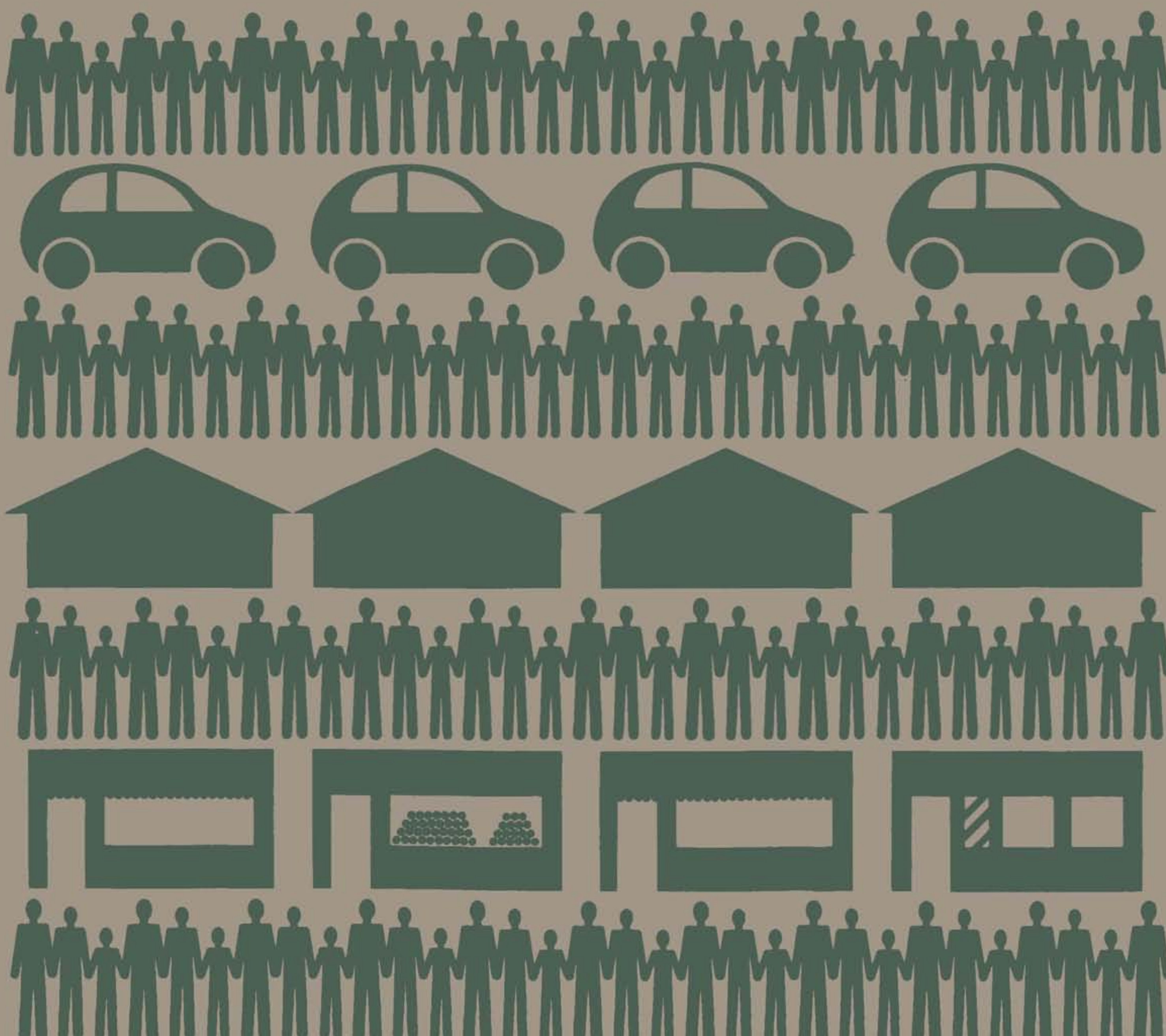
The Community Analysis Model

HUD-000803

U.S. Department of Housing and Urban Development
Office of Policy Development and Research

ORIGINAL

The Community Analysis Model



THE COMMUNITY ANALYSIS MODEL

by

David L. Birch, Principal Investigator
Eric S. Brown
Richard P. Coleman
Dolores W. Da Lomba
William L. Parsons
Linda C. Sharpe
Sheryll A. Weber

Prepared for the office of Policy Development and Research, U.S. Department of Housing and Urban Development, by the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University under contract #2165R.

The research and studies forming the basis for this report were conducted for the Department of Housing and Urban Development (HUD). The sponsoring office was the Office of Policy Development and Research. The statements and conclusions contained herein are those of the contractor and do not necessarily reflect the views of the U.S. Government in general or HUD in particular. Neither the United States nor HUD makes any warranty, expressed or implied, or assumes responsibility for the accuracy or completeness of the information herein.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Stock Number 023-000-00497-6

FOREWORD

The process of neighborhood change is a complex and multifaceted subject which has engaged the attention of scholars from a number of diverse disciplines. Much has been written and a variety of approaches to the subject have been tried. One approach has taken the form of large-scale computer-based mathematical models which attempt to explain the behavior of the various actors who make up urban neighborhoods. The Community Analysis Model is the latest effort in a history of urban spatial model building that covers two decades.

This report is a detailed description of the Community Analysis Model. This model is the product of an ambitious research effort that has been carried out by a group of scholars at the Massachusetts Institute of Technology under the direction of David L. Birch. Professor Birch and his research team have done an excellent job in preparing the report, as well as in developing the model.

This report provides all readers with an interesting and thought-provoking analysis of neighborhood change, and it serves to introduce an important stepping-stone for future urban model builders.

Donna E. Shalala

Donna E. Shalala
Assistant Secretary
for Policy Development
and Research

PREFACE

The Community Analysis Model is a description of a behavioral model of the processes by which neighborhoods change. This model is the product of an ambitious and wide-ranging research effort that has been carried out by a research team at the Massachusetts Institute of Technology. This report and a companion report, also published by the Office of Policy Development and Research, provide a detailed presentation of the structure and uses of the Community Analysis Model. The companion volume, entitled The Behavioral Foundations of Neighborhood Change, summarizes the results of the detailed field work and empirical analyses upon which the theory contained in The Community Analysis Model is based. The latter volume presents an overview of the model, a detailed presentation of the theoretical foundations of the model, and a discussion of the model's accuracy. An independent technical evaluation and summary of the model is provided in another HUD publication entitled A Critical Evaluation of the Community Analysis Model.

The Community Analysis Model is a very large and ambitious model. It simulates the behavior of individuals, households, employers, builders, homeowners, landlords, and zoning board officials as they make decisions regarding where to live, where to work, how many people to hire, what to build, how to maintain a housing unit, whether to alter a zoning ordinance, and so forth. The results of these decisions are accumulated to predict aggregate changes in neighborhoods over time.

The model utilizes a variety of important published data sources, as well as a large amount of data from field surveys in the six metropolitan areas studied. An impressively large data collection and storage effort has gone into the estimation of the model, and the computer method for solving the model is modern and efficient.

The model is primarily the work of non-economists and is multi-disciplinary in approach, as theory and previous research from a variety of academic fields are utilized. This differentiates the Community Analysis Model from most other large scale urban spatial models which rely more exclusively on economic theory.

Table of Contents

Section 1	Introduction	1
Section 2	Overview of the Model Trends in Model Development Relationship of Our Model to Historical Antecedents Tying the Pieces Together Remaining Structural Issues Transferability and Use	3
Section 3	Neighborhood Definition Neighborhood Typologies - A Summary Defining Neighborhood Units for Study Appendix 3-1: Tract Aggregation Procedure	34
Section 4	Natural Increase and Migration Stratification Natural Increase Flows To and From the Region Mapping Between Households and Population Mathematical Statement	49
Section 5	Residential Location Mobility Choice of New Location The Mover Submodel Clearing the Market Mathematical Statement Appendix A: Derivation of Mobility Rates Appendix B: Consistency of Micro and Aggregate Models of Journey-to-work Behavior Appendix C: Probability Function Generator or PFG Appendix D: PFG Equations for 27 Household Types in New Haven	61

Section 6 Dynamics of Local Real Estate Markets 114
Setting Prices on Homes and Apartments
Conversions and Mergers
Denying Loans and Insurance
Mathematical Statement

Section 7 Construction and Demolition of Housing Units 136
New Construction
Exogenous Influences on the Housing Stock
Mathematical Statement

Section 8 Investment in Maintenance of the Housing Stock 179
Deterioration of the Stock
Investment in Maintenance
Abandonment
Simulation of Deterioration, Maintenance, and Abandonment
Mathematical Statement

Section 9 Land Use and Zoning 198
Land Use Restrictions
Land Use Changes
Mathematical Statement

Section 10 Employment Location 216
The Classification of Firms
Location of Jobs or Location of Firms
Level of Geographic Aggregation
Location of Jobs Over Time
Mathematical Statement

Section 11 Schools 253
School Characteristics
School Administrators
Mathematical Statement

Section 12 Validation 284
Measurement Accuracy
Experience with the 1960's
Validation in the 1970's
Validation in the Future
Appendix A: The Creation of Land Use Maps and Tabulations from Landsat Imagery

Section 13 Conclusion 322

Section 1

INTRODUCTION

Neighborhood change is a complex process. Much has been written in a general way about what happens. Very little has been written about why. This report represents the findings of a project whose charter was to explain neighborhood evolution and decline in terms of the behavior of the actors causing it. We wanted to understand why people do what they do, and how their actions cause change over time.

Our explanation takes the form of a computer model. Different parts of the model simulate the behavior of different actors -- households, individuals, landlords, homeowners, builders, employers, school superintendents, zoning board members. Each year the model predicts what these people will do and accumulates the results of their individual actions, thereby simulating aggregate change in neighborhoods.

The project was funded by the U.S. Department of Housing and Urban Development. Its goal is to develop and test this model (called the Community Analysis Model) in six areas across the United States: New Haven, Worcester, Dayton, Rochester, Houston, and Charlotte. Unlike many modeling efforts, this one makes no effort to test policies. Our purpose is not to recommend programs to anyone. We simply wish to understand neighborhood change and to construct a model that captures that understanding in a realistic way.

In pursuing this purpose we have sought and received the participation of planning agencies in each of the six regions for which models were developed. These agencies have been actively involved in the gathering of data, the testing of propositions on which the model is based, and the validation of the model. Through their participation, these groups have kept the model close to the world it is intended to replicate. The groups are now becoming consumers of the product they helped to create.

The report begins with an overview of the model, describing its structure and raising some of the fundamental issues we faced and resolved in designing it. The bulk of the report presents our theory of actor behavior, the heart of the model. A series of sections sets forth the rules that we think different kinds of people follow in going about their urban business. These rules are the underlying explanation of neighborhood change. We conclude with a discussion of the results of our extensive efforts at validation in all six cities.

A companion volume, The Behavioral Foundations of Neighborhood Change, summarizes much of the detailed empirical work and findings that lie behind the formulation of the Community Analysis Model.

Section 2

OVERVIEW OF THE MODEL

In its short life, urban modeling has already developed traditions and streams of development. The best way to begin a description of our approach is to relate it to other efforts, pointing out similarities and points of departure.

Trends in Model Development

Surely the main line of formal model development has been the gravity model. Its fundamental premise is that people are highly sensitive to travel time and cost and will incur such costs only when a suitable residential location cannot be found near their places of work. The "gravity label originated as a result of the resemblance of the equation used to express cost-distance relationships to the gravity formula in physics.

The basic gravity model has undergone several mutations, each one attempting to enrich the theory by increasing the number of considerations entering into a household's choice of residence. Steven Putman [16] has documented these developments. Parts of his review can be summarized as follows:

Model Name	Contribution
Lowry	Pure gravity -- the grandfather of the clan.
TOMM	Incremental instead of "instant city" allocation Disaggregated people types Introduces measure of amenities
TOMM-III	Introduces ten independent variables in the demand function in addition to distance from work place Never calibrated
PLUM	Introduces a new distance-related function to the Lowry model proportional to an exponential function of the negative reciprocal of distance as well as inversely proportional to distance squared
IPLUM	An incremental version of PLUM
BASS	For the first time, explicitly, incorporates the supply side of the housing market
NBER	Develops the supply side further and introduces a linear programming algorithm to clear the market
Herbert-Stevens	Introduces the bid-rent concept of household budget allocation and a linear programming algorithm to maximize bid-rent paying ability

Most models in this stream acknowledge antecedents other than their immediate predecessors on the list. The NBER model [10], for example, finds close ancestral ties to the models of Alonso [1], Muth [15], and Mills [14], each of whom has shared NBER's strong concern with the relationship between housing and transportation costs.

Related to and building on the transportation and land use focus of the gravity models have been several large-scale efforts centered around particular regions -- the Penn-Jersey Model [5] and the Susquehanna River Basin Model [8] being good examples.

Taking another tack, Jay Forrester and his systems dynamics group [7] have modeled urban evolution in terms of the balances between rates and flows in a system of differential equations. The model has little or no spatial dimension; it portrays a somewhat abstract city without subzones. As such, it pays no attention to access or land use, but focuses instead on the aggregate relationships among jobs, housing, and people.

Most social scientists who have conceptualized models of urban change have not formulated them mathematically. Their contributions have been important, however, and their starting points have been quite different. The sociologists, for example, usually conceptualize neighborhood change in terms of social status and changes in neighborhood reputation over time. Writers such as Firey [6], Suttles [17], and Wolfe and LeBeaux [18] attribute much of the change taking place in neighborhoods to shifts in attitudes toward dwelling unit style, race, privacy, social prestige, and so forth.

Sociologists have added concreteness to their theories not by writing equations for the most part, but by interviewing people to determine how they actually behave. Researchers like Lansing et al. [12], Butler et al. [4], and Coleman [2] have used a variety of survey instruments ranging from large-scale structured probability samples, to panels, to in-depth interviews, to construct behavioral models of residential choice. Their emerging models differ significantly from the one postulated by the gravity modelers. For one thing, they find that journey-to-work costs are important only if the journey is particularly long -- forty minutes according to Butler et al. [4]. Also, they find that people choose neighborhoods first -- paying great attention to the qualities the neighborhood has to offer -- and then find a unit in a satisfactory neighborhood. This is in direct contrast with the economists' focus on the housing unit as the primary basis for selection, which in many cases ignores neighborhood surroundings altogether. Eric Moore [13] sums up the difference regarding access quite nicely:

> Much has been written, especially by economists, regarding the importance of the journey to work in selection of residence. In particular, it is argued that cost of the journey to work relative to other household expenditures generates a residential pattern of increasing socio-economic status with increasing distance from major workplace locations. One would expect, on the basis of this literature, to find that accessibility to work and also amenities such as shopping centers, playgrounds, schools, and medical services strongly influence movement decisions. However, in almost every study which has made observations on individual decisions to move, the impact of accessibility is apparently negligible.

As a geographer, Moore represents still another school of thought. As one might expect, geographers tend to emphasize spatial patterns and the spatial organization of activities. They are concerned with rings and sectors and central places. Recently, they have begun to relate space to human perception, noting that people are not indifferent to different sectors or quadrants of the same region, and that they are sensitive to the proximity of other neighborhoods and to topography and terrain generally and that they act on these perceptions. Moore [13] and Johnston [11] have developed an interesting concept of mental maps and the relationship of behavior to these maps.

