

Spatial Price Indices: The Florida Experience

GARY D. COOPER*

Price indices such as the Consumer Price Index (CPI) produced by the Bureau of Labor Statistics (BLS) receive widespread attention and play an important role in a host of economic activities such as economic stabilization programs and wage contract negotiations. These activities are concerned with the change of prices over time and the inflationary impact of such changes.

A second important type of problem also exists: the distributional impact of prices on public and private policy. Many programs funded by Federal and State legislatures call for spending over a wide geographical area. If funding formulas do not allow for differences in price level in the areas covered, low-price areas will benefit at the expense of high-price areas. In an effort to overcome this problem in its educational funding program, the Florida Legislature authorized in 1972 the compilation of an index to indicate the relative price level in each of Florida's 67 counties. This paper describes the Florida Price Level Index (FPLI), the spatial index developed to meet this need. Construction of the FPLI is an ongoing responsibility of the Florida Department of Administration, which has produced annual indices since 1972.¹ Results of the 1975 survey, which are used as illustrations throughout the paper, are reported more fully in [3].

The desirability of having spatial price level indices makes the lack of data in this area suprising. The BLS Family Budgets program [8] has been the only exception; it provides dollar amounts for three normatively selected budgets for 40 cities each Fall. However, because of limited coverage, the Family Budgets are of little use for intra-region and intra-state analysis. The CPI, also produced by BLS, provides indices for 56 cities on a monthly basis, but these indices measure price changes over time. Since different market baskets of goods are priced in each city, the BLS cautions explicitly against making intercity comparisons.

I. CHARACTERISTICS OF THE FLORIDA PRICE LEVEL INDEX

The FPLI is used to equalize the purchasing power of state educational funds distributed to Florida's counties. It is similar to the CPI in that it does not measure differences in cost of living but rather measures changes in prices, the most important cause of cost of living changes.

*Florida Department of Administration. I wish to thank Tony Huggins for valuable suggestions and comments on an earlier draft of this paper.

To measure these price differences, a market basket, defined by a set of items and the quantities of each purchased, is chosen and priced in each county. The index value for a county is the ratio of the market basket price in that county and in a base county. The market basket is selected to represent the purchasing patterns of a representative individual in the base county, so the index answers the question: How much would it cost a representative base county individual to buy his market basket of goods in each county? It does not measure how the base county representative individual would actually spend his money if he were transplanted to each county. It is the well-known fixed base or Laspeyres index:

$$PLI(c) = \frac{\sum_{t=1}^T p(c,t) q(b,t)}{\sum_{t=1}^T p(b,t) q(b,t)} \quad (1)$$

where
 PLI(c) is the price level index for region c,
 p(c,t) is the average price in region c of item t,
 p(b,t) is the average price in the base region of item t,
 q(b,t) is the quantity of item t purchased in the base region b,
 T is the number of items priced.

The FPLI does not measure differences in the cost of living as experienced by the representative resident in each county. This would require using a different standard in each county and the resulting index would reflect both price and standard differences, a result inconsistent with the purposes of the index.

The choice of a base county can have an effect on the range of the index as well as the ranking of individual counties. This can be more easily seen if equation 1 is rewritten:

$$PLI(c) = \frac{\sum_{t=1}^T p(c,t) q(b,t) \frac{p(b,t)}{p(b,t)}}{\sum_{t=1}^T p(b,t) q(b,t)}$$

or

$$PLI(c) = \sum_{t=1}^T \left[\frac{p(c,t)}{p(b,t)} \left(\frac{p(b,t)q(b,t)}{\sum_{t=1}^T p(b,t)q(b,t)} \right) \right] \quad (2)$$

The expression in the inner brackets is the item weight or "relative importance": the percent of total expenditure accounted for by the item.

