A Test of Some First Generation Residential Land Use Models

CARL N. SWERDLOFF and JOSEPH R. STOWERS, Highway Engineers, U.S. Bureau of Public Roads

•THIS paper reports on a comparative evaluation of five operational residential land use forecasting techniques, four of which have been previously used in urban transportation planning studies. These techniques are representative of the earliest of efforts in the development of operational urban activity simulation models and continue to serve, either in their original or in modified form, a great number of transportation planning organizations. Urban activity simulation models currently under development, while in most cases considerably more complex and, hopefully, more accurate, in many instances draw upon notions and fundamental concepts which either originated with or were adapted to these early techniques. Improvements being introduced in these later, second generation models include more complex statistical estimating procedures, the stratification of residential locators into several distinct groups, and the incorporation of behavioral relationships in the model formulation. These newer techniques may require several years of research, evaluation, and refinement before they become fully operational. Meanwhile, the less sophisticated approaches evaluated in this report should continue to be useful to smaller metropolitan areas lacking the resources for developmental research.

The primary objective of this project was to compare the relative accuracy of these approaches through a series of ex post facto tests, holding all conditions constant except the interrelationships among variables, so that differences in "forecasts" would be a function only of inherent differences in models.

There is a temptation to interpret a study of this nature as a contest of sorts and to turn to a table of results for the proclaimed "winner." Any such evaluation of the results is unwarranted for several reasons. First, the contestants are not all of the same class. Some are more truly "forecasts," and some are merely data fitting problems. The latter involve fitting different numbers of parameters. More information is used in some than in others. Perhaps most important, the results represent a sample of one, out of a rather large universe of possible test conditions. Entirely different results might occur in other cities, at other time periods, by other forecasters, working with other data problems.

GENERAL PROCEDURES

The five residential land use forecasting procedures are each variants of work done by others. The only innovations introduced here are the authors' simplifications and modifications to suit peculiar test conditions—apologies are made to the progenitors of these models for possible misrepresentations of their original work. In any realistic planning application, more care would necessarily be given to the particular forecasting tool used. Trends would be more carefully analyzed, the forecasters would be more familiar with the area, and output of models would be scrutinized in detail and modified as judgment indicated. In contrast, the authors have applied the models coldly and crudely, accepting the immediate output in an attempt to make objective comparisons.

Paper sponsored by Committee on Land Use Evaluation and presented at the 45th Annual Meeting.

The techniques used were (a) the density-saturation gradient method, (b) accessibility model, (c) regression, and (d and e) two intervening opportunity models.

The density-saturation gradient method (DSGM) is a simplification of the approach used by the Chicago Area Transportation Study (1, 2). Of the five techniques, the DSGM is least computer oriented, more demanding of subjective inputs, and therefore least suitable for objective comparison with other approaches, particularly when the forecasters are not intimately familiar with the area. The method is based essentially on the regularity of the decline in density and percent saturation with distance from the CBD, and the stability of these relationships through time.

The simple accessibility model is based upon the concept formulated by Walter Hansen (3, 4). Growth in a particular area is hypothesized to be related to two factors: the accessibility of the area to some regional activity distribution, and the amount of land available in the area for development. The accessibility of an area is an index representing the closeness of the area to all other activity in the region. All areas compete for the aggregate growth and share in proportion to their comparative accessibility positions weighted by their capability to accommodate development as measured by vacant, usable land.

The third method used in this study, multiple linear regression, is a popular approach because of its operational simplicity and ability to handle several variables (5, 6, 7). The proportion of total regional growth which locates in a particular area is assumed to be related to the magnitude of a number of variables which in some manner are measures of geographic desirability as viewed by those making the locational decision. The procedure is to determine those factors, and their weights, which in linear combination can be related to the amount of growth which has been observed to take place over a past time period. These factors (called independent variables) and their weights (regression coefficients), in linear combination (the regression equation) can then be applied to the individual analysis areas to forecast the magnitude of growth (the dependent variable).

Although more commonly applied to the problem of trip distribution, the intervening opportunities models can be used in simulating the distribution of urban activity. Two separate and distinct formulations were applied in this study, both based upon the general notion that the probability that an opportunity is accepted decreases as some function of the number of opportunities ranked closer to a central distributing point. The Stouffer formulation was originally applied to intra-urban migration (8). A related formulation has more recently been investigated as a trip distribution technique (9). Schneider's formulation was originally applied to trip distribution (10) and is currently being used in distributing urban activity (11, 12).

The test area used in this study was Greensboro, North Carolina. This city was chosen for a number of reasons. First and most important, a rather extensive information file on a small area basis for two time periods (1948 and 1960) was available. Secondly, it was felt that Greensboro was representative of the kind and size city for which forecasting techniques of the kind being examined would still be most appropriate after the development of more sophisticated models in the largest metropolitan areas.

The data for the study came from two major sources. The data obtained from the University of North Carolina contained a wide variety of information for the Greensboro area coded to 3,980 grid cells, each one 1000 ft square, for a circular area of about 7-mi radius. These data included quantitative measures of land use, population, residential density, proximity to various activities and to the CBD, and certain environmental measures (13). With certain exceptions, these data were available at the grid level for two time periods, 1948 and 1960.

The data supplied by Alan M. Voorhees and Associates included 1960 population, employment, accessibility to shopping, and accessibility to employment, for each of about 250 zones. These latter accessibility measures were computed from zone-to-zone traveltimes over the highway network.

A number of problems were encountered in combining the data from these two sources in a form suitable for testing of the models. Principal among these were the following.

1. The aggregation of grids to zones. Since it was felt desirable to work at a level of aggregation more typical of transportation studies, it was necessary to define new zone boundaries following grid lines approximating the irregular old zone boundaries. No important error was introduced since only accessibility scores from the original zone file were used in subsequent analyses—all extensive quantities used were grid aggregates.

2. Estimation of 1948 dwelling units. Consideration of all data sources and the purpose of the study led to the decision to use dwelling units as the item to be predicted. However, 1948 dwelling unit data were not directly available. Estimates were made and various checks applied by using 1948 land area, a 1948 USGS map for suburban areas, 1950 census block statistics for the central city (changes were not large for the inner area from 1948-1960), and the 1960 land area and dwelling unity densities.

3. Estimation of accessibility measures for 1960 for certain zones at the fringe. The area covered by the zone file did not extend to the boundaries of the grid coverage area in all directions. Rather than eliminate this area entirely, estimates of accessibility measures were made for about one-half of the outer ring of zones by examining contours of iso-accessibility lines, which follow fairly regular patterns in the fringe area.

MODEL DESCRIPTION AND METHODOLOGY

Density-Saturation Gradient Method

The DSGM is the least formally structured forecasting procedure of the five. No formal theoretical statements or mathematical hypotheses are required, although the staff of the Chicago Area Transportation Study have presented excellent conceptual explanations of their empirical findings and rationale for their projections (1). This theoretical development, however, is not essential to the purpose of this paper.

Before discussion of the actual application of the DSGM to the Greensboro area, mention should be made of certain reservations which existed prior to the testing. The only known previous application of this approach was for the Chicago area. There was some initial fear that the regularities in activity distribution about the central place, which is axiomatic to the method, would not be manifest for a city of the size of Greensboro. The declines in density and percent capacity result from the operation of the competitive land market, a mechanism which might not exert the dominating influence upon spatial organization in a city of Greensboro's size. It will be seen that these fears were unwarranted, and that in fact the distribution of residential activity was markedly structured about the CBD.

Two semi-independent forecasts were made using the DSGM in order to determine the sensitivity of the results to variations in the critical assumptions made. A principal distinction was that the first trial was made using air-line distance from the high value corner (HVC) as the key spatial variable, whereas traveltime to the HVC was used in the second trial. (The HVC is a point representative of the hypothetical activity center of the CBD).

Figure 1 shows the relationship between 1948 dwelling unit density and air-line distance from the HVC. Each point on this plot represents the gross residential density (street area included) for a ring around the HVC. Each ring is defined by the boundaries of all zones whose centroids fall within $\pm 1/2$ mile of the nominal distance of the ring from the HVC with the exception of the first or CBD ring. The plot indicates a surprisingly regular decline in residential densities with distance from downtown in Greensboro in 1948. This was encouraging since the reliability of the DSGM depends greatly on the strength and stability of this relationship.