Relationship of Our Model to Historical Antecedents

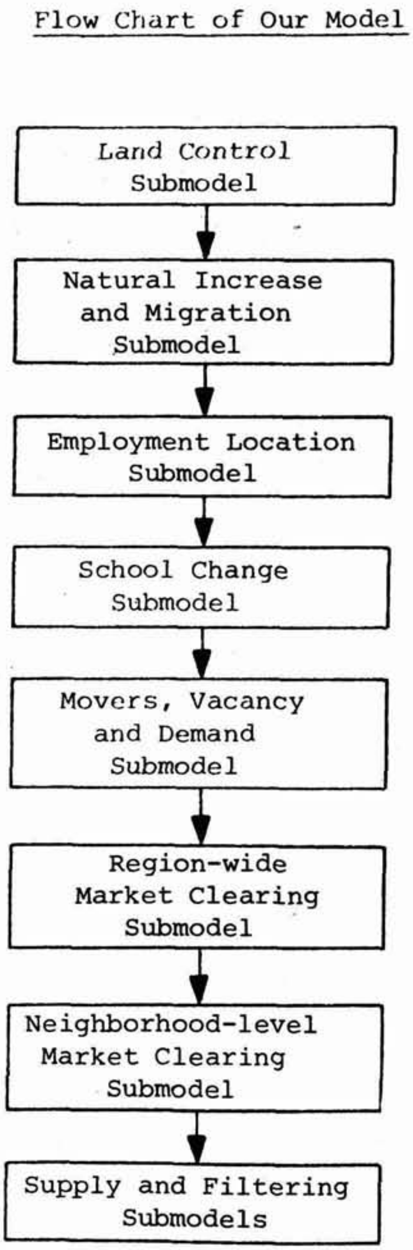
Our model embraces several of the traditions mentioned above. Structurally, it most strongly resembles the NBER model; many of its submodels contain the same labels and perform most of the same functions. A comparison of block diagrams for the two models (Figure 2-1) shows the resemblance. Both models deal with the location of employment within the region. Both treat explicitly a stream of events that begins with movers leaving units, continues through a system of vacancy accounting, a set of demand preferences for available units, a supply response (ours is lagged) both for the existing stock (filtering) and for new construction, and includes a market clearing algorithm. The major differences apparent at this level are the sequence of the submodels (in a few cases) and the fact that we deal explicitly with natural increase and decrease (births, deaths, aging), land control, and changes in schools.

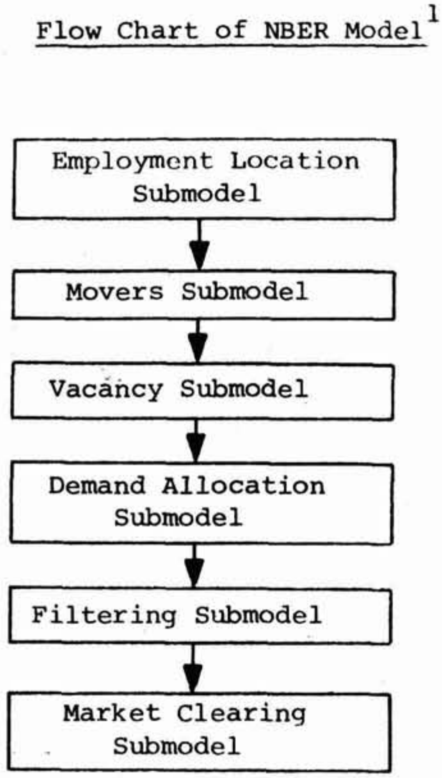
At a more general level, both modeling groups reject the simplicity of Alonso's and Muth's static cross-sections and their featureless-plain cities with all employment concentrated at the center. Moreover, we both reject the concept of long-term equilibrium as an achievable state. The rates of change in transportation, communication, food costs, energy costs, life styles, and housing preferences are at least as rapid as the ability of the region to respond, yielding cross currents and an endless process of adapation. The NBER group puts it as follows.

> Our view is that long-run equilibrium may never be attained in a metropolitan housing market. Over time, as the characteristics of population change, as employment locations change, as the transport system is modified, as new building technology is developed, as real incomes rise, the equilibrium position keeps shifting in response to these forces and many others. Thus the housing market

Figure 2-1

Comparison of Our Model with the NBER Urban Simulation Model





1. G.K. Ingram, J.F. Kain, and J.R. Ginn, The Detroit Prototype of the NBER Urban Simulation Model (New York: National Bureau of Economic Research, 1972).

> is perpetually chasing the moving target and is constantly in disequilibrium. In the NBER model we view the adjustment process as being carried out at the margin during each market period.[1]

We would only substitute "will" for "may" in the first sentence, and generalize the statement beyond the housing market to employment location, land control, school changes, social mobility, and so forth.

The mechanics by which both models clear the housing market are quite different, but it is not a fundamental conceptual difference. In fact, we suspect that with a significant amount of work, we could employ the NBER strategy, and vice versa. We feel that ours is more realistic and, at present, it is considerably less expensive to operate.

Where we part company with NBER and most other models is over the contents of the boxes. The NBER model is built on the premise that economics holds the key to unraveling the urban puzzle. This orientation is made quite explicit:

> Even though large models of urban areas may not be very novel, the NBER model is nearly unique among them because of its economic content. It is deeply rooted in economic theory; the utility-maximizing households and the profit maximizing firms that pervade microeconomics are the basic building blocks of the model.[2]

We reject the notion that any single discipline holds the key. We do not feel, nor does the evidence available to us suggest, that people are predominantly economic people, or sociological people, or geographic people. Their thought processes are not organized the way university departments are.

---

1. _Ibid_.

2. _Ibid_.

They are "real" people. When choosing a neighborhood to live, real people consider costs. They consider prestige. They consider the view. They consider their own image of themselves and their life styles. They consider places they "know" a good deal more carefully than places they do not know, and so forth.

To make matters more complicated, real people do not act like the near-perfect information processers and maximizers that much existing theory implies they are. Real people appear to be dominated by inertia not action. They do not respond instantly and automatically to changes in prices or to class changes or to vacancies. Household members we have interviewed, particularly older ones, put off moving as long as they can. Landlords sometimes do not raise prices in the face of rising demand, either because they wish to control who their tenants are or because they have an unusual concept of return; they only wish to cover costs and live free. Builders do not lay off workers and cut back their levels of construction as soon as vacancies and inventories rise. They believe, at least for a while, that things will get better.

In many cases, these theoretically imperfect people are not able to maximize their returns because they lack the information to do so. Many households, for example, know one sector of a city better than others and limit their search for housing to that sector regardless of opportunities elsewhere.

Whether the reason for imperfect response is indifference, a complex utility function (dominated by noneconomic considerations), or lack of adequate information, the net result can be described better as "inertial adaptation" rather than optimization or maximization. This is a nontrivial distinction if, as we are coming to believe, the lags in response to changing conditions in many cases are of the same magnitude as rates of change in circumstances. In a world of this sort, it is the imperfections, rather than maximizing behavior, that dominate the dynamics of change. Wide swings in construction rates, discontinuities in housing prices, and sudden changes in neighborhood composition are not the natural result of maximizers operating with perfect information.

Our model is constructed of these real people engaged in adaptive behavior. They are not real, of course, in the sense that we purport to describe everything that goes on in their heads. No model can begin to approach such complete description, nor is it useful to try. Rather, we label our people "real" because we view them as being influenced by a rich variety of information and stimuli, and as responding in a complex, not a simple, way to these influences. We do not take the position that the economists or sociologists or geographers are wrong, but that they are partial and tend to emphasize those aspects of behavior that fit nicely into their own theoretical rubrics. We have tried instead to draw on various viewpoints, to confront this set of diverse views with the evidence available to us, and to piece together descriptions of behavior that are consistent with that evidence. We consider the behavior of the people in our model to be neither irrational, imperfect, nor suboptimal. We consider it to be the natural response of normal, complex individuals to changes in the world around them.

As part of our quest for a description of normal, real behavior, we reviewed some 200 different models (conceptual as well as mathematical) purporting to describe the functioning of various kinds of urban actors.[1] The abundance of partial, often conflicting, theories was striking.

Under these circumstances one would expect that urban modelers would have placed great emphasis on testing. That has not been the case. For a variety of reasons, most urban modeling groups have stopped short of testing. The Forrester group [7] was never particularly interested in real cities. More typical, however, is the following comment by the Susquehanna group:

> In the course of the research program, we stressed the use of sensitivity experiments as a means of identifying those parts of the model that are most sensitive and therefore most promising candidates for further research. The results of these sensitivity experiments depend, however, on the structure of the model, and the conclusions drawn from these experiments are therefore based on the assumption that the model structure is a valid representation of the real world. The model's validity can be tested only be comparing its output to real world data. This may be done, for example, by seeing how well the model can reproduce past regional performance. Unfortunately, in attempting to conduct validation runs of this kind, our experience has been that obtaining reliable and comparable data for past years takes a great deal of effort and is extremely difficult. Although several runs of this type were made, because of these difficulties with the data, our original plans for conducting more extensive validation tests were abandoned.[2]

This remark is echoed by the NBER team:

> ...if the only problem had been estimation of the gross price coefficients, we probably could have eventually achieved an adequate calibration of the model for Detroit. Unfortunately, the unavailability of prices for individual housing units also meant we could not estimate accurate housing price surfaces by housing submarket. If the submarket demand functions had been "correct," this problem could have been circumvented by running the model over several time periods to produce a consistent set of housing

---

1. See David Birch, et al., Models of Neighborhood Evolution (Cambridge, Mass.: MIT-Harvard Joint Center for Urban Studies, 1974).

2. H.R. Hamilton et al., Systems Simulation for Regional Analysis: An Application to River-Basin Planning. (Cambridge, Mass.: The MIT Press, 1969).

> prices. Conversely, if submarkets housing prices had been available, we believe we eventually could have calibrated the submarket demand functions by trial and error. But since we had neither a good set of gross price coefficients nor reasonable estimates of the housing price surfaces, the task of achieving satisfactory calibration of the model for Detroit appeared nearly impossible.[1]

In most cases, the thrust is the same: "Testing is important, but difficult, and we will get to it as soon as we can." The result is that we have little idea which competing models and theories are valid. Few have demonstrated that they are capable of replicating the real world.

We feel strongly that before a model can be used with confidence for any purpose, the builder must demonstrate that it behaves the way the world behaves. It is not good enough that the model responds "reasonably" in sensitivity tests or that it looks "reasonable" to some set of observers.

We have placed great emphasis on testing and on the collection of the data necessary to conduct tests. Earlier versions of the model have been tested in New Haven and Houston. The present version has been tested extensively on a neighborhood-by-neighborhood basis for the period between 1960 and 1970 in Dayton, Worcester, Rochester, and Charlotte, and has been tested once again in New Haven and Houston.

For tests beyond 1970, the conventional census sources are unavailable and we rely on acre-by-acre satellite data for land use measures in 1972 and 1975, Dun and Bradstreet data on employment in 1970, 1973, and 1975, R.L. Polk data on the housing stock and its occupants by neighborhood where they exist, and local data series (special censuses, surveys, electrical connections, and so forth) to ensure that each version of the model is

---

1. Ingram et al., NBER Urban Simulation Model.

behaving correctly. Some of these data can be obtained annually and their use is built into the testing procedure.

A natural consequence of our interest in real people and the lags and inertia that dominate their behavior is our concern with time. Most early models ignored time altogether. They created entirely new cities from scratch in each run. More recent models deal with increments of time and incremental change, as they should. Nonetheless, they tend to be somewhat cavalier about it, being indifferent among one-year, five-year, and fifty-year increments. Our closest cousin structurally -- the NBER model -- is the most careful about its use of time, but still is somewhat fuzzy on the issue:

> The number of years covered in the simulation and the period of time represented by each iteration depend on the objectives of the simulation. For both technical and budgetary reasons, the period of time represented by each iteration will usually be greater for longer simulated periods. In order to simulate the effects of particular policies within a ten-year period, each iteration of the model would probably represent a year. However, for replicating the effects of a policy of major public investment over a period of twenty or thirty years, each iteration might represent two years or more.[1]

From our perspective, theory should be time-specific. If people put off making decisions and are adapting slowly rather than rapidly to change, then the "time constant" of their response relative to the time interval in the model is important. The one-year interval is an important aspect of our structure. Our model would not work as well if the increment were changed from one year to two years or five years. We have thought a great deal about how shorter intervals might better capture market dynamics but think we can get by with whole years if we are careful to compensate for multiple events within the years.

---

1. Ibid.

TYING THE PIECES TOGETHER

As we moved away from hypothetical, texbook maximizers into the world of complex people responding to a rich variety of stimuli, we faced two major structural issues: which actors do we include in the model, and how do we relate their behaviors to each other. We addressed these two issues by asking a broader question: How can we best characterize the phenomena governing neighborhood change? The most appropriate conceptualization to us is one of stimulus and response. Our real people are constantly bombarded with information, most of which they ignore and some of which they respond to, causing change. In this regard, we were much taken, by Britton Harris's notion of a decade ago:

> The organs of the body communicate information leading to action by nerve impluses and those maintaining homeostasis by chemical messengers; what are the messengers in a large city or region?...It is quite apparent that the generic name for these messengers will be information and it seems quite likely that some gains in theoretical clarity will be achieved if a systematic application of communications theory can be made to the diffusion of information through and about the systems under study .... Considered in the communications context, there is some merit in combining the study of decision with a priori considerations from different disciplines as to what information is likely to be important and available. At one extreme this type of merger leads to a consideration of the individual's reaction to the visual environment as developed in studies by Lynch and others. At a different extreme, economics suggests that prices are the messengers by which important economic information regarding, say, the housing market is transmitted. Between these extremes lie many combinations of phenomena which are observable, influential in behavior, and to some extent predictable as consequences of other developments.[1]

---

1. Britton Harris, "The Uses of Theory in the Simulation of Urban Phenomena." Journal of the American Institute of Planners. September, 1966.

Our design problem boiled down to picking a set of actors whose behavior generated most of the changes in the region and deciding which of the changes were worth recording and passing on as "information." Over time we have evolved a set of actors and a set of decisions made by those actors, the consequences of which are important to others, and the sum of which constitutes most of the important information flow in the regional system.

In choosing actors, we encountered a secondary issue: stratification. How are we to categorize households or business people or builders? Theories of intra-group behavior are worthless if the groupings are improper. A theory of behavior is simultaneously a statement of stratification and of intra-group behavior. Of the two, stratification is the more important in the sense that if it is improperly done, everything thereafter is nonesense.

If we really understood what goes on inside people's heads as they choose neighborhoods, we would find such distinctions as age, race, income, or education superficial. However, we do not have such an understanding. As a consequence, we tend to rely on observed variations from group to group, and choose subgroups in such a way that variations within groups are minimal. The NBER has struggled with this same issue and sums up the state of the art:

> The households which appear in the simulation model are defined by household class much in the way that dwelling units are defined by type. A household may fall within any of seventy-two household classes depending on the age and educational level of the household head, the household's income, and its family size....The development of these seventy-two household classes was basically arbitrary, although the classes generally reflect differences in housing consumption patterns that were sustained in empirical work.[1]

---

1. Ingram *et al*. *NBER Urban Simulation Model*.

After deliberating about the stratification problem for some time, we have decided, as have others, that many actors must be stratified jointly along at least two, and sometimes three or more dimensions. For instance, all young households do not behave in the same way; households that are young and well educated behave differently from those with less education. Furthermore, young, educated, black household follow different patterns than their white counterparts. To describe household behavior accurately we must know age, education, and ethnic/racial background.