If the base county has both a low (high) relative importance and price for an item, then the index will tend to minimize (maximize) price differences for the item. No county can be lower (higher) in relative importance and price for all items, so the effects have to be netted out across all items. Because of the large number of items in an index such as this, it is difficult to develop apriori notions of what the net effect would be.

The market basket used in the FPLI represents the consumption patterns of the average household headed by a wage earner or clerical worker in Orlando, Florida, as determined by an in-depth consumer expenditure survey. The survey, which was conducted by the BLS, is the only one currently available for a Florida city. No income or family size requirements were set for inclusion in the sample. The survey is the same one used by BLS as the basis for their CPI price collection program in Orlando. The consumer expenditure survey enumerates all items purchased over the period of a year—about 1800—and the percent of total expenditure accounted for by each item; it's "relative importance."

The BLS market basket for Orlando includes over 400 items in 52 expenditure classes. Examples of expenditure classes are bakery items, men's footwear and tobacco products. In coming up with the final Orlando sample, BLS divided the items in each expenditure class into two categories: certainty items—the most important weightwise—and the remainder, known as probability items. All of the certainty items are included in the final sample. Only a sample of probability items are chosen, however, the ones chosen being representative of all other expenditures in that class.

The FPLI market basket is a subset of the final BLS one. Items are selected based on their suitability for spatial pricing (as explained below) and their relative importance. Certainty items are the first choice for inclusion. Some of them, however, do not meet the suitability requirement and have been replaced with probability items from the same expenditure class. Additional items from the probability category are also chosen, and in 1975 a total of 128 items were priced.

Suitability for spatial pricing is determined by the answers to two questions. First, can a specification of the item's characteristics be writ-

ten that will uniquely delineate the item, but still be useful? Dozens of interviewers across the state must be able to find identical items based on the specification. If the specification is not reasonably precise, different items, and consequently different standards of living, will be priced. Second, given the specification, will the item be generally available? Some items in each expenditure class meet these often conflicting requirements better than others.

The analagous quality-change problem in creating a temporal index is solved by going back to the same outlet and pricing the identical item each time period. Thus, even if a slightly different quality of the same generic item is priced in each outlet in the city sample, the composition of the item sample will remain constant over time, and meaningful price comparisons can be made. This solution is not appropriate in constructing a spatial index since it is precisely the outlet-to-outlet comparisons that are important and the item must be identical in all outlets. Preserving this strict comparability is a major problem in spatial indices, and great care must be taken in writing the pricing specifications for each item.

II. ESTIMATING THE INDEX

The FPLI is calculated using formula 2. Since the relative importances or weights in the formula are available quarterly from BLS, the only additional information required is a set of county average prices for the items in the market basket.

The county average prices are to be interpreted as the average or typical price paid for that item in a county. Prices for goods are determined based on a sample of outlets drawn from the set of all retail outlets in each county. The sampling frame is provided by the Florida Department of Revenue from their records of outlets collecting the Florida Sales Tax. Examination of this population shows a severe skew in the distribution of outlets by annual sales. The top ten percent of each kind of store accounts for more than 50 percent of sales, while the smallest 50 percent accounts for only 10 percent of sales. Since different size stores are very likely to have different price structures, a simple random sample of outlets will not yield average prices that reflect the typical price paid for the items. Thus, the outlet sample for an item needs to be stratified based on some measure of item sales volume. Since item sales volume data by outlet is not available, total sales by outlet is used as a substitute. This is done on the assumption that the percent of sales accounted for by an item is constant across all outlets of a given kind. More information on how the outlets were selected for pricing is provided in Zahn [9].

For a few items—mostly services which are not subject to the state sales tax—sales or volume of business measures are not available and a simple random sample is drawn. It is not expected that this will bias

the results because the providers of these services are much less likely to exhibit the skewed distribution by sales volume that retail outlets do.

