

Prototype Development of Statistical Land-Use Prediction Model for Greater Boston Region

DONALD M. HILL, Senior Research Analyst, Traffic Research Corporation;
DANIEL BRAND, Project Engineer, Traffic Research Corporation; and
WILLARD B. HANSEN, Program Planning Analyst, Boston Regional Planning Project

Traffic Research Corporation has developed and tested a prototype model for the subregional distribution of population and employment growth in a metropolitan area. The model, called the EMPIRIC model, requires externally specified regional growth totals for population and employment categories to be projected. The development work was carried out in conjunction with the efforts of the Boston Regional Planning Project to prepare a comprehensive development plan for the Boston region.

Validity checks have been carried out by applying the model to forecast five categories of population and employment from 1950 to 1960 for 29 subregions and then comparing the observed and calculated subregional activity levels for 1960. Root-mean-square error ratios obtained with prototype EMPIRIC Model Application III were about 1 and 3 percent for total population and total employment, respectively.

The results seem to confirm the hypothesis that urban land use may be predicted on a subregional level, using an "associative" or "statistical" model of the EMPIRIC type, with sufficient accuracy for land-use and transportation planning purposes. The use of such a model enables planners to study systematically the effects of transportation facilities, land-use controls, and other policies on urban development and to produce staged plans and policies which take these interactions into account.

•THE UNDERLYING concept of the EMPIRIC model is that the development patterns of urban activities are interrelated in a systematic manner which provides a reasonable basis for their prediction. The model provides the formal mathematical mechanism for evaluating the extent of these interrelations between activities. The only restriction imposed by the model is that the interrelationships be expressed so that the influences of variables are additive. Accordingly, the model assumes a linear form. Any desired combination or transformation of variables may be introduced to describe the urban activities whose locational pattern we wish to measure and predict. The model requires exogenous specification (i. e., external predictions) of regional growth totals for all urban activities to be projected.

To describe the model, it is convenient to define a number of quantities as follows:

1. The region is divided into a number of small areas called subregions (Fig. 1).
2. The purpose of the model is to predict the amounts of several urban activities in each subregion at the end of a given forecast period. These activities are called

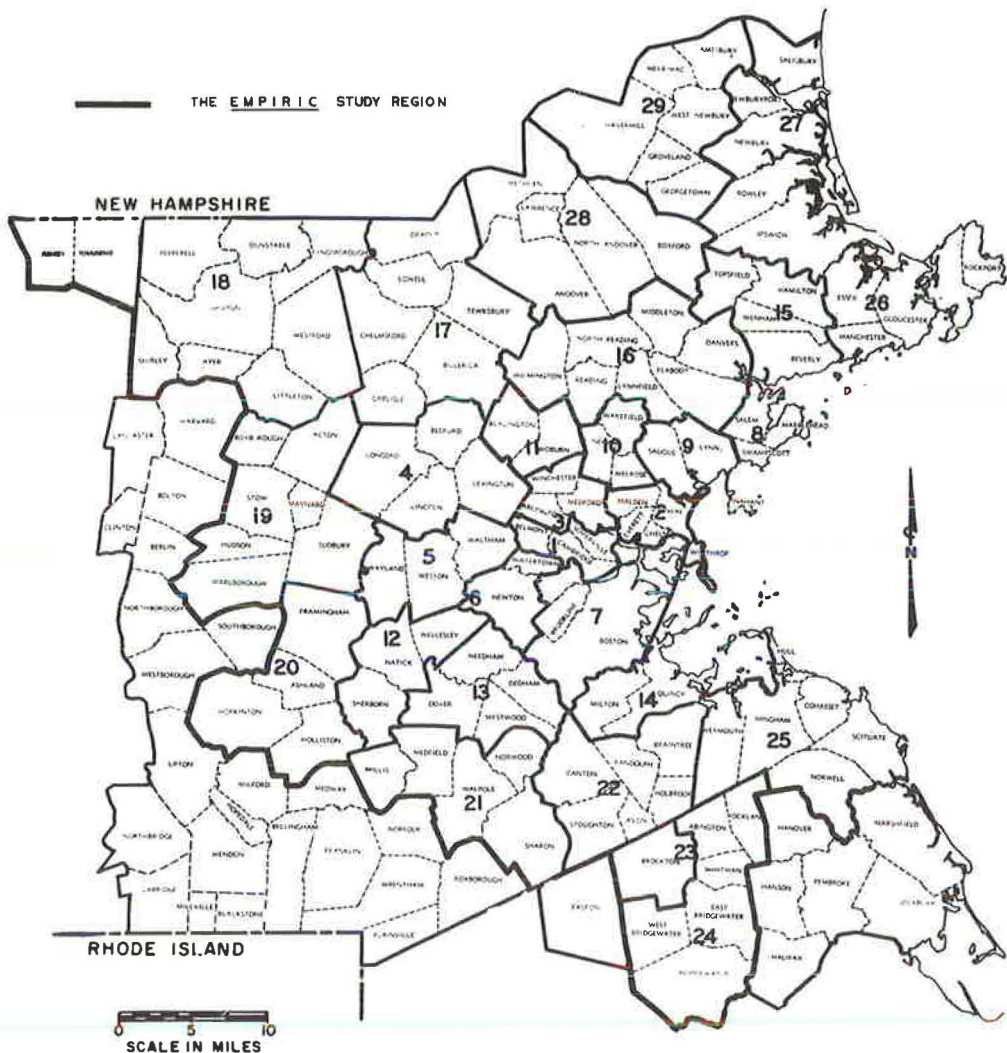


Figure 1. Subdivided EMPIRIC study region, Boston Regional Planning Project.

located variables, signifying that the task of the model is to allocate given regional totals of these variables at the end of the forecast period to the subregions comprising the region.

3. It has been found that the locations and intensities of several urban activities are related to development patterns of one or more variables in a casual manner, that is, whose presence or absence in a subregion, or whose ease of accessibility to the subregion, may be said to influence the amounts of one or more located variables in each subregion. These influencing variables are called locator variables.

The model is formulated to explain changes in activity levels of urban subregions over one or more time periods. Accordingly, the concept of the model may be stated as follows: the change in the subregional share of a located variable in each subregion is proportional to the change in the subregional share of all other located variables in the subregion, the change in the subregional share of a number of locator variables in the subregion, and the value of the subregional shares of other locator variables.

The concept of model may be stated by:

$$\begin{aligned}
& \frac{R_{i\ell}^{(t+1)}}{\sum_{\ell=1}^L R_{i\ell}^{(t+1)}} - \frac{R_{i\ell}^{(t)}}{\sum_{\ell=1}^L R_{i\ell}^{(t)}} = \sum_{\substack{j=1 \\ j \neq i}}^N a_{ij} \left[\frac{R_{j\ell}^{(t+1)}}{\sum_{\ell=1}^L R_{j\ell}^{(t+1)}} - \frac{R_{j\ell}^{(t)}}{\sum_{\ell=1}^L R_{j\ell}^{(t)}} \right] + \\
& \sum_{k=1}^{M-m} b_{ik} \left[\frac{Z_{j\ell}^{(t+1)}}{\sum_{\ell=1}^L Z_{k\ell}^{(t+1)}} - \frac{Z_{k\ell}^{(t)}}{\sum_{\ell=1}^L Z_{k\ell}^{(t)}} \right] + \\
& \sum_{k=M-m+1}^M b_{ik} \left[\frac{1}{L} - \frac{Z_{k\ell}^{(t)}}{\sum_{\ell=1}^L Z_{k\ell}^{(t)}} \right] \quad (1)
\end{aligned}$$

where

- $R_{i\ell}$ = level of located variable i in subregion ℓ ;
 $Z_{k\ell}$ = level of locator variable k in subregion ℓ ;
 L = number of subregions, $\ell = 1, 2, \dots, L$;
 N = number of located variables, $i = 1, \dots, i, j, \dots, N$;
 M = number of locator variables, $k = 1, 2, \dots, M$;
 $(t+1), (t)$ = (located and locator) variables at end and beginning of forecast or calibration interval, respectively; and
 a_{ij}, b_{ik} = coefficients expressing interrelationships among variables.