In Exhibit 2-1 we have summarized the actors whose behavior we are simulating, our stratification of them, their decisions which we are simulating and recording, and the information upon which we assume those decisions to be based. Our notation indicates a joint stratification descriptor; for example, a household would be described on the basis of age by education by race.

We simulate information flow by recording what each actor does so that other actors can "learn" about it if our theory of behavior suggests that they consider the change important. The structure of the memory in which events are recorded is thus an important aspect of the design of our model. In fact, another way to conceptualize the model is in terms of the information it records and passes on to actors. When the model is used as a forecasting device, the contents of Memory are its output. Memory contains the net results of everything all the actors have done to the region's neighborhoods up to and including the end of the past period simulated. If, for instance, that period is the year 1980, Memory is a forecast of what neighborhoods will look like in 1980. In Figure 2-2, we summarize the contents of Memory. The model maintains a considerable amount of information about

Figure 2-2

Main Items in Memory

<u>For Each Neighborhood</u>

1. Number of households, by type of household by type of housing unit
2. Number of people, by type
3. Number of households, by type
4. Average household income
5. Number of jobs, by type
6. Unemployment rate
7. Number of occupied housing units, by type
8. Number of vacancies, by type
9. Annual construction, by type of unit
10. Average condition of the housing stock
11. Annual maintenance of the housing stock
12. Abandonment of housing units, by type
13. Number of households dissolved (primarily through death) per year, by type of household head, by type of housing unit
14. Number of new households per year created through natural increase, by type
15. Number of households in-migrating per year, by type
16. Number of households out-migrating per year, by type
17. Number of local movers arriving, by type of household
18. Number of local movers leaving by type of household
19. Excess demand for housing units per year, by type of household, by type of unit

20. Acres of land, by use

21. Quality of schools, by characteristic, by school

22. Maximum density at which builders are permitted (through zoning) to build

23. Access time to every other neighborhood

For the Region as a Whole

1. Unemployment rate

2. Interest rates

3. Unemployment rate, by occupation

4. Unemployment rate, by industry

5. Unemployment rate, by type of person

6. Anything known at the neighborhood level

each neighborhood for use by the actors. Our theory postulates that this information is the set necessary to stimulate and predict actor behavior, and that the actors and decisions we have chosen are sufficient to generate it. With some exceptions, the model is thus a self-contained system. The central information flows and their relationship to the updating of memory are summarized in Figure 2-3.

The sequence of events within a given period is important to the design of a model that depends on information flow. If one event precedes another in the real world, that event should take place first in the model. Figure 2-1 shows the sequence now in use. We placed the employment location decision early on the assumption that business location decisions affect subsequent household behavior. However, our research indicates that, to an increasing extent, people are selecting business locations to suit their residential convenience's. Thus, we may alter the sequence in the future to reflect this shift.

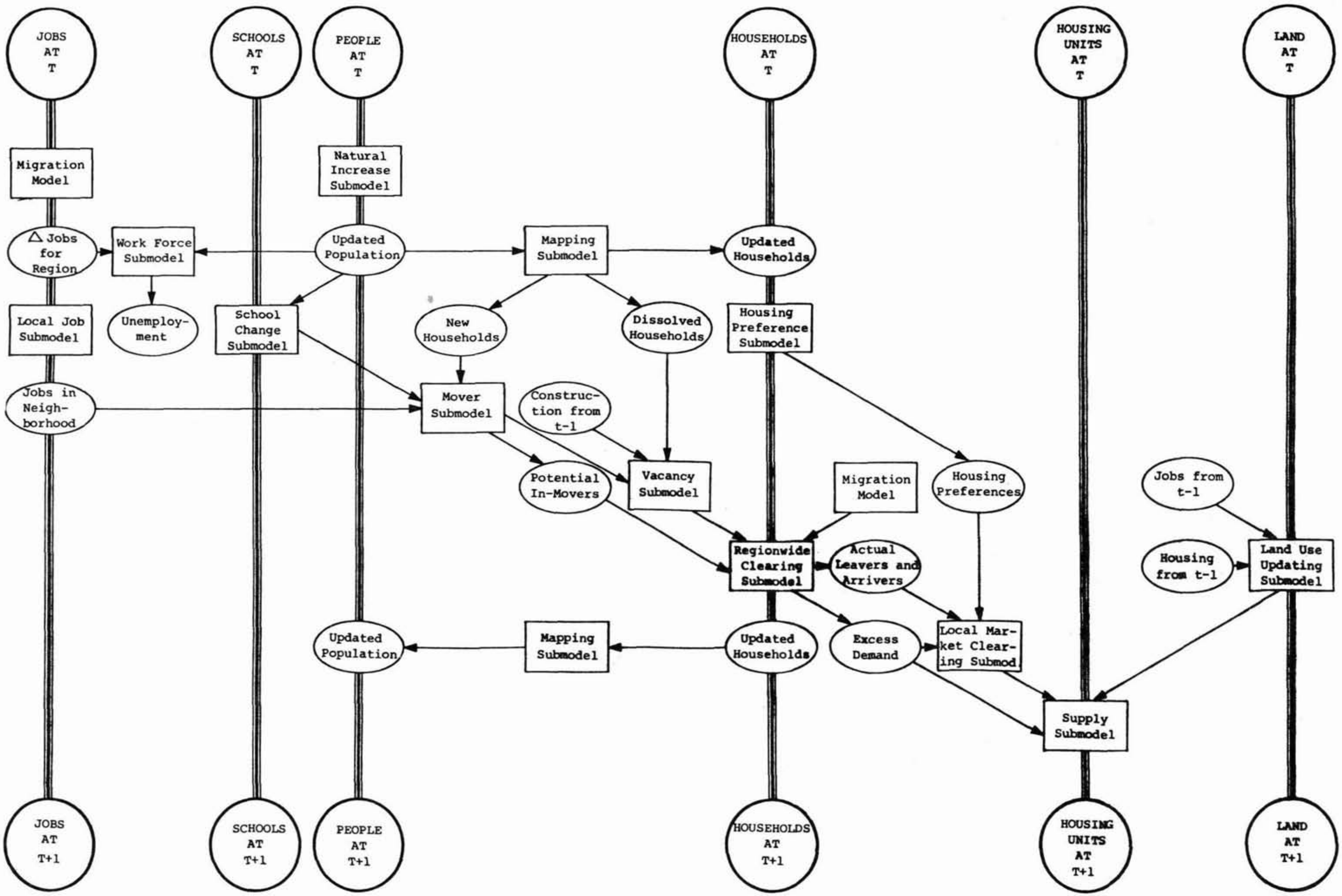
Natural increase and migration come next to introduce new households into the region and to identify vacancies caused by death and outmigration. schools change, and then local movers move, further defining the number and type of available vacancies and the number of buyers in the marketplace.

The behavioral work we have reviewed and our own interviews suggest that people first choose a neighborhood and then a unit. We thus clear the housing market across neighborhoods within the region before working out the local details. Supply responses take place near the end of the period and affect the market in the subsequent period.

In a nutshell, it is a demand-driven system, and households are the key demanders.

Figure 2-4

Summary of Flows Between Submodels and Resulting Changes in Memory



Remaining Structural Issues

With a well-developed theory of information flow and a defined set of actor-response submodels, three major structural decisions remain: (1) whether to model individual or household behavior, (2) how to clear the housing market, and (3) how general to make the model's structure in attempting to model six quite different cities. We will take them up in that order.

Households vs. Individuals

Certain events in a metropolitan area can best be described in terms of the behavior of individuals: bearing children, aging, dying, joining the labor force, gaining an education, and so forth. Other events, particularly the occupation of housing units, can best be described in terms of household behavior because households move as units.

We have tried both methods of description and have concluded that neither is satisfactory by itself. As a result, we maintain both and have developed an accounting mechanism for mapping back and forth between the two. Whenever something happens to individuals, we assess the consequences for household formation and dissolution. Conversely, whenever households move, we map the changes back into population terms.

Clearing the Housing Market

More fundamental than our system of household accounting is our procedure for clearing the housing market. For any given period, the mover submodel estimates the number of movers, the vacancy submodel determines how many vacancies are available, and the demand submodel estimates where movers would most like to locate given their preferences for units, neighborhoods, and access to jobs and amenities. In general, there is a mismatch between the supply of units (by type and neighborhoods) and the demand for these units. The clearing algorithm must resolve the mismatch, allocating people to units in a realistic manner.

Different modeling groups have experimented with several different ways to handle the clearing problem. At one extreme, mismatched households are put up in tents or hotels until the next period, i.e., the problem is ignored. Some groups have tried iteration -- solving and resolving the demand equations until demand matches supply within some arbitrary limit. Others, including NBER and the Herbert-Stevens model, have adopted a formal allocation system -- linear programming -- to do the job.[1]

We have employed still another method which we feel is realistic and, fortunately, quite a bit less expensive to simulate than many others. It is a bidding system. Each year we conduct an "auction" in each neighborhood. Bidders' success in the auction is a function of their preferences, the nature of the available vacancies, and the relative bidding strength of others seeking the same units. If our estimates of bidding strength are correct, each subgroup of households will acquire the right number of units. If not, some will have too many and some too few. Under these circumstances, we adjust the bidding strengths in a way that should come much closer to clearing the market, and try again.

The procedure is efficient and, we think, realistic. People do, in fact, bid for units, either directly or through a broker, and some are able to

---

1. In addition to being expensive (the NBER algorithm accounts for 90 percent of the running cost of the entire model), the linear programming approach poses special problems when the dual variables are interpreted as prices. Any time a housing supply constraint is changed (through significant construction, conversions, or demolitions) or when demand falls off sharply, the dual variables can oscillate beyond what might be expected in reasonable price behavior. It is an inherent limitation of the method and is usually overcome by smoothing prices over several periods to dampen the effect of eratic behavior.

outbid others in the process.

The procedure becomes unrealistic only if the initial estimates of demand are highly inaccurate. In that case people end up in units or neighborhoods they might not normally choose. Thus far, however, our initial demand estimates have been close enough, and the algorithm has introduced minimal distortion.

Commonality

We are building models for six different metropolitan areas and must face the issue of commonality. Can we build one model that will suffice for all six, or do we need six different models?

It is more than merely a question of mechanics. Ideally, a model should not require recalibration when applied to different areas. If we understood human behavior well, all we would need to know for a new city is a description of its history, its topography, and its location relative to other areas in order to simulate behavior within it. Surely no one would argue that Texans are fundamentally different from residents of Connecticut. Each survives, even flourishes, when moved from one place to the other.

The fact is, of course, that we do not come even close to the kind of fundamental understanding necessary to avoid recalibration. Since that understanding is what we are striving for, however, we have created only one model (literally, in the code sense) and have forced all variations from area to area to take the form of parameters in data sets. In this manner, we can make explicit comparisons across regions to gauge the nature of the variations and to begin to home in on the underlying causality.

Transferability and Use

This same model is now running with considerable accuracy in all six cities. We have demonstrated to our own satisfaction that its structure is general and could be adopted elsewhere without significant internal changes. Transfer is thus primarily an issue of time, cost, and the mechanics of transferring the code.

Time

Based on our experience to date, it will take about one year to build and test the model in a new area. About half of that time is devoted to data gathering, and the rest to debugging the data and code and testing and validating the model. The finished products include: (1) detailed population, employment, housing, school, and land use data for each neighborhood, (2) routines for reducing these data to useful form, (3) a model for projecting them into the future, and (4) a retrieval system for displaying past data or model results.

Costs

The total cost of a model is the sum of the costs of three sets of tasks: (1) that which the model builders do to create it, (2) that which the user organization does to create it, and (3) manintenance by the user organization.

We have not yet built enough models outside of a research environment to determine the long-term cost curves. They are a function of the size of the region, the initial availability of data, the skill, experience, and interest of the user organization, and our own learning curve over time. Roughly speaking, however, to develop a complete data set and an operating, tested model for an interested, medium-sized city would cost the builders about $70,000 to $100,000 over the one-year development period and would cost the user organization the equivalent of about 1.5 years of one person's time. Once the model is developed, the annual operating and maintenance costs should average $40,000 to $50,000, including computer time, staff time, and overhead.

Technical Transfer

At present, models are built in Cambridge, Massachusetts, and transferred to users' facilities at the end of the development phase. All the software is programmed in standard FORTRAN IV, which makes transfer a relatively simple matter. The transfer time is on the order of two to three weeks, requiring one skilled programmer during that period.

Application

Once transferred, the model and its attendant software packages have generally been used for four different purposes:

1. Data retrieval: The ability to retrieve detailed data over time, for a consistent geographical area, over a wide range of subject areas is viewed as an asset by many planners.
2. Projection: Projections of households and housing units, jobs, and land use are essential to the planning process of some users, and the model's output is directly useful as it stands.
3. Translators: For others, the models output must be translated into terms that relate directly to an organization's functions: health care, school enrollments, tax rates, budget

line items, millions of cubic feet of natural gas, likely sites of fires, and so forth. Auxiliary software ("translators") to estimate such specific items must be developed by the user organization. Once developed, this software can be transferred from one city to the next because the model's output is standard. Several translators have been developed already.

4. Program testing: Many federal, state, and local programs are directly related to the model's structure (building houses, demolishing houses, building roads, subsidizing income, insuring or not insuring properties, etc.) and can be tested before implementation by altering the model and observing the results.

User groups in all six cities are now in the process of transferring the software to their cities. In the meantime, using telephone lines, they are working with the packages in Cambridge. Retrieval has been possible for several months. Users have reviewed projections carefully during late 1976, and are now putting them to use in some areas. The second round of translator development is just getting underway. Each of the six participating cities will develop one or more translators and will share its final products with the others. Program testing is also in its infancy, and will become a major activity over the next six to eight months.

