Introduction

This paper describes a study recently conducted at the Research Triangle Institute for the U.S. Department of Transportation. The primary objective of this research was to identify spatial patterns of urban development strongly correlated with rates of gasoline consumption across existing U.S. cities. Both statistical and graphical analyses were conducted. Several new spatial measurement methods were developed by the author. All statistical analysis was done using SAS. The GMAP procedure of SAS/GRAPH was used (along with other mapping software) to display observed urban land development patterns.

Study Rationale

Why is it that people in Houston, Atlanta, and Grand Rapids use more gasoline than people in Buffalo, New Orleans, and Milwaukee? To explain such differences we must consider numerous factors such as local economic conditions, public transportation availability, climate, terrain, and urban spatial structure. By urban spatial structure we mean the physical layout of cities, i.e., how close people's homes are to their work places, schools, shopping centers, restaurants, and recreation areas. Since one of the goals of urban planning is to guide city growth along socially desirable courses, uncertainties surrounding long-term gasoline supplies make it important that planners understand relationships between urban spatial structure and gasoline consumption.

Since the 1973-74 fuel crisis, a number of studies have attempted to explore these relationships statistically. (For examples, see Watt, 1978; Stewart and Bennett, 1975; Keyes and Peterson, 1977.) However, all of these studies have encountered two sets of difficulties: (1) problems stemming from the lack of comparable data across cities describing non-residential spatial structure, and (2) problems related to the difficulty of measuring numerous important aspects of spatial structure in general.

Regarding data problems, some hope is possible for future studies as a result of the place-of-work data now being assembled in comparable format for many U.S. cities as a part of 1980 Census tabulations. These data will detail spatial distributions of daytime employment by industry and occupation codes for numerous urban areas as of 1980.

Regarding the effectiveness of available spatial structure analysis methods, the future is less clear. How would we go about analyzing differences in spatial structure across cities if data were available? How would we quantify across cities such things as land use mixture, home-place/work-place separation, racial residential segregation, urbanized area density, dispersion, polynucleation, and shape and a host of other things that come to mind when we compare mentally the spatial structures of a Boston and a Houston?

Convinced of the need for more powerful methods of urban spatial analysis, this study investigates a new method of spatial analysis that appears better equipped for investigation of relationships between urban form and gasoline consumption (Ray, 1977). The method defines urban spatial structure in terms of mathematical measures characterizing distance relationships between distributions.

More specifically, the method uses mathematical concepts from trip distribution modeling theory (e.g., gravity models, entropy maximizing models, linear programming (Wilson, 1970)) to formulate several new measures of spatial congruence and proximity between areal distributions. Despite the use of trip distribution concepts, however, these measures make no assumptions concerning traffic flows between distributions. In fact, all of the measures proposed are purely mathematical constructs and, thus, are independent of any assumptions concerning travel behavior. Because of this, they may be used to measure distance between black and white household distributions within urban areas as well as distance between population and employment distributions. Furthermore, the measures proposed appear superior to more conventional measures since they do not require that distributions be of certain types (bivariate, normal, circular, etc.). And, unlike conventional methods, they appear to be highly insensitive to the spatial resolution at which measurements are made (census tracts, blocks, zones, etc.).

Data

To evaluate the usefulness of these measures for study of relationships between urban form and gasoline consumption, we constructed a database using a number of Census Bureau data files. Data describing U.S. gasoline consumption patterns were available from 1972 Census of Retail Trade files. Since the only unit of geography common to these and other Census files was U.S. county, we aggregated county-level gasoline sales figures to the level of 1970 SMSAs and conducted all city comparisons at the SMSA level.

For 145 1970 Census SMSAs for which digitized census tract boundaries were avail-
able, we tabulated tract-level distributions of:

1. total population
2. white population
3. black population
4. blue-collar workers
5. white-collar workers.