The method depends equally upon the relationship between distance and percent saturation. To compute the latter, residential capacity must be defined. Mathematically capacity is defined as existing dwelling units plus the product of vacant available, suitable land, and expected residential density. A decision had to be made at this juncture as to the density values to be used in the computation. Theoretically this should be the anticipated average density at which all future residential development will occur.

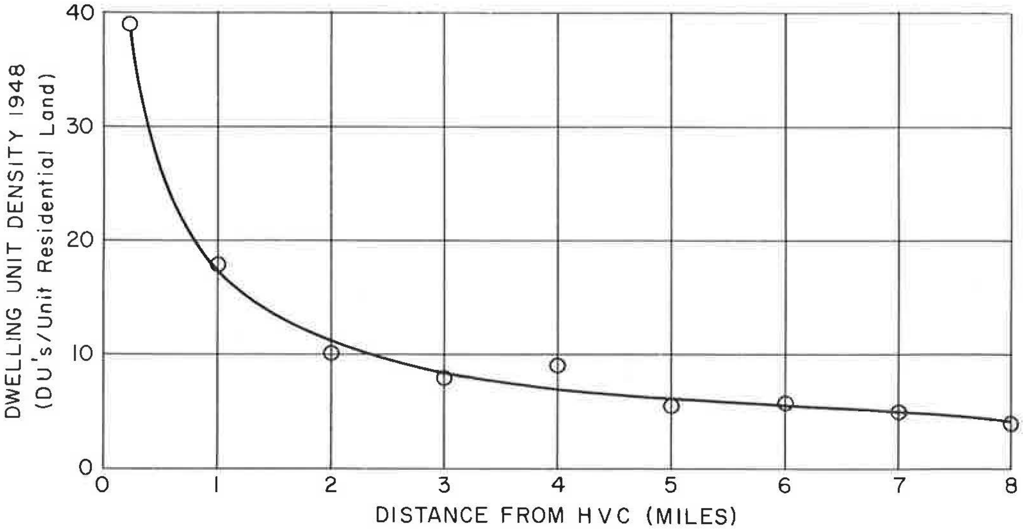

Figure 1. Dwelling unit density by distance bands—1948.

These values should be developed from an intensive analysis of trends in residential density patterns and zoning policies. For purposes of this study, however, future densities for each zone were assumed to be those given by the smooth hand-fitted curve of Figure 1. Prior to the acceptance of this single curve for the density gradient, gradients were plotted for each of five sectors. Although these plots exhibited less regular relationships, no significant variation between sectors was noted.

Vacant, suitable land for residential development was estimated by subtracting marginal land and land zoned for nonresidential uses from 1948 nonurban land. A systematic, but subjective procedure was used in the treatment of zoning: land was weighted by factors ranging from 0 for grids zoned only for industry to 1.0 for grids zoned only for residential use; land in grids zoned for mixed uses and other nonindustrial uses was weighted subjectively on a scale from zero to unity.

Having future residential development densities and vacant available land, it was possible next to compute both the residential saturations, in dwelling units and existing percent saturation, for each distance ring from the HVC. The latter values, resulting from the division of saturation into 1948 dwelling units, were then used to construct the percent saturation gradient. Figure 2 conforms very well with the plot expected for an urban area. The rather distinct and sharp transition between the 3½- and 4½-mi points indicates a transition from the area of urban character into the predominantly rural portions of the study region. The almost negligible slope of the curve beyond the 4½-mi point is indicative of agricultural development and the absence of any strong competition for location with reference to central Greensboro.

The next step involved the 1960 projection of the percent saturation curve, also shown in Figure 2. (Percent saturation gradients by sector for 1948 were also plotted; however, as in the case of the density gradient, there was some additional scatteration of points, but no basis for using sector-specific gradients.) This is the most critical and subjective step in the forecasting process, the only restraint on the projected curve being that the area under the new curve must account for the projected regional growth. The number of dwelling units in the study area grew from a 1948 total of 27,191 to 41,250 in 1960 or a growth of 52 percent. One can proceed in almost an infinite number of ways insofar as establishing an acceptable projection of the percent saturation gradient. It was, however, found useful to first develop a feeling for the overall scale of the problem, that is, the area under the final curve which would be commensurate with the required final regional population. As a first approximation to the 1960 gradient each ordinate value was raised a distance equivalent to 52 percent of the 1948 value.

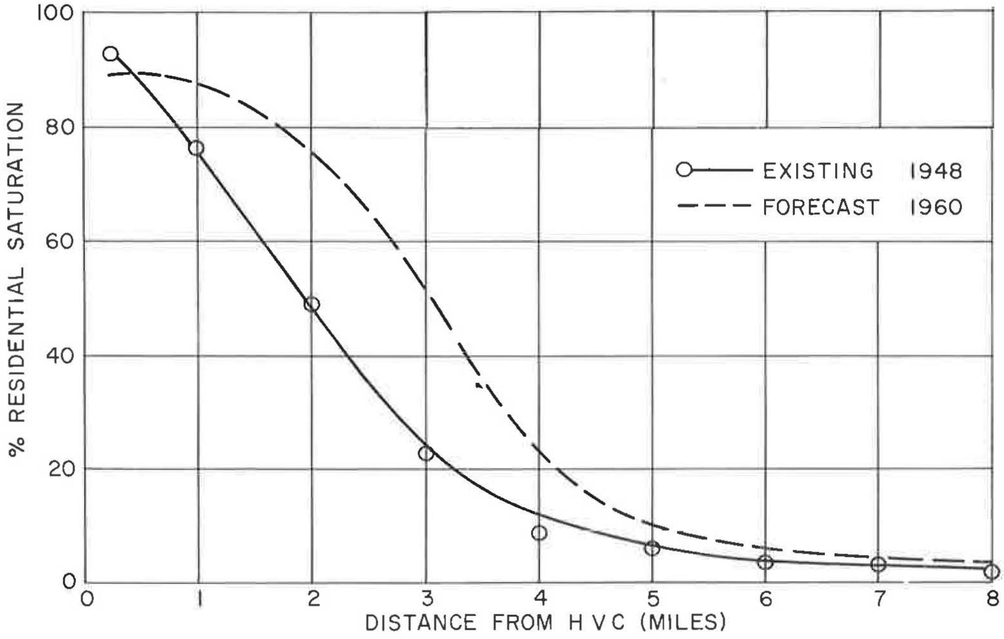

Figure 2. Residential saturation by distance bands.

The resultant curve then approximated the forecast condition under the assumption of uniform growth over the entire region. The following general criteria were then introduced to modify the naive first approximation of the shape of the gradient in 1960:

1. The bulk of the residential growth would occur in the 2-, 3-, and 4-mi rings.
2. The inner ring would suffer a slight decline.
3. The shape of the gradient would tend to bow out in the 1- to 3-mi range.
4. The sharp transition in slope of the 1948 saturation gradient observed at about the 4- to 5-mi point would become less abrupt in 1960.
5. The areas 5 miles and beyond would show some exurban growth, but the general flat slope would remain.

Relatively few attempts were necessary to arrive at a solution which was of satisfactory shape and which conformed with the actual 1948-1960 increase in total dwelling units.

Multiplying the appropriate ordinate value from the forecast percent saturation gradient (Fig. 2) by the ring saturation quantities established the forecast dwelling unit totals by analysis ring.

The projected growth of each ring was distributed to zones in a two-step process following the logic of CATS. The allocation to districts (defined by ring-sector boundaries) was handicapped by a lack of historical data. Ideally the trends in land use composition and growth rates between sectors should be studied in detail. For trial one, however, the simple assumption was made that sectors would share growth in proportion to available residential capacity.

The final distribution to zones was based on a systematic, but subjective linear weighting of the following factors:

1. Distance to convenience shopping,
2. Available residential capacity,
3. Distance to the major street system,
4. Percent of industrial development in the zone, and
5. Percent of residential development in the zone.

Trial two, which was conducted independently of trial one, differed from the above procedure in two principal ways:

1. Traveltime to the HVC was substituted for airline distance as the major independent variable. Zones were aggregated into 1-min interval rings for all analyses.
2. Ring growth was allocated to sectors (i.e., the district-level forecast) in proportion to the product of each sector's available residential capacity and the number of existing (1948) dwelling units.

Otherwise, the process followed that of trial one, including the method of estimating density and holding capacity, the sector definitions, and the allocation of growth from districts to zones.

