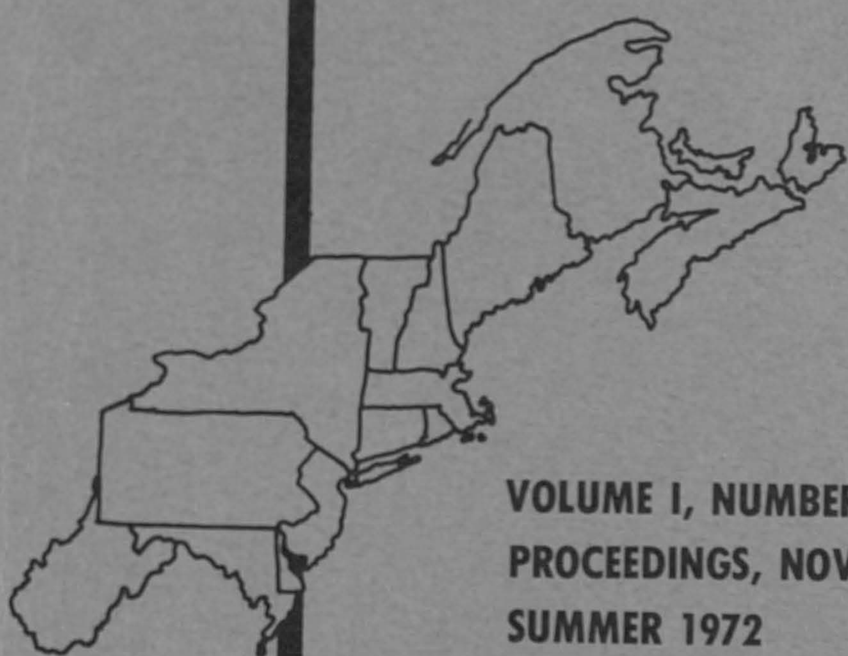
# JOURNAL OF

# Northeastern Agricultural Economics Council



VOLUME I, NUMBER I
PROCEEDINGS, NOVA SCOTIA
SUMMER 1972

# ECONOMIC, ENVIRONMENTAL, AND LAND-USE PLANNING UTILIZING THE EMPIRIC MODEL

John W. Green
Regional Economist
Economic Research Service, USDA
Upper Darby, Pa.

and

Ronald J. Glass
Resource Economist
Economic Research Service, USDA
Durham, New Hampshire

## Introduction

The New England River Basins Commission (NERBC) has placed first priority on implementation of its responsibilities for comprehensive planning. Its enabling legislation resulted in a resolution specifying that NERBC participate vigorously in the prosecution of a framework study for the region (The North Atlantic Study), and shall initiate promptly comprehensive studies of major subregions as elements of the comprehensive coordinated joint plan. The Southeastern New England (SENE) is the first such major subregion to be studied.

Briefly stated, it is the purpose of the SENE Comprehensive Study to identify and recommend actions to be taken by all levels of government and by private interests to secure for the people of the region the full range of uses and benefits which may be provided by balanced conservation and development of water and related land resources. The study is intended to lead to action by private, local, state and federal interests to solve specific problems and realize specific opportunities.

Economic Research Service (ERS) is a member of the United States Department of Agriculture (USDA) team and is Chairman of the Social, Environmental and Economic Framework Study element. ERS will also make major inputs to study sections concerned with Water Resource Program Elements and Alternatives; Related Land; Land Use Patterns, Allocations and Management; Special Environmental Factors; and Water Supply.

ERS is also participating in the Massachusetts Type IV (MASS IV) Water Resource Study, a cooperative Federal-State study designed to provide the data needed by the State to formulate a coordinated plan to meet water and related land resource needs through 1980, and to indicate potentials to meet the needs through 2020. Since the SENE and MASS IV Study Areas overlap extensively, it would be advantageous to utilize the same or similar methodology to formulate planning alternatives.

