1968

Inter-city Variations in Retail Sales: Some Hypotheses Revisited

Roy W. Bahl
Georgia State University, rbahl@gsu.edu

Lawrence X. Tarpey

INTER-CITY VARIATIONS IN RETAIL SALES: SOME HYPOTHESES REVISITED

by
L. X. TARPEY
University of Kentucky
and
ROY W. BAHL
Post-doctoral Fellow
Syracuse University

INTRODUCTION

In general, marketing theory is not well formulated. This is especially true in certain areas such as the spatial and temporal dimensions of shopper behavior. An important question which, to date, has not been fully answered is why do retail sales vary between cities. While population accounts for a substantial amount of inter-city differences in the aggregate level of sales, the observed variability in *per capita* retail sales necessitates an explanation in terms of non-population factors. The authors undertook a study to identify and measure some of these factors of variation in an attempt to replicate and extend the earlier research of Mrs. V. K. Russell and Dr. Robert Ferber. In a 1957 article, Mrs. Russell statistically tested the "closed system" hypothesis.[1] Her findings revealed a positive correlation between per capita income and per capita sales *only* when a city was a closed system.[2] Generally speaking, SMSA's and individual communities are *not* closed systems and therefore little or no systematic relationship exists between income and sales.

Ferber applied a multivariate model in an attempt to identify factors affecting differentials in retail sales among 51 Illinois cities and to measure the relative importance of each factor.[3] He found income to be of little importance in explaining inter-city sales' variations, while city population and distance to the next largest accounted for a substantial proportion of explained variation.

A primary objective of this study was to determine whether there are measurable economic or non-economic factors, in addition to those suggested by Ferber and Russell, which significantly affect intercity variations in retail sales. Further, the present study, by extending Ferber's technique and interpretation to a consideration of the implications of multicolinearity,[4] may well provide a more adequate conclusion as to why retail sales vary among cities.

The Approach. In order to gain a more comprehensive picture of retail sales variations among cities, the present analysis is accomplished on two levels. The first involves regressing total retail sales on ten

independent variables. The second involves measuring the effects of these ten independent variables on *per capita* sales for six different types of retail stores (i.e., census establishments). The types of stores included are: (a) general merchandise, (b) food, (c) automotive groups, (d) furniture and appliances, and (e) drug. The sample is comprised of 136 cities in the four states of Alabama, Mississippi, Kentucky and Tennessee. Only cities with populations between 5,000 and 25,000 are included. Retail sales data are taken from the 1958 Census of Business, while data for most explanatory factors are taken from the 1960 Census of Population.

The Model

The present study makes use of a single equation linear model ($Y = a + \Sigma b_i X_i$) in which the respective b_i are net regression coefficients. Ferber used a multiplicative model ($Y = \Sigma x_i b_i$) where the exponents (b) yield estimates of the elasticities of the respective independent variables. Corresponding to Ferber's use of elasticities, the explanatory variables in the present study are ranked in importance by the magnitude of their standardized regression or beta coefficients.[5] In both Ferber's and in the present study, the data are a cross-section sample of cities at a fixed point in time, and in that respect, both analyses are static. This means that while the results of the correlation analysis may contribute much to an understanding of inter-city *differences* in retail sales, little may be inferred as to the *changes* or dynamics of retail sales patterns.

Assumptions. The present research is based on three important assumptions. (1) Intercity variations in retail sales are a function of more than one variable. (2) The relative importance of each variable differs with the class or type of retail commodity being considered. (3) Factors affecting among-city differentials in total retail sales are not identical with those which explain variations in per capita retail sales.

Independent Variable. For convenience of analysis the ten independent variables used in this study may be grouped into three classes. These are: (a) demand factors, (b) supply factors, and (c) socio-economic factors. Each of the ten independent variables is placed in one of these categories as shown below.