Figure 3 shows the dwelling unit density gradient as determined from the ring analysis for trial two. As expected the same general shape is observed as for trial one. Figure 4 shows both the percent saturation curve calculated for the 1948 base period, and the forecast of the 1960 percent saturation curve. The shape of the latter gradient is quite similar to that for trial one except for a slight decrease in the growth allocated to the inner rings, resulting in a lessening of the bowing effect and a reduction in the slope of the gradient in the intermediate areas.

Accessibility Model

The generalized form of the accessibility model is as follows:

$$G_i = G_t \frac{A_i^a V_i}{\sum_i A_i^a V_i}$$

where

G_i = the forecast growth for zone i;
G_t = total regional growth $= \sum_i G_i$;

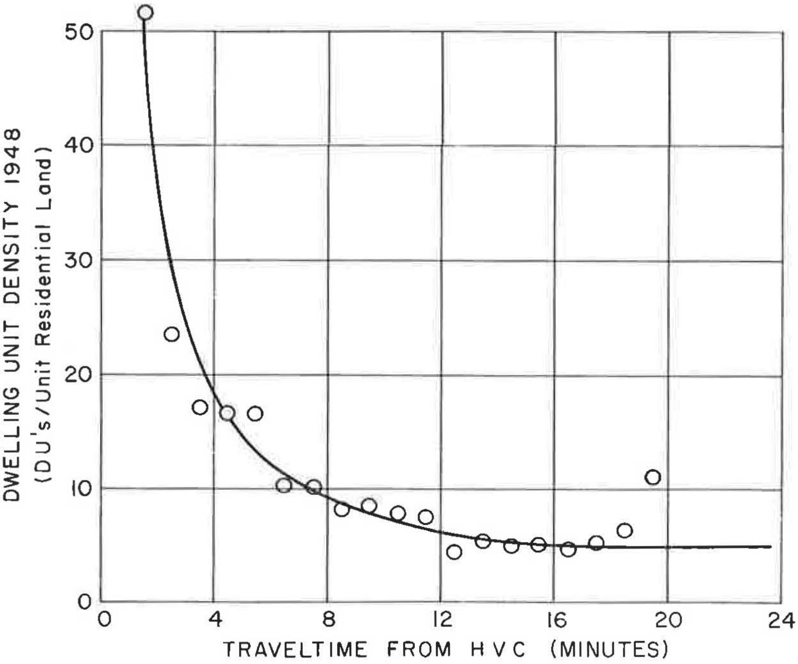

Figure 3. Dwelling unit density by time bands—1948.

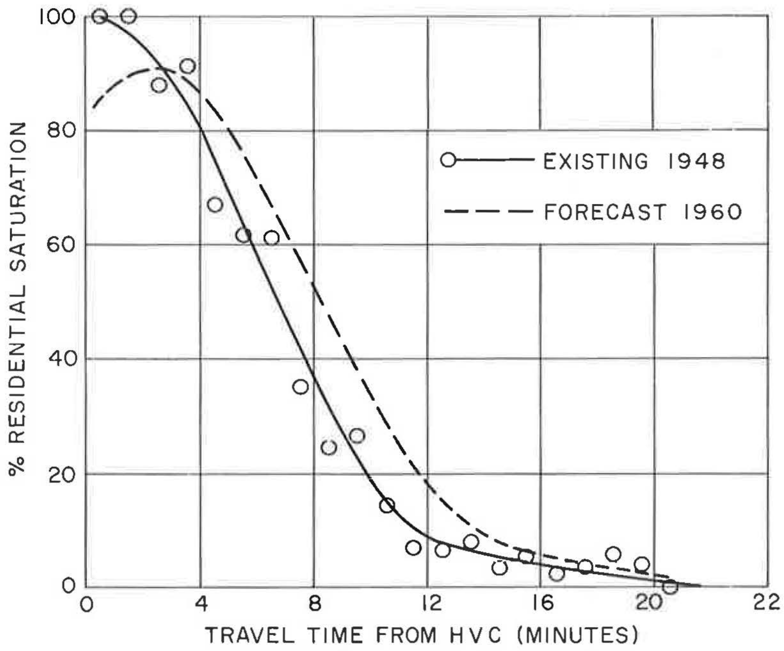

Figure 4. Residential saturation by time bands.

A_i = accessibility index for zone i;
V_i = vacant available land in zone i; and
a = empirically determined exponent.

The computation of the accessibility index traditionally is as follows:

$$A_i = \sum_j \frac{E_j}{T_{ij}^b}$$

where

E_j = a measure of activity in zone j (total employment used in this study);
T_{ij} = traveltime from zone i to zone j; and
b = an empirically determined exponent.

However, "friction factors" developed in the gravity model calibration by Alan M. Voorhees and Associates were actually used in the computation of accessibility:

$$A_i = \sum_j E_j F_{ij}$$

where F_{ij} is the friction of time separation of zones T_{ij} minutes apart. The F_{ij} values are approximately proportional to the actual number of trips T_{ij} minutes long per trip-end in each pair of zones T_{ij} minutes apart. In practice the computation of F_{ij} is considerably complicated by a desire to have the F_{ij} values form a smooth monotonic relation to T_{ij} yet maintain approximate equality between the resulting mean trip length and the actual mean trip length.

With the above definition of the model only one parameter, a, need be estimated to make the forecast. Two options were open:

1. Make a judgment of the value of a from previous work in other cities, and forecast 1960 zonal growth to have an independent test of the model; or
2. Fit a "best" value for a using the actual 1948-1960 changes in dwelling units.

Both options were actually used. For option 1 a value of 2 was assumed for a. (Hansen found that a value of about 2.7 was optimal for Washington, D.C.; the presumption that accessibility would have less influence in shaping growth in a smaller city is substantiated by the subsequent results in fitting values for a.) Methods used in fitting a to the 1948-1960 data are described in the Appendix.

Regression

For several reasons it was felt desirable to express the dependent variable of the multiple regression formulation as some function of the 1948-1960 growth rather than as some function of the absolute amount of cumulative development at a single point in time. The latter option was open, and has been used by others (13, 14); however, it was rejected to maintain comparability with the dependent variables of the other models, as well as to conform to standard practice in transportation planning models. As has been pointed out by the Traffic Research Corporation (15), there is good reason to expect greater accuracy for relatively short-range forecasts when predicting increments of growth.

Using change in dwelling units, or some function thereof, as the dependent variable, it was not possible with the available data to produce an independent forecast to check against the 1960 data. The equation parameters had to be estimated from the full 1948-1960 data files. Hence, accuracy results are shown in the next section only for a fitted model, and not for a forecast, in contrast to the other 4 methods. Dwelling unit data for a third point in time would be required to examine the forecasting reliability of the calibrated regression equation.

The usual regression approach differs from the other models used in this study in two additional important ways:

1. Many, rather than one or two independent variables may be incorporated, and
2. Variables are related to growth only in linearly weighted combinations, although variables may be transformed prior to regression.

The latter restraint is imposed by the use of a standard regression program (the BIMD 34 stepwise multiple regression program developed by the UCLA Bio Medical Center for the IBM 7090/7094 was used in this work). Of course nonlinear regression equations may be developed, but different normal equations must be solved and standard regression programs may not be used.

Numerous equations were developed, each involving the testing of various hypotheses regarding the functional relationships between variables. A total of 44 independent variables plus certain selected nonlinear transformations were examined in all, including:

1. Measures of zone size and amount of land in different uses;
2. Accessibility to employment;
3. Time and distance to HVC;
4. Zonal employment, total and by major type;
5. Densities for 1948;
6. Vacant available land;
7. Zoning protection;
8. Land value; and
9. Proportions of total land and developed land in each major use.

Four definitions of the dependent variable were tested:

1. Increase in dwelling units (DU);
2. Log DU;
3. DU per unit of available land (DU/L); and
4. Log [DU/L].

The logarithmic transformations were employed to test certain hypotheses regarding exponential relationships, as for example, are expressed in the accessibility model. The growth-per-unit-of-available-land transformations were employed in an attempt to remove all measures of zone size from the equations, and thereby, to avoid the possibility of distorted relationships due to the pecularities of area definitions.