Exhibit 2-1

Actors, Decisions, and Major Determinants

Actor	Decision	Stratification	Major Determinants
Household	To Move Within Region	Age by E/R[1] by Education	Life Cycle Race Education Racial Change in Neighborhood Forced Moves
	Choice of Neighborhood		Life Cycle Race Education Available Units Social Class of Neigh. Job Access. of Neigh. Location of Neigh. Racial Transition of Neigh.
	Choice of Unit		Housing preferences Financial status Availability of mortgage credit
	Migrate in and out of Region		Employment opport. Unemployment Income Levels in Reg. Educational Mix in Reg. City Size Proximity to other Areas
Individual	Have Children	Age by E/R by Education	Life Cycle Race Education
	Obtain Education		Ethnic/Racial Background
	Join Work-force		Age Ethnic/Racial Background Education Growth rate of Local economy

1. E/R = Ethnic/Racial

Exhibit 2-1 (con't)

Actor	Decision	Stratification	Major Determinants
Homeowner	Setting selling price of home	Age by E/R by Education by Price of home	Potential demand relative to available units
	Investing in home maintenance		Characteristics of homeowner (e.g., housing preferences) Characteristics of housing unit (e.g., age of unit) Characteristics of neighborhood (e.g., average housing condition)
Landlord	Setting rent levels on apartments	Rent level of apartments	Potential demand relative to available units
	Investing or disinvesting in maintenance for apartments (including abandoning apartments)		Characteristics of tenants (e.g., age, education, ethnicity) Characteristics of apartment (e.g., age of unit Characteristics of neighborhood (e.g., ethnic composition)
Builder	Constructing single-family homes under contract	Contract vs. Speculative Type of unit (tenure and price)	Vacancy rate in submarket and region Availability of suitable vacant land Availability of credit
	Constructing apartments under contract		Absorption rate in submarket Excess demand in submarket Vacancy rate in submarket and region Zoning restrictions Availability of suitable vacant land Availability of credit

Exhibit 2-1 (con't)

Actor	Decision	Stratification	Major Determinants
Builder	Constructing single-family homes and apartments speculatively		Assessment of prospects for future demand in the neighborhood according to demand and rate of growth in nearby areas. Availability of suitable land Availability of credit
Lender	Lending for new construction and home mortgages	None	Security of loans: depends on characteristics of loan applicant (e.g. age, ethnicity, education), characteristics of housing unit (e.g., condition), and characteristics of neighborhood (e.g., ethnic composition, average housing condition) Availability of funds
	Refusing to grant loans in certain neighborhoods		Risk attached to loans: depends on characteristics of neighborhood (e.g., ethnic composition, average housing condition)
Insurer	Insuring homes and apartments	None	Expected profit (excess of premium revenue over insurance payments): depends on characteristics of insurance applicant (e.g., age, ethnicity, education), characteristics of housing unit (e.g., condition), and characteristics of neighborhood (e.g., ethnic composition, average housing condition)

Exhibit 2-1 (con't)

Actor	Decision	Stratification	Major Determinants
Insurer	Refusing to write insurance in certain neighborhoods		Expected loss (excess of insurance payments over premium revenue): depends on characteristic of neighhood (e.g., ethnic composition, average housing condition)
Zoning board member	Determining permissible land uses, granting variances	None	Density preferences of residents of neighborhood Extent of commercial activity in a neighborhood
Employer	Location of a new firm Expansion of employment Relocation Contraction of employment Closing a firm's operations	Major SIC code divisions	Vacant Land in a neighborhood Characteristics of the population of a neighborhood Population density Job concentration: (Proportion of jobs in a given industry located in a given area) Job specialization: (Proportion of jobs in a given area which are in a given industry) The tenure and value of occupied housing units
School Superintendent	Modify School Characteristics	None	Residents' reactions to school characteristics Court decisions
Resident	Support (or not) existing school policies	Age E/R Education	Percentage minority in school Teachers' qualifications in school Public Image of school Curriculum in School Class Size in School Resident's Attitude toward educa. Social Class of Resident

REFERENCES

1. Alonso, W. Location and Land Use. Cambridge: Harvard University Press, 1964.

2. Birch, Atkinson, Clay, Coleman et al.: America's Housing Needs: 1970-1980. Cambridge: MIT-Harvard Joint Center for Urban Studies, 1973.

3. Birch, Atkinson, Coleman, Parsons, Rosen and Solomon. Models of Neighborhood Evolution. Cambridge, Mass.: MIT-Harvard Joint Center for Urban Studies. Report to the U.S. Department of Housing and Urban Development, 1974.

4. Butler, Edgar W.F., Chapin, Stuart; et al.; Moving Behavior and Residential Choice. Highway Research Board, 1969.

5. Fagin, Henry. "The Penn-Jersey Transportation Study: The Launching of a Permanent Regional Planning Process." Journal of the American Institute of Planners, February, 1963.

6. Firey, Walter. Land Use in Central Boston. Cambridge: Harvard University Press, 1947.

7. Forrester, Jay. Urban Dynamics. Cambridge, Mass.: MIT Press, 1969.

8. Hamilton, H.R. et al. Systems Simulation for Regional Analysis: An Application to River-Basin Planning. Cambridge, Mass.: The MIT Press, 1969.

9. Harris, B. "The Uses of Theory in the Simulation of Urban Phenomena," AIP, September, 1966.

10. Ingram, G.K., Kain J.F.: and Ginn, J.R. The Detroit Prototype of the NBER Urban Simulation Model. New York: National Bureau of Economic Research, 1972.

11. Johnston, R.J. "Activity Spaces and Residential Preferences: Some Tests of the Hypothesis of Sectoral Mental Maps." Economic Geography, April, 1972.

12. Lansing, J.; Mueller, E.; and Barthin. Residential Location and Urban Mobility. Ann Arbor: Survey Research Center, Institute for Social Research. The University of Michigan, 1964.

13. Moore, Eric. Residential Mobility in the City. Washington, D.C.: Association of American Geographers, 1972.

14. Mills, Edwin. An Aggregative Model of Resource Allocation in a Metropolitan Area." American Economic Review, May 1967.

15. Muth, R. Cities and Housing. Chicago. University of Chicago Press, 1969.

16. Putman, S. "Urban Land Use and Transportation Models: A State-of-the-Art Summary." Paper presented at the Second Intersociety Conference on Transportation; Denver, Colorado; September, 1973.

17. Suttles, Gerald D. The Social Construction of Communities. Chicago: University of Chicago Press, 1972.

18. Wolf, Eleanor and Charles LeBeaux. "Class and Race in the Changing City: Searching for New Approaches to Old Problems." in Urban Research and Policy Planning, L.F. Schnore and H. Fagin (eds). Sage Publications, 1967.

Section 3

NEIGHBORHOOD DEFINITION

A first step in any effort to model neighborhood change is to define "neighborhood." To this end, we have developed a general typology for describing neighborhoods and used that typology to identify surrogate variables and draw actual boundaries on maps. As a logical starting point, therefore, we will present a summary of our general typology. We will then describe how we used the typology to draw neighborhood boundaries.

Neighborhood Typologies -- A Summary

Our basic proposition is that what happens in neighborhoods is the result of a complex interaction of actors and neighborhood characteristics. We have already identified nearly fifty neighborhood characteristics that we would like to take into account in our model. We have come across these in various ways, many, of course, in our previous modeling experience. Our literature search yielded more.[1] Some of the most interesting variables came to our attention in the course of a pilot experiment with in-depth interviews of Houston residents concerning neighborhood behavior; still others, usually reaffirmed by our Houston interviews, we had encountered in

---

1. See Models of Neighborhood Evolution [2].

previous research on cross-sections of Boston and Kansas City residents.[2]

All fifty of these neighborhood characteristics are things that we think interest people who are acting in the system. It would be ideal if all these characteristics could be measured and thus incorporated into our formal computer model. We will of course fall short of this ideal, since for some, the data are not readily available, and for a few, devices for their measurement have not yet been developed. Those that are left out may still prove valuable in explaining the outcomes, by supplying qualitative insight to our quantitative analyses.

Our neighborhood characteristics fall into eight main groups: (1) characteristics of residents, (2) characteristics of the housing stock, (3) land use characteristics, (4) proximity to employment, (5) residential and social characteristics of adjacent neighborhoods, (6) public services and governance, (7) history of neighborhood and related reputational factors, and (8) the natural environment. The order in which these eight categories are listed indicates the relative extent of their inclusion in our data files. Nine variables that can be considered characteristics of residents are included while none that are direct measures of the natural environment have yet been part of our simulation. The explanation for this is simple: the 1960 and 1970 U.S. Census Reports provide us with detailed statistics for the former on a tract-by-tract basis. There is nothing comparable on a neighborhood basis for the latter -- it is not a census concern. We would be derelict, however, were we to take this Census disinterest as a cue or an excuse to dismiss the eighth category from consider-

---

2. See especially Chapter 5, "The Influence of Consumer Preferences on Housing Markets" in America's Housing Needs: 1970 to 1980 [1].

ation, since our interviews with Boston people, Kansas Citians, and Houstonians have shown it to be an important factor in neighborhood choice and satisfaction. People care about the terrain, the vegetation, the elevation, and presence or absence of bodies of water for residential siting and/or recreational use.

The characteristics of neighborhood residents that are most crucial to our modeling are age, race, ethnicity, and educational attainment. It is essential that we be able to type neighborhoods on these variables in order to understand whom the neighborhoods attract and who rejects them. We are also typing neighborhood residents by the occupations of those who are in the work force, by the industries in which they are employed, and by the levels of their household income. Two characteristics of special interest are the percentage change in total number of households within a neighborhood between 1960 and 1970 and the percentage change in racial and ethnic composition of these households.

The housing stock characteristics which interest us most in our neighborhoods are tenure (i.e., percentage owner- versus renter-occupied), market value of owner-occupied units, and gross monthly rents of renter-occupied units. Rates of demolition, rates of addition to the housing stock, and percentage of vacancies also are currently in the model and serve as variables by which neighborhoods may be typed.

Land-use characteristics under examination include residential density in 1970, increase in residential density from 1960 to 1970, and percentage of neighborhood acreage in residential versus other uses (manufacturing, commerce, parks, cemeteries, institutions), and two classifications of vacant: easy-to-build and hard-to-build. Two characteristics that are important to residents and potential residents, but for which we are currently lacking measurement or objective classification, are concentration of housing, and something that, for want of a better expression, we call "the feel of the neighborhood." For example, if the neighborhood is one of high residential density, is it composed of high-rise or old-style, low-rise residences? Or, to take another example, if it is an area of light density, does it "feel" to the public "like a small town," or would it be characterized as a "badly developed, ugly urban fringe?"

We can type neighborhoods according to their location vis-à-vis employment in these ways: driving time to center city, job accessibility throughout the metropolitan area, employment by industry, total number of people employed in the immediate neighborhood and in adjacent neighborhoods, and projected future job accessibility.

This last is facilitated by our research into employing establishments and their potential birth, death, contraction, expansion, and prospective locations.

To residents and potential residents of any given neighborhood, the residential and social characteristics of adjacent neighborhoods are commonly of considerable interest. Recognizing this, we have developed the means for quickly characterizing each neighborhood by the ethnic and racial distribution of the adjacent neighborhoods (and, for that matter, by any other of the demographic or socioeconomic variables we might wish to examine). Perhaps most important in this regard is the geographic relationship a neighborhood bears to areas undergoing change in racial or ethnic composition; here our interest is not in proximity per se as much as in the direction of racial change and a neighborhood's relationship to corridors of change.

The public service which our interviews show to be most important in determining neighborhood satisfaction is schools. In Section 11, which treats school officials as actors in the neighborhood process, we discuss possible ways of simulating the behavior of these officials and measuring its effect on school quality. Crime and fear of crime are concerns of neighborhood residents that we bracket together with public service and governance. We are now working on the problem of assembling crime rate data on a neighborhood base and hope ultimately to be able to type neighborhoods on this variable.

The history of a neighborhood's role in the life of a city helps explain why that neighborhood is as it is, where it may go in the future, and how people feel about it. Walter Firey's Land Use in Central Boston [3] is almost scriptural in its illustration of this principle. Our pilot interviewing in Houston has reaffirmed our desire to introduce an historical variable into our characterization of neighborhoods. Of particular interest in this connection is the effect of the aging process on neighborhoods, such as Montrose and Houston Heights in Houston, toward which a considerable amount of positive sentiment is expressed, compared with the effect of aging on those toward which current residents, former residents, and the citizenry at large seem relatively indifferent or even negative.

Defining Neighborhood Units for Study

In developing a scheme for categorizing and aggregating neighborhoods for study, we have looked to the United States Census Bureau's prior efforts at neighborhood definition for guidance. All neighborhood units in this research are, in fact, either 1970 census tracts or aggregations of 1970 census tracts. This rule was established for our definition of neighborhoods for one simple reason: much of the data relevant to this study had already been gathered, taped, and published by census tract. That enormous mass of available data simply precluded the use of any other approach to neighborhood definition.

On occasion, we aggregated 1970 census tracts instead of treating each as a separate neighborhood unit, for one or more of the following reasons:

(1) to reduce the total number of neighborhood units in an area and thereby affect certain economies of computer time;

(2) to unite adjacent tracts that are so similar in demographic, socioeconomic, and housing stock characteristics that they are not functionally distinguishable in modeling of household choice behavior; and

(3) to make our 1970 neighborhood units conform with aggregated 1960 census tracts when 1960 and 1970 tract boundaries were not identical.

For the most part census tracts, or aggregations of tracts, are satisfactory neighborhood units for the purposes of our study. This is especially true when tract boundaries have been drawn along major arteries, rivers, railroad tracks, or other "natural" dividing lines. In such cases, the tract boundaries almost certainly parallel the neighborhood boundaries recognized by community residents themselves. But sometimes tract boundary lines, which may have reflected community feelings in an earlier time, do not adequately represent current social realities. The realization of this phenomenon accounts in part for our willingness to aggregate tracts on occasions when aggregates appear to reflect present social realities more accurately.

There are many other respects in which census tracts and even aggregates fall short of what might be ideal for our neighborhood units. For one thing, they are rarely coterminous with school district lines, yet much of neighborhood choice, as well as the sense of neighborhood identity, historically has revolved around schools. This may be less important today, when, in so many communities, the school attended is no longer the one within closest walking distance. Second, census tracts, even when aggregated for certain purposes, do not necessarily correspond with municipal subdivisions such as wards, police districts, and traffic control areas; the disadvantage in this case is that some of the data relevant to our study were gathered by these units and are not easily mapped onto tracts.

Another respect in which census tracts are less than fully satisfactory is that more often than not they embrace two or three subneighborhoods of differing status and housing characteristics. This is apt to be the case especially in a metropolitan area whose older sections were built up when neighborhoods of homogeneous character tended to be physically smaller than they are today. Some of the tracts most satisfactory for our purposes are those which are coterminous with single-builder suburban developments of the post-World War II era. Tracts at the very periphery of an urban area are, on the other hand, apt to be large and heterogeneous.

Our primary goal in drawing neighborhood boundaries for this study has been to identify homogeneous neighborhoods. Pursuit of this goal has determined, insofar as it was possible, tract aggregation. Similarly, this goal was an uppermost consideration when we decided to leave any tract in its original state, integral and unaggregated. We consistently drew neighborhood boundary lines with the intention of ensuring that the phenomena encompassed were similar, and that all nearby areas which fell outside a boundary differed from those within it.

Identifying distinct homogeneous areas is particularly important from a modeling point of view. When adjacent areas (city blocks, census tracts, or other geographic areas) are very similar in characteristics of population, housing, and historical development, it is practically impossible to predict differences in their futures or to simulate past differences. Thus it is important that all like adjacent areas be within one neighborhood area for the modeling.

The goal of homogeneity within the neighborhood areas could not always be achieved. The major constraint, of course, was the pre-existing 1970 census tract boundaries. When these boundaries did not define homogeneous areas, we were forced to define as a neighborhood unit an area which might contain two or more distinctly different types of neighborhoods.

In addition to the dominant rule that all neighborhood units be 1970 census tracts or aggregations thereof, there were three other rules which we applied from time to time. First, neighborhood boundaries were never to cross a county boundary. This rule facilitates aggregation of neighborhood data up to the county level for reporting purposes.

Second, neighborhood boundaries were to maintain the integrity of the central city boundaries where the central city boundaries coincided with census tracts and where the central city was not regularly annexing new territory. This rule was applied in order to facilitate the aggregation of neighborhood data up to the central city level for reporting purposes.