In carrying out its responsibilities in these studies, ERS has chosen to utilize the EMPIRIC Activity Allocation Model. The EMPIRIC Model is one of a family of regional planning models designed to project population, employment, and land use growth for smaller subregions or districts based on specified regional constraints.

The Model is designed to perform three specific functions:

1. To generate small area forecasts of population, employment, and land use based on exogenously specified regional totals, and exogenously specified management policies;

2. To evaluate the impact of alternative public and private management decisions on the future distribution of regional activity;

3. To serve as a mechanism for analyzing, interrelating and coordinating future public management decisions.

The Model was developed originally for the Boston, Massachusetts region, under the auspices of the Eastern Massachusetts Regional Planning Project. It has since been applied successfully in Southeastern Massachusetts; Washington, D.C.; Winnipeg, Canada; and Minneaplis-St. Paul, Minnesota. Further studies using the Model system are currently underway in Denver, Colorado; Washington, D.C.; and Northwestern England.

The Eastern Massachusetts Regional Planning Project was created by the Massachusetts legislature for the purpose of preparing a comprehensive development plan for the Eastern Massachusetts Region. As part of the planning process, it was necessary to produce comprehensive forecasts of land use activities (i.e., population, employment, and other socio-economic activities for which land is used). It was desired that the technique used to produce these land use forecasts be sufficiently sensitive to management inputs so that the technique could be used as a planning tool with which to test alternative sets of public policies, i.e., management sets designed to produce differing future-year patterns of land use.

The EMPIRIC Model is being used in much the same way in the Metropolitan Washington, D.C. area. Alternative projections of future economic activity (employment) within the District are being combined with varying management strategies (transportation networks) to determine future land use settlement patterns in the Metropolitan Region.

The application of the Model in the SENE and MASS IV Studies will also yield future alternative land use patterns. The impact of these settlement patterns on the water resources of the Region can then be determined. This information will facilitate more comprehensive and efficient management of the water resources in each watershed and for the study regions.

General Structure of the Model

The EMPIRIC Model is based on the concept that development patterns of urban activities are interrelated in a systematic manner that provides a reasonable basis for their prediction. These relationships are embodied in a set of linear equations of the form

$$\sum_{j=1}^{N} a_{ij} R_{jh} (\Delta) + \sum_{k=1}^{N} b_{ik} R_{kh} (t-1) + \sum_{k=N+1}^{M} b_{ik} Z_{kh} = 0$$

where $h$ = one of the subregions of districts comprising the study region = 1, 2, ..., H

$i$ = one of the (output) activities to be forecast = 1, 2, ..., j, ..., N

$k$ = one of the variables whose locations and intensities are related to development patterns of the forecast (output) activities in a causal manner = 1, 2, ..., N, ..., M

$R_{jh}(\ \ )$ = the change in the output variable j in district h from the beginning to the end of a forecast interval

$R_{kh}(t-1)$ = the value of the causal variable k (=output variable j) in the district h at the beginning of a forecast interval

$Z_{kh}$ = the value of causal variable k in district h

$a_{ij}$ and $b_{ik}$ = coefficients ($a_{ij}$ = 1 when i = j)

The basic equation thus relates the growth of a single output variable i in district h to (1) the growth of the other output variables j in district h, (2) the present amount of the output variable i in district h, and (3) the amount of the causal variables k in district h. Since there are N output variables, there would be N equations of this form comprising the Model.

The coefficients of the equations are estimated using simultaneous multiple linear regression techniques. There are N x N values of the $a_{ij}$ coefficients and N x M values of the $b_{ik}$ coefficients to be estimated; or, a total of N(N+M) coefficients. All $b_{ij}$'s having i=j are set equal to unity, and, for proper identification of the equation system, at least (N-1) of the $a_{ij}$'s and $b_{ik}$'s in each equation are set equal to zero. (The inclusion of all or nearly all of the independent variables in each equation would lead to major problems of coefficient instability, but theoretical and empirical analyses have shown a sufficient number of the $b_{ik}$'s to be statistically insignificant in the equations.)