Classification of Independent Variables

Demand Factors

Population (X_1)

Median Family Income (X_2)

Families with income less than $3,000 ($X_5$)

Families with incomes over $7,000 ($X_6$)

Percent unemployed (X_9)

Supply Factors

Driving time to next larger city (X_3)

Driving time to nearest SMSA (X_4)
Number of retail establishments per 10,000 residents (X_{10})

Socio-economic Factors
Median school years completed (X_7)
Percent of labor force in agriculture (X_8)

Population (X_1) is included to measure the effect of city size on retail sales, however, intercorrelations with other explanatory variables present difficulties in ascertaining the separate effect of population. Median family income (X_2), percent of families with incomes over $7,000 ($X_6$) and percent of families with incomes under $3,000 ($X_5$) are included to measure the effect of the average income level and the distribution of income. Both the driving time to the next largest city (X_3) and the driving time to the nearest SMSA (X_4) are leakage variables because the distance does provide some measure of the spatial restrictions facing the residents of smaller cities who desire to search for merchandise in the larger shopping areas. The number of retail stores per 10,000 residents (X_{10}) is also a leakage variable in that it reflects the availability of merchandise — a factor which tends to pull shoppers to a city. Percent unemployed (X_9) may provide a sensitive indicator of the general economic climate of the city.

Two of the independent variables are included to measure the influence of socio-economic differences on retail sales levels. Variations in median school years completed (X_7) may suggest something about differing tastes of consumers, in addition to reflecting inter-city income differentials. It could also be a leakage factor as higher educational levels may result in greater shopper migration because of a search for more variety or higher quality merchandise. Percent of labor force employed in agriculture (X_8) reflects a socio-economic diversity between city and rural populations and yields a measure of the effect of differing tastes and buying characteristics.

THE RESULTS

Total Retail Sales

As would be expected, most of the variability in total retail sales among cities is attributable to differences in population. Population alone accounts for 57 percent of the variability in total retail sales and the addition of the stores variable raises explained variation to approximately 80 percent. The remaining eight independent variables contribute very little (less than 2 percent) to explained variation.

The importance of the population variable may be reduced substantially by analyzing a sub-sample of 46 cities having populations between 15,000 and 25,000. The effect of using this more homogeneous sub-sample of cities is a decrease in the importance of the population variable and an increase in inter-city sales variations attributable to differences in the num-

ber of retail establishments per 10,000 residents (X_{10}).

On the surface, Ferber's results seem somewhat better than those of the present study. His six variable model on 51 Illinois cities explained over 95 percent of the variation in total retail sales. Four of his variables, however, did not prove statistically significant at either the .05 or the .10 level, consequently the population and distance variables in his model explained almost 95 percent of the sales variation. By using standardized regression coefficients Ferber was able to show that population is by far more important than distance, having an influence on sales more than eight times as large as distance.

The results of the present research, as well as those of Ferber's, differ substantially from the findings of traditional consumption analyses since the income variables have little influence on the level of total retail sales. This does not mean that income is of no importance in determining the level of retail sales but, instead, it probably indicates that income is a poor proxy for measuring many other social and economic forces which contribute to shopper behavior patterns. Russell's study on the closed system hypothesis clearly shows that cities are not "closed systems." This finding suggests that low cor-
lations between income and sales
to be expected in the present
ysis.
. simple ranking of the inde-
lent variables by their standar-
d or beta coefficients reveals
population (X_1) and the number of stores per capita (X_{10}) to be the most "important" variables in the model — both having beta coefficients five times larger than that of any other explanatory variable.[6] As in Ferber's study, the distance factors (X_3 and X_4) are found to be of little importance in the present analysis. However, by restricting our analysis to the sub-sample of the 46 largest cities the driving time to the nearest SMSA (X_4) becomes quite important and is a better measure of distance than is driving time to the closest largest city (X_3).

If one deletes the per capita stores variable from the model it has the effect of increasing the explanatory power of the percent of families with incomes greater than $7,000 ($X_6$). The explanation for this is to be found in the intercorrelation between per capita stores (X_{10}) and this particular measure of the income distribution ($r = -.51$). Thus in a given city the greater the percentage of families with incomes in excess of $7,000, the lower will be the number of stores per person. In addition, the nearest SMSA will be closer in terms of driving time and consequently total retail sales will be lower because of the presence of these leakage factors. These results confirm the Russell hypothesis which, among other things, states that high income suburbs are not closed systems.