Finally, in order to facilitate comparisons over time, we wished to maintain wherever possible the 1960 census tract boundaries as the basis for neighborhood areas. This latter rule was not always observed, however, and was usually disregarded if, by 1970, the 1960 census tract had been subdivided in a way that produced individual neighborhoods more homogeneous than the original whole. This rule was also not followed if the original 1960 tract boundaries had been moved by the 1970 tracting. Our guideline was as follows: if individual 1970 tracts that had been created as subdivisions of a single 1960 tract were essentially similar in character, i.e., together formed a homogeneous area, they were aggregated into one neighborhood. If the 1970 tracts proved dissimilar, they were established as separate neighborhood areas for study and the 1960 entity was ignored.

The procedure used to ascertain which census tracts should be aggregated into homogeneous neighborhoods is explained in Appendix 3-1. This procedure involved matching 1970 census tracts on a series of criteria designed to determine their similarity. Through experience with this scheme we came to realize that two of the criteria were particularly important and their respective scores were double-weighted. These critical dimensions were the racial composition of the tract (percentage of nonwhite residents) and the percentage of the housing stock that was built between 1960 and 1970. The effect of doubling the comparison scores for these two criteria was to guarantee that if marked differences appeared between neighboring tracts on either, such tracts were not aggregated regardless of how similar they might have been on all other criteria. Marked differences here could mean as small a difference as 5 percent, depending on the dimension and location on the continuum; more commonly, however, a difference was not considered "marked" unless it was at least 10 percent.

Our method for determining homogeneous neighborhood areas has its limitations. All our criteria for determining neighborhood homogeneity (within the guidelines for for a "homogeneous neighborhood," as previously set forth) came from only two points in time, namely the 1960 and 1970 censuses. Thus, any processes of neighborhood change taking place since 1970 would not show up. For advice on this, as well as for insights on variables and dimensions not included in our formal procedure (such as ethnicity and changes in racial composition for noncensus year periods) we turned to the people in our field organizations who, because they live within the particular regions, have on-the-spot knowledge about what is taking place there. Their guidance proved the key from time to time in resolving questions about where neighborhood boundaries should be drawn.

Application of these several procedures for determining when and where to aggregate tracts resulted in our designating as individual neighborhoods about 60 percent of all 1970 census tracts within the six regions we are studying. The other 40 percent have been aggregated into larger neighborhoods, usually in combinations of two tracts per neighborhood, but on occasion in combinations of three or four. The procedures have now been implemented to our satisfaction, and to the satisfaction of the field organizations with which we are working in each of the six regions.

Districts

For certain research purposes -- especially communicating our findings to the field organizations in the six cities -- we have aggregated neighborhood units into districts and have created district boundary lines. Districts on the average include four or five neighborhoods but they may contain as few as one or as many as twelve. Our purpose in drawing district boundaries was to provide a means for capturing broader sections of the region primarily for reporting purposes. People tend to be much more familiar with and know more about our district areas than they do the finer areas encompassed by our neighborhoods. We also sometimes use districts for modeling purposes when only a moderate level of geographic resolution is required, as in simulating where a firm might look for a location within a region. A neighborhood would be much too small an area for this purpose. The procedure used to determine the precise district boundaries was essentially the same as that used to determine neighborhood boundaries. The only modification was that greater differences were tolerated when aggregating areas in order to capture broader sections of the region. All districts are either neighborhoods or aggregations of neighborhoods. District boundaries never cross a county boundary. Finally, district boundaries maintain the integrity of the central city boundaries where the central city boundaries coincide with neighborhood boundaries, and where the central city is not regularly annexing new territory.

We again turned to the local field organizations for their insight into where the district boundaries should be drawn. It was helpful, in most cases, to draw the district boundaries to coincide with existing planning districts in the region. This generally imposed few extra constraints, since planning districts tended to be based on the same general criteria we had chosen.

The result of this method was that most regions were divided into about 26 districts. Since the size of districts was not the critical determinant, the larger regions tended to have roughly the same number of districts as the smaller ones.

REFERENCES

1. Birch, Atkinson, Clay, Coleman, Frieden, Friedlaender, Rainwater, and Teplitz. America's Housing Needs: 1970 to 1980. Cambridge, Mass.: Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, 1973.

2. Birch, Atkinson, Coleman, Parsons, Rosen, and Solomon. Models of Neighborhood Evolution. Cambridge, Mass.: Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, 1974.

3. Firey, Walter. Land Use in Central Boston. Cambridge: Harvard University Press, 1947.

Appendix 3-1

Tract Aggregation Procedure

In order to determine which census tracts should be combined, we compared each 1970 census tract with all its neighbor tracts on each of the following criteria (borrowed from our neighborhood typology):

% of nonwhite residents in tract

% of residents under 25 years of age

% of residents 65 years of age or older

% of residents 25 years of age or older with less than high school education

% of housing units occupied by renters

% of housing stock built between 1960 and 1970

% of housing considered low in market value or monthly rent

% of housing considered high in market value or monthly rent

All tracts were examined for degree of similarity to adjacent tracts on each of the above criteria. The similarities were then scored and summed. Assume, for example, we wish to look at tracts X_2, X_3, X_4,...X_n which are adjacent to Tract X_1. We wish to determine if any of these adjacent tracts is so similar to Tract X_1 that it should be aggregated with X_1 into one neighborhood. The first step in making this determination is to calculate the mean and standard deviation for all of the tracts in question for each criterion.

The second step in this process is to implement a scoring scheme which determines the precise degree of similarity each tract has to Tract X_1 on each criterion. This is done as follows. The standard deviation is divided into thirds. Then all tracts which fall within one third of the standard deviation around X_1 (those that differ the least from Tract X_1 on a particular criterion) are given a score of 3 on that criterion. All tracts which fall between one third and two thirds of the standard deviation from X_1 on a criterion are given a score of 2 on that criterion. All tracts which fall between two thirds and one standard deviation from Tract X_1 are given a score of 1 for that criterion.

All scores for each tract are summed. If the summed similarity scores for adjacent tracts are high enough, suggesting that it would be difficult to distinguish between them (in predicting actual behavior), they are aggregated into a single neighborhood. For example assume the following scores were calculated for tracts adjacent to Tract X_1:

Neighbor	Score
X_2	18
X_3	15
X_4	28
X_5	21
X_6	10
X_7	6

(maximum possible score = 30)

In this case, Tract X_4 is an excellent candidate for combination with Tract X_1 into one neighborhood. In cases in which a particular tract does not score particularly high on summed similarities with any of its neighbors, it is not aggregated with them, but is left as a neighborhood unto itself.

Section 4

NATURAL INCREASE AND MIGRATION

People make many of the important decisions that shape cities as members of households. Certain decisions, however, are individual decisions, and some events -- like aging and dying -- are the inevitable consequence of being an individual. For this reason, we keep parallel track of individuals and households (as described by the characteristics of the household head).

This section deals with several of these individual processes -- birth, death, aging, obtaining an education -- and with one household-related acticity -- migration. As we mentioned in Section 2, a major step in the study of individual and household behavior is stratification. It is particularly important for households because it affects our definition of their entry into and actions within the housing and employment markets, as well as their behavior as individuals.

Stratification

If we knew a great deal more about how people think and make decisions, and if we had measures of these thought processes, we would undoubtedly classify poeple along psychological dimensions. Such measures are not available for the population at large, nor do we have the knowledge to employ them if they were. Instead, we and others have tended to rely on observed variations in behavior among different groups and to use these variations as the basis for stratification. For example, we observe that older people relocate less frequently than younger people. There undoubtedly are good psychological reasons for this, which we do not attempt to articulate. We simply select age as a dimension along which to subdivide the population, capitalizing on an observed pattern.

In selecting a set of dimensions, we learned very early that single characteristics taken one at a time are not very meaningful. All old people do not behave in the same way, any more than all young people do, or all white people, or all people without high school diplomas. We need a system of joint classification that identifies individuals along several dimensions simultaneously, and thereby begins to reveal some of the underlying causes of their behavior. Since we cannot keep tract of an unlimited number of combinations, parsimony is of the essence -- we must isolate those dimensions which, in combination, carry the greatest amount of information about a person's or household's likely behavior with respect to the surrounding urban environment. Age and education, for example, would be a good set of dimensions because they describe different aspects of a potential migrant or homeowner. Age, education, and income would be a poor set because income and education are highly correlated; knowing one, the other is, for the most part, redundant. The income dimension "takes up space" and adds very little new information.

Two characteristics of individuals that seem immediately obvious as candidates for separate dimensions are age and ethnic/racial background. It is well known that housing needs and preferences vary significantly over the life cycle. Young people experiment with apartment living and "exciting" neighborhoods. Families with school-aged children pay far more attention to schools and social status. Older people want to be near relatives and want a minimum of residential up-keep. Mobility rates likewise vary greatly with age.

No careful observer of an urban area in the United States can ignore ethnic/racial differences. The clear separation of different groups, and the tensions at the boundaries, are major influences in determining who lives where, and such differences must be accounted for.

Beyond age and ethnic/racial composition, the problem of classification becomes more difficult. Ideally, we would prefer to maintain several additional dimensions, such as income, occupation, education, marital status, and family size. Practically speaking, however, we cannot afford much more than one additional dimension within the limits of present-day computing equipment and our data sources. Initially, we thought income would be the most important additional piece of information to record. Some experiments with census data have persuaded us, however, that education is a more powerful third dimension. A middle-aged carpenter without a high school diploma may well earn as much as a young lawyer, but the two are likely to make quite different residential location decisions. Groups with the same educational background, on the other hand(particularly when age and ethnic/racial background are controlled), are far more likely to behave in the same way. Education serves as a proxy for expected life-time income, and it is this expected, rather than current, income that influences location choices.

As a consequence, we have divided groups of people (and household heads) along the following three dimensions:

Dimension	Categories
Age	1. 0-19 2. 20-39 3. 40-64 4. 65 and Over
Racial/Ethnic Group	1. Native (non-minority) 2. Dominant, non-minority ethnic group 3. Minority
Educational Attainment	1. Less than 12 years 2. 12 years (high school) 3. Over 12 years

Thirty-six population types are defined jointly from three dimensions by considering all the combinations. These types are:

		Native			Foreign			Minority		
	Education:	< HS	HS	> HS	< HS	HS	> HS	< HS	HS	> HS
Age	0-19	1	2	3	4	5	6	7	8	9
	20-39	10	11	12	13	14	15	16	17	18
	40-64	19	20	21	22	23	24	25	26	27
	65	28	29	30	31	32	33	34	35	36

The dominant, non-minority ethnic group varies from region to region. In New Haven, for instance, it is foreign-born Italians. In Houston, it is Mexican-Americans. Charlotte appears to have no such group -- at least none of sufficient size to warrant its identification in the context of our model.

Natural Increase

The processes by which people are born, age, obtain an education, and eventually die, are of great importance to neighborhoods, particularly older ones. In many cases, natural increase is the only source of population growth, barely compensating for the out-movement of more mobile persons. We thus cannot afford to ignore it in our modeling work. Four processes are particularly important: birth, death, aging, and upward educational mobility.

Birth

The number of children born in each of the nine ethnic/educational categories is estimated by assuming: (1) a type-specific birth rate for each of the 36 cells, and (2) a 50 percent female population in each cell.

Death

Deaths are simulated in much the same way as births. Knowing 36 type-specific death rates, we obtain the number of deaths in each cell in each tract by simple multiplication.

Aging

Aging is a little more difficult to simulate. In any model, there is a constant trade-off between the level of detail and size of the model. It is prohibitively expensive to store the year-by-year age of each individual in each tract. Instead, we store the number in each age group, and assume that a certain percentage of them are in the last year in that group, i.e. that they are 19 or 39 or 64, based on historical age distribution curves for the area.

Educational Mobility

The aging of children into the adult category -- from 19 to 20 -- poses a problem. One cannot assume, for example, that the child of a foreign-born parent with less than a high school education will not aspire to more education. Quite the contrary, all the evidence suggests that he or she will. Furthermore, in almost all cases, the child of a foreign-born parent is not foreign born, but native, according to our definitions. Young adults must thus be pooled, and then moved upward educationally and into the native category from foreign born.

Defining the educational attainment of children before they are 20 years old poses another problem. Clearly, they do not have the same educational attainment as their parents. All those age 15 or younger have less than a high school education, a few are high school graduates, and virtually none has completed much college by the age of 19. On the educational dimension, therefore, children are pooled by ethnic group and assigned to educational categories on the basis of historical experience.

Flows To and From the Region

Our initial efforts to simulate neighborhood change were constantly confounded by our lack of knowledge about who was or was not flowing across the metropolitan area's boundaries. In that initial work, we were forced to make assumptions, based on past trends, about which people and which businesses would migrate in and out. Recent evidence suggests that past trends will not necessarily repeat themselves in the future. Thus, it is important to have a consistent, reasonably up-to-date picture of migration flows.

In response to the need for such information, we undertook a separate project to measure and model the flows of people and businesses from place to place. A recent report [2] describes its methods and products. In brief, the model provides an estimate for each year of the number of people (by type) and the number of jobs (by type) that will move in and out of the area for each of the regions we are examining.

There remains the job of deciding where within the region arriving households will settle, and which areas out-migrants will leave. Our evidence on those arriving suggests that they find out about neighborhoods primarily through friends or relatives already there, and hence tend to settle in areas where people like themselves are located. We have them locate in the places that local movers of the same type choose for themselves. The way those choices are made is described in the next section.

The location of vacancies created by households that leave is more difficult to determine. We do know that out-migration streams tend roughly to equal in-migrant flows, and that almost half of the out-migrants are "chronic movers" -- that is, they tend to move out within two or three years of the date they moved in. They are more likely to be apartment dwellers than homeowners. Beyond that, we can only assume that they are distributed across neighborhoods in proportion to the number of people of their same type.

Mapping Between Households and Population

Having decided to maintain a parallel accounting of households and the people forming them, we are left with the problem of mapping any simulated change in the number or composition of households (through migration, for example) into corresponding changes in the population, and vice versa. This is largely an accounting exercise, capitalizing on our knowledge of household structure. It is described in detail elsewhere and will not be repeated here. The interested reader is referred to Birch et al. [1] for a description of how it is done.

Mathematical Statement

Most of the calculations related to births, deaths, aging, and educational mobility are done by age group. For this reason, the 36-element population vector is transferred into a 4 x 9 array (dimensioned J by L), each of whose rows represents the ethnic/racial and educational breakdown of each age group, as follows:

	L	1	2	3	4	5	6	7	8	9
		Native			Foreign			Minority		
J	Age	< HS	HS	> HS	< HS	HS	> HS	< HS	HS	> HS
1	0-19	1	2	3	4	5	6	7	8	9
2	20-39	10	11	12	13	14	15	16	17	18
3	40-64	19	20	21	22	23	24	25	26	27
4	65+	28	29	30	31	32	33	34	35	36

Births produced by each ethnic-education Type L in each neighborhood are:

$$\text{Births (L)} = \sum_{K \in 1} \text{POP(K)} * .5 * \text{FERT(K)}$$

where:

POP(K) = the population of Type K in Neighborhood I and

FERT(K) = is the fertility rate of Type K.

We assume a constant sex ratio of .5, and suppress the neighborhood index for simplicity.

Similarly, deaths for each J, L combination are:

Deaths(J,L) = Pop(K) * Mort(K)

where Mort(K) is the mortality rate of persons of Type K.

Newborn children obviously have no formal education, nor have many of the persons younger than 20 years of age attained their final levels of education. For these reasons, we pool children into the three ethnic/racial groups. We assign those who remain children for the next period (all those currently under age 19) to education levels based on historical experience. Most, of course, have less than a high school education; a few have completed high school.