There is one Eq. 1 for each located variable i . The coefficients a and b are determined by simultaneous regression analysis of the data from two past points in time (i.e., the model is calibrated).

After determining the coefficients, the equations are used to estimate future subregional shares of each located variable by substituting into each equation the pertinent values of the locator variables for that subregion and solving the equations simultaneously for the subregional located variables. To obtain the forecast in absolute rather than relative values, the subregional shares at the end of the forecast interval are multiplied by the exogenous (i.e., externally forecast) control figure for the total of each located variable in the study region.

DEVELOPMENT OF EMPIRIC MODEL

Development of the EMPIRIC model required detailed analyses of cause and effect relationships between development patterns of all land-use categories, as well as detailed analyses of the independence and interdependence of locational groupings of urban activities at the subregional level. An associative or statistical model rather than a true behavioral model was the goal, since it was felt that existing theories of urban

development and data sources were not far enough advanced to permit the development of a suitable behavioral type of land-use model (2). Four methods of study were followed to achieve the necessary appreciation of urban development:

1. A graphical analysis of a large number of relationships between subregional growth rates of population and employment, and a large number of causal (locator) variables which were thought to be important;
2. An investigation of the changes and trends in population and employment that have taken place during the 1950-1960 decade;
3. An analysis of population and employment data (using factor analysis techniques) to determine to what extent variables can be grouped according to their tendency to locate in proximity to one another, or to exhibit similarities in their influence on the locational tendencies of other variables; and
4. Development of multiple regression equations which demonstrate the relationships between growth rates of population and employment and a large number of variables determined to be significant in their effect on locations of population and employment.

The findings of these studies provided the necessary insight to formulate the model for use as a prediction tool (3).

The method selected for expressing growth rates in Eq. 1 (subregional shares or normalized values) was found to be most suitable in that the predictive model incorporating shares demonstrated the highest coefficients of determination in explaining development during the calibration period (1950-1960). In all, three measures of growth were tested: (a) change in absolute value of an activity, (b) change in intensity of activity on land (density), and (c) change in proportion (share) of activity in each subregion (normalized value of an activity). The latter measure was chosen. There are definite mathematical conveniences associated with this measure. The aggregate of growths forecast for all subregions will always match the control figure for the region. Accordingly, iterative procedures are not necessary to obtain a forecast where the aggregate equals some control figure.

CALIBRATION OF MODEL

Calibration of the EMPIRIC model proceeded with 1950-1960 data for the 29 subregions comprising the Greater Boston region. The specified subregions, each comprising several towns or cities, are shown on the map in Figure 1.

Selected Located Variables

The set of located variables, R_i , selected for inclusion in the model was as follows:

1. White collar population (PW), the resident population participating in the white collar labor force (workers and families), including professional, technical, managerial, clerical and sales workers;
2. Blue collar population (PB), the resident population participating in the blue collar labor force, including craftsmen, foremen, operatives, service workers, laborers and occupation not reported;
3. Retail plus wholesale employment (RW), covered by the Massachusetts Division of Employment Security;
4. Manufacturing employment (M), covered by the Massachusetts Division of Employment Security; and
5. All other employment (OE), including employment not covered by Massachusetts Division of Employment Security plus all others.

Selected Locator Variables

Three sets of locator variables, Z_k , were selected for calibration of the prototype EMPIRIC model (Table 1). The first set of 12 locator variables was used in Model Application I and comprised densities of urban activities and automobile accessibilities to urban activities. The prototype EMPIRIC Model Application II is defined by the

TABLE 1
LOCATOR VARIABLES FOR CALIBRATION OF EMPIRIC MODEL

No.	Name	Symbol of Normalized Variable
(a) Application I		
1	Difference between standardized ^a residential population density (at practical holding capacity) and gross residential population density at time t (serves as a zoning variable).	$\frac{STN\ PLR}{\sum_t STN\ PLR} - \frac{PLR(t)}{\sum_t PLR(t)}$
2	Difference between standardized manufacturing employment density and the gross density at time t.	$\frac{STN\ MLR}{\sum_t STN\ MLR} - \frac{MLR(t)}{\sum_t MLR(t)}$
3	Difference between standardized non-manufacturing employment density and the gross density at time t.	$\frac{STN\ NMLR}{\sum_t STN\ NMLR} - \frac{NMLR(t)}{\sum_t NMLR(t)}$
4	Value of subregional gross residential population density at time t.	$\frac{1}{L} - \frac{PLR(t)}{\sum_t PLR(t)}$
5	Value of subregional gross manufacturing employment density at time t.	$\frac{1}{L} - \frac{MLR(t)}{\sum_t MLR(t)}$
6	Value of subregional gross non-manufacturing employment density at time t.	$\frac{1}{L} - \frac{NMLR(t)}{\sum_t NMLR(t)}$
7	Change over the time period in automobile accessibility to white collar residential population at time t.	$\frac{A^V(t+1)PW(t)}{\sum_t A^V(t+1)PW(t)} - \frac{A^V(t)PW(t)}{\sum_t A^V(t)PW(t)}$
8	Change over the time period in automobile accessibility to blue collar residential population at time t.	$\frac{A^V(t+1)PB(t)}{\sum_t A^V(t+1)PB(t)} - \frac{A^V(t)PB(t)}{\sum_t A^V(t)PB(t)}$
9	Change over the time period in automobile accessibility to total employment at time t. (TE(t)).	$\frac{A^V(t+1)TE(t)}{\sum_t A^V(t+1)TE(t)} - \frac{A^V(t)TE(t)}{\sum_t A^V(t)TE(t)}$
10	Value of subregional automobile accessibility at end of period (t+1) to white collar residential population at time t.	$\frac{1}{L} - \frac{A^V(t+1)PW(t)}{\sum_t A^V(t+1)PW(t)}$
11	Value of subregional automobile accessibility at end of time period (t+1) to blue collar residential population at time t.	$\frac{1}{L} - \frac{A^V(t+1)PB(t)}{\sum_t A^V(t+1)PB(t)}$
12	Value of subregional automobile accessibility at end of period (t+1) to total employment at time t.	$\frac{1}{L} - \frac{A^V(t+1)TE(t)}{\sum_t A^V(t+1)TE(t)}$
(b) Application II		
1-12	Same as above.	Same as above.
13	Change over the time period in transit accessibility to white collar residential population at time t.	$\frac{A^Q(t+1)PW(t)}{\sum_t A^Q(t+1)PW(t)} - \frac{A^Q(t)PW(t)}{\sum_t A^Q(t)PW(t)}$
14	Change over the time period in transit accessibility to blue collar residential population at time t.	$\frac{A^Q(t+1)PB(t)}{\sum_t A^Q(t+1)PB(t)} - \frac{A^Q(t)PB(t)}{\sum_t A^Q(t)PB(t)}$
15	Change over the time period in transit accessibility to total employment at time t.	$\frac{A^Q(t+1)TE(t)}{\sum_t A^Q(t+1)TE(t)} - \frac{A^Q(t)TE(t)}{\sum_t A^Q(t)TE(t)}$
16	Value of subregional transit accessibility at end of time period (t+1) to white collar residential population at time t.	$\frac{1}{L} - \frac{A^Q(t+1)PW(t)}{\sum_t A^Q(t+1)PW(t)}$
17	Value of subregional transit accessibility at end of time period (t+1) to blue collar residential population at time t.	$\frac{1}{L} - \frac{A^Q(t+1)PB(t)}{\sum_t A^Q(t+1)PB(t)}$
18	Value of subregional transit accessibility at end of period (t+1) to total employment at time t.	$\frac{1}{L} - \frac{A^Q(t+1)TE(t)}{\sum_t A^Q(t+1)TE(t)}$
(c) Application III		
1-18	Same as above.	Same as above.
19	Change in quality of water supply over the time period.	$\frac{W(t+1)}{\sum_t W(t+1)} - \frac{W(t)}{\sum_t W(t)}$
20	Change in quality of sewage disposal service over the time period.	$\frac{S(t+1)}{\sum_t S(t+1)} - \frac{S(t)}{\sum_t S(t)}$
21	Value of subregional quality of water supply at time t.	$\frac{1}{L} - \frac{W(t)}{\sum_t W(t)}$
22	Value of subregional quality of sewage disposal service at time t.	$\frac{1}{L} - \frac{S(t)}{\sum_t S(t)}$