Some items are sold in more than one kind of outlet, so the sample must be further stratified by kind of outlet, based on the proportion of item sales in each kind. This information is compiled from data in the *1967 Census of Business* [7]. The estimate for the county average price of an item is thus built up in three stages. The stratum average prices of the item within each kind of outlet are computed, then the stratum average prices are combined to form kind average prices, and finally these kind average prices are combined to form the county average price. The formula for the county average price is shown below as equation 3.

$$p(c,t) = \frac{\sum_{k=1}^{K(c,t)} W(t,k) \left[\frac{\sum_{h=1}^{H(c,t,k)} S(c,t,k,h) \left\{ \frac{\sum_{o=1}^{N(c,t,k,h)} p(c,t,k,h,o) S(c,t,k,h,o)}{\sum_{o=1}^{N(c,t,k,h)} S(c,t,k,h,o)} \right\}}{H(c,t,k) \sum_{h=1}^{H(c,t,k)} S(c,t,k,h)} \right]}{K(c,t) \sum_{k=1}^{K(c,t)} W(t,k)} \quad (3)$$

where: $p(c,t,k,h,o)$ is the price collected in outlet o in stratum h , kind k for item t in county c ,
 $S(c,t,k,h,o)$ is the annual sales of outlet o in stratum h , kind k , where item t is priced in county c ,
 $N(c,t,k,h)$ is the number of outlets in stratum h , kind k , where item t is priced in county c ,
 $H(c,t,k)$ is the number of stratum in kind k , where item t is priced in county c ,
 $K(c,t)$ is the number of kinds where item t is priced in county c , and
 $W(t,k)$ is the percent of state-wide sales of item t in kind k .

Equations 2 and 3 provide a procedure for estimating the index of a county where all prices are collected. Because of budget constraints, however, a complete set of prices is not collected in each county. Certain items—which include housing, utilities, and services—can be priced efficiently from the central office, and they are priced in all 67 counties. For the remaining items, which are priced in the field by visits to retail outlets, prices are collected in a sample of counties. In 1975, 41 of the 67 counties were surveyed. Field-priced items constitute the bulk of

the item sample, about 75 percent of the items and 60 percent of the relative importance.

The PLI for counties where a full set of prices is not collected cannot be estimated in the manner described above. Rather than building the index up from individual prices, the PLI is estimated in these counties using a regression model developed for the surveyed counties. The literature does not provide a theory of spatial cost of living differences which can be used as the basis for this estimation procedure. Indeed, the only discussion of the determinants of spatial cost of living differentials is in Haworth-Rasmussen [4], but their model is not sufficiently precise to use as a forecasting tool. As the need for predictive accuracy is more important than theoretical specification, a strictly forecasting strategy is developed. Thus, the purpose of the regression is to develop the best estimate for the non-surveyed counties, rather than to find the equation which best describes the surveyed counties.

To develop the forecasting model, a stepwise regression is run on a bank of 100 economic variables for each county. If the counties to be estimated belong to the same model that those in the regression regime do, the standard stepwise technique, which uses a measure of goodness of fit as the variable selection criterion is appropriate. Since there is no underlying causal model, there is no basis for making this assumption. It is possible, because of the great heterogeneity of the counties, that the county sample will not capture the salient features of all the counties, and that a different model will prevail in each set of counties. To overcome this problem, a new technique was developed. Since the goal is a set of index values that are as precise as possible, at each step in the regression the criterion is to choose the variable that results in the lowest estimated total standard deviation. The resulting model then reflects characteristics of all the counties, not just those of the surveyed counties. Killion, Zahn, and Cooper [5] gives a more detailed description of this technique.

The sample of counties where all prices are collected is selected using a statistical technique called cluster analysis. Counties are split into clusters based on variables or attributes thought to be related to, but not necessarily determinants of, county price levels. These variables are the ones employed in estimating the index values for the non-surveyed counties the previous year. The number of clusters is determined by the sample size desired, and one county is selected from each cluster. To the extent possible, counties that were not surveyed in the previous year are selected. It is important that the sample be representative of all counties, because the data collected in the sample will form the basis for estimation of the index values for the non-surveyed counties.