Using these tract-level spatial distributions we computed for each SMSA indicators of population density and spatial dispersion as measured by distance variance (Neft, 1966). As a traditional measure of residential racial segregation, we also computed for each SMSA the index of dissimilarity between black and white population distributions (Duncan, Cuzzort and Duncan, 1961). In addition to these spatial structure indicators, we also had available for each SMSA a number of aggregate SMSA indicators (economic conditions, transit ridership, climate, through traffic, and tourism, etc.) for use as control variables.

For a sample of 27 SMSAs covered by the Annual Housing Surveys (AHS) of 1975 and 1976, a more complete description of non-residential spatial structure was possible. Using the data of the DOT-sponsored Travel-to-Work Supplement to the AHS, it was possible to derive tract-level spatial distributions of daytime employment by mode of travel. Thus, for this smaller sample of 27 AHS SMSAs, we were able to compute numerous measures of distance, congruence, and co-organization between population and employment distributions using our new spatial analysis measures.

Spatial Analysis Measures

Let X (nx1) and Z (nx1) be discrete probability distributions of some two urban variables over some area sampling frame for some SMSA. For example, X might be the distribution of total population by place of residence across census tracts and Z might be the tract-level distribution of total employment by place of work. In this case, x_i is the probability that a randomly selected person lives in tract i, and z_j is the probability that any worker works in tract j.

Let C (nxn) be a matrix whose elements c_{i,j} are squared (straight-line) distances between tracts i and j measured between tract centroids. Then, a measure of composite distance separating the two spatial distributions X and Z may be formulated:

\[ \sum_{i=1}^{n} \sum_{j=1}^{n} q_{i,j}^2 c_{i,j} \]

subject to:

\[ \sum_{i=1}^{n} q_{i,j} = x_i, \quad i,j=1,\ldots,n \]

Equations 1-4 will be recognized by many as a special case of linear programming known as the transportation problem (Dantzig, 1963). If X and Z are given in terms of absolute supply and demand quantities instead of probabilities, and C is defined in terms of network travel costs (distances, times), then the problem is to determine a set of flows from supply sites to demand sites that minimizes aggregate travel costs. Here, we use the transportation model to define an abstract measure of distance between spatial distributions. Elsewhere, we have called this measure pattern distance (Ray, 1982).

A general family of mean squared distances measuring composite separation between spatial distributions may be formulated as

\[ \text{MSD} = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{i,j}^2 \]

where the matrix Q (an n x n joint probability matrix) is selected in some manner from the set of all Q's satisfying 2-4 above. A convenient way to determine legitimate Q matrices of interesting properties is via trip distribution modeling methods (e.g., gravity models and entropy maximizing models) developed for transportation network simulations (Wilson, 1970). The square roots of MSD's computed via 5 with Q's satisfying 2-4 are called root-mean-square distances (RMSD's).

One MSD of special interest results if we determine Q in the form

\[ q_{i,j} = x_i z_j \]

Note that, in this case, Q is chosen to reflect an assumed statistical independence of the probability vectors X and Z. We call this MSD unconstrained mean squared distance (UMSD). Its root is called unconstrained RMS distance (URMSD).

When gravity model methods are used to determine the Q matrix, we call the resulting MSD measures constrained mean squared distances (CMSD's) and constrained RMS distances (CRMSD's). In our study, we used a CMSD associated with a Q matrix solving the set of equations:

\[ \sum_{j=1}^{n} q_{i,j}^2 c_{i,j} \]

subject to:

\[ \sum_{i=1}^{n} q_{i,j} = z_j, \quad j=1,\ldots,n \]

\[ \sum_{i=1}^{n} x_i = 1 \]

\[ \sum_{j=1}^{n} z_j = 1 \]

\[ q_{i,j} \geq 0 \]

When gravity model methods are used to determine the Q matrix, we call the resulting MSD measures constrained mean squared distances (CMSD's) and constrained RMS distances (CRMSD's). In our study, we used a CMSD associated with a Q matrix solving the set of equations:
For any \( C, X, \) and \( Z \), Equations 7-9 may be solved iteratively to determine unique \( Q, U, \) and \( V \) satisfying 2-4.