^aThis is the maximum net density of each pertinent activity in each subregion observed over the calibration period.

addition of transit accessibilities as locator variables to the previous 12 input variables for a total of 18 input variables. Prototype Model Application III is defined by the addition of quality of water supply and sewage disposal service (4 locator variables) to the 18 previous input variables.

The gross density variables employed above were calculated by dividing the appropriate subregional activity levels by the subregional effective area. This was the area suitable for development by any of the located variables. The areas selected comprised the land in use in 1960 for residential, commercial, manufacturing and wholesaling, plus the agricultural or vacant land which was either suitable (having less than 15 percent slope and not swampy) or which in 1960 was under development for commercial or industrial uses. Data availability was a prime factor in the selection of this definition of effective area.

The standardized densities were based on development trends during the decade 1950-1960 and were set equal to the maximum of the 1950 or 1960 activity levels in each subregion divided by the appropriate area in use in 1960 in that subregion. The areas used were residential land for standardized residential population density, commercial land for nonmanufacturing standardized density, manufacturing-wholesaling land for standardized manufacturing employment density.

The accessibilities were based on the formulations:

$$A_{\ell k}^V = \sum_{\ell=1}^L Z_{k\ell} e^{-BT^V_{\ell p}} \quad (2a)$$

and

$$A_{\ell k}^Q = \sum_{\ell=1}^L Z_{k\ell} e^{-BT^Q_{\ell p}} \quad (2b)$$

where

$A_{\ell k}^V$ or Q = accessibility of subregion ℓ to variable k in all L subregions ($\ell = 1, 2, \dots, \ell, p, \dots, L$; $k = 1, 2, \dots, M$), by highway (V) or transit (Q);

$Z_{k\ell}$ = level of locator variable k in subregion ℓ ;

e = base of natural logarithms (2.71828...);

$T_{\ell p}^V$ = auto travel time plus terminal time between subregions ℓ and p ($p = 1, 2, \dots, \ell, p, \dots, L$);

$T_{\ell p}^Q$ = transit travel time plus terminal time between subregions ℓ and p ; and

B = parameter measuring degree to which propensity for interaction between urban activities decreases with increasing travel time between them. (Graphical and bivariate correlation analysis of the interactions between population and employment in the Boston region indicated that the value $B = 0.05$ was most representative of this effect.)

Travel times were calculated by tracing shortest time paths between pairs of subregions and adding the subregional terminal times at both ends. Shortest time paths were derived from networks representing highway, street and mass transportation systems for 1950 and 1960. The procedure followed was similar to that employed in area transportation studies where networks are constructed and travel times between subregions are obtained by tracing minimum time paths (in this case for peak hour traffic loadings).

The 1950 and 1960 ratings of quality of water supply and sewage disposal service ratings were based on the following scale:

- 1 = Metropolitan District Commission (MDC) system.
- 1.5 = partial MDC system.
- 2 = supplied by community.
- 3 = not publicly supplied.

When a combination of these four possibilities occurred for some subregion, intermediate values to the 1, 1.5, 2 or 3 ratings were calculated, weighted by zonal areas.

FORECASTING WITH EMPIRIC MODEL

All three applications of the calibrated model were used to project 1960 subregional activity levels from corresponding 1950 land-use data, 1960 regional totals of the projected activities and 1960 values of the variables coming under external control (policy variables). The projected (located) variables were the two classes of population (white collar and blue collar population) and three classes of employment (retail plus wholesale, manufacturing and all other employment). The input (locator) variables for the three model applications may be summarized as follows:

Application I—1950 subregional levels and densities of the two classes of population and two groupings of employment; changes in the highway transportation system between 1950 and 1960 and the system in 1950.

Application II—locator variables of Application I; changes in the mass transportation system between 1950 and 1960 and the system in 1950.

Application III—locator variables of Application II; changes in the subregional quality of water supply between 1950 and 1960 and the quality in 1950; changes in the subregional quality of sewage disposal service between 1950 and 1960 and the quality in 1950.

The subregional values of the two classes of population and three classes of employment projected were aggregated by subregion to form the additional located variables, total population and total employment, respectively.

Figures 2 and 3 present the observed and projected 1960 subregional values of population and employment obtained with Model Application II. The accuracy with which the model simulated subregional development in the Boston region in the decade 1950 to 1960 is shown by the close correspondence of the subregional values presented for each subregion in these figures.

Three numerical indices were calculated to measure the correspondence attained between observed and predicted 1960 subregional values. These indices, defined in the following paragraphs, summarize the accuracy obtained over all subregions and result in a single reliability measure for each output variable.

Root-Mean-Square Error

The root-mean-square (RMS) error expresses the deviation from corresponding observed values of the predicted values produced for each located variable. Statistical theory indicates that for about 67 percent of the subregions, the observed value will not differ from the predicted value by more than plus or minus the RMS error. The RMS error is computed by:

$$\text{RMS error} = \sqrt{\frac{\sum_{i=1}^L [R_i(O) - R_i(C)]^2}{L}} \quad (3)$$