III. MISSING DATA

Because there are no restrictions on size for inclusion in the county sample, problems of item unavailability occur in the smaller counties in

the sample. There are 11 counties in Florida with populations less than 10,000 and 12 counties have no city of population 2,500. In small counties, a market simply does not exist for some items and as many as one-third of the items may be unavailable in extreme cases. For items that can be purchased elsewhere, prices are obtained in the places residents of these counties actually purchase them. Information on purchasing patterns is obtained through a survey of residents in each of the small counties. Selected residents are asked in what counties they shop for 25 groups of items. The substitute price for a missing item is estimated as a weighted average of county average prices from these counties. Weights are the number of respondents going to each county to shop for the item.

Additional transportation expenses for this shopping are assumed to be zero. Extra expenses would normally be incurred for large durable items or groups of other items bought in one shopping trip. In either case, the extra costs will be a small portion of total expenses and will be at least partially offset by a tendency to concentrate shopping into fewer trips.

It should be noted that this substitution procedure is used only when no market exists for an item in a county. If an item is available within a county, its price is used; the FPLI is a measure of prices prevailing in a county, not of the purchasing habits of its residents.

For items that cannot be reasonably purchased elsewhere, other techniques must be used. For example, in some small counties, people rely on private wells and septic tanks because there are no municipal water or sewer facilities. In these cases, the weights for items such as water and sewer charges are distributed proportionately over other items in the same expenditure category.

IV. ESTIMATING HOUSING PRICES

Collecting prices based on a standard item specification is not feasible for apartment rent and monthly house payments. There is a bewildering array of different features and qualities available, and trying to find a sufficient sample of units of one specific type is exceptionally difficult. To overcome these problems, hedonic estimation techniques are employed. Although application of this technique to housing has been studied by BLS personnel [2], the FPLI is the only index using it on an ongoing basis [1]. The following discussion for apartment rent illustrates the technique which is used for both rent and house value. The technique is to collect data on a large number of apartments with different characteristics and to adjust them later to fit a standard specification. A sample is drawn in each county from the set of all apartment complexes with three or more units. Information on monthly rent as well as characteristics of apartments in these complexes is collected through a survey of apartment owners.

Monthly rent is regressed on the apartment characteristics and several control variables to determine the contribution of each characteristic to the total rent. The control variables are included to capture variations in rent which are due to location and quality. They are included not because we wish to adjust rents for them, but to insure that the characteristic variables represent just these characteristics and are not serving as proxies for unincluded variables. The regression coefficient for each characteristic shows the dollar influence of the characteristic on final rent.

By using the regression coefficients, the rent of each surveyed apartment may now be adjusted to what it would be if the apartment met a standard specification. For example, if the standard calls for air conditioning and the regression coefficient for this characteristic were \$45.00, then rent values for all apartments without air conditioning would be increased by \$45.00 to show what the rent would be if they had air conditioning. This procedure is repeated for all other characteristics.

For single family housing, prices and characteristics of new units are collected and adjustments made as with apartments. The purchase price is then converted to a monthly principal and interest payment using average interest rates for each county.

V. RESULTS

Construction of the index is a large undertaking. For the 1975 index, 175 interviewers went to 5,073 retail outlets in 41 counties to collect 34,173 prices on 97 items in a period of five days. In addition, central office personnel collected 10,309 prices for another 31 items in all 67 counties. PLI's for the 41 field surveyed counties were estimated using equations 2 and 3. The remaining 26 PLI's were estimated using a regression model, as described earlier. The central component is by far the most important variable in the forecasting equation, as is shown below. Numbers in parentheses are t-ratios.

$$\text{PLI} = 35.51 + 1.21X_1 - 2.57X_2 + .65X_3 - .64X_4$$

$$(30.32) \quad (15.58) \quad (3.44) \quad (2.52) \quad (1.70)$$

$$R^2 = .96 \quad \text{std error} = .99$$

where: X_1 = portion of index contributed by centrally priced items,
 X_2 = percent of county sales in department and general stores,
 X_3 = number of apartments per capita, and
 X_4 = percent of county sales in drug stores.