Finally, note that the \( D^2 \) defined above by 1-4 (pattern distance) is also a MSD. It may be called least mean squared distance (LMSD) since it represents the minimum MSD possible over all \( Q \) satisfying 2-4. The root of LMSD is then LRMSD or least root-mean-squared distance (LRHSD).

MSD measures may be standardized by dividing them by UMSS. In our study, we found useful a standardized CMSD of the form:

\[
(10) \quad \text{COORG} = (1 - \text{CMSD}/\text{UMSS}).
\]

COORG seemed to measure the extent to which two areal distributions were spatially co-organized, i.e., both distributions consisting of multiple modes or nuclei with the modes (nuclei) of the other. This measure of inter-distribution co-organization was particular useful for analysis of urban spatial structure in multimodal (poly-nucleated) SMSAs (e.g., the Allentown-Bethlehem-Easton, PA.-N.J. SMSA).

At this point it becomes clear that, given the variety of measures and spatial distributions available to our study, we had more spatial structure indicators than we had observations (SMSAs). Some of these measures correlated strongly with SMSA per capita gasoline consumption; others correlated only weakly. However, between spatial structure measures, for the most part, correlations were high. Thus, we factor analyzed spatial measure correlations across all SMSAs. This factor analysis revealed measurement of only four independent dimensions of spatial structure:

1. employment/population separation
2. auto employment/population co-organization
3. auto employment/transit employment co-organization
4. black/white population separation.

Factor scores for these four composite spatial structure indicators were used in subsequent statistical analyses of gasoline consumption across the 27 ABS SMSAs.

Stagewise Regression Analysis

To assess the value of these spatial structure indicators as predictors of SMSA gasoline consumption, a two-stage, multiple regression analysis was performed (Droper and Smith, 1964). In stage one, we used PROC GLM to regress SMSA per capita gasoline sales against all independent variables available for the large sample of 145 SMSAs. These independent variables included not only basic SMSA spatial characteristics (e.g., total population, density, transit ridership, per capita income, and unemployment) but also several spatial summary variables as well (e.g., population distribution distance variance, land area, density, and black/white population dissimilarity).

In stage two, we used PROC STEPWISE to regress stage-one residuals against the more detailed spatial structural indicators determined for our sample of 27 ABS SMSAs. Note that such a two-stage regression procedure allowed us to use the larger sample of 145 SMSAs to control for a large number of SMSA characteristics, and the smaller sample of 27 SMSAs to investigate the additional effects of the more detailed spatial structure indicators (i.e., employment/population separation, black/white population separation, etc.) available for the ABS SMSAs.

Stagewise Regression Results

The stage-one regression explained 58% of the variance of 1972 SMSA gasoline sales across the 145 SMSAs. Regression results indicated significant effects on 1972 SMSA per capita gasoline consumption for several spatial independent variables including SMSA unemployment (negative), percentage of workers commuting by transit (negative), and per capita receipts of eating and drinking establishments (positive). It was of greatest interest, however, to find that SMSA density and population dispersion (distance variance) were not significantly related to SMSA gasoline consumption.

Regressing stage-one residuals against the spatial structure indicators determined for the 27 ABS SMSAs, stage-two results indicated significant effects on SMSA gasoline consumption for two of the four composite spatial structure indicators. These significant spatial factors were: (1) population/employment separation and (2) black/white population separation.

The relationships between SMSA per capita gasoline consumption and SMSA population/employment separation was positive, as hypothesized. The stronger the congruence of daytime employment and nighttime population distributions, the smaller the rate of gasoline consumption. Note that this relationship proved statistically significant even after controlling for more than two-dozen independent variables included in the stage-one regression.

