIN_k + \gamma_j CATEGORY_j + \beta_1 IPL + \beta_2 \left(IPL \times ESL \right) \\ &\quad + \beta_4 City\_Density + \beta_5 City\_MedFamInc + v_k + \varepsilon \end{aligned} \tag{4}$$

$$\begin{aligned} P &= \alpha + \theta_k CHAIN_k + \gamma_j CATEGORY_j + \beta_1 IPL + \beta_2 \left(IPL \times ESL \right) \\ &\quad + \beta_4 Zip\_Hhlds + \beta_5 Zip\_MedFamInc + v_k + \varepsilon \end{aligned} \tag{5}$$

$$\begin{aligned} P &= \alpha + \theta_k CHAIN_k + \gamma_j CATEGORY_j + \beta_1 IPL + \beta_2 \left(IPL \times ESL \right) \\ &\quad + \beta_4 County\_Density + \beta_5 County\_MedFamInc + v_k + \varepsilon \end{aligned} \tag{6}$$

To maintain symmetry with the earlier analyses, we ran the model with and without heteroskedastic error correction (corresponding to Regressions 4 and 4-Modified in the results reported in page 7, Analysis 3 of the earlier analyses). The results are reported in the following table.

	Regression (6)			Regression (6, Modified)		
Variable	Estimate	t	Sig.	Estimate	t	Sig.
Intercept	3.141	10.418	0.000	3.141	10.418	0.000
[chain2=Stop & Shop]	-0.117	-0.860	0.390	-0.117	-0.860	0.390
[category=Condiments]	-0.892	-10.513	0.000	-0.892	-10.513	0.000
IPL	0.278	2.402	0.017	0.278	2.402	0.017
IPLxESL	-0.029	-0.156	0.876	-0.029	-0.156	0.876
County_Density	0.000	2.535	0.012	0.000	2.535	0.012
County _MedFamInc	0.000	-0.905	0.366	0.000	-0.905	0.366

Dependent Variable: price
NOTE: The “Chain=Other” and the Household category dummy variables are excluded from the regression because of their redundancy.

The results are encouraging with regards to our main findings. The IPL variable is still significant in both specifications (which return largely identical estimates anyways). The current

[1] Recall that we used Zip_Hhlds in the zip level regression because we did not have the density figures at the zip level. However, the density variable was available with us at the county level.

specification returns a coefficient of 0.278 for the IPL variable (significant at $p<0.05$) suggesting that IPL store prices are 27.8 cents higher than non-IPL store prices on average. Note that this is within the 21-28 cent range of such differences we have observed in our earlier analyses.

Interestingly, the County_Density variable is significant just as the City_Density variable was in Regression 4. However, the magnitude of the coefficient is 5.8×10^{-5} suggesting much smaller contribution to the observed price variation relative to the IPL regime.

The IPLxESL coefficient is -0.029 which is of the right sign but is not significant, with a $p>0.8$.

We also intended to conduct a similar analysis corresponding to Analysis 4 (pages 6-8). Recall that this data is restricted only to Connecticut. Unfortunately, there was no variation in the County variable (only the Fairfield county was included in the sample). Consequently we could not conduct a similar analysis here.

In conclusion, our inability to look at county differences in the CT sample notwithstanding, the results of the new analyses offer <u>further support to our belief that trade area differences are an unlikely explanation for our observed price variation</u>.

Notes on Our Data Collection Efforts

(1) We looked at publicly available census data.
(2) We studied the relevant corporate websites.
(3) We dredged the published popular press articles for any relevant articles.
(4) We contacted the individual stores for more information.

1. Economic and Demographic Data from U.S. Censor Bureau

The census data was publicly available online on the US Census website. The gateway to this information is: http://quickfacts.census.gov/qfd/index.html. A detailed compilation of data for the grocery stores in the tri-state area was collected. The US Census website provided variables such as population, number of households, average family and individual income, and lastly percentage of families and individuals below the poverty line. The data were extracted at the county level, city level, and at the Zip Code level.

2. Store/Chain Level Data from Corporate Websites

The next step in the research process was to collect data at the specific chain and store level by visiting their websites. These sites provided variables such as store location, store phone number, total number of employees in the chain, total number of stores, etc.

3. Store/Chain Level Data from Popular Press

We used Lexis/Nexis academic database to conduct a search of press releases from the stores; however the articles found were at the chain level and contained no useful information.

4. Store Level Data from Direct Personal Phone Interviews

We individually called each of the grocery stores to get more detailed information. We wanted to collect data on the number of employees at the store, square footage of the store, significant changes in store or store management in the past five years, and weekly sales figures. The store managers were contacted on August 10th and 11th in the morning and early afternoon. However, the majority of the store managers felt that they did not have the authority to answer our questions because the head office made most decisions. The managers felt the best way to get our questions answered would be to speak with a representative from the head office. Other responses by managers were very abrupt and dismissive and felt their time was too precious to take part in such a telephone interview. Head office managers were even less helpful and patient with us.

Websites that Were Referred to in the Course of Our Data Collection Process

U.S. Census website: http://quickfacts.census.gov/qfd/index.html

Grocery Retailer's Websites:

http://www.stopandshop.com/about/ - STOP AND SHOP
http://www.ctownsupermarkets.com - CTOWN
http://www.apsupermarket.com/index.asp - A&P
http://www.aptea.com/company.asp - A&P
http://www.pathmark.com/ - PATHMARK
http://www.shoprite.com/Default.aspx - SHOPRITE
http://www.shaws.com/ - SHAWS

DEFINITIONS OF THE VARIABLES

Variable	Definition
pop	Population
Dens (/km^2)	Density per square km
hhlds	Number of households
Med Hhld Inc ($)	Median Household Income (in $)
Med Fam Inc ($)	Median Family Income (in $)
Per capita Inc ($)	Per Capita Income (in $)
Fam below pvrty (%)	Percentage of Families below poverty line
Pop below pvrty (%)	Percentage of Population below poverty line