Technique for Relating Transportation Improvements and Urban Development Patterns

DANIEL BRAND, Peat, Marwick, Livingston and Co.;
BRIAN BARBER, Metropolitan Area Planning Council (Boston); and
MICHAEL JACOBS, Peat, Marwick, Livingston and Co.

•THIS paper describes a land-use forecasting model which embodies desirable features not heretofore available for use in planning transportation facilities. In the process of developing and calibrating the model, it was assumed that there are limited controls available at the regional scale for guiding the development of an urban area. One of these controls is the transportation system. It is hypothesized that there exists a partial ability to influence the development of a region by means of the transportation system. This is an ability which the planner should utilize, both for the promotion of a more desirable region in which to live, and for planning the transportation system in a more complete and efficient manner.

In the past, the required plan or forecast of the future pattern of land uses has normally been prepared somewhat independently of the planned transportation facilities. An important missing link in the overall urban plan-making process has been a systematic measurement of the effect that future transportation facilities themselves have in shaping the land-use pattern. This is an effect which generally leads to higher usage of transportation facilities than would otherwise be expected, since transportation facilities often attract land uses which require such facilities. It is, therefore, imperative that the planner and engineer plan transportation facilities to accommodate not only those land-use activities already in place and those expected owing to urban expansion, but also those activities which will be induced by the proposed facilities to redistribute themselves.

In this paper, attention is focused primarily on the information which the calibration of the EMPIRIC model reveals on the relative and absolute effect of transportation and community facility improvements on land development patterns. Secondary attention is focused on some recent results of production forecasts with the model. The (production) EMPIRIC model, to date, has been structured and the equations estimated, for three data sets involving two different urban regions. Production forecasts have been carried out for the two different urban regions for which the model was calibrated.

The remainder of this paper describes (a) the formulation of the EMPIRIC model, (b) the estimation of coefficients for the equations comprising the model, (c) generalized equations reflecting knowledge gained to date with the model on the forces underlying urban development patterns, and (d) some results of forecasting with the EMPIRIC model.

FORMULATION OF THE MODEL

The EMPIRIC land-use forecasting model is a technique, programmed for the computer, which was designed for use in the planning process. It does not apply optimization techniques nor does it restrict freedom of choice; rather, it attempts to make planning a more meaningful procedure by forecasting one important consequence of a set of alternative policies and plans: namely, the future distribution of population, employment and other socioeconomic activities in the region.

Paper sponsored by Committee on Land Use Evaluation.

The model was formulated such that it would satisfy several criteria, some of which were felt to be important theoretical constraints, and others of which were the operational realities of applying the model in the Eastern Massachusetts region. These criteria,[1] which are largely applicable to any North American metropolitan area, were the abilities:

1. To recognize the simultaneous and interacting nature of metropolitan development;
2. To take as direct input, planned changes in the transportation system (both highway and transit);
3. To output important categories of population, employment, and automobile ownership (i. e., the model must provide data for forecasting trip origins, destinations, and modal splits);
4. To provide forecasts for areas sufficiently small to allow meaningful forecasting of trip origins, destinations, and modal splits; and
5. To be applied recursively (in steps) over relatively short time intervals to allow inputting new values of staged construction of facilities (i. e., the model should produce information directly useful for public works programming).

Criteria of a second order were:

1. The model should accept other important non-transportation policy decisions as inputs. In effect, its output should be a systematic estimate of how a region would develop under the influence of regional growth rates and planning policies relative, not only to transportation, but also to utilities, zoning, open space, etc.
2. The model should allow for reasonable budget limits on operating costs of the model.
3. Input and output to the model should be compatible with other needs; e. g., input transportation networks should be the same as those needed for traffic work.

The framework decided on for the EMPIRIC model consists of a set of simultaneous linear regression equations. That is, more than one output variable is contained in a single equation, and the relationships embodied in the model between the input and output variables are linear and additive. The simultaneous nature of the model (the coefficients of the equations are estimated using simultaneous regression techniques) is a major innovation, getting around the problem of having to decide which activities to locate or forecast first.

All variables in the equations are expressed as shares of regional totals, and the model forecasts changes in shares of activities, between base year and forecast year, in each of the zones or subregions into which the region is divided. Mathematically, a change in subregional share may be expressed as

$$\frac{R_{ih}(t)}{\sum_{h - 1}^{H} R_{ih}(t)} - \frac{R_{ih}(t - 1)}{\sum_{h - 1}^{H} R_{ih}(t - 1)}$$

where R_{ih} is the level of activity i in zone h, H is the total number of zones in the region, (t) indicates the forecast year, and (t - 1) indicates the base year.

Data from two points in time are used to calibrate the model. The formulation of the variables enables both growths and declines of activity levels to be easily handled. Having forecasted changes in shares, the model adds these changes to the shares at the beginning of the forecast interval to obtain the new zonal shares, and then multiplies the new shares by regional totals at the end of the forecast interval to obtain the actual activity levels in each traffic zone. The regional totals are forecast independently of

---

[1] This list is similar in many respects to the list of criteria presented by Lathrop and Hamburg (10).