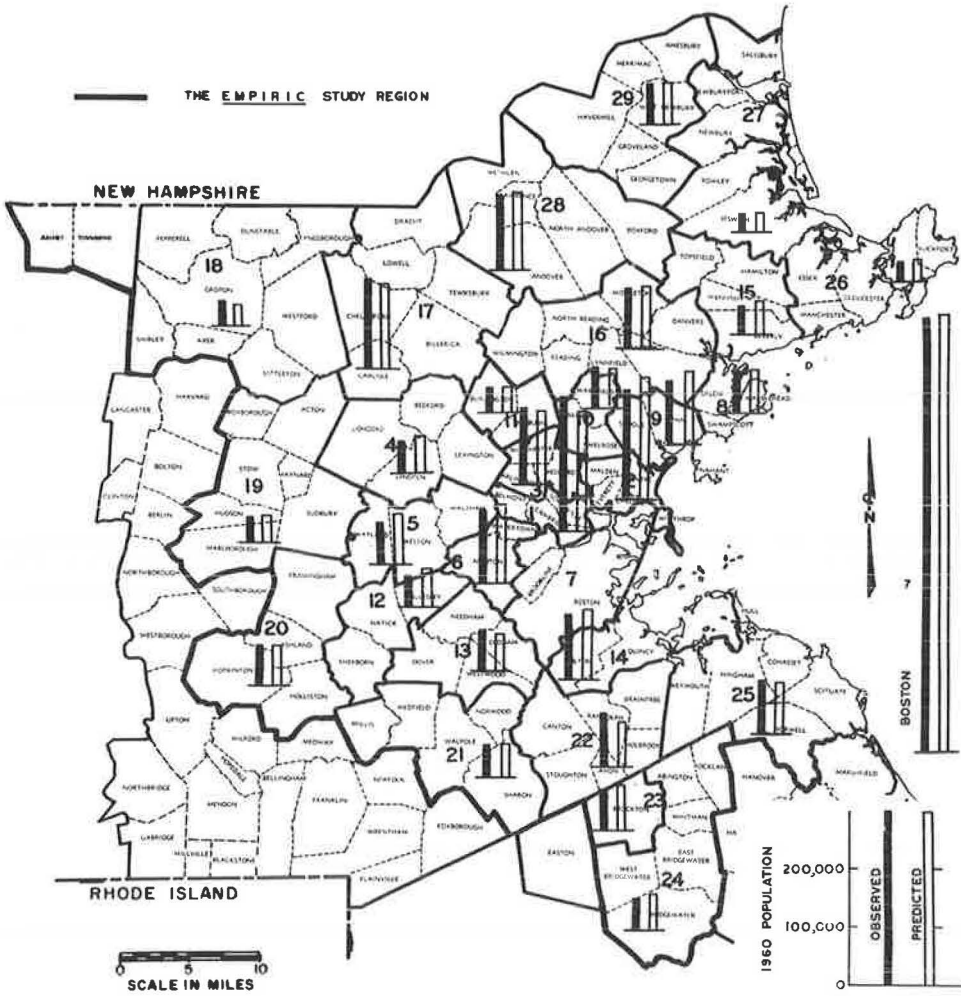


Figure 2. Comparison of observed 1960 population levels with those predicted by EMPIRIC land-use model (Application II).

where

- $R_{i\ell}(O)$ = observed value of located variable i in subregion ℓ , and
- $R_{i\ell}(C)$ = predicted (calculated) value of located variable i in subregion ℓ .

RMS Error Ratio

The RMS error ratio is the ratio of the RMS error to the arithmetic average of the observed output activity values, $\bar{R}_i(O)$. The RMS error ratio is computed as follows:

$$\text{RMS error ratio} = \frac{\text{RMS error}}{\bar{R}_i(O)} \tag{4}$$

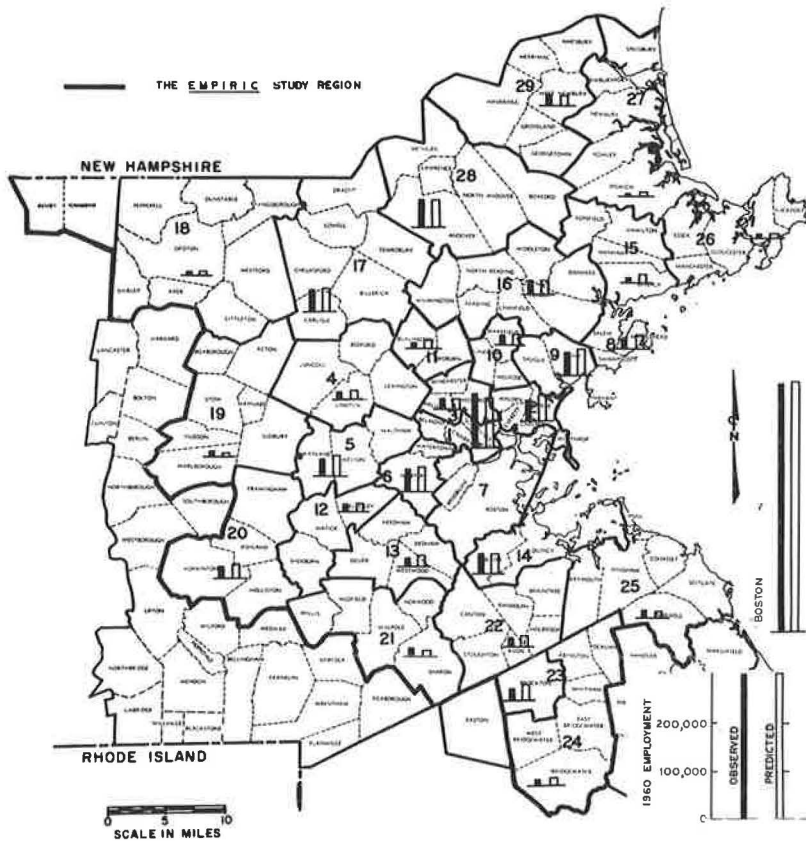


Figure 3. Comparison of observed 1960 employment levels with those predicted by EMPIRIC land-use model (Application II).

where

$$\bar{R}_i(O) = \frac{\sum_{\ell=1}^L R_{i\ell}(O)}{L}$$

Coefficient of Determination

The coefficient of determination, R^2 , a third summary measure for each located variable, represents the proportion of the sum of squared deviations of the observed subregional located variable levels and the mean subregional level, that is explained by the model, and is computed as follows:

$$R^2 = \frac{\sum_{\ell=1}^L [R_{i\ell}(O) - \bar{R}_i(O)]^2 - \sum_{\ell=1}^L [R_{i\ell}(O) - R_{i\ell}(C)]^2}{\sum_{\ell=1}^L [R_{i\ell}(O) - \bar{R}_i(O)]^2} \quad (5)$$

As R^2 approaches 1, the reliability of the model is regarded to be quite high; conversely, as R^2 approaches 0, the reliability is considered to be quite low. There are, however, instances where R^2 is a poor measure of reliability and, therefore, should not be regarded critically. One of these instances occurs as one of the observations (activities for subregion 7) is much larger than the mean $\bar{R}_1(O)$. Accordingly, the first two summary measures more accurately reflect the reliability of the model.

The summary measures, calculated using the subregional values obtained with Model Application II and presented in Figures 2 and 3, are contained in Table 2.

INVESTIGATIONS OF EFFECT ON MODEL FORECAST ACCURACY OF VARYING INPUT CONDITIONS

Several tests have been conducted with the EMPIRIC model to evaluate its accuracy as a forecast tool. Some of the tests involved the investigation of effect on forecast accuracy of varying the number of locator variables or the size and number of subregions. The tests involved both recalibrations and forecasts with the model.

Varying Number of Locator Variables

A comparison of the results of the three model applications illustrate the effect on model forecast accuracy of increasing or decreasing the number and types of input variables. The conditions under which the three applications were made are the same except for numbers and types of input variables. As described in the previous section, Model Application I deletes transit accessibilities as input variables, whereas Model Application III adds water supply and sewage disposal services as input variables to those used in Model Application II. The effect of this decrease and increase in the number of input variables is reflected in the three summary measures calculated using 1960 subregional values predicted with Model Applications I and III. These results are contained in Table 3.

The results presented in Table 3 indicate that greater forecast accuracy is obtainable when increased information becomes available on transportation systems and other planned variables. Additional tests will be carried out incorporating additional information on planned variables into sets of locator variables. Such tests will enable an evaluation of the effect on forecast accuracy of increased information per variable, as well as of the effect of increased total information.

Varying Sizes and Numbers of Subregions

In an attempt to measure the model's error sensitivity to an increase in the number of subregions and a consequent decrease in subregion size, the prototype Model Application III was calibrated with data for the 123 cities and towns in the specified study region. The model was calibrated again with data for the 123 cities and towns, but the city of Boston was divided into 12 areas giving a total of 134 subregions.

Table 4 presents values of RMS errors for total population and total employment. It appears that the errors increase as subregions become smaller. Further, it is noted that with more uniform distribution of subregional activity levels (R_i) in the case of the 134 subregion test, the fore-

TABLE 2
SUMMARY RELIABILITY MEASURES OBTAINED WITH
MODEL APPLICATION II

Output Variable	RMS Error ^a	RMS Error Ratio	R ²
Total population:	3,439	0.03	0.99
White Collar	3,031	0.06	0.99
Blue Collar	1,686	0.03	0.99
Total employment:	1,512	0.03	0.99
Retail and wholesale	564	0.06	0.99
Manufacturing	1,128	0.08	0.99
Other	955	0.05	0.99

^aPeople or jobs.