The final equation accepted after comparing the accuracy and reasonableness of all trials was

$$Y = -2.3 + 0.061\,X_1 + 0.00066\,X_2 + 1.1\,X_3 - 0.11\,X_4 - 0.0073\,X_5$$

where

Y = logarithm of growth in dwelling units 1948-1960 per unit vacant land;

X_1 = zoning protection, 1948;
X_2 = percent of total land area in residential use, 1948;
X_3 = logarithm of accessibility to employment, 1960;
X_4 = dwelling unit density, 1948; and
X_5 = percent of total use land in industrial use, 1948.

The coefficient of correlation is 0.61. Table 1 contains the t and beta (β) values (standardized regression coefficient) for each of the independent variables in the equation. All regression coefficients are significantly different from zero with 95 percent confidence. Having the greatest β value, the transformed accessibility variable is shown to exhibit the most influence upon the estimate of the dependent variable. Percent of urban land which is in industrial use has the lowest β values and, therefore, contributes least to the total equation estimate.

The zoning code was a value from 0 to 9, where a higher value indicated zoning control closer to single family residential only, and lower value marginal-to-no zoning control. The positive relationship then indicates the positive environmental influence of strict residential zoning policy. The positive contribution of accessibility to work areas is self-explanatory. Also, the positive contribution of percent of total area devoted to residential development is interpreted as a measure of residential clustering. The tendency for slow growth or even decline in the residential stock of the close in, old city areas, coupled with the rapid increase in the fringe and newly settled locations accounts for the negative coefficient for dwelling unit density. The negative contribution of percent industrial land is indicative of the restraint on new residential development in areas immediately adjacent to industrial areas.

Because the estimation was couched in both logarithmic and intensity units, several operational difficulties were introduced. The estimating equation was incapable of either accepting negative values for the dependent variable or estimating decline in any zone. All zones which suffered dwelling unit decline over the calibration period were approximated to have shown no change. An additional problem was encountered for several zones which experienced dwelling unit growth, but which had no vacant land available in 1948. Without some adjustment the growth intensity value becomes infinite. These few cases were handled by substituting large arbitrary values of growth intensity. Finally, there is no built-in provision, as there is for other models, to assure that the accumulated zonal estimates obtained from the regression equation solution will equal the actual total regional growth. All regression estimates had to be factored up to sum to the actual regional growth.

TABLE 1

RELATIVE SIGNIFICANCE AND EXPLANATORY POWER OF VARIABLES IN REGRESSION EQUATION

Independent Variable	t	β
Log accessibility to employment, 1960	4.30	0.321
Zoning code, 1948	2.89	0.213
Percent of total land residential, 1948	2.70	0.187
Dwelling unit density, 1948	3.28	0.177
Percent of urban land industrial, 1948	2.98	0.159

Two Intervening Opportunity Models

Although the two opportunity models tested are based on quite different initial assumptions and take on dissimilar mathematical form, nevertheless, both can be reduced to a simple general hypothesis. In the context of this problem, the probability that a suitable residential opportunity (a unit of available capacity) is accepted for development is hypothesized to be a monotonically decreasing function of the number of intervening opportunities, opportunities being ranked by time from the HVC.

Some improvement in these models could undoubtedly be made by allocating increments of growth from more than one point, perhaps from all major centers of employment in proportion to the amount of employment in each center. This would make the test of the intervening opportunities models more comparable to the accessibility model procedure.

Stouffer Formulation. The Stouffer model may be defined in the following manner:

$$g_p = \frac{k\,O_p}{O}$$

where

g_p = number of dwelling units forecast to be located in a particular area p;
O_p = opportunities in interval p;
O = total number of opportunities from central distribution point through interval p; and
k = constant of proportionality to assure that the total number of dwellings located equals the actual total growth.

As stated, the Stouffer formulation can be applied without the need for assuming any parameter values. However, it is an operational requirement that the study area be structured into a number of discrete geographic units which are then ranked from a central distribution point, the HVC in this case. One method of aggregating areas, which Strodtbeck has shown to have some appealing properties, is to delineate a small number of rings containing approximately equal numbers of opportunities (16). For the initial application of the Stouffer model to the allocation of residential growth, the Greensboro study area was divided into 10 rings, each of which was composed of a whole number of zones and an approximately equal number of opportunities. Zones were assigned to rings according to their ranking in time from the HVC.

It was then possible to determine g_p, the forecast number of dwellings in ring p by direct substitution in the formula. The ring forecasts were then proportioned among the constituent zones on the basis of opportunities.

For an explanation of the fitting of the Stouffer equation to 1948-1960 data the equation must be converted into its continuous differential form as follows:

$$d(G_p) = \frac{kd(O)}{O}$$

By integrating

$$G_p = k \ln O + C$$

where

G_p = the total number of dwellings allocated to all opportunities from the central point up to and including opportunity interval p;
$d(G_p)$ = dwellings allocated to opportunity interval p;
$d(O)$ = opportunities in interval p; and
C = constant of integration.

This equation plots as a straight line of slope k where the ordinate, total allocated dwellings, is in linear form and the abscissa, total accumulated opportunities, is a logarithmic scale. As a test of the appropriateness of the Stouffer formulation in describing the spatial distribution of residential growth in Greensboro, the actual accumulated zonal dwelling unit growth 1948-1960 was plotted against accumulated 1948 opportunities, the zones being ranked by traveltime to the HVC. If the Stouffer model is valid the resulting plot should follow a straight line. It was immediately obvious that a single straight line could not be adequately fitted to the points, but rather that two distinct straight lines were necessary (Fig. 5). After hand fitting the two lines, 1960 growth estimates were made to the individual zones from the straight lines and the error computed. These results and those computed from the initial, noncalibrated test of the Stouffer formula are discussed later with the results of the other four models.

Schneider Formulation. As applied to the distribution of residential activity, the Schneider model takes the following form:

$$d(G_p) = g_t \left[e^{-\ell O} - e^{-\ell (O + O_p)} \right]$$

where

G_p = total number of locations in opportunity interval from the central point up to interval p.
g_t = total growth to be allocated;
ℓ = model parameter expressing the probability of an opportunity being accepted for location;
O = total number of opportunities ranked from the central point up to interval p.

As a necessary condition for applying the model the parameter ℓ must be stipulated. For the first trial of the model for a 1960 forecast without benefit of the 1948-1960 data, ℓ was estimated from the assumption that the actual dwelling unit increase within the study boundaries was 99 percent of the aggregate Greensboro oriented growth. (The theoretical model is based on a distribution to an unbounded area; applicaton to a finite area requires specification of the number of accepted opportunities being outside the

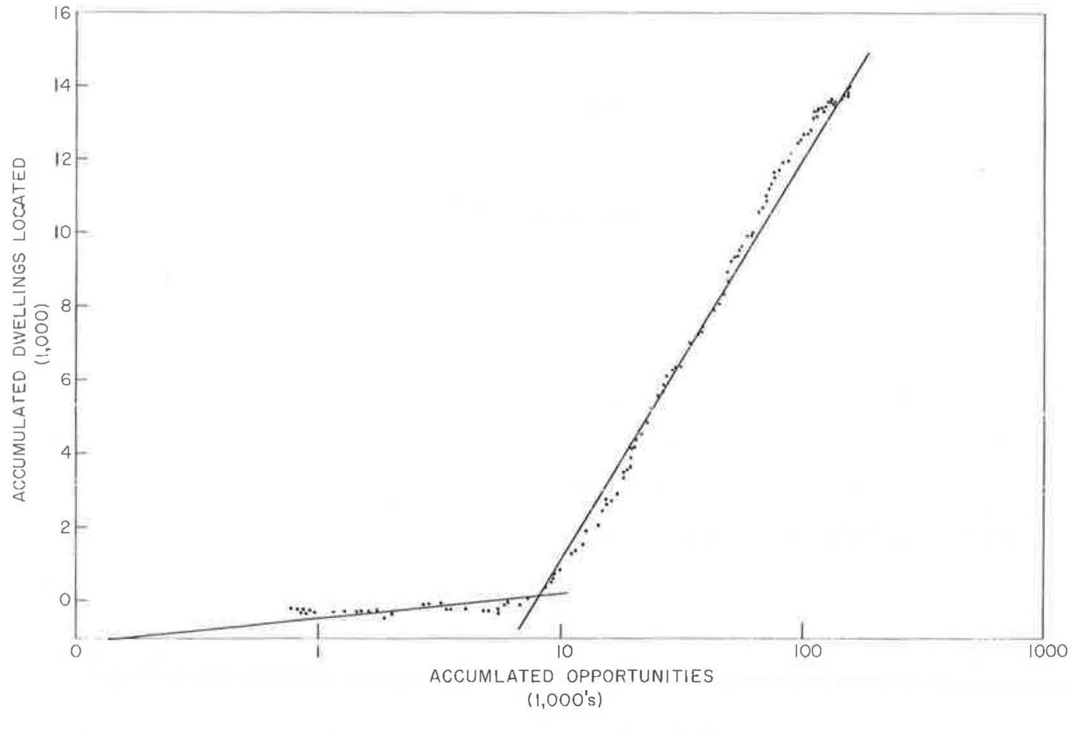

Figure 5. Test of Stouffer's formulation.

boundary, or equivalently, the percentage accepted up to the boundary.) The ℓ resulting from this assumption was 12.76×10^{-6}.