For those who do "graduate" into adulthood (i.e., become 20 in our model) we estimate how many will end up in each of the education groups as adults, and assign them accordingly. In addition, the "graduates" whose parents were foreign-born will become natives themselves. An adjustment is made to reflect this shift also.

Aging of those over 20 is handled computationally by assuming that a constant fraction (based on detailed census data for the region) of each ethnic/education group is in the final year of the group (i.e. is 39 or 64) and moving that fraction into the next group:

Agers(J,L) = Pop(J,L) * Tran(J,IR)

where:

Agers(J,L)	=	is the number of people of ethnic-education Type L aging out of Age J, and
Tran(J,IR)	=	is the proportion of people of Ethnic IR in the final year of Age Group J.

Graduates (those moving from age 19 to 20) are a special class of agers because they adopt their final education and ethnic classification in the process of aging.

The accounting is straightforward. For children:

$$\mathrm{Pop}(1,L)_t = \mathrm{Pop}(1,L)_{t-1} - \mathrm{Deaths}(1,L) + \mathrm{Births}(L) - \mathrm{Agers}(1,L)$$

For adults:

$$\mathrm{Pop}(J,L)_t = \mathrm{Pop}(J,L)_{t-1} - \mathrm{Deaths}(J,L) - \mathrm{Agers}(J,L) + \mathrm{Agers}(J-1,L).$$

As a final step, the J,L array is transformed back into the 36-element population vector for each neighborhood.

REFERENCES

1. Birch, Atkinson, Sandstrom, and Stack. The New Haven Laboratory: a Testbed for Planning. Lexington, Massachusetts: Lexington Books/D.C.Heath, 1974.

2. Birch, D., P. Allaman, and E. Martin. A Model of Population and Employment Change for Metropolitan and Rural Areas in the United States. Cambridge: MIT-Harvard Joint Center for Urban Studies, 1977.

Section 5

RESIDENTIAL LOCATION

Our review of the literature [1] suggested that the behavior of households is the key determinant of urban form. The planner, the builder, the banker, the landlord, and the school superintendent are all quite sensitive to the household's needs. The shape of our urban regions appears largely demand rather than supply-driven. Suppliers of housing, land, and services may experiment with new ideas at the margin, but only at the margin, and more often than not, under duress rather than of their own volition.

Our review of the literature also revealed a considerable divergence of opinion about the way residential location decisions are made. The field is dominated not so much by disagreements as by a collection of assertions, few of which have been tested, and most of which either ignore or do not acknowledge the others. We designed our field work to force confrontation and, we hope, resolution of some of the more important differences. We have postulated a model that is consistent with the facts before us, and that replicates behavior as we observe it in the six regions.

Mobility

While there is considerable disagreement over the way households choose neighborhoods, modelers generally tend to agree on how households arrive at decisions to move. Virtually every incremental model we have investigated[1] separates the decision to move from the choice of destination. We continue in that tradition. We assume that each household has some notion of the kind of environment that is suitable for it at any point in time, that each household is constantly comparing its present circumstances against its own ideal standard, and that the household is measuring and talking over dinner about the degree of mismatch.

For the most part there will always be a mismatch. Any change in the household's circumstances will affect the match, as will any change in the present neighborhood. If there were no financial costs of relocation and no interpersonal ties with neighbors and nearby relatives, many households would be moving every few months to achieve a closer match. The fact is that thresholds impeding relocation are real -- moving is widely perceived as "a pain in the neck," being close to neighbors and relatives is important to many, changing schools is difficult for some children, leaving a house into which one has put much effort is sometimes difficult, and so forth.

In practice, then, it takes a fairly significant change of circumstances to cause a move. The impetus might come from a change in the neighborhood (development of adjacent land, change in racial or economic balance in the area, construction of a big industrial plant nearby), a significant change in personal circumstances (change or relocation of job, substantial change in income and hence aspirations, new child, departing child, departing wife or husband, transfer to another region, racial integration of nearby schools, retirement, graduation from school), or a dramatic change in options (attrac-

---

1. See, for example, Butler and Chapin [4] and Ingram, Kain, and Ginn [6].

tive new condominiums requiring little personal maintenance work, a new highway or rapid transit line bringing outlying areas within commuting range, a new trailer park near a better school system).

Significant changes in personal circumstances tend to occur during younger years. The threshold at which households decide to relocate increases with age; hence the observed statistical propensity for younger households to have higher mobility rates (see Table 5-1). Similarly renters, who have a much smaller stake in their units, are far more likely to move than are those who own their homes. The statistics do not imply, however, that the odds of a given type of household moving are constant over time and over space. If we are to explain the variations in mobility rates at the neighborhood level, we must recognize that the propensity to move is a continuously changing function of the conditions at origin, conditions at destination, and personal circumstances.

Choice of a New Location

One can safely predict that a majority of moves will be to another housing unit in the same town, and probably in the same neighborhood. Staying nearby minimizes the pain of breaking human and institutional ties.[1]

When some stimulus forces a wider search, the traditional analyst usually employs some variant of the "attractiveness score" concept to predict the final destination. Using this approach, the analyst assumes that a household rates neighborhoods in terms of some weighted average of desirable

---

1. See Volume 4, Section 2 of this report (The Behavioral Foundations of Neighborhood Change) for evidence of the narrowness of the search procedure.

Table 5-1

Percentage of Households Moving Each Year, for the New Haven SMSA in the 1960's

Owners	Native			Foreign Born			Minority		
Age	LHS	HS	GHS	LHS	HS	GHS	LHS	HS	GHS
20-39	3.9	4.0	4.0	6.1	4.0	4.2	4.4	3.9	3.9
40-64	1.3	1.3	1.3	1.5	1.5	2.8	1.5	2.8	2.8
65 and over	.8	.6	.4	.6	.6	.6	2.0	2.0	2.0

Renters	Native			Foreign Born			Minority		
Age	LHS	HS	GHS	LHS	HS	GHS	LHS	HS	GHS
20-39	20.7	23.0	23.3	20.9	29.1	29.1	23.7	20.9	20.9
40-64	11.1	8.6	9.8	12.9	10.0	13.7	10.9	20.9	20.9
65 and over	10.6	9.1	7.3	10.4	2.8	5.9	6.6	6.6	6.6

These rates cannot be obtained directly from census data because of the manner in which census data are presented. Our procedure for deriving these rates is described in Appendix A at the end of this section.

1. The three ethnic/racial categories are: (1) Native (everyone not foreign or minority), (2) Foreign Born (everyone born abroad), and (3) Minority (non-white or Puerto Rican). The three education levels are: "LHS" = less than a high school degree, "HS" = a high school diploma, and "GHS" = greater than a high school education.

properties, and chooses the neighborhood with the highest score. Economists call the score (or common denominator) utility, and the procedure, utility maximization. The implication is that the household is capable of, and interested in, conducting an exhaustive search and is constantly re-evaluating a function of many variables before making a decision.

To date, all the evidence from our field work suggests that this is not how the process works. First, the process appears to be hierarchical, and the first step in the hierarchy is suppression, not selection. Before a serious search begins, the household almost automatically eliminates from consideration a very high percentage of all neighborhoods. The reasons are many. First, the time and cost of seriously "looking" in many neighborhoods are prohibitive. Second, a resident of a region is better acquainted with some areas within that region than with others. Apparently people's images of a region are quite sharp in the vicinity of their own neighborhoods and along the major artery by which they travel to and from the city, but fuzzy at best for other parts of the area.[1] The search for a new location is concentrated, for the most part, in the high-resolution section of the region. When charted, the high-resolution portion is frequently limited to a pie-shaped wedge radiating out from the center of the city in the general direction of the household's present location. As a check on this proposition, we determined in our field work the location of the respondent's previous and intended next residences. These locations were coded to four quadrants defined by the Census Bureau, running out from the center of the region. A strong preference for staying in the same, or an adjacent, sector emerged.

---

1. See Moore [11] and Johnston [7] for a thorough investigation of mental maps and their role in residential location decisions.

	Percentage Previous Move	Percentage Next Move
Same Sector	70.5	70.3
Adjacent Sector	19.6	19.8
Distant Sector	10.9	9.9

Clearly, households do not view all places in the region as equally plausible places to relocate. Within their own sectors, however, they appear to be aware of trends taking place outside of their immediate neighborhoods. Of particular importance in this regard is transition of race and/or class. Households are aware of racial transitions and general downgrading of housing values due to "lower-class folks moving in" to adjacent neighborhoods. Such transitions appear to cast a "shadow" beyond the immediate location in which they are taking place, adding a reason to leave for those who reside nearby.

Access to work has long concerned urban modelers. In fact, in the 1960's, it dominated the thinking of many modelers such as Lowry [10], Kain [8], Muth [13], and Alonso [1], as well as the whole stream of gravity-model development described in Section 2.

Recent, more behaviorally oriented work by Butler et al. [4], Moore [11], and others suggests that access to jobs is indeed important, but not as originally postulated. The initial formulation was that households regard every additional minute of travel time as a drawback which influences the evaluation of residential sites. The recent work cited above, borne out by our findings, suggests instead the existence of a threshold phenomenon.

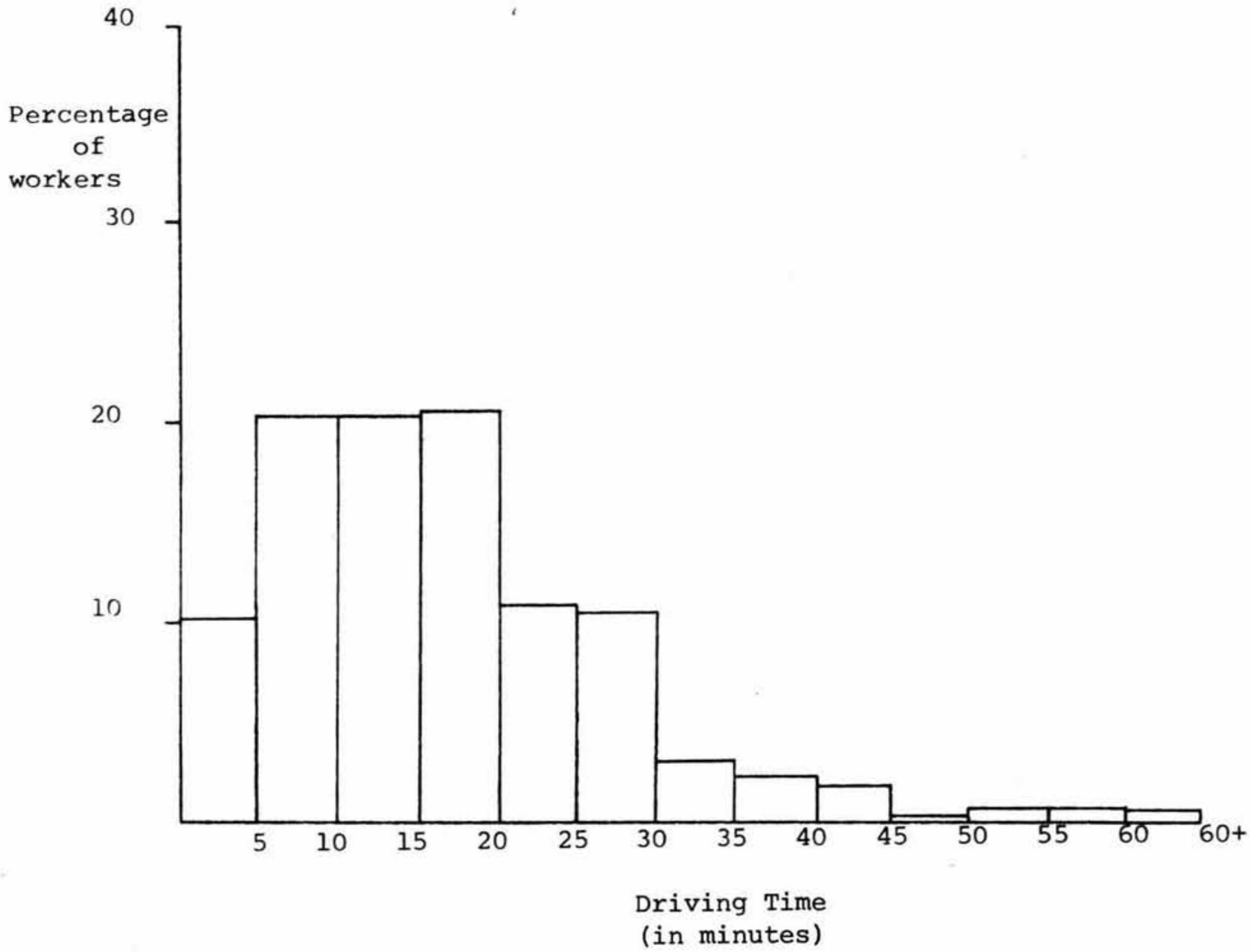
People are willing to travel up to 35 or 40 minutes without much complaint. Beyond that point, however, they become quite averse to additional time spent in transit. In Figure 5-1 we plot the actual work-trip travel times of our respondents and find a sharp dropoff around 30 minutes. These results are consistent with NBER's [6] survey work, which finds relatively little residential movement associated with small shifts in job location, but a significant effect when the job location changes greatly. In Appendix B of this section, we show that this threshold model of individual behavior is consistent with the exponential decrease in aggregate density from the center as observed and modeled by the gravity modelers.

By the time a household has suppressed all neighborhoods outside the high-resolution part of its map of the region, more than 35 minutes from work, and unacceptable because of nearby racial or class transition, very few neighborhood choices remain.[1] The situation is diagrammed in Figure 5-2. The choice process now becomes one of selection and seems to hinge on some of the more traditionally cited determinants -- social status (the "appropriateness of the address"), quality of the schools, availability of a unit that meets the household's specifications for space, yard, appliances, cost, safety, and so forth. Tiebout [14] and others have argued that households weigh tax-service packages carefully in making their neighborhood choices. On the contrary, in our surveys, we have found that households pay very little attention to differences in levels in public services, with the exception of schools. Friedman's study [5] "Housing Location and the Supply of Local Public Services," was specifically designed to test the proposition as set forth by Tiebout and others. This group at Berkeley

1. In Volume 4, Section 2 we show that 90 percent of households looking for a new residence search in three or fewer neighborhoods.