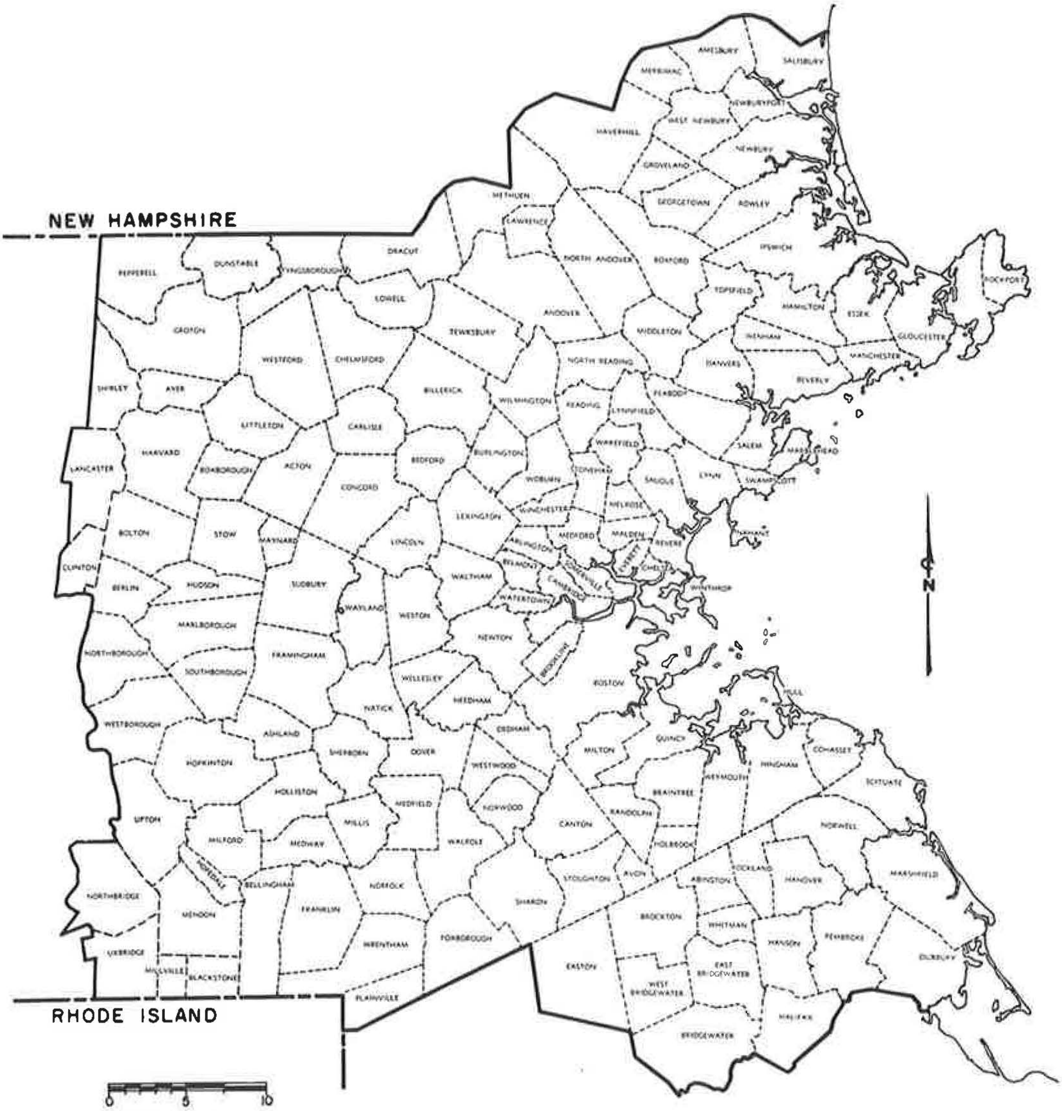

Figure 1. The Eastern Massachusetts region.

the model so that, with this formulation of output variables, the model is strictly a distributional model.

There are two classes of input variables used in the EMPIRIC model: policy variables, and non-policy variables. Policy variables may be manipulated or preset by planners, and therefore they enter the model as terminal or forecast year data. Examples are the transportation system (in the form of accessibilities) and sewage disposal and water supply service levels. Non-policy variables are base-year data, such as families-by-income categories and employment-by-industry categories. Also defined as non-policy variables are various measures of the capacity of a zone to house development of the various types of activities. These, however, could be used as policy variables by reserving land in zones in accordance with recreation and/or open space policies.