For an explanation of the fitting of the Schneider formulation to 1948-1960 data, the formula can be restated after integration as

$$G_p = g_t \left[1 - e^{-\ell O} \right]$$

Subtracting g_t from both sides and rearranging,

$$g_t - G_p = g_t e^{-\ell O}$$

or

$$\ln (g_t - G_p) = \ln g_t - \ell O$$

This relationship plots as a straight line where the ordinate, $(g_t - G_p)$, is in logarithmic scale and the abscissca, total accumulated opportunities from the central point (O), is in linear scale. The slope is ℓ and the intercept g_t.

If the Schneider formulation effectively replicates the spatial distribution of residential growth in Greensboro then plotting the actual quantity $(g_t - G_p)$ versus accumulated opportunities (O), in semilogarithmic forms, should yield a straight line (Fig. 6).

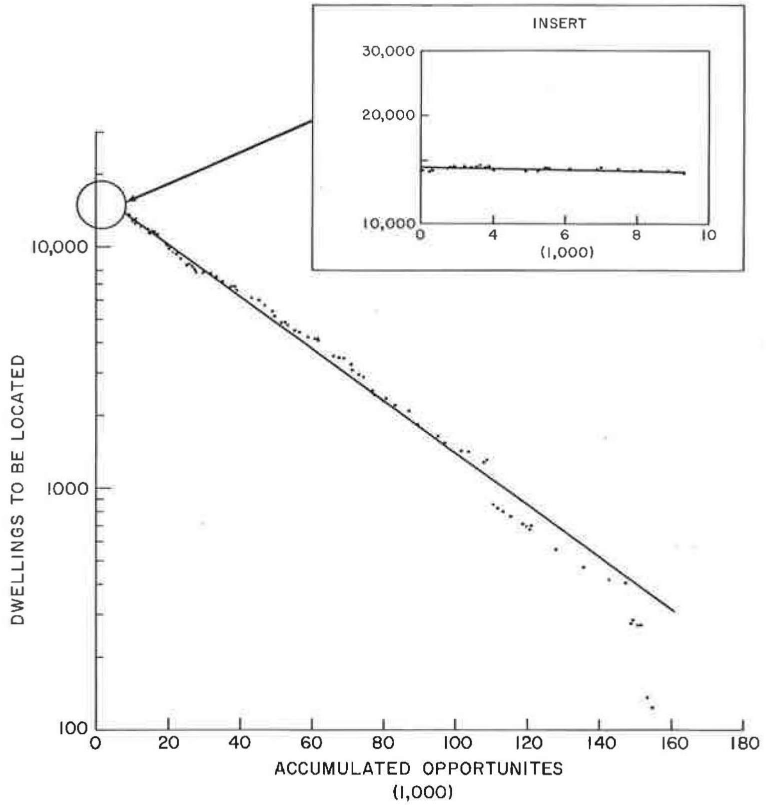

Figure 6. Test of Schneider's formulation.

As with the Stouffer formulation, the Greensboro data appear to exhibit two distinct straight line segments, rather than one, as required by the initial model formulation. The zones comprising the transition area between the two straight line segments (Fig. 6) are the same ones as those at the juncture of the two line segments for the Stouffer formulation (Fig. 5). The slopes of the fitted lines can be loosely compared to the short and long trip ℓ's which have become standard practice in applying the Schneider formula as a trip distribution model. The slope for the central city line segment is 1.707×10^{-6}, and that for the outer, suburban area is 10.9×10^{-6}.

The distribution of residential growth in Greensboro from 1948 to 1960 did not adequately conform to either of the intervening opportunities formulations over the complete range of opportunities. It is noteworthy, however, that the data plot as two straight lines in both Figures 5 and 6. It was also pointed out that the transition points in the vicinity of the intersection of the fitted straight lines in both figures were the same data points representing the same zones. Although a detailed examination of these zones has not been attempted it does appear that they approximate a transition ring in Greensboro which separates the "inner city," marginal growth area from the suburban, rapid expansion area. This band encircles the HVC at a radius of 1½ to 2 miles. For a city the size of Greensboro, which in 1948, exhibited a leveling off in the percent saturation gradient at 3½ to 4½ miles from the HVC, the area circumscribed by this transition band probably was characteristic of similar areas in most cities—old and perhaps showing signs of blight with little available residential capacity.

The inner area straight line slopes drawn to the two plots are both very close to the horizontal. In contrast, there are quite steep slopes for the plots representing suburban areas. A hypothetical locator viewing the opportunity surface from the HVC in accordance with either of the two plots apparently assesses himself a greater penalty in passing up suburban opportunities as opposed to inner-city ones. That is, the inner-city opportunities are a less desirable subset of the total as evidenced by the significantly lower slope on the plots, hence a lower probability of accepting individual opportunities. One may conjecture that location choices from the inner-city opportunity subset are responsive more to the individual living qualities of the opportunities other than its accessibility, which may be extended to the notion that the inner-city opportunities are viewed more or less as of homogeneous access in opposition to the suburban subset where opportunity access is of greater import in the locational choice.

Of interest from a purely forecasting viewpoint is the question of the stability of the handfitted lines in Figures 5 and 6. Do the slopes remain more or less constant over time and how does the transition area behave in relation to the total opportunity surface? One may speculate, for example, that the straight line relationships fitted to the data will hold over time and that the diffusion in residential location observed in the past is merely a reflection of the diffusion in the opportunity surface; that is, a physical dispersion outwards occasioned by the filling in of less distant areas, rather than of an alteration in the location function. On the other hand, it is possible that over time the slopes of the plots may be flattening out which is symptomatic of a society less restrained by the impedance of travel. Clearly, answers to speculations of this nature are required before one can estimate the applicability of the fitted lines to forecasting to a future time point.

PERFORMANCE AND INTERPRETATION OF RESULTS

Performance

The single accuracy measure which was calculated for all trial forecasts was the sum of squares of dwelling unit forecasting error. These measures were computed at four levels of geographic aggregation: sector, ring, district, and zone, for all trials. A sixth forecast was made using the naive assumption of equal growth for all zones. The error sum of squares computed under this assumption, which will be referred to as the naive model, is (n - 1) times the variance in actual zonal residential growth. It will serve as a benchmark in evaluating the results of the five techniques listed.

Table 2 gives the computed error sum of squares for all of the forecasts and calibrations at each level of aggregation. For sake of complete comparisons, the results of zone level forecasts for each of the models (not for the DSGM) have been aggregated to districts and rings defined both by time and distance from the HVC. Trial one of the DSGM was based on analysis at the level of district as defined by distance from the HVC; therefore results are not shown for districts as defined by time to HVC, and vice versa for trial two of the DSGM.

The sums of squares of differences between estimated and actual are analogous to "unexplained" variances of a statistical model. However, since valid statistical inferences obviously cannot be drawn, this terminology should not be used. The error measurements of Table 2 do provide an index which can be used to compare results in any single column, that is, for the same level of aggregation. Comparisons between columns are meaningless, since different numbers of areas and different variances from mean growth rates are involved at different levels of aggregation.

To provide some degree of comparison between levels of aggregation, as well as between forecast techniques, Table 3 gives the ratio of each error to that for the naive model.

There are rather poor results at the zone level for all five methods. In some instances the naive model, assuming equal growth for all zones, actually exceeds the level of accuracy of forecasts. The particularly discouraging results of the DSGM at the zone level are evidence of poor choice of criteria by the authors in distributing growth from districts to zones. As pointed out earlier, this method requires historical data that were not available and requires intimate familiarity with the area, which the authors lacked. The technique itself should not be blamed.