Figure 5-1

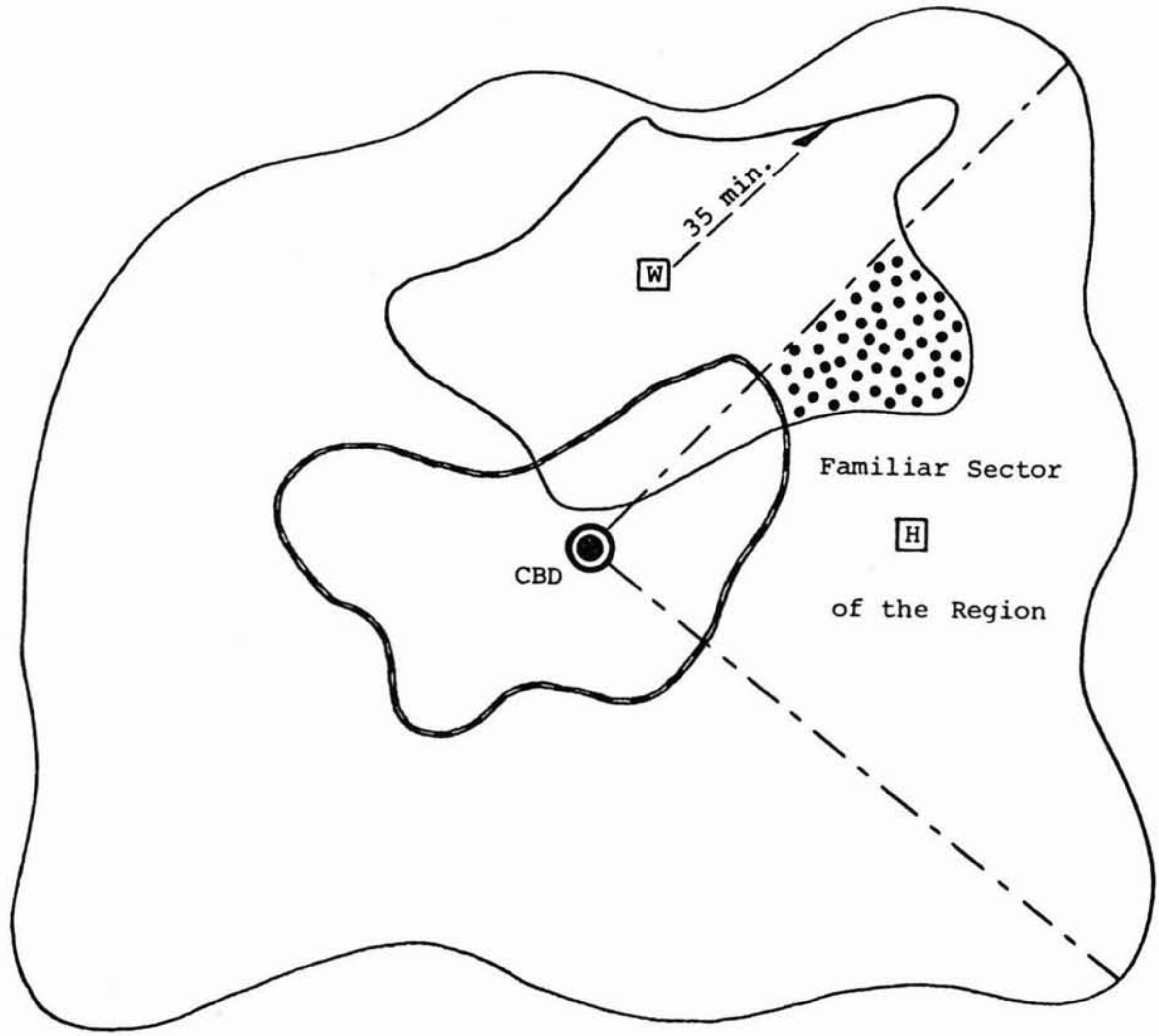
Distribution of Driving Times
for Workers in
Houston, Dayton and Rochester



See Volume 4, Section 2 of this report for very slight variations from city to city in this function.

Figure 5-2

Neighborhoods Remaining After Suppression



Legend

[W] Work Place

[H] Present Home

Socially or Racially Unacceptable

Set of Likely Next Moves

reaffirms our conclusions:

> The empirical results allow us to conclude that the local public services play only a minor role in affecting the residential choice.

We have thus decided for the moment to focus our attention on public schools and to exclude other public services from consideration.

The Mover Submodel

As indicated above, we have separated the decision to move from the choice of destination, a strategy suggested and followed by many. In our Mover Submodel, the decision to move is initially a function of age, race, education level, and tenure (homeowners vs. renters). Within each age-race-education-tenure group, we begin with a constant rate throughout the region similar to those presented in Table 5-1. We modify these rates to reflect local pressures caused by racial transition and forced moves caused by demolitions. A final, implicit adjustment is made as a result of the market clearing process. Those unable to find suitable housing elsewhere end up living in the same neighborhood and, quite probably, the same unit in which they began the period.

The second major step in the Mover Submodel is to determine where movers of each type will look for new units. As our theory of neighborhood choice evolved, it became increasingly clear to us that its emphasis on rejection, its thresholds and discontinuities, and its potential reliance on interactive terms would push existing techniques for specifying and fitting equations to, and perhaps beyond, their intended limits. Our early experiments confirmed this suspicion. We began, therefore, to develop a method of equation specification and fitting that was better suited to our purpose -- estimating probability density functions to express choices among neighborhoods in a world filled with discontinuities, thresholds, and interactive terms.

What evolved is a package we have called the Probability Function Generator (PFG). It bears some generic resemblance to regression analysis (it fits functions to data by minimizing an error term), but it operates differently, and is designed specifically to estimate probability choice functions.

The equations generated by the PFG require a notation system of their own, which is described, along with the package itself, in Appendix C. The equations will look unusual to a reader unfamiliar with the PFG. Social scientists have spent decades learning how to read regression results, and it is most inconvenient to have to learn a new system. We feel, however, that regression analysis is inadequate in many instances. We have tried to make the PFG notation a mnemonic system, and we have retained the R^2 measure of fit (along with our own measure) to permit comparison.

Like any other system, the PFG package requires that the user specify the structure of the equations to be fitted. And, as always, the problem is selecting the best set of surrogate variables to capture as much of the underlying theory as possible. As present we are using the following variables:

Variable	Definition and Rationale
Slots	The number of available vacancies vacated by people of the same type. The Slots variable reflects the tendency for particular groups to concentrate in different sectors, and the tendency for people to seek out others of their own type.

Contig	A measure of racial transition in contiguous neighborhoods. While varying from ethnic group to ethnic group, Contig is used as a binary filter to separate neighborhoods that are acceptable (Value = 1) from those that are cast under some sort of shadow due to events or conditions in adjacent neighborhoods (Value = 0). For white natives and foreign stock, Contig goes to zero when racial transition is taking place nearby. For minority groups, the picture is reversed -- Contig goes to 1 when it seems that racial transition is possible.
Pmin	Percentage minority. Pmin describes the minority concentration in the particular neighborhood under consideration. Different groups are obviously sensitive to this measure in different ways. For many natives and foreign stock, and for quite a few minority households, high minority concentrations are a cause for rejection. Lower concentrations, on the other hand, are frequently a sign that racial transition is possible.
PFor	Percentage foreign. PFor describes the amount of foreign stock present in a neighborhood. In some areas, like Charlotte, this measure has little significance. In others, like Houston, where the Spanish-speaking population is classified as foreign stock, it is quite important. In particular, the Spanish and black populations tend to occupy different territories, and the natives tend to avoid both.
Jobacc	Job accessibility. Our finding of a tolerance for 35 minutes of travel time places less importance on defining the exact travel time from each neighborhood to each work place, and places more emphasis on identifying those neighborhoods that are generally remote for most people. Thus far, we have found it sufficient merely to identify in Jobacc the total number of jobs in the region that can be reached from a particular neighborhood within 35 minutes as a measure of its accessibility. Our initial formulation took account of the occupation of each household head and the occupational distribution of jobs in each neighborhood, on the theory that a concentration of one type of employment in one sector of the region might influence the residential choices of a subset of the population

in that sector. Elaboration beyond the original calculations did not increase the explanatory power, however, even when tested in the Houston, the most segmented city in our set. We have therefore reverted to the total job accessibility measure.

Class	A measure of social class, defined on a scale from 1 to 5, 5 being high. Class is a composite index reflecting the education level, incomes, and occupations of the neighborhood. It serves as a surrogate for the cost and prestige associated with living in a particular area.

Different types of households, of course, attend and react to each of these variables in their own ways. The PFG package is designed specifically to determine how each type of household uses each filter, or decision rule, defined by the variable. The end product is a set of 27 equations (one for each household type) that identify the filters used, the nature of their uses, and the sequence in which theyare used.

Several conclusions can be drawn from these equations. First, they explain a great deal of the variance in moving behavior in most cases where sample sizes are significant, suggesting that our theory is on the right track.

In general, job accessibility is not a major factor in New Haven, as our theory would suggest, since it is possible to drive from any point in New Haven to any other point in less than 35 minutes. The equations for the much more expansive Houston region, in contrast, rely somewhat more on the accessibility measure to explain location behavior.

Slots (the Base variable) appears to be a good starting filter. Households generally seem to fill vacancies created by the departure of similar kinds of households. It is the variations on this theme that are interesting, however. In general, class and race create the main exceptions to this replacement hypothesis, and there are distinct variations across ethnic/racial groups.

The native households tend to vary considerably among themselves according to education. Households with lesser educated heads tend to reject areas with any substantial minority concentration, while the younger, college-educated group selects areas with more than a five-percent minority concentration. There is also a noticeable tendency for natives to select areas with a lower social class than the areas where replacement opportunities exist. This is a curious phenomenon, and suggests that, in the face of rising construction, energy, and mortgage costs, households are finding ways to utilize the existing stock more effectively -- perhaps even revitalizing it.

The foreign-born households systematically reject areas with high concentrations of minority groups and areas that are on the upper end of the class spectrum. They appear to tolerate, however (perhaps against their will), small minority concentrations. In New Haven, mixed neighborhoods may be the only ones in which Italian and Irish immigrants can afford to rent or buy a unit.

The minority households act according to the replacement hypothesis to a certain extent. They select areas with existing minority concentrations and, in general, stay away from neighborhoods in the high-class bracket. Two exceptions to this pattern are apparent however. First, minority households systematically reject areas with minority concentrations that are very high (generally over 50 percent). Second, better-educated, minority households seek out areas contiguous to neighborhoods with existing minority concentrations. In other words, minority households are avoiding excessive minority concentrations almost as much as the native and foreign-born households are, preferring more mixed neighborhoods. The "opening up" of such neighborhoods is, apparently, being undertaken by the younger, better-educated segment of the minority population.

These patterns are quite consistent across all six regions. The weights assigned to different factors vary slightly from place to place, but the same filters and rules crop up over and over again for each respective group of households.

Clearing the Market

The final product of the demand equations described above is a set of 27 probability density functions, each one of which describes the odds that a particular household type will search for a unit in each neighborhood. Generally, the demand for units does not match the supply. We must, therefore, devise some method of simulating the bargaining that goes on among households for units to create a proper match. We call this bargaining "clearing the market."

We clear the market in two steps: (1) balance the supply and demand within each of six housing submarkets[1] region-wide, and (2) clear each submarket across all neighborhoods and household types.

The region-wide balancing is accomplished by first estimating how many of each kind of unit would be purchased or rented if all households in the market for a unit could have their first choices. This set of first choices

---

1. A detailed description of the six markets is provided in Section 6. Briefly, they are made up of three price categories (high, medium, and low) for two tenure classes (owned and rented).

is based on the historical propensities for each household type to live in a particular unit type,[1] modulated by the tendency for renters to shift to owned units and for all movers to move up in price each move. The first choices are then adjusted using a standardization technique described by Mosteller [12]. This technique adjusts the preferences the smallest amount necessary to achieve a proper balance between supply and demand.[2]

The second step is to sort out who will live where. Using the revised preferences described above and the likelihood that each group of households will choose each neighborhood, we can estimate in a straightforward manner how many units of each type would be demanded in each neighborhood if every household could have its first choice. As in the region-wide case, however, a perfect match is rarely achieved. In general, there will be more (or less) demand for, say, medium-priced owned units in Neighborhood 10, than there are units of that type in Neighborhood 10 because no one "orchestrates" the house hunting each year. Again, we turn to the standardization technique to resolve the difficulty (with a minimum of distortion) and match demanders to units.

When this two-step process is complete, we are assured that each kind of household has found a unit in a neighborhood that comes as close as possible to satisfying its initial preferences. All that remains is the accounting to record the final decisions.

---

1. See Volume 4, Section 2 of this report.

2. See Birch et al. [3], Chapter 4, for a description of the technique and its use.

Mathematical Description of the Mover Submodel

The model simulates, in sequence: (1) the number of movers by type, (2) the destination chosen by each mover group, and (3) the market clearing process.

Mobility

The number of households of Type K moving from Neighborhood I is:

$$\text{Mover}(K,I) = HH(T,K,I) * MOB(T,K) * ADJ(I) + \text{FMove}(K,I)$$

where:

Mover(K,I) =	The number of movers of Type K moving from Neighborhood I.
HH(T,K,I) =	The number of households of Type K living in a unit with Tenure T in Neighborhood I.
MOB(T,K) =	The expected mobility rate of households of Type K living in units with Tenure T.
ADJ(I) =	Adjustment to reflect stronger than average pressures in Neighborhood I due to racial transition nearby. ($ADJ(I) \in [1,0, 2.0]$)
FMove(K,I) =	The number of households Type K which are forced to move from Neighborhood I because demolitions exceed the expected turnover in the stock.

Choice of Destination

An initial set of neighborhood destinations for each household type is derived from the PFG[1] as follows:

---

1. A description of the PFG package and how it works is presented in Appendix C. Appendix D presents the equations derived for New Haven as an example of the final PFG product.

$$X(K,I) = [LM(K) + Imig(K)] * Prob(K,I)$$

where:

X(K,I)	=	The number of households of Type K arriving in Neighborhood I.
LM(K)	=	The number of local movers of Type K. LM is the sum of Mover(K,I) over I.
Imig(K)	=	The number of households of Type K coming from outside the region.
Prob(K,I)	=	The probability that a household of Type K will choose Neighborhood I as the site for a new residence. Prob is derived from the PFG package.

Clearing the Market

The first step in clearing the market is to estimate each household's initial preference for a unit type. We assume that these preferences are reflected in existing housing occupancy patterns, after adjusting for the known tendencies of renters to move more often than owners and the tendency for all movers to move up in price. Explicitly,

$$BP(K,JH,JP) = Tendm(JH,K) * Tpdm(JH,JP,K)$$

where:

BP(K,JH,JP)	is the likely bidding preference of a Type K household for a unit of Tenure JH and Price JP.
Tendm(JH,K)	is the odds that a Type K household will choose a unit of Tenure JH.
Tpdm(JH,JP,K)	is the odds that a household of Type K looking for a unit of Tenure JH will select a unit of Price JP.

Tendm is initially estimated as follows:

$$Tendm(JH,K) = \frac{\sum_{JP} HH(JH,JP,K)}{\sum_{JH}\sum_{JP} HH(JH,JP,K)}$$

where HH(JH,JP,K) is the number of households of Type K presently living in units of Tenure JH and Price JP. Tendm is then adjusted to reflect the greater propensity for renters to move.

$$\text{Tendm}(1,K)' = \text{Tendm}(1,K) * a$$

$$\text{Tendm}(2,K)' = \text{Tendm}(2,K) + (\text{Tendm}(1,K) * (1-a))$$

where $a \in (0,1)$.

Similarly Tpdm is calculated initially as follows:

$$\text{Tpdm}(JH,JP,K) = \frac{HH(JH,JP,K)}{\sum_{JP} HH(JH,JP,K)}$$

It is then adjusted to reflect upward price mobility:

$$\text{Tpdm}(JH,1,K) = \text{Tpdm}(JH,1,K) + (\text{Tpdm}(JH,2,K) * a)$$

$$\text{Tpdm}(JH,2,K) = \text{Tpdm}(JH,2,K) - (\text{Tpdm}(JH,2,K) * (1-a)) + \text{Tpdm}(JH,3,K) * b)$$

$$\text{Tpdm}(JH,3,K) = \text{Tpdm}(JH,3,K) - (\text{Tpdm}(JH,3,K) * (1-b))$$

where a and $b \in (0,1)$.