The equations for the Model are initially structured after analysis of the existing pertinent literature, with theoretical reasoning and the exposition of hypotheses as to the proper interrelationships of economic-environmental activities, and after examination of the results of statistical analyses performed on the data. The parameters of coefficients for the equations are then estimated using regression techniques.

The estimation techniques programmed as part of the EMPIRIC Model provide not only the estimated coefficients, but also measures of fit with the data for each equation (e.g., standard error of the equation, multiple coefficients of determination ($R^2$), and measures of signficance of each of the variables in each equation of the Model (t-values)). For the number of sampling points used to calibrate the Model, a t-value of 2.00 or greater is indicative of a variable which is said to be significant at a level of confidence of 95 percent or greater. This is the significance level strived for in the Model, and it is felt to be an extremely stringent standard, at which the independent variable is said to be exerting a systematic influence on the dependent variables. (A very low t-value, on the other hand, would be indicative of a variable which is probably having a random effect on the dependent variables.)

After the application of the regression techniques, reliability checks are performed on the Model by using the estimated parameters (coefficients) and the base year data for the calibration interval, to "forecast" activity growths to the terminal year of the calibration interval (i.e., with the forecasting program block of the EMPIRIC Model).

The reliability program block of the EMPIRIC Model is then run to compute statistical comparisons between the calculated ("forecast") activity levels and the observed or actual activity levels. Indication is thus provided of the accuracy with which the Model is able to reproduce the growths (and declines) of the output activities occurring in the subdivisions (towns, in this application) during the calibration interval (1960 to 1970).

The results of the regression analyses and reliability checks are then scrutinized. Following this, the equations comprising the Model are restructured on the basis of the logic of the estimated parameters (i.e., their signs and magnitudes), the significance levels of the variables, and the goodness of fit obtained with the data. The process is continually repeated: application of the regression technique to estimate coefficients for the newly specified equations, use of the estimated parameters to perform reliability checks by attempting to recreate activity growths over the calibration interval, analysis of the results, and respecification of the Model structure. The process ends ehen it is felt that further restructuring will not result in appreciable improvements over the last estimated version of the Model.

Once the Model has been calibrated, it is operated in a recursive manner for forecasting purposes. There is one equation for each of the output variables in each district, and the system of equations is solved separately for each district. At full utilization, therefore, the Model comprises N equations per district of the form outlined above. The simultaneous solution for a given forecasting interval will provide growths of district activity levels during this interval.

## Application of the Model

In the SENE and MASS IV studies, it will be necessary to distribute the broad regional OBERS[1/] projections of economic activity to watershed planning areas, which are groups of towns. Since OBERS projections start with national expectations and break them down to functional areas consisting of groups of counties, it is necessary to secure a further subdivision into towns to meet the objectives of these studies. The towns will then be aggregated into the previously discussed watershed planning areas. The Model will be used not only to allocate the subregional shares of economic activity, but also to translate these shares into terms of land use and environmental quality.

Previous applications of the EMPIRIC Model have not considered quantitatively the environment of the areas with which they dealt. In this application of the Model, we intend to do this in three ways. First, the land use projections themselves will be indicative of one type of future environmental quality. The initial planning effort of the Southeastern New England Study lists ten types of land use as critical indicators of environmental quality.

The second strategy involves the incorporation of water and air pollution data in the multiple regression equation system. Data, dealing with point locations of public and industrial water and air polluters and with public waste disposal systems, are available from the Environmental Protection Agency (EPA), the States, and the Corps of Engineers. All these data are computerized for easy access and manipulation, but they are complete only for very recent time frames, i.e., since 1968. This will present problems in our attempts to calibrate the Model for the 1960 to 1970 period. We may be forced to calibrate using data from only one time frame (1970), or to develop an environmental submodel which would be calibrated and run independently, using the results as inputs to the EMPIRIC Model.