The resulting set of PLI's is based on Orange County (Orlando), since the Orlando consumer expenditure survey defines the standard. This is not a particularly meaningful basis for comparison, however. A more appropriate base is the average price paid by residents of the state to purchase the Orlando market basket. The index is rebased by dividing each county's PLI by a population-weighted average of all the PLI's. These transformed PLI's are known as the Florida Price Level Index.

The final set of FPLI values are shown in Table I. There is a wide dispersal of values, with a range of nearly 20 points; nevertheless, three-fourths of the counties are within eight points of each other. The spatial distribution of the index values is shown in Figure I. The lowest indices are for small, rural counties, usually in the northern part of the state. The high indices are mainly in the southern, highly urbanized counties, where in-migration and tourism have had a large impact on the local economy. These counties are not necessarily the largest ones, however. There are twelve counties with indices over 100.00; of these, six have less than 60,000 population, an additional two have less than 500,000, and the remaining four have over 500,000.

In looking for the sources of the variation in index values, it is useful to disaggregate the FPLI into several major components and to see how each behaves. Table II gives some basic information on the components while Table III shows the R^2 's from simple regressions of various parts of these variables. There are actually two breakdowns; one among the five components, and one between the field and centrally priced items. The striking thing about these data is the behavior of the Housing component. It has a range nearly twice that of the total index, and accounts for .93 percent of the variation in the total index. Food and Apparel on the other hand have very low explanatory powers of .28 and .23 percent respectively. Examination of the R^2 matrix shows that each of the components varies from county to county in a fairly independent fashion. There is very little relation between Food or Apparel and any other component. Health and Transportation are each fairly strongly related to Housing, and to a lesser extent, to each other. Variations in the non-housing components, which are of about the same magnitude, largely cancel out and thus the large variations and weight of the housing component dominate the index. The fact that centrally priced Housing items account for 70 percent of the central weight and 80 percent of the housing weight explains the large explanatory power of the central component in the estimating equation for the non-surveyed counties.

VI. CONCLUSIONS

Florida's experience in the past four years with the FPLI shows that spatial indices are viable and useful. Since 1973 the Florida Legislature has used the FPLI as part of its formula for the state-wide disbursement of educational funds. There has been considerable private demand for the index as an aid in determining geographic pay differentials. For research purposes the index and its underlying prices provide a useful source of data for work on determinants of spatial price differentials. The index is designed explicitly for spatial comparisons and provides intensive price information for a large and varied area. The only other source for this type of information is the BLS Family Budget data which covers only 40 cities nationally.