Table 1 shows the simple correlation coefficient between all combinations of two independent variables. These intercorrelations indi-

cate the nature of the multicolinearity problem referred to earlier (see footnote 4). It may be seen from this table that the three income variables are highly interrelated; however, by making these interrelationships explicit, a clearer insight into the nature of intercity variations in retail sales may be obtained. The importance of the population variable may also be analyzed within a context of these relationships. These inter-correlations indicate that cities with larger populations are also characterized by: (a) higher median incomes, (b) greater driving times to larger cities, (c) smaller proportions of families in the less than $3,000 income bracket, (d) lower rates of unemployment, and (e) smaller numbers of stores per capita.

It has been shown that when the analysis is restricted to the subsample of 46 cities having populations of 15,000 to 25,000, the importance of the population size variable is diminished substantially, though it remains significant as does per capita stores. This reduction in variance of the population factor results in increased relative importance of percent of families in the upper income bracket and driving time to the nearest SMSA. Thus, when only the larger cities are considered, distance to larger retail centers is shown to be an important explanatory factor, thereby bringing the findings of this analysis into closer agreement with Reilly's law of retail gravitation. In each of these models, however, population size is of greater importance than driving time to larger cities — thereby contradicting the basic axiom of Rielly's law.[7] This is understandable since Reilly's model includes only two variables (population and distance) and is concerned solely with shopping goods. This apparent contradiction with Reilly's law does not relate to his use of distance and population as proxy variables for potential explanatory factors[8] but instead to the relative importance which he attaches to each.

The high association between per capita stores and total retail sales presents a hazy cause-effect relationship. Findings of this analysis indicate that much of the explanatory importance of the per capita stores variable lies in a series of interrelationships with demographic, economic, and social characteristics of the cities in the sample. Table 1 shows that cities having a greater number of stores per resident have: (1) greater percentages of families in the lower income bracket, (2) smaller percentages in the upper income bracket, (3) greater proportions of the labor force employed in agriculture, and (4) longer distance (trip time) the nearest SMSA. Each of these relationships suggests a minimum of leakage and some amount of shopper immigration, and, therefore, a higher level of retail sales. Hence given equal populations, the estimating equation will yield higher levels of sales for cities having a greater number of retail establishments and vice versa (i.e., stores breed sales). The negative correlation between population and stores per capita is explained in terms of capacity and shopper mi-

TABLE 1

INTERCORRELATIONS AMONG INDEPENDENT VARIABLES[a]

		1	2	3	4	5	6	7	8	9	10
Population	= 1		.20*	.35*	-.05	-.18*	.29*	.05	-.22*	.01	-.39*
Median Family Income	= 2			-.23*	-.35*	-.83*	.91*	.11	-.37*	-.39*	-.49*
Driving Time to Next Larger City	= 3				.17*	.27*	-.19*	-.09	-.01	.02	.09
Driving Time to Nearest SMSA	= 4					.41*	-.34*	.03	.21*	.16	.29*
Percent of Families with Incomes less than $3,000	= 5						-.79*	-.15	.47*	.41*	.47*
Percent of Families with Incomes greater than $7,000	= 6							.12	-.35*	-.39*	-.51*
Median School Years Completed	= 7								-.11	.03	-.03
Percent of Labor Force Employed in Agriculture	= 8									.13	.27*
Percent Unemployed	= 9										.02
Number of Retail Establishments per 10,000 residents	=10										

[a]*Asterisk denotes significance of the simple coefficient at the .05 level.*

gration. Large towns have stores which are equipped to serve customers from a wider trading area than the city limits but this is not generally true for small towns which are not equipped as "trading centers."