The third strategy incorporates environmental constraints. These constraints are being developed by a private consulting firm[2/] and involve environmental holding capacities, unique environmental characteristics, and environmental projections. Each constraint set has been developed for each town in the study area and will be used to limit the Model's allocation activities for projection time frames.

---

1/ A joint project between the Bureau of Economic Analysis, U. S. Department of Commerce and the Economic Research Service, U.S. Department of Agriculture to develop historical and projected measures of population, income, employment, and production for the United States, 50 States, 20 Water Resources Regions, and 206 Water Resources Subareas.

2/ Research Planning and Design Associates, Inc., University of Massachusetts, Amherst, Mass.

Each environmental holding capacity is limited by one of several constraining factors:

1. the amount of developable land left in the town;
2. the amount of developable land that can be developed to criteria maximums;
3. the amount of forest and open land that can be developed before violating criteria minimums;
4. other criteria for farm or forest-towns, i.e.,
   (a) population densities for each town type, or
   (b) population that can be added to developable land without changing developed land density

Other criteria considered in this constraint set include minimum limits on open, agricultural, and forest land. With this constraint set, it is possible to hypothesize the characteristics of land where development should take place, without being point specific.

A unique characteristic constraint set has also been developed and mapped for each town in the study area. The environmental projections constraint set quantifies architects' and planners' concepts of the desirable environment in each town in each future time frame.

Each of these constraint sets can be used to define and limit development alternatives. The alternatives will be used to compare Model outputs and results of other projection methodologies (OBERS, etc.).

## Data Requirements

Before discussing the implications of specific management strategies, a better understanding of the data requirements of the Model is appropriate. The EMPIRIC Model essentially requires two sets of input data. For Model calibration, information is required for two separate points in time on the distribution of activities and the levels of each management variable to be included in the Model. For forecasting, data are required on the base-year distribution of activity, on the base-year level of each management variable, and on the values of each management variable, together with appropriate regional activity totals for each forecast year.

In structuring a model for forecasting the impacts of alternative future development strategies, we need to know the subregional distributions of population (broken down by income, household size, and/or age) and employment (broken down by aggregated SIC codes), as a function of region-wide growth in transportation, public utilities, major developments, zoning, density, and restricted land use policies. The set of required data includes:

i. Subregional population[3/]
ii. Subregional employment[3/]
iii. Subregional land area data[3/]
iv. Subregional transportation management
  Minimum transport network derived from district-level highway and transit networks[4/]
v. Other subregional management data[4/]
  Water service measures
  Sewerage service measures
  Location, size, and character of major commercial, residential, and industrial developments
  Zoning controls; maximum and minimum activity levels and rezoning propensity indices
  Acreages in restricted uses
  Permissible densities of development
vi. Regional and jurisdictional control totals
  External district population and employment estimates for each forecast year
  External transportation service measures -- e.g., travel times to interval networks[5/]
  Regional control totals for all activities, to be allocated by the Model for each forecast year

Most of the data requirements listed above are available from secondary sources in various State and Federal agencies and regional planning commissions. Data related to environmental quality are more difficult to find. However, a number of leads are being pursued that may be fruitful. Data on water and air pollution were found on the STORET System maintained by the Environmental Protection Agency. The System contains point measures of water and air quality in both raw and summarized form, municipal waste treatment inventories and needs, industrial waste treatment inventories and needs, and up-to-date industrial pollution compliance schedules. In cooperation with the Environmental Protection Agency, the Corps of Engineers maintains four-digit industrial data on pollution dischargers who are required under existing law to register. The Environmental Protection Agency information was largely obtained from the individual state agencies.

Using the above data sources, polluters can be identified with towns. The pollution test points can also be located by town. Using correlation analysis with pollution data and data for other socio-economic characteristics, we hope to relate individual residential and industrial location decisions to levels of water and air pollution and to system capabilities for reducing pollution levels.

---

3/ Data required for two points in time for Model calibration (e.g., 1960 and 1970).