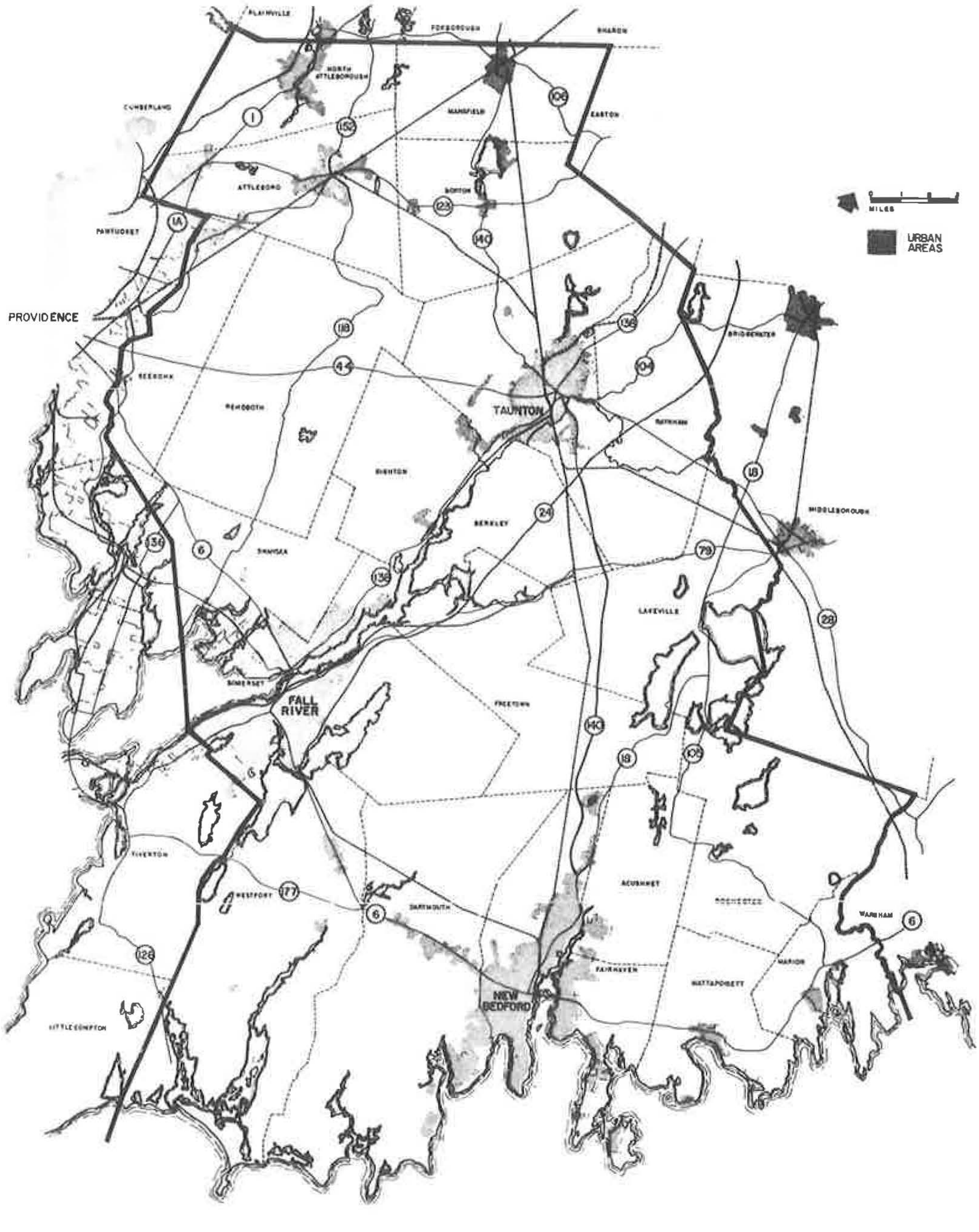

Figure 2. Southeastern Massachusetts planning region.

ESTIMATION OF EMPIRIC MODEL EQUATIONS

To date, the EMPIRIC model has been calibrated for three data sets. Two of these data sets were for the 3.4 million population (in 1960) Eastern Massachusetts region (Fig. 1). The first involved the region disaggregated into 626 traffic zones (i.e., observations for each variable), whereas the second divided the region into 97 subregions. The third data set divided the 400,000 population (in 1964) Southeastern Massachusetts Regional Planning District (Fig. 2) into 71 land-use forecasting districts.

The regression analyses used to estimate the coefficients of the models were preceded by intensive theoretical studies as to the proper structure of the model. These studies, based on a priori reasoning as well as on knowledge gained from prototype EMPIRIC model development work (2) and the literature, resulted in initial or preliminary structuring of each model.

The theoretical studies were augmented by data-analysis techniques programmed as part of the EMPIRIC model, notably factor analysis, which provides insight into the proper grouping of data categories to form model variables that are as independent of one another as possible, and bivariate correlation analysis, which provides insights into the nature and strengths of the correlations or relationships between pairs of model variables. Using these analyses and the theoretical studies, coefficients for several models for each area (data set) were estimated, each successive model showing improvement over the preceding one. The improvements exhibited were not so much in the "goodness of fit" of the data, but in the stability, the conformance with theory, and the improved significance levels of the variables in the model. These factors are important criteria which must be met if the model is to be a valid and reliable forecasting tool.

The estimated coefficients for the most disaggregated version of the model will be described in detail here. This version of the model comprises a set of nine simultaneous equations, and forecasts four categories of population and five categories of employment for a set of 626 traffic zones. The categories are:

1. Families with less than \$5, 000 annual income (1959 dollars);
2. Families with between \$5, 000 and \$9, 999 annual income;
3. Families with between \$10, 000 and \$14, 999 annual income;
4. Families with greater than \$14, 999 annual income;
5. Manufacturing and construction employment (Standard Industrial Classification codes 15-39);
6. Wholesale, transportation, communication, utilities, government, and other employment (SIC codes 01-14, 40-50, 91-99);
7. Retail employment (SIC codes 52-59);
8. Service employment (SIC codes 70-89); and
9. Finance, insurance, and real estate employment (SIC codes 60-67).

The estimated equations in this calibrated model are described in detail in the Appendix.

Data for two points in time (1950 and 1963) were used to calibrate this model. Due to insufficient data for the earlier year, the model was calibrated using data from 453 of the 626 traffic zones (representing about 80 percent of the 1960 regional population of about 3. 4 million persons). Forecasting, however, is being done for all 626 zones.

The statistical significance of each of the variables in the equations of this model is measured with the t-test, which provides an index of the degree to which the effect of a variable upon an output variable is either random or systematic. The t values were computed, for all input and output variables, by the same programs which applied the regression techniques for the estimation of the coefficients in the equations. For 453 sampling points (i. e., traffic zones), a t value of 1. 96 or greater is indicative of a variable which is significant to the 95 percent confidence level—a level which is felt to be a very stringent test of significance. Fifty-one of the 63 input and output variables in the nine equations of the model met this high standard. Of the other 12, ten are significant to a level of confidence of 75 percent or greater (the two exceptions being significant at the 56 percent and 58 percent levels).

The few variables which were significant to less than the 95 percent confidence level were still felt, therefore, to be statistically acceptable, and were retained in the final model structure because they, along with the other variables in the model, had regression coefficients whose signs and relative magnitudes satisfactorily expressed the hypothesized relationships between the variables.

An additional test of the model was its "goodness of fit" over the calibration period. That is, an indication of the model's reliability as a forecasting tool was obtained by using the calibrated model and the calibration base year (1950) data, and "forecasting" to the terminal year of the calibration period (1963) to see how well the model reproduced

TABLE 1

STATISTICAL SUMMARIES OF OBSERVED VS CALCULATED POPULATION AND EMPLOYMENT LEVELS

Category	453 Zones			104 Districts		
	RMS Error	RMS Error Ratio	R^2	RMS Error	RMS Error Ratio	R^2
Families, < \$5, 000	108	0. 249	0. 951	232	0. 123	0. 990
Families, \$5, 000-\$9, 999	209	0. 269	0. 906	685	0. 203	0. 950
Families, \$10, 000-\$14, 999	82	0. 380	0. 793	233	0. 250	0. 915
Families, ≥ \$15, 000	61	0. 578	0. 826	150	0. 328	0. 946
Mfg and construction employment	1, 031	1. 23	0. 549	2, 301	0. 630	0. 862
Wholesale, TCU[a] Govt., and other employment	412	0. 782	0. 876	969	0. 422	0. 982
Retail employment	310	0. 781	0. 860	846	0. 490	0. 949
Service employment	677	1. 43	0. 500	1, 958	0. 949	0. 880
FIR[b] employment	224	1. 33	0. 953	260	0. 352	0. 997

[a] Transportation, communication and utilities.
[b] Finance, insurance, and real estate.

the activity growths occurring during the calibration interval. Statistical summaries were then prepared comparing observed and calculated (forecast) 1963 zonal values of the output variables.