TABLE 3
SUMMARY MEASURES, OTHER APPLICATIONS

Model Application	Output Variable	RMS Error ^a	RMS Error Ratio	R ²
I	Population	7,825	0.07	0.99
	Employment	3,631	0.08	0.99
III	Population	1,259	0.01	0.99
	Employment	1,306	0.03	0.99

^aPeople or jobs.

TABLE 4
RMS ERRORS OBTAINED WITH 29,
123 AND 134 SUBREGION PROJEC-
TIONS, 1950 TO 1960 MODEL
APPLICATION III

No. of Subregions	Activity	RMS Error
29	Population	1,259
	Employment	1,306
123	Population	5,673
	Employment	3,715
134	Population	3,479
	Employment	2,755

cast accuracy is greater than the 123 town test but remains less than the accuracy of the 29 subregion tests.

The losses in forecast accuracy with more subregions are expected to be minimal when the EMPIRIC model is used on a production basis with large numbers of small subregions. Although the data analyses preceding the formal calibration of the model will be carried out in detail with the small subregions (600), the data are expected to be available for subregion sets whose levels of population and employment are more uniform than those

used for the tests reported on in Table 4. Further, it is expected that new sets of locator variables which reflect locational decisions at a finer areal basis will be introduced and they will help to improve the forecast accuracy of the EMPIRIC model.

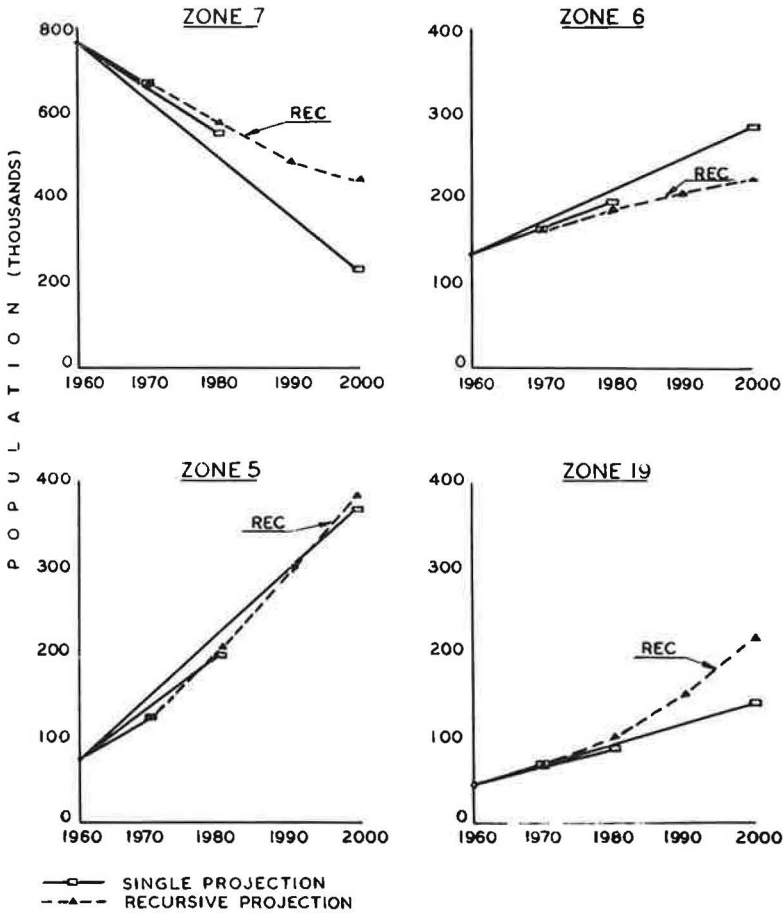


Figure 4. Population projected to 1970, 1980 and 2000 in single projections and to 1970, 1980, 1990 and 2000 recursively in 10-year intervals.

INVESTIGATIONS OF MODEL SENSITIVITY TO VARIATIONS IN FORECAST CONDITIONS

Several tests of the model were carried out for the purpose of investigating the effect on forecast values of: (a) changes in length of forecast period, (b) changes in specification of regional growth rates, (c) changes in zoning policies for suburban "bedroom" communities, and (d) changes in design policies of transportation facilities. These tests did not involve new model calibrations. Each of the tests and their findings are discussed in relation to one of the original 29 subregional model applications (I or II).

Changes in Length of Forecast Period

With any forecast model, errors generally increase as the length of the prediction period increases into the future. This model, as other models, forecasts subregional growths and declines, so that the further into the future one tries to forecast, the greater are these subregional growths (or declines) as a percentage of the original subregional activity levels. Hence, even if the percentage errors in the projected subregional growths remained constant, the percentage errors in the projected subregional values would increase.

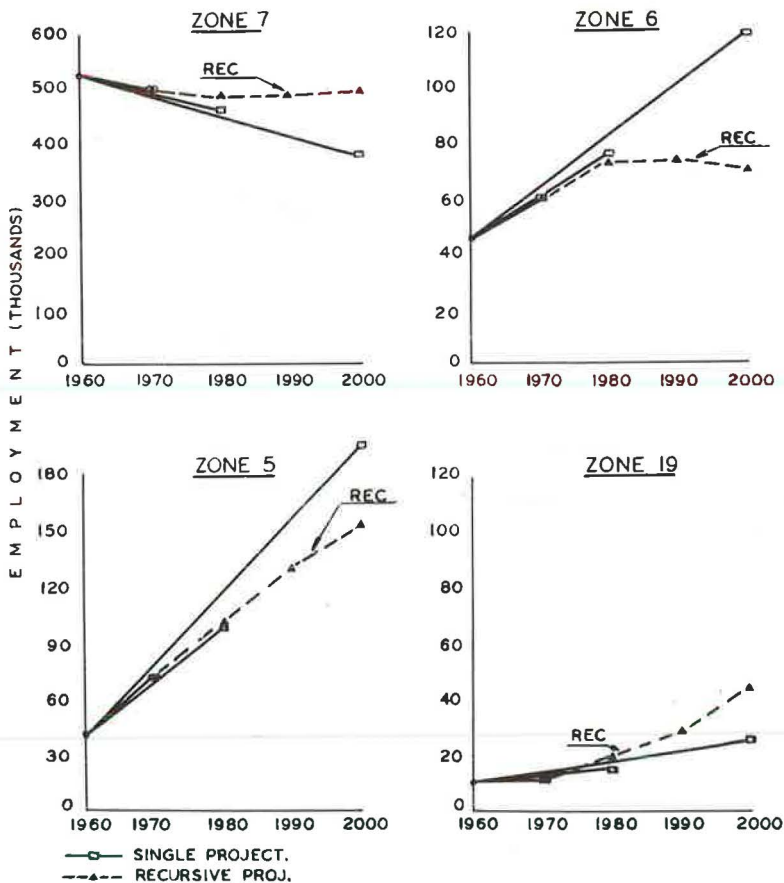


Figure 5. Employment projected to 1970, 1980 and 2000 in single projections and to 1970, 1980, 1990 and 2000 recursively in 10-year intervals.

Three single projections were made with Model Application I, i.e., from 1960 to 1970, from 1960 to 1980, and from 1960 to 2000. The results are demonstrated by Figure 4 for population and Figure 5 for employment. The three forecasts together demonstrate a near-linear change over time (not exactly linear due to compounding of growth). The definite linearity of projected subregional shares is achieved but not shown here.

In addition to the single (nonrecursive) tests, a recursive 10-year projection was conducted. (Recursive forecast means that the model is applied sequentially for relatively short forecasts with the results of each forecast providing input for the succeeding forecast.) At the end of each 10-year projection period, new accessibilities and densities were calculated and input into the model, based on the same 1960 transportation network but on the newly predicted subregional activity levels. The findings of this recursive test are presented in Figures 4 and 5, and the projection demonstrates now a curvilinear trend.