FIGURE I
FPLI VALUES BY FLORIDA COUNTY, 1975

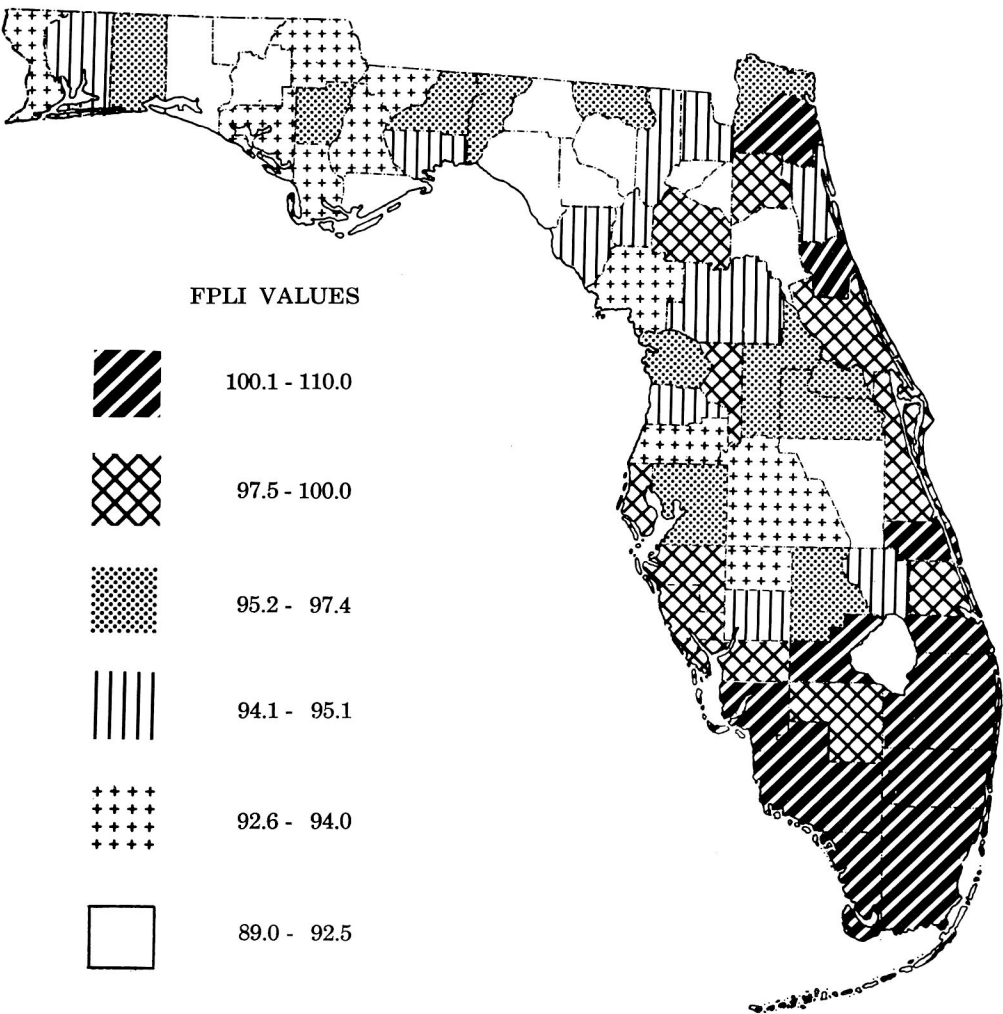


TABLE I
FLORIDA PRICE LEVEL INDEX OCTOBER 1975
(Population-Weighted State Average = 100)

County	1975 Index	County	1975 Index
Alachua	97.62 (21)†	Lafayette	91.24 (64)
Baker	95.09 (35)	Lake	96.53 (27)
Bay	93.84 (46)	Lee	101.24 (9)
Bradford	92.29 (58)	Leon	97.24 (24)
Brevard	97.75 (20)	Levy	93.05 (52)
Broward	104.02 (4)	Liberty	93.09 (51)
Calhoun	96.31 (28)	Madison	92.37 (56)
Charlotte	97.47 (22)	Manatee	98.95 (14)
Citrus	96.13 (29)	Marion	94.61 (42)
Clay	99.55 (13)	Martin	103.48 (7)
Collier	103.56 (6)	Monroe	109.26 (1)
Columbia	94.77 (40)	Nassau	97.10 (25)
Dade	107.39 (2)	Okaloosa	96.07 (30)
DeSoto	94.57 (43)	Okeechobee	94.84 (39)
Dixie	94.19 (44)	Orange	95.22 (33)
Duval	100.21 (10)	Osceola	91.31 (63)
Escambia	93.58 (48)	Palm Beach	104.42 (3)
Flagler	103.74 (5)	Pasco	93.29 (49)
Franklin	92.28 (59)	Pinellas	100.11 (12)
Gadsden	92.80 (54)	Polk	92.76 (55)
Gilchrist	94.98 (37)	Putnam	91.82 (62)
Glades	100.15 (11)	St. Johns	95.09 (35)
Gulf	92.98 (53)	St. Lucie	98.51 (15)
Hamilton	95.71 (31)	Santa Rosa	94.72 (41)
Hardee	93.69 (47)	Sarasota	98.23 (17)
Hendry	98.10 (18)	Seminole	95.22 (33)
Hernando	94.88 (38)	Sumter	98.37 (16)
Highlands	95.50 (32)	Suwannee	90.07 (66)
Hillsborough	96.60 (26)	Taylor	92.15 (61)
Holmes	89.62 (67)	Union	92.35 (57)
Indian River	102.47 (8)	Volusia	97.80 (19)
Jackson	93.12 (50)	Wakulla	94.09 (45)
Jefferson	97.42 (23)	Walton	92.28 (59)
		Washington	90.71 (65)