Undoubtedly, a substantial amount of the error at such a fine level of detail as the zone can be attributed to inaccuracies in data—assumptions made in certain estimates, incompatibility of merged files, differences in definitions between time periods, etc. However, other factors are contributory. The average zone contained only 109 dwelling units in 1948 and increased 56 to 165 by 1960. These values are far too small to hope for reliable predictions with any model. Obviously, differences between zones

TABLE 2

ERROR SUM OF SQUARES FOR ALL TRIALS[a]

Method	Levels of Aggregation					
	Zone	Districts		Rings		Sector
		By Distance Ring	By Time Ring	By Distance	By Time	
DSGM						
Trial I	2.33	6.97	—	8.36	—	9.69
Trial II	2.41	—	4.43	—	4.07	3.02
Accessibility model						
Forecast	1.80	4.16	2.84	3.25	2.33	4.58
Fitted	1.79	3.98	2.76	2.18	1.99	4.46
Regression (fitted)	1.85	4.71	3.14	5.16	2.84	3.71
Stouffer model						
Forecast	2.21	6.45	4.22	5.57	3.48	11.25
Fitted	1.91	4.72	3.07	2.42	1.46	8.84
Schneider model						
Forecast	2.07	6.16	4.13	4.10	3.38	13.92
Fitted	1.95	4.65	3.08	1.91	1.65	10.18
Naive model	2.20	7.66	5.22	20.64	10.54	16.18

[a] All values have been multiplied by 10^{-6}.

TABLE 3

RATIO OF ALL ERRORS TO NAIVE MODEL ERROR

Method	Levels of Aggregation					
	Zone	Districts		Rings		Sector
		By Distance Ring	By Time Ring	By Distance	By Time	
DSGM						
Trial I	1.06	0.91	—	0.41	—	0.60
Trial II	1.10	—	0.85	—	0.39	0.19
Accessibility model						
Forecast	0.82	0.54	0.54	0.16	0.22	0.28
Fitted	0.81	0.52	0.53	0.11	0.19	0.28
Regression (fitted)	0.84	0.62	0.60	0.25	0.27	0.23
Stouffer model						
Forecast	1.01	0.84	0.81	0.27	0.33	0.70
Fitted	0.87	0.62	0.59	0.12	0.14	0.54
Schneider model						
Forecast	0.94	0.80	0.79	0.20	0.32	0.86
Fitted	0.89	0.61	0.59	0.09	0.15	0.63
Naive model	1.0	1.0	1.0	1.0	1.0	1.0

at this level are largely due to random variations not explainable by models. The districts represent a more reasonable level of detail at which to examine and compare accuracies. For the sake of comparison with transportation study practices, the average district (defined by distance rings) used in this study could be expected to have about 8,000 person trip-ends in 1948 (about 660 dwelling units with 3.2 persons per dwelling and 4 trip-ends produced per person).

Table 4 shows the relative accuracy of the accessibility model forecast at various levels in comparison to the size of the values being forecast. In this table the root-mean-square-error (RMSE) is used as the measure of error, since it can be compared with the magnitude of the forecast values: about two-thirds of the errors fall within RMSE values.

The RMSE is roughly half of the average 1960 dwelling units per zone, and about a third of the average 1960 dwelling units per district. Of course, these accuracies must be viewed in relation to the overall growth rate of 52 percent. Intuitively one would expect that the ratios of the RMSE's to the 1960 values might be nearly cut in half if the overall growth rate was half as large.

The accessibility model performed substantially better than other unfitted models at most levels of aggregation (Table 3); but the fitted Stouffer and Schneider models were

TABLE 4

COMPARISON OF ERRORS TO SIZE OF FORECAST VALUES
ACCESSIBILITY MODEL FORECAST

Levels of Aggregation	RMSE $\sqrt{\frac{s \cdot s}{n}}$	Average 1960 DU (per areal unit)	Average Growth 1948-1960	Number of Areas
Zone	85	165	56	249
District[a]	381	1,006	342	41
Rings[a]	600	4,580	1,560	9

[a] By distance.

quite comparable to the fitted accessibility model. Somewhat surprisingly, the addition of several other explanatory variables in linear regression form did not improve the accuracy.

Results at the sector level are of interest because of the implications for forecasting radial corridor movements. Here the intervening opportunity models yield comparatively poor results, perhaps because they were not made sensitive to the distribution of employment, as were the accessibility model and regression equation.

Trial one of the DSGM assumed relative growth by sectors in proportion to available capacity—a weak assumption judging by comparison with the error of trial two. The importance of residential character in attracting additional growth apparently holds at all levels—between sectors as demonstrated by comparison of the two DSGM trials, and as a factor at the zone level as demonstrated by the statistical significance of that factor in the regression analysis.

Examination of Actual Patterns of Growth

All forecasts of 1960 density were based on the assumption that development in any zone would occur at the density indicated by a smooth line drawn through the 1948 density vs distance (or time) from the HVC. Figure 7 compares the actual 1960 density-distance gradient with that for 1948. There appears to have been a rather uniform amount of decrease in density at all distances, except for the core area where the decrease was substantial. This obviously accounts for some error in the forecasts which required estimates of 1960 density (DSGM and the opportunity models), especially in the core area.

The actual 1960 and 1948 percent saturation gradients are compared in Figure 8, along with the forecast curve used for trial one of the density-saturation gradient method. Not surprisingly, the actual 1960 curve does not follow as smooth a curve as for 1948, since the plot represents percentage of 1948 capacity rather than 1960 capacity. The most significant errors in the forecast appear to be due to the unexpectedly large decline in the core and the amount of growth that occurred in relatively remote portions of the area, ring 5 and 6. However, the general shape of the forecast curve is appropriate.

Figure 9 shows the same comparisons for the results of the accessibility and regression models. The agreement with the actual 1960 gradient is quite good, except for the obvious inability of these techniques, as used in this study, to predict decreases in the core.

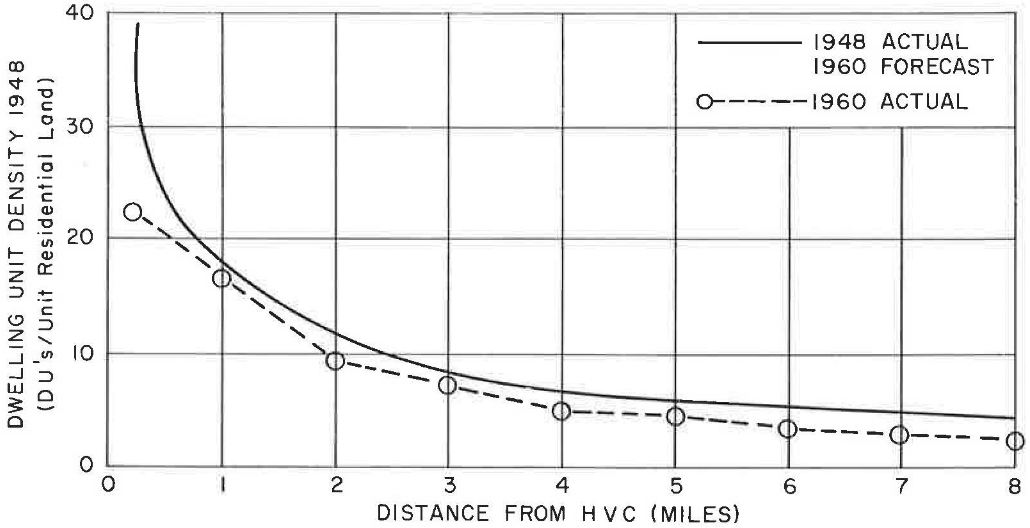

Figure 7. Dwelling unit density by distance bands.