The recursive projection is anticipated to predict subregional growths and declines more accurately because changes in policy measures, the utilization of vacant land and accessibilities are input more frequently. Accordingly, the divergence of the recursive projection from a linear trend probably closely reflects future development rates and, hence, exhibits a generally lower forecast error than the single projections.

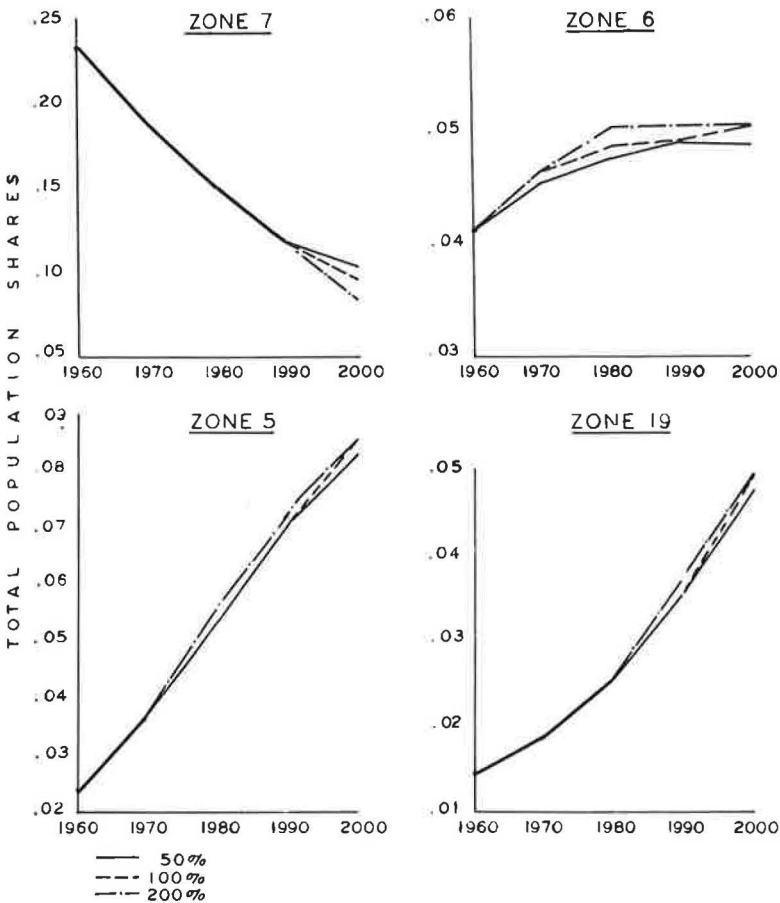


Figure 6. Population shares projected to 1970, 1980, 1990 and 2000 in recursive 10-year projections based on growth rates of 50, 100 and 200 percent of 1950-1960 growth.

Changes in Regional Growth Rates

Recursive forecasts to the year 2000 were conducted with Model Application I, setting the annual growth rates at 50, 100 and 200 percent of the average annual growth rates of each located activity experienced during the 1950-1960 period. (The same average annual growth rate was applied in the preceding forecast tests.)

Absolute subregional population and employment levels were found to increase in all cases as regional growth rates increased. However, subregional shares of regional growth were found to vary on a more selective basis. Subregional population shares for suburban and exurban subregions increased with increased estimated regional growth rates, but Boston (subregion 7) decreased slightly its share of increases in regional population growth. For employment shares, the pattern was reversed. Boston and inner ring subregions increased their share of regional employment as rates of regional growth increased, whereas the exurban subregions participated relatively less in an increased regional employment growth rate until the latter stages of the 40-year forecast interval.

These results for typical subregions are plotted in Figures 6 and 7. Subregion 7 (Boston) is typical of a core subregion. Subregions 5 and 6 are suburban subregions, the latter much older and more densely developed. Subregion 19 is typical of the exurban Boston subregions which only in the 1960's are beginning to undergo suburbanization. Their positions in the region may be seen in Figure 1.

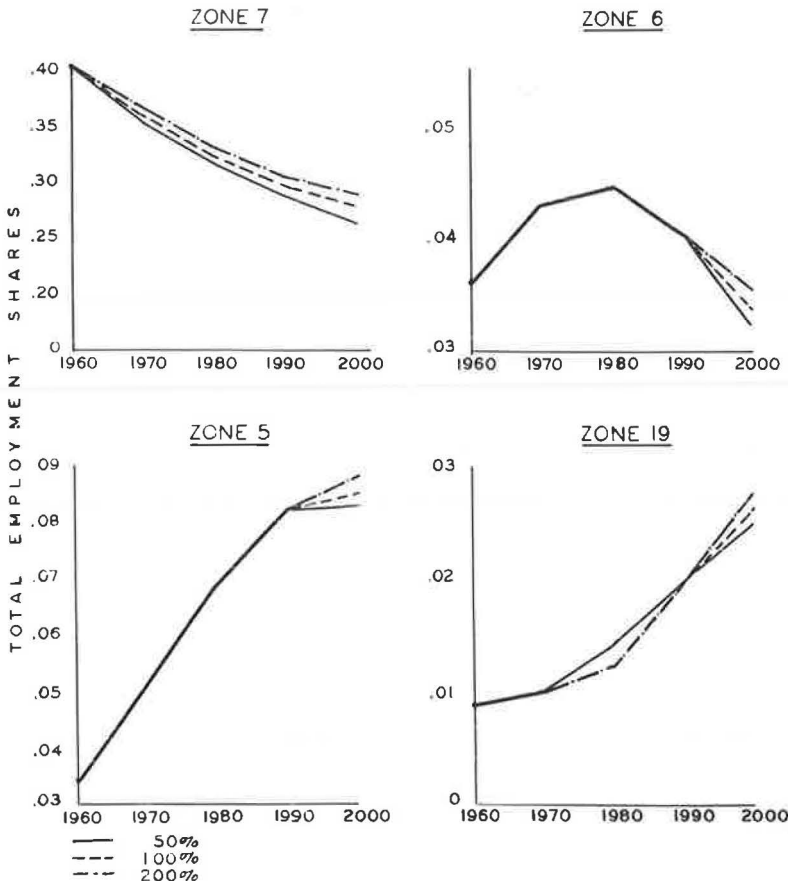


Figure 7. Employment shares projected to 1970, 1980, 1990 and 2000 in recursive 10-year projections based on growth rates of 50, 100 and 200 percent of 1950-1960 growth.

The results presented in Figures 4 through 7 can be interpreted only quantitatively at this stage, since they are intended to illustrate a test of the model's sensitivity to a change in a given forecast condition. They do not necessarily represent accurately future growth in the Boston region (which will be based on new policy decisions).

Changes in Suburban Zoning Policies

A 29-subregion projection from 1950 to 1960 was conducted with Model Application I, in which the 1960 standardized (allowable) population density, STNPLR, of one suburban subregion (Zone 4) was increased slightly and then greatly. This test was intended to determine the effect of such a zoning change on forecast values for the particular subregion and surrounding subregions.

Figures 8 and 9 show that forecast values of population and employment in subregion 4 increased as the STNPLR variable increased from 6.3 to 8.0 and lastly to 80.0. The increases are very pronounced and are demonstrative of the model's sensitivity to the zoning variable. Close examination of the results revealed that subregions with larger original amounts of population and employment tended to lose larger amounts of these activities to subregion 4 than did those subregions with smaller original values.