These summaries include the root-mean-square (RMS) error, the RMS error ratio, and the coefficient of determination (R^2). The RMS error is computed in the following manner:

$$\text{RMS error} = \sqrt{\frac{\sum_{h=1}^{H} \left(O_{ih} - C_{ih}\right)^2}{H}}$$

where O_{ih} is the observed value of variable i in zone h, C_{ih} is the calculated value of variable i in zone h, and H is the total number of zones in the region. Assuming normality, the observed value does not differ from the calculated value for about 67 percent of the zones by more than plus or minus the RMS error. The RMS error ratio is the ratio of the RMS error to the mean or arithmetic average of the observed output variables ($\bar{O}_i$).

The coefficient of determination (R^2) is computed as follows:

$$R^2 = \frac{\sum_{h=1}^{H} \left(O_{ih} - \bar{O}_i\right)^2 - \sum_{h=1}^{H} \left(O_{ih} - C_{ih}\right)^2}{\sum_{h=1}^{H} \left(O_{ih} - \bar{O}_i\right)^2}$$

As R^2 approaches unity, the reliability of the model is regarded to be quite high, and conversely, as R^2 approaches zero, the reliability is said to be quite low. These summaries, for the 453 traffic zones in the calibration area, are given in Table 1 for the nine equations in the final calibrated model. In addition, the corresponding statistics have been recomputed following the aggregation of the 453 traffic zones into 104 calibration analysis districts. This procedure was designed to provide some indication of the sensitivity of these reliability statistics to zonal aggregation.

It can be seen that the model fits the population data better than the employment data. This is to be expected, since a statistical model fits large numbers of small locating units (e. g., households) better than the "lumpier" activities which typify the employment locating units. The fit to the geographically small 453 zones appears highly satisfactory, and compares favorably with similar error measures calculated for home interview survey origin-destination data, and for various types of traffic models, e.g., gravity models (11). In addition, the model in the Appendix appears quite sound from the standpoints of statistical significance (high t values), and logic (conformance with hypothesized relationships).

GENERALIZED LAND-USE FORECASTING EQUATIONS

The true measure of the EMPIRIC model's worth as a forecasting and plan-making tool is in the empirical and logical reliability of the regression coefficients. Because the variables are formulated as zonal shares or changes in zonal shares, these coefficients may be interpreted as indicators of the relative effects of the variables in influencing relative growths or declines of an output variable at the zonal level. The sign of the coefficient (positive or negative) indicates whether the variable induces or hinders the growth in zonal share of the output variable, while the magnitude of the coefficient indicates the importance of this influence relative to the influence of the other variables in the equation on the growths of the output variables. Coefficient stability, therefore, becomes an important indicator of the success achieved in producing true relationships in the model; relationships from which one may learn about influencing the shape of metropolitan development, and, consequently, the usage of transportation facilities.

Two types of coefficient stability may be described. The first is coefficient stability as successive model structures are estimated in the model calibration process using a single data set. The coefficients in the model described in the Appendix behaved extremely well in this regard over the successive equation estimations (8). In the few instances when coefficients in the final model exhibited appreciable changes from the corresponding coefficients in earlier models, it was almost always attributable to a problem of collinearity between independent variables in the earlier models. The situation was remedied by the deletion in the final model of all but one of the related independent variables, or by the substitution of a single variable for the complete set of collinear independent variables.

The second type of coefficient stability pertains to the similarity of the relationships expressed by the coefficients, as different data sets for the same region or for different regions are used to estimate the same structural equations. The three calibrations of the EMPIRIC model just described did not use the same structural equations because of the purposes for which the models were developed, and because of the differences in the data available for calibration. It is hoped that future work will allow the estimation of the same EMPIRIC model structural equations for different data sets.

Nevertheless, the three models all distributed classes of population and employment to relatively large numbers of small areas. And the types of independent variables used in each model were similar. The results indicate that there is enough coefficient similarity between corresponding input and output variables for the differing data sets and areas to warrant an attempt to generalize the results of the three models. In recording these results, it is recognized that there should indeed be different relationships between variables with differing zone sizes. Also, different urban areas have different regional growth rates and different compositions of activities comprising the urban development pattern. In fact, if the coefficient set were completely stable it would not be necessary to recalibrate the model for different zone systems and areas.

The generalized equations are written out completely below. The following notation is employed:

(Δ) = change in subregional share over the time interval
(t) = subregional share at the end of the time interval
$(t - 1)$ = subregional share at the beginning of the time interval

(All variables are formulated as shares or as changes in shares.)

POPL, POPM, and POPU = lower-, middle-, and upper-income population
MFG, RTL, SVC, and OTH = manufacturing, retail, service, and other employment
UTIL = measure of utilities service

CAPP, CAPM, and CAPR = measures of the capacity and propensity of a zone to house new population (i. e., residential), manufacturing, and retail development (the measures are defined in the Appendix)

VACC and QACC = measures of vehicle (automobile) and transit accessibility (accessibility is defined in the Appendix)

The magnitudes of the coefficients are indicated as s, m, or b—small, medium, and big (<0.1, 0.1 to 0.4, and >0.4).