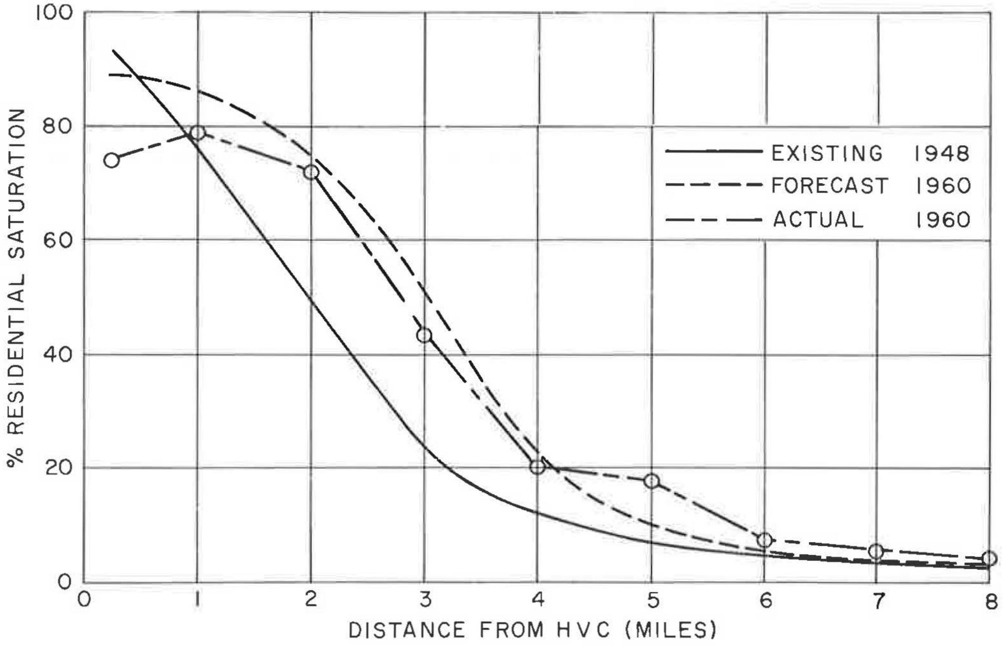

Figure 8. Percent residential saturation by distance bands.

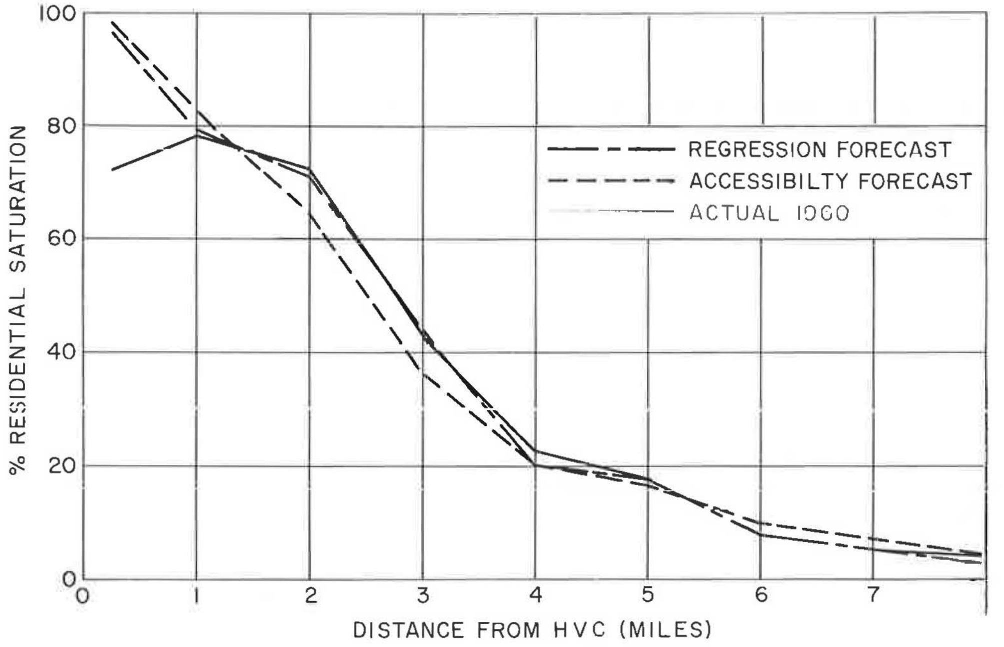

Figure 9. Percent residential saturation by distance bands.

In an attempt to picture how the residential density structure of the study region changed, Figure 10 was drawn. Using the data for total dwelling units and residential land area from the distance to HVC ring analysis, cumulative percent of total regional dwellings was plotted against cumulative percent total residential land area on a ring aggregate basis, proceeding outwards from the core ring. The plots for the actual conditions in 1948 and 1960 are shown. If smooth curves were drawn the slope at any point would represent the inverse of density for the marginal dwelling unit. A diagonal line drawn on Figure 10 would represent uniform residential density for the entire study area. The bowing of each of the curves below the diagonal indicates the decline in density as one proceeds outwards from the HVC. If densities in the inner area were to decline along with an increase in the dwelling unit densities in the outer rings, the region as a whole would be approaching a state of uniform density, and the curve would shift toward the diagonal. On the other hand, if the difference between inner and outer area densities were to increase substantially, then there would be a shifting of the plot down and to the right. Understanding that the plots in Figure 10 represent an overall increase from 1948 to 1960 of 52 percent, the rather minute change in the density structure of the study area as described by these plots is outstanding.

Although the two plots (Fig. 10) appear to coincide almost exactly, they should not be misread as indicating no change in the geographic distribution of dwelling units from 1948 to 1960. Each of the data points representing a distance ring has shifted downward and to the left from its 1948 position to 1960. That is, inasmuch as the majority of residential growth occurred in the suburban rings, the dwelling stock of the inner rings in 1960 represents a smaller proportion of the total region stock than in 1948 and also utilizes a smaller proportion of total residential land; hence, the shifting of the data points downward and to the left.

An interesting question is whether similar plots for other urban areas exhibit this same constancy as found in Greensboro. If this is found to be so, such plots could be quite helpful in residential forecasts.

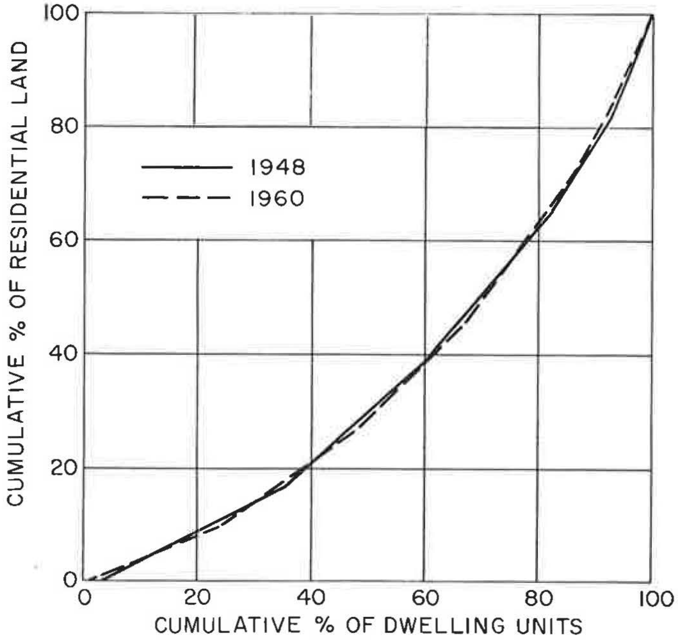

Figure 10. Cumulative dwelling vs cumulative residential land.

CONCLUSIONS

1. Simple, nonbehavioral residential land use forecasting models, which do not discriminate between the locational patterns of different types of households, are sufficiently accurate to be recommended for use in relatively small metropolitan areas of 100,000 population or larger. The Greensboro area's spatial structure and pattern of growth clearly demonstrates a degree of organization warranting analytical treatment in the planning process.

2. Land use forecasting with simple first generation models produced reasonably accurate results for levels of geographic aggregation where the average areal unit contained a population of about 2,000 persons. Efforts to forecast growth for much smaller areas may prove unjustified. At zone levels of about 300 population, these models appeared to offer little or no assistance in forecasting.

3. Differences in accuracy among the five forecasting methods are not large enough to warrant a strong recommendation for any single one in preference to others. Any of the methods would appear to be preferable to forecasting without the benefit of analytical techniques.

4. The simple accessibility model yielded the most accurate forecast of all methods used without benefit of calibration to time series data, for this one test. Errors in fitting were relatively insensitive to small changes in the exponent of accessibility.

5. None of the multiple linear regression models tested offered improvement over two-variable fitted models despite the fact that five or more factors were included in the regression equations.