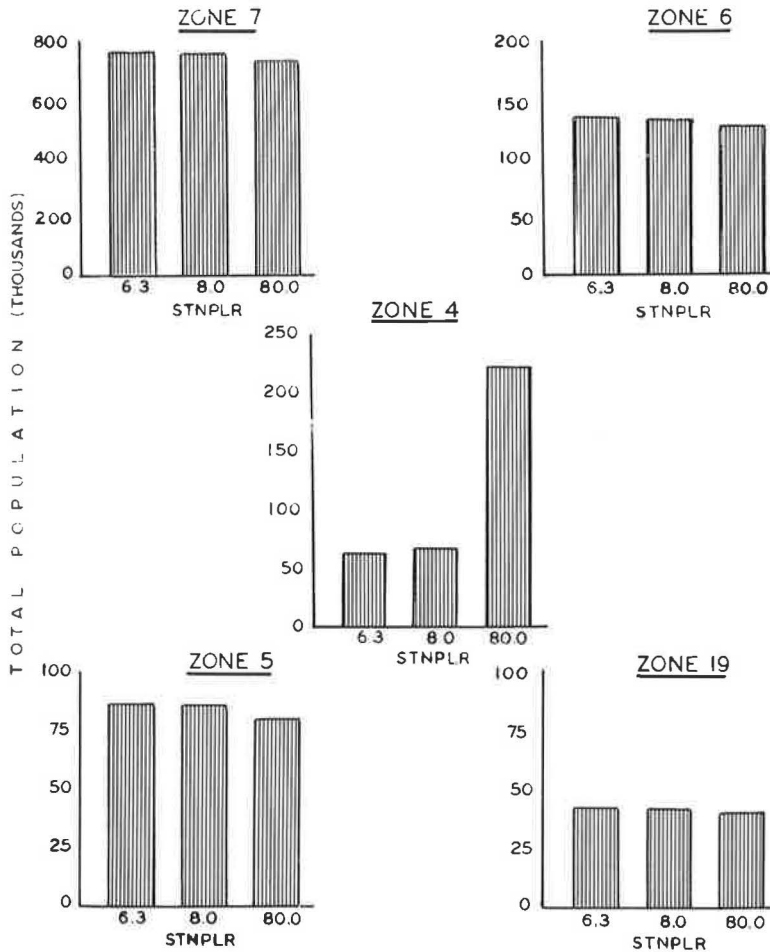


Figure 8. Growths or declines of population by changing STNPLR values of subregion 4.

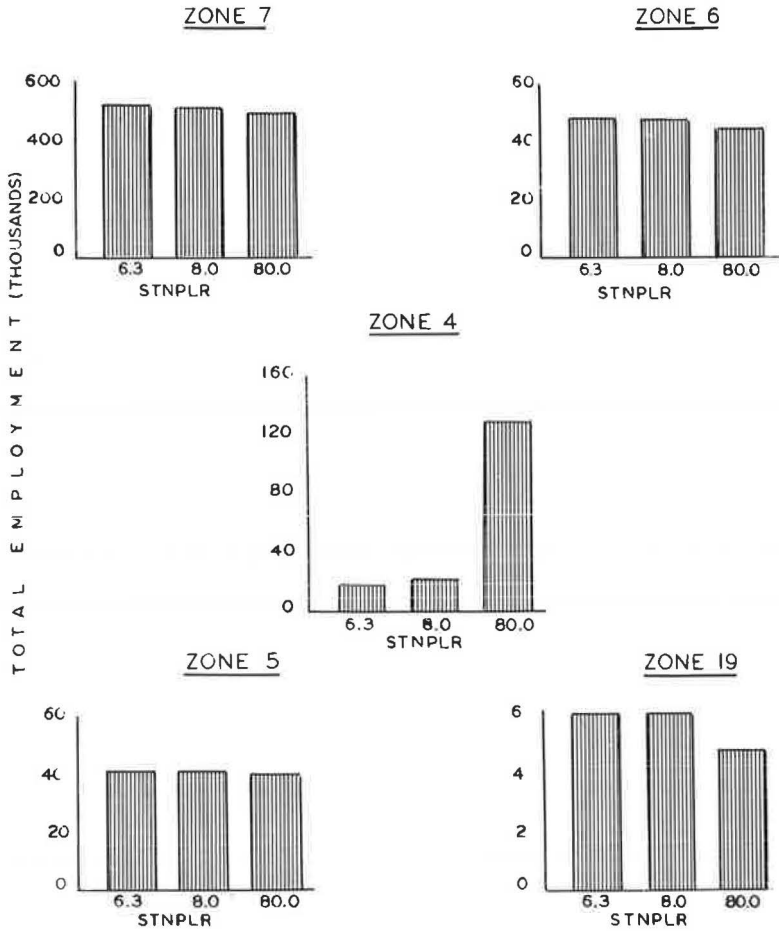


Figure 9. Growths or declines of employment by changing STNPLR values of subregion 4.

Changes in Design Policies of Transportation Facilities

As a final test of the sensitivity of the model, Model Application II, described previously, was used to simulate the effect on the locational patterns of population and employment in the Boston region for two different design policies of transportation facilities over the 1950-1960 decade. The first design policy simulated was exactly that which took place in the Boston region between 1950 and 1960 insofar as highway and mass transportation improvements or closures were concerned. The second simulated design policy was premised on no changes in the highway and mass transportation system between 1950 and 1960. The regional growth of population and employment was assumed to be the same for both design policies.

Figure 10 shows the major expressway segments built in the Boston region between 1950 and 1960. Major mass transportation improvements and closings during the decade are also shown. It should be noted that the transportation improvements consisted primarily of radial expressway sections, plus the major circumferential expressway, Rte. 128, which passed through a tier of suburban communities.

Figures 11 and 12 contain the 1960 subregional values of population and employment predicted by the model with the two simulated transportation policies. Examination of the results shown in these figures indicates that the policy of radial and circumferential transportation improvements result in the expected benefits of increased

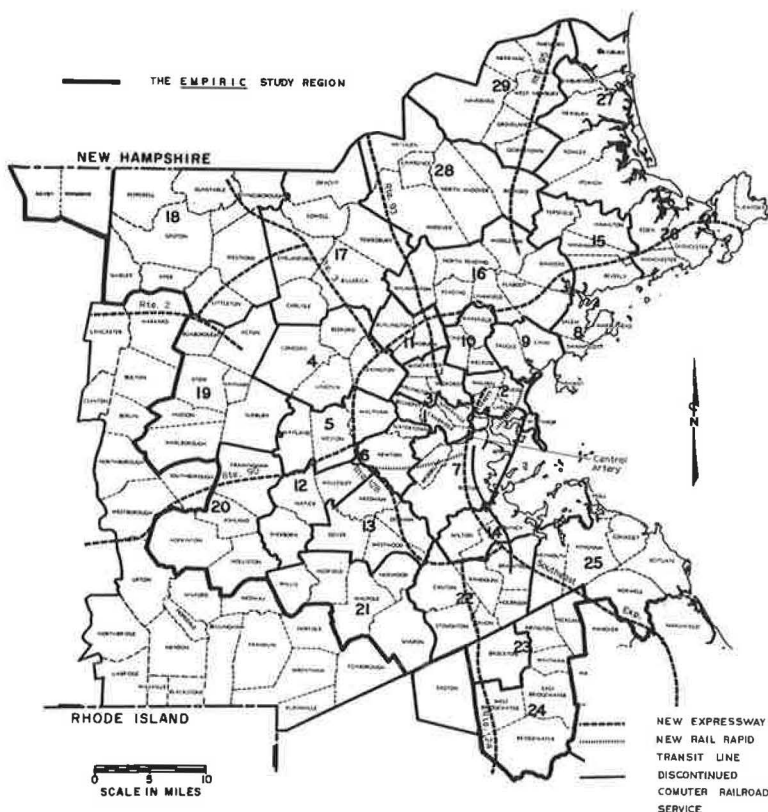


Figure 10. Major transportation facilities built or closed between 1950 and 1960, Boston metropolitan region.

population and employment in the third and fourth tier of suburban subregions. Also, it is interesting to note the increases in population, and employment in the older core cities of Boston, Cambridge and Somerville resulting from the policy of transportation improvements. This predicted redistribution of population and employment as a result of a transportation improvement is illustrative of the way in which a land-use model may be used to evaluate alternative design policies for public facilities.