Per Capita Retail Sales

One possible method of abstracting from the effects of population size on intercity sales variation is to express retail sales on a per person basis. Retail sales of the 136 cities and 46 city sub-sample are analyzed on a per capita basis for total retail sales and for sales of six merchandise categories. In most instances, more variation in per capita sales of the 46 larger cities may be explained than in the 136 city model. Ferber was able to explain a maximum of 64 percent of the variation of total retail sales for his 51 cities. The models employed in the present research resulted in explained variations of 67 and 72 percent for the 136 and 46 city models respectively. (See Table 2.)

In the 136 city model the number of retail stores per person accounts for about 61 percent of per capita retail sales variation, thus the addition of the remaining nine variables explains less than 10 percent. In the 46 city model, per capita stores accounted for 70 percent of sales variations — the other nine explanatory variables contributing only 2 percent to explained variation. These results confirm Ferber's findings which show the number of retail stores per person to be the most important variable in explaining variations in total per capita retail sales.[9]

More of the variance in per capita general merchandise, apparel, and automotive sales is explained when the 15,000-25,000 population group is considered separately. The finding that driving time to the nearest SMSA is significant for both furniture and apparel sales is not unexpected if one considers these items to be in the shopping goods category. This finding parallels Reilly's law of retail gravitation. In the 46 city model the importance of per

TABLE 2

COEFFICIENTS OF MULTIPLE DETERMINATION[d] ON PER CAPITA RETAIL SALES: A COMPARISON OF THE TEN VARIABLE MODELS OF THE PRESENT STUDY WITH FERBER'S MODEL

Sales by Type of Store	Model 1[a] R^2	Model 2[b] R^2	Ferber's[c] R^2
Total	.67	.72	.64
Food	.19	.40	.21
Drug	.05	.01	.51
General Merchandise	.05	.12	.39
Furniture	.30	.24	.64
Automotive	.07	.20	.50
Apparel	.42	.46	.64

[a]Model 1 includes all 136 cities in the four states.
[b]Model 2 is a sub-sample of 46 cities in the four states with populations between 15,000 and 25,000.
[c]Ferber, *op cit.*, Table 4, p. 301.
[d]Adjusted for sample size.

capita stores is diminished substantially.

The leakage factors and the measures of general economic climate assume roles of increased importance in the 46 city analysis. Especially noteworthy is the inverse relation of both percent of families with incomes under $3,000 and percent of families with incomes over $7,000, with per capita furniture sales. A greater proportion of families in the lower income brackets indicate a lower effective demand for new furniture therefore a lower sales level, while a higher proportion in the upper brackets probably indicates greater leakage and consequently lower sales.

The percent of families with incomes of $3,000 or less can be considered a non-leakage factor. That segment of the population represented by this low income class has the following characteristics which affect their shopping behavior: (1) They are usually locked in by local credit buying arrangements for most shopping goods and even for necessities such as food. (2) They have little or no discretionary income or discretionary credit which prevents them from purchasing significant amounts of fashion or luxury items. (3) Family size, availability of transportation, and in general the costs of search (i.e., shopping) all work to constrain their "going to the big city" too often. This income variable is strongly intercorrelated with education level, percent unemployed, and agricultural employment which suggests that it does not tend to be a leakage factor.

A ranking of the independent variables by the size of their beta coefficients supports this logic. For per capita apparel sales (a shopping goods category), the percent of families with incomes less than $3,000 may be treated as a non-leakage factor because it is of minor importance in the model. Education, unemployment, and agricultural employment are also of minor importance in explaining apparel sales. The population factor appears to be unimportant as per capita stores and the percent of families with incomes greater than $7,000 assume dominant roles in explaining variations in the sales of shopping goods.

As would be expected, leakage factors contribute little to explain variation in per capita sales of the convenience items such as general merchandise, food, and drugs. The number of stores per capita is significant in every case and non-leakage factors, especially percent of families with incomes less than $3,000, rank high in terms of beta coefficients. The nature of the results for convenience goods suggest the need for a more individualistic approach in each case and more refined data concerning the type of good in each category.

CONCLUSION

The present analysis lends support to the research findings of both Russell and Ferber. Relative to the "closed system hypothesis" the authors have sought to identify some of the variables which explain why shoppers migrate from their places of residence into distant trading areas. These variables are designated leakage and non-leakage

factors. It was found, for example, that income distribution variables tend to play an important role in affecting leakage in the case of shopping goods.