The equations follow:

$$(\Delta)\,\text{POPL} = b\,(\Delta)\,\text{POPM} - m\,(\Delta)\,\text{POPU} + s\,(\Delta)\,\text{SVC} + m\,(t-1)\,\text{POPL} - s\,(t-1)\,\text{POPU} + s\,(t)\,\text{UTIL} - m\,(\Delta)\,\text{VACC}$$

$$(\Delta)\,\text{POPM} = -\,s\,(\Delta)\,\text{POPL} + m\,(\Delta)\,\text{POPU} + s\,(\Delta)\,\text{RTL} + s\,(\Delta)\,\text{SVC} - m\,(t-1)\,\text{POPM} + m\,(t)\,\text{UTIL} + m\,(\Delta)\,\text{VACC} + s\,(\Delta)\,\text{QACC}$$

$$(\Delta)\,\text{POPU} = -\,m\,(\Delta)\,\text{POPL} + m\,(\Delta)\,\text{POPM} - m\,(t-1)\,\text{POPU} + m\,(t)\,\text{UTIL} + s\,(t-1)\,\text{CAPP} - m\,(\Delta)\,\text{VACC} + s\,(\Delta)\,\text{QACC}$$

$$(\Delta)\,\text{MFG} = -\,m\,(\Delta)\,\text{POPM} - b\,(\Delta)\,\text{POPU} + m\,(\Delta)\,\text{OTH} - b\,(t-1)\,\text{MFG} + m\,(\Delta)\,\text{CAPM} + m\,(t)\,\text{VACC} + m\,(\Delta)\,\text{QACC}$$

$$(\Delta)\,\text{RTL} = m\,(\Delta)\,\text{OTH} - s\,(t-1)\,\text{POPU} - m\,(t-1)\,\text{RTL} + m\,(t-1)\,\text{CAPR} + m\,(\Delta)\,\text{VACC}$$

$$(\Delta)\,\text{SVC} = -\,s\,(\Delta)\,\text{OTH} - m\,(t-1)\,\text{SVC} + m\,(\Delta)\,\text{UTIL} + m\,(t)\,\text{VACC} + m\,(\Delta)\,\text{QACC}$$

$$(\Delta)\,\text{OTH} = m\,(\Delta)\,\text{MFG} + s\,(\Delta)\,\text{RTL} - m\,(t-1)\,\text{OTH} + s\,(\Delta)\,\text{QACC}$$

It must be reemphasized that these equations are for discussion and theory building purposes only, and are abstracted from only three models calibrated for two areas: the relatively slow-growing Eastern and Southeastern Massachusetts regions.

The equations generalize the interrelationships among activities in this type of area for this scale of zonal disaggregation (i. e., for an average zonal population of from about 5,000 to about 35,000), and for this type of model (linear and share). Each of the equations describes hypothesized relationships designed to explain the growth of a particular output activity. For example, the first equation states that growth of lower-income population in a zone is induced by a simultaneous growth of middle-income population but hindered by the growth of and presence of (at the beginning of the time interval) upper-income population. It is also induced by the simultaneous growth of service employment, by the presence at the beginning of the time interval of low-income population (the ghetto effect), and by the presence at the end of the time interval of utilities services. It is hindered by the (relative) growth of vehicle accessibility of the zone (since they compete for more accessible land with higher-income groups, as explained later).

An examination of the equations indicates that the accessibility variables are the most important of the policy variables for forecasting the location of population and employment. However, the non-policy variables, over which the planner has no direct control, are generally stronger determinants of locational patterns than are the policy variables. In particular, growths in the various population-by-income groupings are strongly related to growths in the adjacent population-by-income groupings. It is also observed that in the employment equations among the strongest variables are one or more of the other output variables. These observations provide evidence of the realism of this type of simultaneous model.

In all equations, one of the more important determinants of growth is the "lagged" variable, i. e., the value of the output variable at the beginning of the forecast interval. In every instance but one, the lagged variable carries a medium or large negative sign. The single exception is important in that it is in the (first) equation for the low-income population. In only that instance does the presence of the (same) activity at the beginning of the time interval induce increased growth in the zone in the regional share of the activity. This is striking statistical evidence of the increasing ghettoism of the low-income family, about which there is much discussion today.

Many of the coefficients capture other relationships worthy of examination. In the low-income population equation again, the medium-sized negative coefficient modifying growth in vehicle accessibility indicates that these low-income families do not have the resources to take their full share of the advantages of improvements in the regional highway system. However, it may also be noted that the highest income group (in the third equation) exhibits the same medium-sized negative sign for this variable. This appears to indicate that they would rather pay increased transportation costs to enjoy the other residential amenities which they desire. The very large middle-income group, on the other hand (in the second equation), exhibits the concern for improved highways with which we are familiar.

It is also of interest to note that the middle- and high-income groups take advantage in a small but noticeable way of transit improvements, which in this case were rapid transit and commuter railroad service changes.

The position of the accessibility variables as the most influential of the policy variables is especially significant because there seems to be considerably greater control at the regional level over the transportation system than over any of the other policy variables relating to the development and physical arrangement of land patterns. This is partly because most land development policies are determined at the local level by the citizens of the localities affected. Transportation policies, on the other hand, cannot be so isolated at the local level. The function of transportation is to connect places (which may have differing transportation desires), and major transportation policies must be decided on a broader (e. g., regional) level. At best, planners can plan and promote transportation improvements which reinforce development decisions made at the local level.

FORECASTING WITH THE EMPIRIC MODEL

The capabilities of the EMPIRIC land-use forecasting model to manipulate data, to reproduce significant parts of the environment, and to quickly simulate complex relationships between the forces which shape the environment, provide the model with the ability to predict the future distribution of land-use activities with varying sets of input public works policies. This ability is essential for providing information for judging alternative plans, i. e., for determining (a) how well each plan functions, (b) how well each plan achieves its desired set of values, and (c) whether a particular programming (scheduling) strategy has been successful. A means of using the model in conjunction with travel forecasting techniques for evaluating alternative transportation policies and programs is outlined as follows:

1. The model is calibrated (i. e., the equations structured and the coefficients estimated) using data from two historical time points; say, time t and time t + x, where time t is x years earlier than time t + x. (The x-year forecasting interval would normally be about 5 or 10 years.)
2. Estimates of regional growth for an x-year period would be made for each activity to be predicted, and regional forecasts of these activities would be made for time t + 2x.
3. The land-use model would be applied for an x-year forecast from time t + x to time t + 2x. The input data required for forecasting would include base year (time t + x) values of activity levels, and base year and forecast year (time t + 2x) travel times (the latter times being based on the anticipated or proposed completion of new transportation facilities and the closure of old facilities). Also input would be base year and forecast year values of other policy variables, such as utilities service.
4. The traffic model would be applied to forecast for time t + 2x, traffic flows, times and costs, based on the predicted land-use pattern and the travel facilities scheduled for completion at time t + 2x.
5. The procedure outlined in steps 3 and 4 would be repeated if the travel times and costs found in 4 differed substantially from final year values used in 3.
6. The procedures outlined in steps 2 through 5 would be repeated for successive x-year periods, using activity levels estimated by the land-use model at the end of each

TABLE 2

FUTURE REGIONAL CONTROL TOTALS (IN THOUSANDS)

Year	Population	Mfg Employment	Non-Mfg Employment
1963	3, 540. 5	426. 8	870. 0
1975	3, 924. 0	433. 5	1, 073. 2
1990	4, 733. 0	478. 7	1, 322. 4

period as starting levels for forecasting the next period. This step would be continued until the final target year had been reached.