Thus, our study yields some evidence to support the hypothesis of a direct, causal relationship between urban spatial structure and urban gasoline consumption: the greater the distance of separation between urban population and employment distributions, the greater will be urban gasoline consumption. To lessen gasoline demands of cities in the future, planners might encourage urban growth and redevelopment policies that result in stronger spatial integration of residential and livelihood land uses.
RALEIGH SMSA
1970 POPULATION DENSITY BY CENSUS TRACT

LEGEND: POP70

Figure 1. SAS/GRAPH GMAP Plot of Raleigh SMSA 1970 Population by Census Tracts
(Source: 1970 Census)

RALEIGH SMSA
1976 EMPLOYMENT DENSITY BY CENSUS TRACT

LEGEND: EMPDEN76

Figure 2. SAS/GRAPH GMAP Plot of Raleigh SMSA 1976 Employment by Census Tracts
(Source: 1976 Annual Housing Survey)
The meaning of the significance of the black/white population separation factor is less clear. Should we infer from our study that SMSAs use gasoline in proportion to the extent that they are racially segregated? Do U.S. cities pay for racial segregation through additional expenditures for gasoline? Or is the effect we see in our analysis simply due to the fact that both variables (SMSA gasoline consumption and racial segregation) vary in accordance with some other unmeasured variable, for example, the degree of suburbanization experienced by U.S. cities in the decades just prior to 1972? We have little hope that such issues can be resolved within the scope of our present study which was restricted to a sample size of only 27 SMSAs. Armed with 1980 Census data, future analyses employing larger samples of SMSAs may be able to clarify some of these issues.
Graphical Spatial Analysis

To learn more about our measures, our data, and the substantive questions of interest, we turned to computer graphics.

Using the SAS/GRAPH GMAP procedure and other computer mapping systems developed by the author, numerous maps of several distributions of many SMSAs were generated (see Figures 1-4). These maps were useful in three different ways. First, they allowed a convenient check on the overall validity of data defining spatial distributions. In general, the maps confirmed that our data were reasonable and free of major errors. Second, they allowed us to confirm that our spatial measures were quantifying specific aspects of spatial structure of interest. Our measures seemed to be behaving as planned. Third, they provided us with good illustrations to be used for study presentations and reports.

Of course, our computer-generated maps could not increase the statistical significance of any relationship between variables measured by our study. Nor could they be used as quantitative evidence for any causal relationship hypothesized between variables. Nevertheless, they served a valuable purpose by reinforcing statistical results with qualitative, visual presentations easily grasped by research audiences. With such maps it is much easier to convince ourselves and others that our spatial measures have merit and that our study results are meaningful. Given the complexity of real-world urban spatial structure (and the problem of characterizing this structure quantitatively), this by itself is no mean accomplishment.

Summary

We have explored the usefulness of several new measures of distance between spatial distributions as instruments for quantifying differences in spatial structure across U.S. cities. Using SAS regression procedures, we have related spatial structure differences measured to per capita rates of gasoline consumption across a sample of cities. We found that, according to the distance measures used, gasoline consumption increases with distance between nighttime population and daytime employment distributions. For some reason, urban gasoline consumption appears also significantly related to distance between black and white population distributions.

Since we used spatial structure measures that were new and untested, it was important to verify that all measures behaved as intended, i.e., that they quantified specific aspects of urban spatial structure of interest. Computer mapping techniques became an indispensable aid to our research for these purposes. The GMAP procedure of SAS/GRAPH could be used to display spatial distributions illustrating certain features measured by our indicators. It is difficult to conceive of such research being conducted without some computer mapping capability.

Author's Address

Questions concerning this research should be addressed to the author at:

Research Triangle Institute
Hill Building, Room 310
Research Triangle Park, N.C. 27709
Telephone: (919) 541-6339

Acknowledgement

This work was supported at the Research Triangle Institute by the U.S. Department of Transportation through Contract DTRS-57-80-C-0037. The views expressed in this report are those of the author and do not necessarily represent the views of the Research Triangle Institute or the U.S. Department of Transportation.

References