†Number in parentheses is rank

TABLE II
FPLI COMPONENT CHARACTERISTICS

Component	Number of Items	Weight	Range of Vales
Food	32	23.62	14.92
Housing	24	33.52	34.45
Apparel	17	9.39	21.88
Transportation	16	13.06	19.88
Health, Recreation, & Personal Services	39	20.41	17.59
Total	128	100.00	19.64
Field	97	62.36	5.27
Central	31	37.64	15.82

TABLE III

SIMPLE REGRESSION R²'S AMOUNT FPLI COMPONENT INDICES

Food	.282						
Housing	.931	.138					
Apparel	.233	.013	.183				
Transportation	.587	.256	.430	.082			
Health, Recreation, and Personal Services	.591	.073	.487	.047	.349		
Field	.546	.699	.347	.243	.397	.228	
Central	.939	.125	.962	.176	.517	.607	.305
	Total	Food	Housing	Apparel	Transportation	Health, Recreation, & Personal Services	Field

FOOTNOTE

¹The current index is the cumulative result of many persons efforts. The initial design of the study was under the direction of Dr. James C. Simmons, the first project director and Dr.

Douglas A. Zahn, statistical consultant. A chronology of their efforts is found in Simmons. [6]

REFERENCES

- [1] Bellante, D. and Killion, R. A., "The Adjustment of Spatial Price Indices for Geographical Differences in Standard of Consumption: An Application of the Hedonic Method", *Review of Regional Studies*, Vol. 6, No. 1, Spring 1976, 109-113.
- [2] Gillingham, Robert, "Place to Place Rent Comparisons", *Annals of Economic and Social Measurement*, Winter, 1975, 153-73.
- [3] Florida Department of Administration, *Florida Price Level Index 1975*, Tallahassee, Florida: March 1976.
- [4] Haworth, C. T. and Rasmussen, D. W., "Determinants of Metropolitan Cost of Living Variations", *Southern Economic Journal*, October 1973, 193-201.
- [5] Killion, R. A., Zahn, D. A., and Cooper, G. D., "The Total Standard Deviation as a criterion for choosing variables in a Multiple Regression, with Applications to the Florida Price Level Index", *Proceedings of the American Statistical Association, Statistical Computing Section*, 1975, 124-27.
- [6] Simmons, James C., ed., *Florida Cost of Living Study: Florida Counties Price Level Index (FPLI)*, Tallahassee, Florida State Board of Regents, June 1973.
- [7] U.S. Bureau of the Census, *Census of Business, 1967; Retail Trade: Merchandise Line Sales, Florida*, Bulletin No. BC67-MLS-11, 1970.
- [8] U.S. Department of Labor, Bureau of Labor Statistics, *Handbook of Methods*, BLS Bulletin 1711, 1971.
- [9] Zahn, D. A., "Statistical Problems Relating to the Florida Cost of Living Survey", *Florida State University Statistics Report*, July 1973.