This process thus provides a systematic representation of the anticipated sequential stages of development of a region under the influence of a set of public policies relating to the transportation system, utilities service, etc. Repeating the process for different sets of policies will produce different anticipated development patterns. The planners and decision-makers can study these various patterns, analyze their relative merits and costs, and can more knowledgeably make decisions as to which sets of policies will be most effective in furthering the social and economic goals of the region. Especially valuable would be the exploration of alternative public works programming strategies. This process allows the program to be developed as an integral part of, and at the same time as, the overall plan.

At this writing, the EMPIRIC model has been used to make four sets of production forecasts at the 97 subregion level for the Eastern Massachusetts region and one set of forecasts at the 71 district level for the Southeastern Massachusetts region.

SUBREGION FORECAST RESULTS

For the purpose of exploring patterns of urban growth which are considered feasible for the future development of the Eastern Massachusetts region, an application of the 97 subregion EMPIRIC model was made. Forecasts were prepared for four regional growth alternatives. Each alternative pursued different basic physical objectives for structuring future urban growth. The alternatives are called (a) the composite plan, (b) the radial corridor plan, (c) the spread city plan, and (d) the nucleated plan. In this application the EMPIRIC model is viewed as a design tool; i. e., the designer is able to determine the consequences of selected programs. This in turn enables him to choose which program (set and schedule of actions) to propose for implementation or to subject to more detailed analysis.

Values for each of the policy variables were altered in this model application. For each of the four plans, appropriate "test" future highway, transit, water and sewer networks were designed. There were differences between the test networks only for the period 1975 to 1990 owing to the region's strong commitment to the 1975 programs for highway and transit networks. Identical regional "control" totals of population and employment were used for each plan. These are listed in Table 2.

Forecast results for 1990 showed an average difference between the highest and lowest subregional values among the four plans of 9 percent for population, 42 percent for manufacturing employment and 13 percent for non-manufacturing employment. The range of differences between the high and low 1990 forecasts was 1 percent to 46 percent for population, with 14 of the 97 subregions having differences over 15 percent. The corresponding figures for manufacturing employment were from 2 percent to 500 percent with 11 subregions over 50 percent, and for non-manufacturing employment, from 1 percent to 89 percent with 7 subregions over 30 percent. However, certain patterns are common to each of the four forecasts. First, the regional core area continues to decline, although at a slower rate than during the model calibration period 1950-1963. Second, each geographic sector retains an almost constant share of regional population and employment. Third, change in share by ring is greater than change in share by sector as would be expected (growth is moving outward from the regional center or core).

By identifying and comparing subregions in which only the highway or transit network input data have been changed, it is possible to measure the impact of transportation facilities. It appears in some cases that good highway connections will result in 10 to 15 percent more population than poorer highway connections. Similar observations are possible with respect to employment. Many such observations would have to be made and investigated before any verified generalizations could be made. Sufficient differences existed between plans to warrent exploration of alternatives at a more detailed level (i. e., with the 626 traffic zone EMPIRIC model).

That sufficient differences occurred was not surprising. The hypothesis that the design of the transportation system plays a large and important role in the shaping of metropolitan development was borne out by a test carried out with a prototype version of the EMPIRIC model (11). This model was used to simulate the effect on the locational pattern of population and employment in the Eastern Massachusetts region of 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 region between 1950 and 1960 insofar as highway and mass transportation improvements or closures were concerned. The second simulated design policy was that no changes were made in the highway and mass transportation systems between 1950 and 1960.

The major transportation improvements consisted of radial expressway sections plus Route 128, a major circumferential expressway which passes through a tier of suburban communities. The simulated policy of transportation improvements resulted in expected increases in population and employment in the third and fourth tiers or rings of subregions, relative to results with the simulated policy of no transportation improvements. However, it is interesting to note that relative increases in population and employment were also obtained in the older core cities of Boston, Cambridge, and Somerville, due to the new radial expressways and the extension of the rapid transit system to the periphery of Newton (i. e., to Route 128).

CONCLUSIONS

From the results obtained thus far with the three versions of the EMPIRIC land-use forecasting model, several observations may be drawn. The model has in each instance been satisfactorily calibrated in terms of logical relationships expressed by the variables and their coefficients (i. e., conformance with hypotheses), high statistical significance (as measured with t values), good fit with the data, and stability of the coefficients within each model (as observed by tracing variables through the successively estimated models).

The model thus far has been successfully used for forecasting to relatively large numbers of zones in two instances: (a) with the 97 subregion version calibrated for the (Boston) Metropolitan Area Planning Council, and (b) with the 71 district version calibrated for the Southeastern Massachusetts Regional Planning District. It is expected that the model will be able to successfully forecast to very large numbers of zones, as will be soon tested when forecasts are made using the 626 zone version of the model calibrated for the Eastern Massachusetts Regional Planning Project.

In addition, it appears that the model is properly sensitive to varying public policy inputs. The four sets of forecasts produced with the 97 subregion version of the model were intended to reflect widely ranging transportation policies, and the results displayed substantial and logical differences in the forecast values of population and employment. It is felt that this is in large part due to the fact that the model deals primarily with growths of activites rather than with absolute levels of activities at one point in time.

While these substantial findings have been made from the research and development work completed to date, further research into and with the EMPIRIC model would be useful. Moreover, future calibrations and applications of the model are warranted. Such calibrations and applications, with data from other metropolitan areas, would contribute to a better understanding of land-use development patterns in urban areas.