APPLICATION OF EMPIRIC MODEL IN CONJUNCTION WITH TRAFFIC MODEL

The EMPIRIC land-use model may be applied in conjunction with a travel forecasting model in the following iterative manner:

1. Inventory would be made for the initial year (e. g. , 1960 for the first forecasting period) of all travel flows, times and costs, and the level of each pertinent activity in each subregion. Estimates of regional growth factors for a 5-year period, for example, would be made for each activity to be predicted.
2. The land-use model (calibrated on the basis of past data) would be applied for a 5-year forecasting period starting with the initial values of activity levels in each subregion and using travel times and costs for the initial year and travel times and costs at the final year, based on the completion of new transportation facilities and the closure of old facilities. Values of other planned variables, such as utility coverage, would also be for the initial year and the final year.

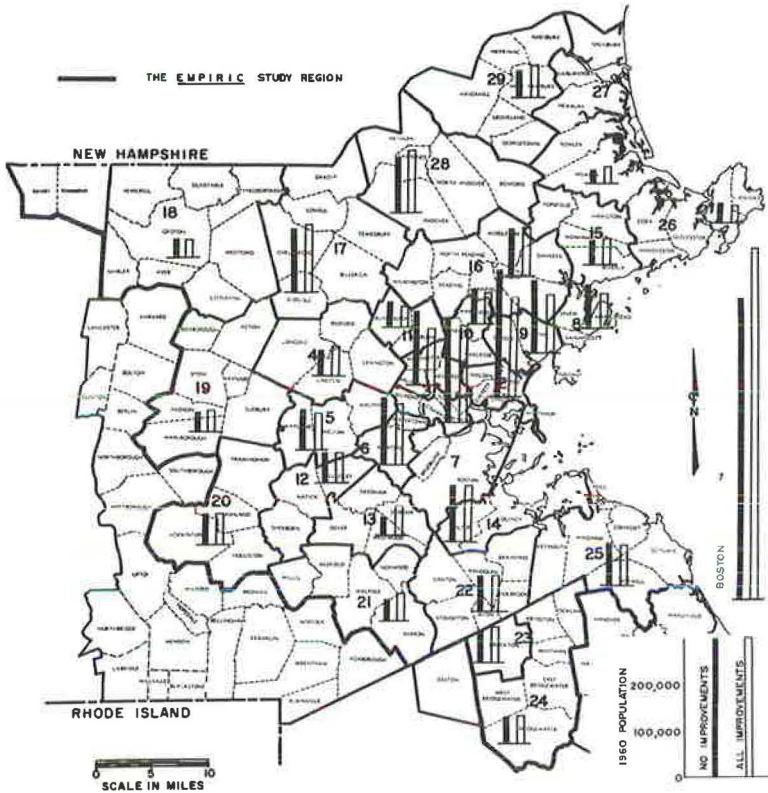


Figure 11. Results of two simulated transportation design policies using EMPIRIC land-use model (Application II): predicted 1960 population levels with no transportation improvements or closures between 1950 and 1960 vs predicted levels with all improvements and closures.

3. Based on the predicted land-use pattern of step 2 and the travel facilities scheduled for completion, the traffic model would be applied for the target year (every fifth year) to determine new traffic flows, travel times and costs.

4. The procedure outlined in steps 2 and 3 would be repeated if the travel times and costs found in step 3 differed substantially from values used in step 2.

5. The procedure outlined in steps 1, 2, 3, and 4 would be repeated for successive 5-year periods, using activity levels estimated by the land-use model at the end of each period as starting levels for forecasting in the next period.

This process would be repeated until the final target year had been reached. The sequence of forecasts of 5-year activity levels, travel flows, times and costs, etc., thus produced, would represent a systematic estimate of how the region would develop under the influence of regional growth rates and planning policies relating to transportation and utilities.

Steps 3 and 4 could be elaborated on if the flows and travel times produced initially were such as to indicate inadequacies in the proposed transportation system. More adequate transportation facilities could be proposed in this case and evaluated by the traffic model to provide a basis for further forecasts. Similarly, iterative application of the land-use model could be carried out in step 2 if the initial urban growth pattern for a particular period is found to be undesirable. In this case, the land-use model would be rerun using the same initial activity levels in each subregion but different proposed values of transportation and/or planned activities, chosen to correct, if possible, the undesirable qualities of the original growth pattern.

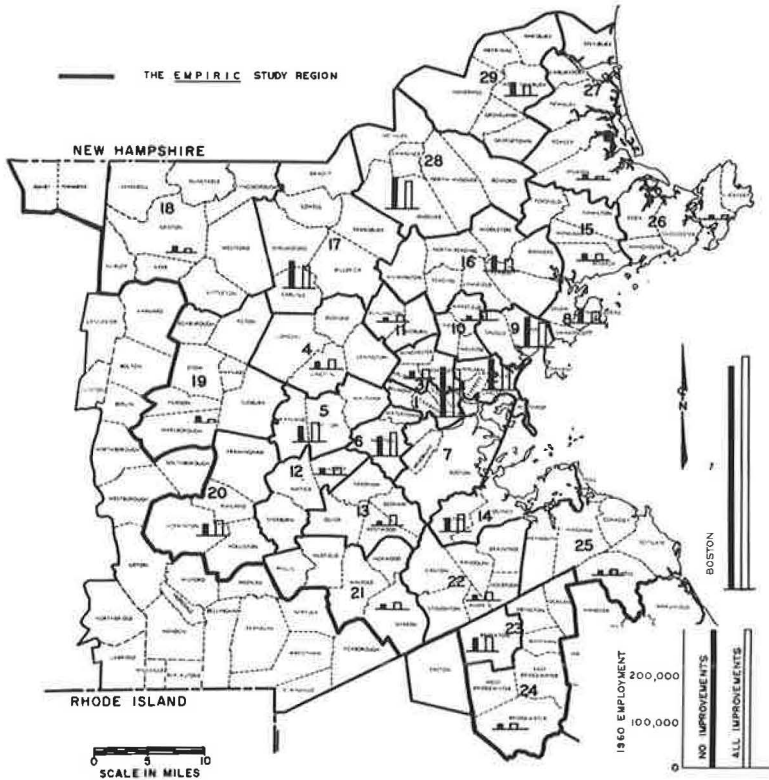


Figure 12. Results of two simulated transportation design policies using EMPIRIC land-use model (Application II): predicted 1960 employment levels with no transportation improvements or closures between 1950 and 1960 vs predicted levels with all improvements and closures.

In this manner, possible alternative staged plans would be evaluated in an effort to produce a master development plan for the region.

CONCLUSIONS

The results demonstrated by the prototype EMPIRIC model give rise to the conclusion that the model shows promise of bringing the planner the ability to simulate a chain of events in urban development, starting with variables he can control (such as the transportation system and open space regulations) and ending with a pattern of residential and industrial development. This process is amenable, efficient and desirable.

ACKNOWLEDGMENTS

During the course of this study and preparation of this paper, cooperation and assistance were received from many quarters. In particular, the authors wish to acknowledge the generous assistance given by R. S. Bolan, H. G. von Cube, N. A. Irwin and K. H. Dieter. Their advice was invaluable in the formulation of the model and during its testing.

REFERENCES

1. Bolan, R. S., Hansen, W. B., Irwin, N. A., and Dieter, K. H. Planning Applications of a Simulation Model. Paper presented to New England Section of Regional Science Assoc., Oct. 1963.
2. Review of Existing Land Use Forecasting Techniques. Special Report prepared by Traffic Research Corp. for Boston Regional Planning Project.
3. Reliability Test Report: Empiric Land Use Forecasting Model. Report prepared by Traffic Research Corp. for Boston Regional Planning Project.