4/ Data required for the two calibrated years (e.g., 1960 and 1970) plus each forecast year (e.g., 1980, 1990, 2000, etc.).

5/ See footnote number 4.

If time and money permit, we would like to utilize air and water diffusion models to determine concentrations of individual pollutants in each town. We would also like to relate the pollution impacts to population subgroups in each town. We would be especially interested in any research results definitively relating individual, residential, and industrial location decisions to levels of visable and nonvisable pollution. It might then be possible to determine for each town an index of "desirability" for location of home or factory.

Policy Implications

The main emphasis of the use of the EMPIRIC Model will be directed towards examining the impacts of various proposed management strategies on the geographic location of economic activities, land use patterns, and environmental quality indicators. Proposed development strategies may be introduced into the Model either as constraints or as variables. In other words, certain proposed programs may affect resource distribution by means of an established relationship within the distribution process or by the imposition of minimal or maximal constraints.

In the initial stages of the SENE Study, two alternative futures will be rather superficially defined. The first of these, the Economic Resource Future, will consider the allocation of the OBERS projection among the subregions on the basis of socio-economic and physical variables. The althernative, the Environmental Resource Future, will differ in the following ways:

(1) OBERS regional control projections will be altered to show a smaller share of the Northeast regional projected population in the Southeastern New England area.

(2) Current zoning regulations by towns for commercial-industrial and residential areas will act as constraints on future development.

(3) A population constraint will be set for each town so that it may not exceed the maximum holding capacity permissible to retain a given functional environment. In the Economic Resource Future, the effects of changes in the functional environments of towns must be assessed.

As the study continues, greater emphasis will be placed on the distinction between the effects that proposed strategies will have on the alternative futures. For example, the Environmental Resource Future would include strategies that might promote more public ownership of resources, more stringent zoning regulations, stronger antipollution laws, and tax policies favoring elements of the private sector that would implement measures to improve environmental quality. By contrast, the Economic Resource Future would include strategies designed to promote economic growth, with limited concern for environmental problems.

The particular value of the EMPIRIC Model is that it is capable of simulating many of the complex interrelationships that determine the location of economic activities and their effect on land use patterns and environmental quality. The introduction of management constraints or variables can give us a good indication of the effects of such measures on the patterns of development.

In the MASS IV Study, we will give particular emphasis to the probable aggregate effects of urbanization and industrialization on traditional rural land uses under alternative resource development strategies. We will also examine the effects of various resource policies on maintaining or developing a desirable rural land use mix, particularly as it relates to a quality environment. This approach should also indicate the sacrifices to economic growth that might be expected if environmental considerations are favored in policy development and implementation.

In both the SENE and MASS IV Studies, the development strategies examined will be provided by the various action agencies involved. Once the Model is operative, it will be relatively simple to introduce additional strategies that act as constraints. However, adding management variables can be quite time consuming and costly. For this reason, whether or not we introduce additional management variables in the later stages of the study will depend on the available funding. In any case, the EMPIRIC Model promises to give us a more sophisticated means of assessing management inputs in our New England river basin planning.

Limitations of the Model

Critics of regression analysis point out that (1) linear relationships must exist between the dependent variable and the independent variables, (2) the effects of the independent variables are additive and the independent variables must not be interrelated, and (3) the errors of estimate must be normally distributed with mean zero and constant variance. In view of these restrictions, it is argued that the advantages of regression analysis are soon canceled by the violation of one or more of the above restrictions in using a particular data set.

But the above restrictions are not insurmountable obstacles in the development of a model. If any of the restrictions are violated, the model can be reformulated to avoid such violations, Nonlinear relationships can be linearized by breaking the independent variable into smaller components. It is also possible to use transformations to achieve linear relationships. Factor analysis techniques can create adjusted variables which are independent of one another.

Past applications of the EMPIRIC Model have been critisized for being insensitive to small changes in management strategies. This is certainly true if the geographical area being examined is too large. This should not be a problem at the town level if the management strategies are formulated in sufficient detail.