There are at least five major areas of research. First, the questions of coefficient stability could be investigated. Second, possibilities for designing optimal sets of inputs (e. g., accessibility variables) to produce desired plans could be undertaken through mathematical reformulation of the model. Third, the potential for developing programs for public investment using the staging capabilities of the model could be investigated. Fourth, further application of the model as a design tool is worth exploring. Fifth, the possibility of joining the EMPIRIC computer programming system to a plan evaluation system should be investigated. Such a joint or tandem system would be extremely desirable since it would increase our capacity for exploration of alternative policies and programs.

Finally, a more intensive analysis of the forecast results produced by the (Boston) Metropolitan Area Planning Council may yield more support for generalizations of the type attempted in this paper.

REFERENCES

1. Hill, Donald M., and Brand, Daniel. Methodology for Developing Activity Distribution Models by Linear Regression Analysis. Highway Research Record 126, pp. 66-78, 1966.
2. Hill, Donald M., Brand, Daniel, and Hansen, Willard B. Prototype Development of a Statistical Land-Use Prediction Model for Greater Boston Region. Highway Research Record 114, pp. 51-70, 1965.
3. Irwin, Neal A., and Brand, Daniel. Planning and Forecasting Metropolitan Development. Traffic Quarterly, Oct. 1965.
4. Bolan, Richard S., Hansen, Willard B., Irwin, Neal A., and Dieter, Karl H. Planning Applications of a Simulation Model. Paper presented to the New England Section of the Regional Science Association, Boston College, Oct. 1963.
5. Hill, Donald M. A Growth Allocation Model for the Boston Region. Jour. Amer. Inst. of Planners, Vol. 31, No. 2, pp. 111-120, May 1965.
6. Seidman, David R. The Present and Futures of Urban Land Use Models. Paper presented to Fourth Annual Conf. on Planning Information Systems and Programs, Univ. of Calif. at Berkeley, Aug. 1966.
7. Calibration Report: EMPIRIC Land Use Forecasting Model, 97 Subregion Version. Report presented by Traffic Research Corporation to the Commonwealth of Massachusetts, Metropolitan Area Planning Council, Nov. 1966.
8. Calibration Report: EMPIRIC Land Use Forecasting Model for 626 Traffic Zones. Report presented by Traffic Research Corporation to the Eastern Massachusetts Regional Planning Project, Dec. 1966.
9. Calibration Report: EMPIRIC Land Use Forecasting Model for 71 Forecasting Districts. Report presented by Traffic Research Corporation to the Southeastern Massachusetts Regional Planning District, Jan. 1967.
10. Lathrop, George T., and Hamburg, John R. An Opportunity-Accessibility Model for Allocating Regional Growth. Jour. Amer. Inst. of Planners, May 1965.
11. Reliability Test Report: [Prototype] EMPIRIC Land Use Forecasting Model. Presented by Traffic Research Corporation to the Eastern Massachusetts Regional Planning Project, Jan. 1964.

Appendix

THE 626 ZONE EASTERN MASSACHUSETTS EMPIRIC MODEL

The following variables are used in the model:

Population variables (All income figures are given in terms of 1959 dollars.)

$F_{<5k}$ = Number of families with an annual income less than $5,000.
F_{5-10k} = Number of families with an annual income between $5,000 and $9,999.
F_{10-15k} = Number of families with an annual income between $10,000 and $14,999.
$F_{\geq 15k}$ = Number of families with an annual income equal to, or greater than, $15,000.

Employment variables (All employment variables are measured at the zone of employment.)

M&C = Manufacturing and construction employment (SIC codes 15-39).
Other = Wholesale, transportation, communication, utilities, government and other employment (SIC codes 1-14 , 40-50, 91-99).
Ret = Retail employment (SIC codes 52-59).
Svc = Service employment (SIC codes 70-89).
FIR = Finance, insurance, and real estate employment (SIC codes 60-67).

Land developability variables (The nomenclature used to define these variables includes: NAP = net residential area; NAM = net manufacturing area; NAR = net retail area; UA = total used area of a zone = NAP + NAM + NAR + other developed area; and GA = gross area = UA + developable area.)

CI Pop = Capacity or land developability index for population = (NAP/GA) (GA-UA).
CI Mfg = Capacity or land developability index for manufacturing = (NAM/GA) (GA-UA).
CI Ret = Capacity or land developability index for retail = (NAR/GA) (GA-UA).

Utilities service variables

Water = Index, from 1 through 7, of water supply service, multiplied by UA.
Sewer = Index, from 1 through 5, of sewage disposal service, multiplied by UA.

Accessibility variables (The accessibility of zone g to activity i is equal to

$$\sum_{h=1}^{H} R_{ih} e^{-\beta t_{gh}}$$

where R_{ih} is the quantity of activity i in zone h, H is the total number of zones, e is the base of natural logarithms, t_{gh} is the travel time between zones g and h, and β (the beta factor) is an empirically derived factor. All accessibilities were then multiplied by UA for use in the model.)

VaccTF = Vehicle accessibility of a zone to total families.
QaccTF = Transit accessibility of a zone to total families.
$\mathrm{VaccF}_{\geq 10}$ = Vehicle accessibility of a zone to total families with an annual income equal to, or greater than, \$10,000 (1959 dollars).
$\mathrm{QaccF}_{<10}$ = Transit accessibility of a zone to families with an annual income less than \$10,000 (1959 dollars).
VaccTE = Vehicle accessibility of a zone to total employment.
QaccTE = Transit accessibility of a zone to total employment.
VaccM&C = Vehicle accessibility of a zone to manufacturing and construction employment.
VaccR&S = Vehicle accessibility of a zone to retail and service employment.