The statistical analysis suggests that population and distance have strong effects on inter-city variations in total retail sales. Population, contrary to Reilly's Law, is found to be much more important than distance, but one measure of distance — driving time to the nearest SMSA — is more important when larger cities are considered separately, and when the proportion of families in upper income bracket is high. The number of stores per capita (a supply factor) assumes a role of almost equal importance with population in explaining intercity variations in total retail sales — its importance also tending to increase with the city size. The income variables have little effect on total sales in the statistical models of this study primarily because of strong intercorrelations with other independent variables, i.e., distance and stores per capita.

The analysis of per capita sales emphasizes the importance of the per capita stores variable for most categories and reveals that income level and distribution factors (all of which are highly correlated) are of considerable importance, particularly with respect to shopping goods. The socio-economic variables are of little importance in explaining sales variations for shopping good items such as furniture and apparel. The income distribution variables are shown to be primarily measures of the degree to which the system is "closed" and secondarily, measures of the level of effective demand in the city. The amount of variation explained differs considerably by merchandise category indicating that, in some cities, relatively high sales of one type of goods are accompanied by lower sales of another type. Thus in some cases a more individualistic approach is needed to study the determinants of intercity differences in per capita retail sales.

Many of the specific findings must be considered as tentative until they can be further tested using a larger sample (perhaps for a different region) and a more homogeneous group of cities. It does not seem reasonable to say that a theory of retail sales variation exists but each piece of research brings a little closer the achieving of an adequate explanation of this phenomenon.

FOOTNOTES

1. The "closed systems" hypothesis related to the idea that there is a positive correlation between income per capita (or per family) and retail sales per capita only when the areas under consideration are relatively closed systems — that is, where the inhabitants both earn their incomes and spend them within the area. "The Relationship Between Income and Retail Sales in Local Areas," *Journal of Marketing*, XXII (January, 1957), pp. 320-32.
2. Russell found the coefficient of correlation between retail sales per capita and median income per family in 78 communities (25,000 to 49,999 population) to be -.06.

3. "Variation in Retail Sales Between Cities," *Journal of Marketing,* XXII (January, 1958), pp. 295-303. His sample included all Illinois cities with a population exceeding 25,000 plus a random systematic sample of one-half of the cities between 10,000 and 25,000.
4. If the separate effect of each explanatory variable is to be considered, the independent variables must not be intercorrelated, or multicolinearity exists. Multicolinearity constitutes a serious threat both to the proper specification and effectual estimation of the type of structural relationship commonly sought through the use of regression techniques. For a technical explanation of this problem and its solution see Donald Farrar and R. R. Glauber's article "Multicolinearity in Regression Analysis: the Problem Remsited." *The Review of Economics and Statistic,* Vol. XLIX, No. 1, (Feb. 1967), pp. 92-107
5. The calculation of the beta coefficient, $\beta_i = b_i \frac{\sigma i}{\sigma y}$ Merely involves standardizing the units of net regression coefficients.
6. The authors recognize that although statistics offers methods for testing "significance" it offers no method for ascertaining the "importance" of a variable. To the extent the independent variables are intercorrelated, neither the test of significance (because the variance of the regression coefficients is affected) nor the beta (β) coefficient (because its calculation involves using the net regression coefficients) yields true measures of the "importance" of independent variables. Because there appear to be no feasible alternatives for this study, the independent variables will be ranked in accordance with the size of their beta coefficient.
7. Ferber also found population to be considerably more important than distance in explaining intercity variations in total retail sales. Ferber, *op. cit.,* p. 298.
8. See G. Schwartz's "Laws of Retail Gravitation: An Appraisal," *University of Washington Business Review,* Vol. XXII, No. 1 (October, 1962).
9. *Op. cit.,* pp. 300-301. Ferber does not report a partial correlation coefficient for each variable — only elasticities and beta coefficients. The beta coefficient for stores was .757 and for income it was .605.