6. Multiple regression models possess certain drawbacks. If the dependent variable is expressed as an extensive quantity (e.g., increase in dwelling units) then measured relationships with independent variables are influenced by pecularities of area definition and size, and may not conform satisfactorily with logical hypotheses regarding the land development process. Nonlinear transformations on the dependent variable such as logarithms or fractional power functions are unsatisfactory because the usual least squares criterion tends to bias the parameter estimates to produce good fits to small values and poor fits to large values. Expression of the dependent variable as an intensive quantity (e.g., dwelling unit increase per unit area) may be the most satisfactory operational solution except that relationships which are actually nonlinear may not be properly represented. Perhaps this might be handled by treating certain independent variables as sets of dummy variables.

7. Although the two intervening opportunity models performed satisfactorily as used in this study, some evidence pointed to the possibility of improvement by allocating growth from all major centers of employment rather than from just a single point, the CBD. In addition, each of the two models implies a different straight line plot on different semilogarithmic coordinates which did not hold true for Greensboro over the entire study area. Apparently the hypotheses are valid, but separate functions may be necessary for the built up, inner-city area, and the developing suburban area.

8. The forecasting approach used by CATS differed from the other models in important respects. It forces the analyst to become intimately familar with the study area before attempting to forecast. This is probably the strongest feature to recommend it. The graphical analyses that the method is based on represent excellent descriptions of the key spatial relationships of a metropolitan area—even for relatively small areas. The methods of analysis are useful tools regardless of the forecasting technique used. They can serve as checks on the reasonableness of forecasts made by less subjective models.

However, as applied in this study, the method is time-consuming, requiring considerable hand work and far more data manipulation. The method is less adaptable to the computer, and hence would be cumbersome for testing of alternative land use policies, or for recursive use in combination with other submodels.

9. The five techniques examined admittedly are far from representative of the extent of current land use forecasting research. They do represent the initial attempts and as such lack the sophistication and elegance of later thinking. These are descriptive models in that they do not involve themselves with the behavior of decision-makers;

nor do they possess any real theoretical content. It is highly probable that the key to increased forecasting accuracy for small subareas lies in the ability of the analyst to simulate the decision process of subpopulations of the region.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to both Professor F. S. Chapin, Jr., of the University of North Carolina and to A. M. Voorhees of Alan M. Voorhees and Associates for making available the data utilized in this study.

Appendix

CALIBRATION OF ACCESSIBILITY MODEL

Two procedures were used in the attempt to estimate the optimal exponent of accessibility: linear regression on transformed variables and an iterative, nonlinear least squares fit of the untransformed dependent variable.

Linear Regression on Transformed Variables

Three transformed versions of the standard accessibility model were tested:

$$\log G_i = \log a + b \log V_i + c \log A_i \tag{1}$$

which, in nonlogarithmic form is

$$G_i = a V_i^b A_i^c \tag{2}$$

$$\log \left(\frac{G_i}{V_i} \right) = \log a + b \log A_i$$

or in nonlogarithmic form

$$G_i = V_i A_i^b$$

$$\log G_i - \log V_i = \log a + b \log A_i \tag{3}$$

which is the same as Eq. 2 in nonlog form.

The nonlogarithmic forms of Eqs. 2 and 3 are essentially equivalent to the standard form of the model as stated in the body of this report. They would be identical if the normal equations contained the condition that

$$a = \frac{G_T}{\sum_i V_i A_i^b}$$

Since a standard regression program was used, this condition may be violated, and equation estimates must be factored to sum to actual total growth. This holds for all three of the transformed versions of the model.

TABLE 5

RESULTS OF THREE VERSIONS OF LINEAR REGRESSION ON TRANSFORMED ACCESSIBILITY MODEL

Item	Eq. 1	Eq. 2	Eq. 3
Accessibility exponent (b)	3.52	1.63	2.29
Log a	-8.0	-3.2	-4.9
Vacant land exponent (c)	1.51	1	1
Sums of squares of error ($\times 10^8$)	2.21	1.89	1.78

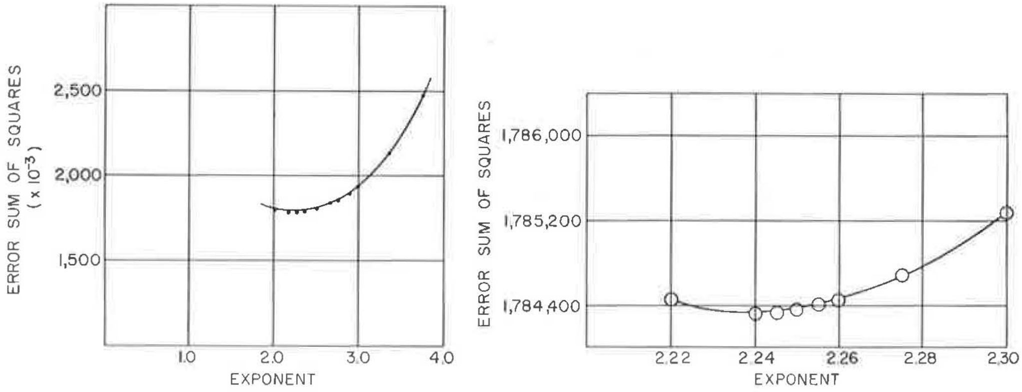

Figure 11. Accessibility model error vs accessibility exponent.

Eq. 1 also expresses vacant land as a power function in contrast to its linear form in the standard formula.

The basic problem, however, is that the least squares criterion is different for each version (the minimization of unexplained variance in the dependent variable) since the dependent variable is different for each. None is the correct criterion. The log transform tends to produce a bias toward better fits for small values of the untransformed dependent variable. Table 5 summarizes the results of the three versions.

The fairly wide variation in the accessibility exponent, as well as in the error term leads one to be suspicious of regression on transformed dependent variables.

Nonlinear Least Squares Fit of Exponent

A routine was programed to iterate toward the true least squares solution for the standard accessibility model

$$G_i = G_T \frac{A_i V_i^b}{\sum_i A_i V_i^b}$$

Figure 11 shows the results in the form of a plot of the sums of squares of error vs a range of exponents. A smooth curve with a minimum at b = 2.24 is apparent.

It is interesting to compare these results with the b value of 2.7 reported by Hansen for Washington, D.C. One might expect this value to increase with the size of the city.

REFERENCES

1. Hamburg, J. R., and Sharkey, R. H. Land Use Forecast. Chicago Area Transportation Study, 3.2.6.10, Aug. 1, 1961.
2. Hamburg, J. R. Land Use Projection for Predicting Future Traffic. Highway Research Board Bull. 224, pp. 72-84, 1959.
3. Hansen, W. G. Land Use Forecasting for Transportation Planning. Highway Research Board Bull. 253, pp. 145-151, 1960.
4. Hansen, W. G. How Accessibility Shapes Land Use. Jour. of the Amer. Inst. of Planners, May 1959.
5. A Projection of Planning Factors for Land Use and Transportation. Baltimore Regional Planning Council, Tech. Rept. No. 9, March 1963.
6. Graves, C. H. The Multiple Regression Models of Small Area Population Change. Highway Research Record 102, pp. 42-53, 1965.
7. Hartford Area Traffic Study Report. Volume I. Conn. Highway Dept., July 1961.

8. Stouffer, Samuel A. Intervening Opportunities: A Theory Relating Mobility and Distance. Amer. Soc. Rev., Vol. 5, No. 6, Dec. 1940.
9. Tomazinis, Anthony R. The Development of the Penn-Jersey Transportation Study Trip Distribution Model.
10. Chicago Area Transportation Study. Final Report. Vol. II. July 1960.
11. Hamburg, J. R., Lathrop, G. T., and Young, G. F. An Opportunity-Accessibility Model for Allocating Regional Growth. Highway Research Record 102, pp. 54-66, 1965.
12. Hamburg, J. R., and Lathrop, G. T. An Opportunity-Accessibility Model for Allocating Regional Growth. Jour. of the Amer. Inst. of Planners, May 1965.
13. Chapin, F. S., Jr., and Weiss, S. F. Factors Influencing Land Development. Univ. of North Carolina, April 1962.
14. Lowry, Ira S. A Model of Metropolis. The RAND Corp., Aug. 1964.
15. Traffic Research Corporation. Review of Existing Land Use Forecasting Techniques. Boston Regional Planning Project, July 1963.
16. Strodtbeck, F. L. Equal Opportunity Intervals: A Contribution to the Method of Intervening Opportunity Analysis. Amer. Soc. Rev., Vol. 14, No. 4, Aug. 1949.