Variables measured at the forecast year are preceded by (t). Variables measured at the base year are preceded by (t - 1). Variables representing changes between the base year and forecast year are preceded by Δ. All (t) and (t - 1) variables are formulated as subregional shares. The "Δ" variables are formulated as changes in subregional shares. The number in parentheses following the accessibility variables indicates the value of the beta factor used for the calculation of that accessibility. The model, then, is comprised of the following equations:

Equation 1: $\Delta F_{<5k} = 0.637 \Delta F_{5-10k} - 0.295 \Delta F_{10-15k} + 0.018 \Delta \mathrm{Svc} + 0.133\,(t-1)\,F_{<5k} - 0.109\,(t-1)\,F_{10-15k} + 0.044\,(t-1)\,\mathrm{Water} - 0.298\,\Delta \mathrm{VaccTE}\,(0.05) - 0.068\,(t-1)\,\mathrm{VaccTE}\,(0.15)$

Equation 2: $\Delta F_{5-10k} = 0.530\,\Delta F_{<5k} + 0.337\,\Delta F_{10-15k} + 0.022\,\Delta \text{Ret} + 0.060\,\Delta \text{Svc} - 0.101\,(t-1)\,F_{5-10k} + 0.036\,(t-1)\,\text{Svc} + 0.044\,(t)\,\text{Sewer} + 0.025\,(t-1)\,\text{CI Pop} + 0.302\,\Delta \text{VaccTE}\,(0.05) + 0.114\,\Delta \text{QaccTE}\,(0.005)$

Equation 3: $\Delta F_{10-15k} = -0.125\,\Delta F_{<5k} + 0.637\,\Delta F_{5-10k} + 0.294\,\Delta F_{\geq 15k} - 0.224\,(t-1)\,F_{10-15k} + 0.196\,(t-1)\,\text{Sewer} + 0.145\,\Delta \text{Sewer}$

Equation 4: $\Delta F_{\geq 15k} = -0.282\,\Delta F_{5-10k} + 0.603\,\Delta F_{10-15k} - 0.278\,(t-1)\,F_{\geq 15k} + 0.145\,(t-1)\,\text{Water} + 0.118\,(t-1)\,\text{Sewer} + 0.046\,(t-1)\,\text{CI Pop} - 0.384\,\Delta \text{VaccF}_{\geq 10}\,(0.15) + 0.093\,\Delta \text{QaccTE}\,(0.15)$

Equation 5: $\Delta \text{M\&C} = 0.220\,\Delta \text{Other} - 0.302\,(t-1)\,\text{M\&C} - 0.015\,(t-1)\,\text{FIR} + 0.138\,(t-1)\,\text{CI Mfg} + 0.278\,\Delta \text{QaccF}_{<10}\,(0.05) + 0.121\,(t-1)\,\text{VaccTF}\,(0.05)$

Equation 6: $\Delta \text{Other} = 0.456\,\Delta \text{M\&C} + 0.081\,\Delta \text{Ret} - 0.132\,\Delta \text{FIR} + 0.106\,(t-1)\,\text{M\&C} - 0.194\,(t-1)\,\text{Other} - 0.414\,\Delta \text{VaccTE}\,(0.15) + 0.095\,(t-1)\,\text{QaccTF}\,(0.05)$

Equation 7: $\Delta \text{Ret} = 0.440\,\Delta \text{Other} - 0.117\,(t-1)\,F_{\geq 15k} + 0.126\,(t-1)\,\text{Other} - 0.363\,(t-1)\,\text{Ret} + 0.165\,(t-1)\,\text{CI Ret} + 0.213\,\Delta \text{VaccTF}\,(0.15) - 0.064\,(t-1)\,\text{QaccTF}\,(0.05)$

Equation 8: $\Delta \text{Svc} = -0.252\,\Delta \text{Other} - 0.510\,(t-1)\,\text{Svc} + 0.022\,(t-1)\,\text{FIR} + 0.620\,\Delta \text{Water} + 0.240\,\Delta \text{Sewer} + 0.564\,\Delta \text{QaccTF}\,(0.05) + 0.390\,(t-1)\,\text{VaccTF}\,(0.05)$

Equation 9: $\Delta \text{FIR} = 0.614\,\Delta \text{Other} + 0.020\,(t-1)\,\text{Svc} - 0.159\,(t-1)\,\text{FIR} + 0.110\,(t-1)\,\text{QaccTF}\,(0.05)$

THE 626 ZONE EASTERN MASSACHUSETTS EMPIRIC SUB-MODEL

In addition to the nine output variables contained in the model, there were four additional variables for which forecasts were desired: total population (Pop); automobile ownership (Autos); school enrollment in grades K-8 (School, K-8); and school enrollment in grades 9-12 (School, 9-12). These variables were not included in the main model owing either to their being highly correlated with other output variables, or to suitable data being available for only one of the calibration time points.

These variables, consequently, were incorporated into a sub-model which was calibrated using data from only one point in time (1963). The equations comprising the sub-model are written out below. The notation is the same as that used earlier for describing the main model structure, with the additional variables TF (total number of families) and Med FI (median family income in terms of 1959 dollars multiplied by TF).

Equation 1: $(t)\,\text{Pop} = 0.944\,(t)\,\text{TF} + 0.016\,(t)\,\text{Water} + 0.034\,(t)\,\text{QaccTE}\,(0.15)$

Equation 2: (t) Autos = 0.871 (t) Med FI + 0.164 (t) Water - 0.042 (t) QaccTF (0.15)

Equation 3: (t) School, K-8 = 0.918 (t) TF + 0.154 (t) Water - 0.065 (t) QaccTF (0.15)

Equation 4: (t) School, 9-12 = 0.874 (t) TF + 0.095 (t) Sewer + 0.037 (t) QaccTF (0.15)

The sub-model is forecast following forecasts with the main model. These latter forecasts are used to derive (t) TF and (t) Med FI for use in the sub-model. The other input variables required for sub-model forecasting (utility service and accessibilities) represent policy variables.

The reliability check performed on the sub-model (i.e., the comparison of observed with "forecast" 1963 values) yielded the following results:

Category	453 Zones			104 Districts		
	RMS Error	RMS Error Ratio	R^2	RMS Error	RMS Error Ratio	R^2
Total population	643	0.104	0.984	2,477	0.092	0.991
Automobile ownership	410	0.229	0.915	1,179	0.151	0.963
School enrollment, K-8	211	0.220	0.929	632	0.151	0.969
School enrollment, 9-12	71	0.211	0.939	239	0.164	0.966