Before market clearing, the region-wide demand for each submarket is:

$$\text{UCRD}(K,KH) = \text{TAR}(K) + \text{BP}(K,KH)$$

for all K,KH, where:

UCRD(K,KH)	is the uncleared region demand for Unit Type KH on the part of Household Type K, and
TAR(K)	is the sum of all arrivers of Type K throughout the region.

The standardization procedure described above adjusts UCRD to meet the demand constraint

$$\sum_{KH} \text{UCRD}(K,KH) = \text{TAR}(K)$$

for all K (arrivers cannot be created or destroyed) and a supply condition:

$$SUPR(KH) = \sum_{K} UCRD(K,KH)$$

for all KH where SUPR(KH) is the supply of units of Type KH in the region.

SUPR is defined as follows:

$$SUPR(KH) = \sum_{I} AVAC(KH,I) - (SUMS-SUMD)*EXPVAC(KH)$$

for all KH, where:

AVAC(KH,I)	is the available vacancies of Type KH in Neighborhood I,
SUMS	is the total supply of units in the region (defined as the sum of AVAC over all unit types and neighborhoods),
SUMD	is the sum of total demand in the region (defined as the sum of arrivers over all arriver types and neighborhoods), and
EXPVAC(KH)	is total regional vacancies expected for units of Type KH.

This expression adjusts the total supply to allow for a "normal" vacancy rate in each submarket, given the overall tightness of the market (SUMS - SUMD). EXPVAC is defined as follows:

$$EXPVAC(KH) = \frac{STOC(KH)*VSTD(KH)}{\sum_{KH} [STOC(KH) + VSTD(KH)]}$$

where:

STOC(KH)	is the number of units of Type KH in the housing stock, and
VSTD(KH)	is the normal vacancy rate, based on census data, for units for Type KH.

The result of the standardization procedure is a revised, cleared estimate of regional demand -- CRD(K,KH) -- that meets both the region-wide supply and demand constraints. When normalized, this array represents a revised set of housing preferences -- BP'(K,KH).

The next step is to clear each housing submarket across neighborhoods. An initial estimate (now guaranteed to balance across the region as a whole) of arrivers into Type KH units, AR(KH,K,I), is

$$AR(KH,K,I) = X(K,I) * BP'(K,KH).$$

X(K,I) is defined above. The procedure used to find a revised version of AR is similar to the one used to match region-wide supply and demand. In general, AR will not meet the tract level supply condition -- that is, the number of units demanded in Neighborhood I of Type KH will either be greater or smaller than the number available. The number available is simply the number that have been vacated by departing households (after allowing for additions to and removals from the stock and an expected vacancy rate). Conceptually,

$$OccStoc = HH - Dep + Growth$$

where:

OccStoc	is the occupied stock at the end of the clearing process,
HH	is the number of households at the beginning,
Dep	is the number of departing households (including deaths and outmigrants), and
Growth	is the total new households arriving in the neighborhood during the current period.

Rearranging the terms,

$$Growth = OccStoc - HH + Dep.$$

More formally,

$$TGROW(KH,I) = STOC(KH,I) * (1 - VBASE(KH,I)) - \sum_{K} HH(K,KH,I) + \sum_{K} DEP(K,KH,I)$$

where:

TGROW(KH,I)	is the total possible arrivers occupying units of Type KH in Neighborhood I,
VBASE(KH,I)	is the anticipated level of vacancies in units of Type KH in Neighborhood I,
HH(K,KH,I)	is the number of households of Type K living in units of Type KH in Neighborhood I, and
DEP(K,KH,I)	is the number of departing households of Type K leaving units of Type KH in Neighborhood I.

VBASE(I) is the standard vacancy rate (VSTD) adjusted to account for: (1) the "excess demand" for units of Type KH in I, (2) the location of I in the region, and (3) the overall level of vacancies in the KH submarket. Excess demand is measured in terms of the ratio of arrivers before clearing minus leavers, over leavers. If it is positive (more people want to come than have left), the market is expected to be tighter. Conversely, a negative ratio indicates a looser market.

Markets farther from the central city tend to be tighter than those in the midst of the city's problems, and a slight adjustment is made to reflect this difference.

The final adjustment reflects the overall condition of the market and acknowledges the partial substitutability of units across neighborhoods.

In addition to the neighborhood supply constraint TGROW(KH,I), there is a region-wide household demand condition that must be met, i.e.,

DMD(KH,K) = CRD(K,KH).

where DMD(KH,K) is the total demand for units of Type KH by households of Type K. The standardization procedure adjusts the original estimates of arrivers (with a minimum of distortion) so as to meet both conditions.

This process is repeated for each submarket. When all six markets have been cleared, each moving household in the region has found a new unit in a new neighborhood, and we are guaranteed that there are enough units of the right type for each household -- that is, the bargaining is over, second choices have been made, and the market is cleared.

The final step is simply to record what has happened, updating the household and housing stock arrays to reflect the decisions that have been made.

Section 5

Appendix A

DERIVATION OF MOBILITY RATES

The mobility rate is defined as the percentage of households which move within the SMSA during the course of year. As might be expected, there is considerable variation among household types; young households are considerably more mobile than are older households. In addition, renters are far more mobile than owners, by a factor of about 5 to 1. This tenure relationship is observed for similar household types; young renters are considerably more mobile than are young owners. In the model, both household type and tenure-specific mobility factors are used. Specifically, the model uses a 2-by-27 dimension array, M(JH,K), where JH runs through tenure, and K runs through household type.

Unfortunately, there are no direct data on houehold type, tenure-specific mobility rates. Thus an indirect procedure is used to find M, based on data from the 1:100 Public Use Sample Tape [15] and the U.S. Census of Metropolitan Housing [16], as well as model results for the region as a whole. Specifically, the Public Use Sample provides

T(JH,K)	The number of households in the SMSA in 1960 which lived in the SMSA in 1955.
S(JH,K)	The number of households in the SMSA which did not move from 1955 to 1960.

The difference between T and S gives an estimate of the number of households which moved within the SMSA between 1955 and 1960. An initial value of M is obtained from

$$1-M(JH,K) = (S(JH,K)/T(JH,K))**(1/5).$$

However, this initial value does not take household formation and dissolution into account, so corrections are required.

The Census of Metropolitan Housing provides a direct estimate by tenure of the number of units vacated by households during 1959. Under the assumption that 1959 data provide a basis for a reasonable estimate for 1960, we used an SMSA-level model to calibrate the mobility rates -- that is, the sum of units by tenure vacated by metropolitan area movers, out-migrants, and dissolved households is computed. Then mobility factors are scaled (by tenure) so the above sums for 1960 from the model agree with the number of vacated units provided by the Census of Metropolitan Housing.

Section 5

Appendix B

CONSISTENCY OF MICRO AND AGGREGATE MODELS OF JOURNEY-TO-WORK BEHAVIOR

How can we resolve the observed tendency for residential densities to fall off almost exponentially from the central work place (particularly in older cities), with an observed indifference to travel times up to 35 minutes? To satisfy our own curiosity about this question, we constructed a miniature model.

We postulated a simplified city, with employment opportunities concentrated at a point in the center, and with ten residential districts, consisting of ten concentric rings around the central point. Each successive ring is five minutes in travel time farther from the core.

Our simulation of this city traces its growth over a 200-year period from 1770 to 1970, in twenty discrete time periods. In this simulation, we have made the following assumptions.

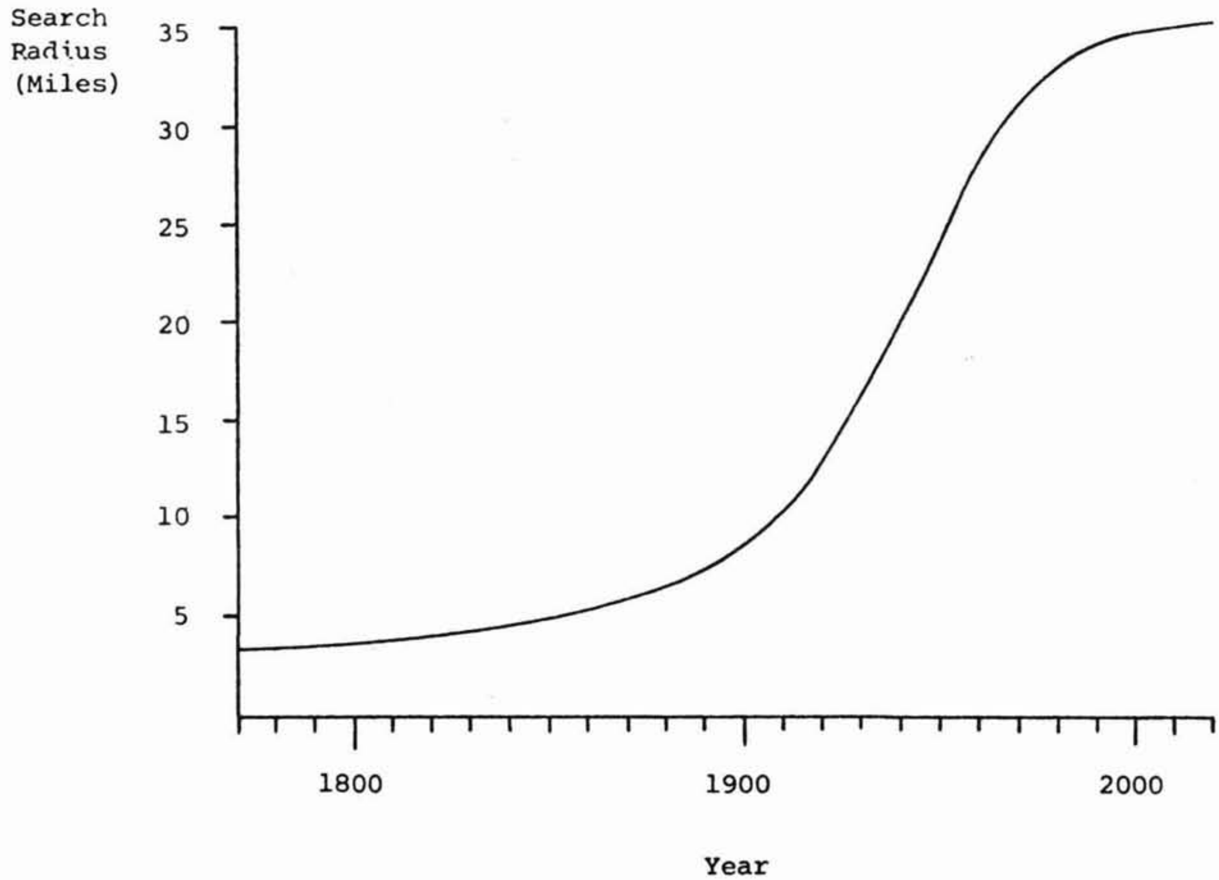
1. The employment remains concentrated at the center.

2. The average distance that can be covered in 35 minutes of travel time varies over history as in Figure 5B-1, which reflects changes in transportation technology, with the advent of the trolley during the Industrial Revolution, the automobile around 1920, and major expressways during the 1950's and 1960's.

3. Travel time to work is the only determinant of residential choice, and the choice function is a discrete variant of that demonstrated in Figure 5-1. In particular, we assume all districts (rings) that fall within the search radius have an equal probability of being selected; all those outside of the radius have no probability of being selected.

Figure 5B-1

Search Radius for Moving Households, 1770-2020



4. In 1770 the regional population is 50,000 households, all concentrated in the central ring.

5. The number of households grows from 50,000 in 1770 to 300,000 in 1970.

The simulation demonstrated that the period in which growth takes place is a major determinant of the shape of the density function. Simulation of an old city, such as Pittsburgh, whose growth took place during the Industrial Revolution (1830-1880), produced a compact city, shown in Figure 5B-2. As is clear from the figure, most of the residential construction takes place when the search radius is relatively small, placing much of the infrastructure in the inner rings. By the time wider access is possible, the growth has slowed, and the shapes of the population and density curves are not greatly affected. The curves have the gravity-like shape analysts have observed, despite the discontinuous search rule.

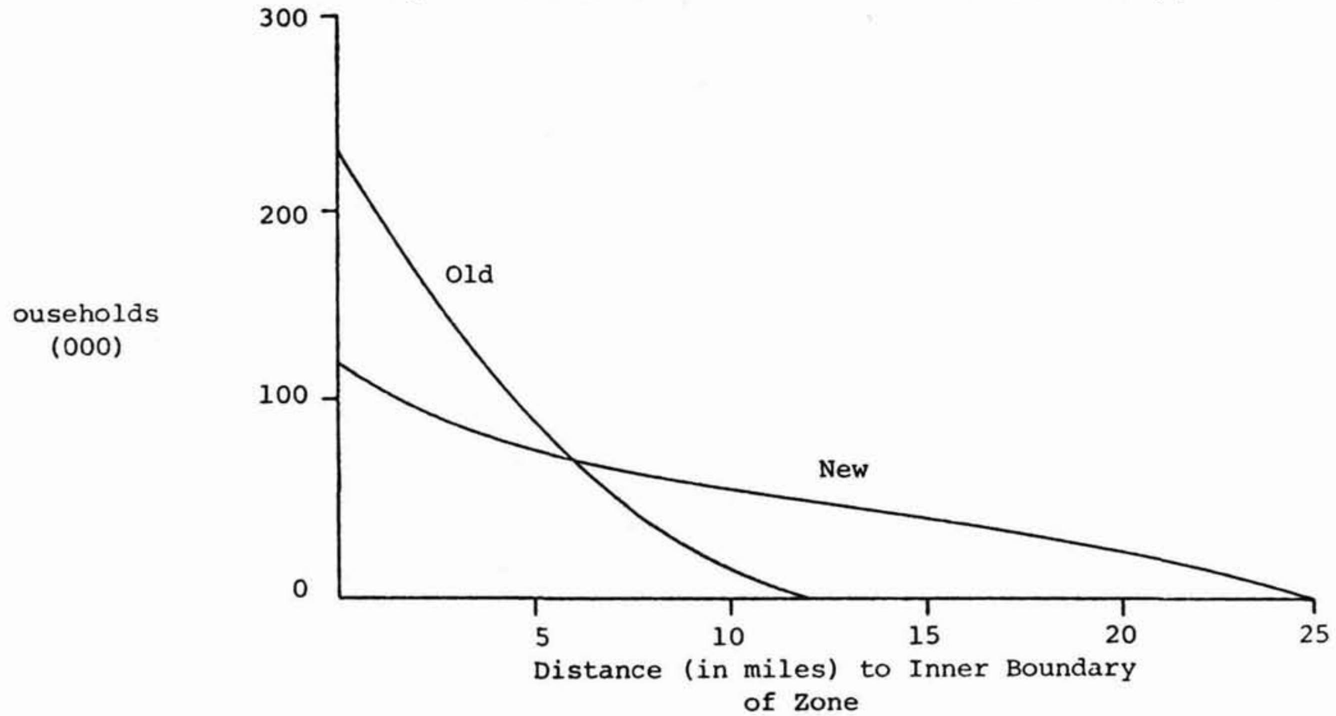
When the growth is predominantly after World War II, as in Houston, the shapes of the two curves are different, as can also be seen in Figure 5B-2. The same 300,000 households are spread out over a much wider area.

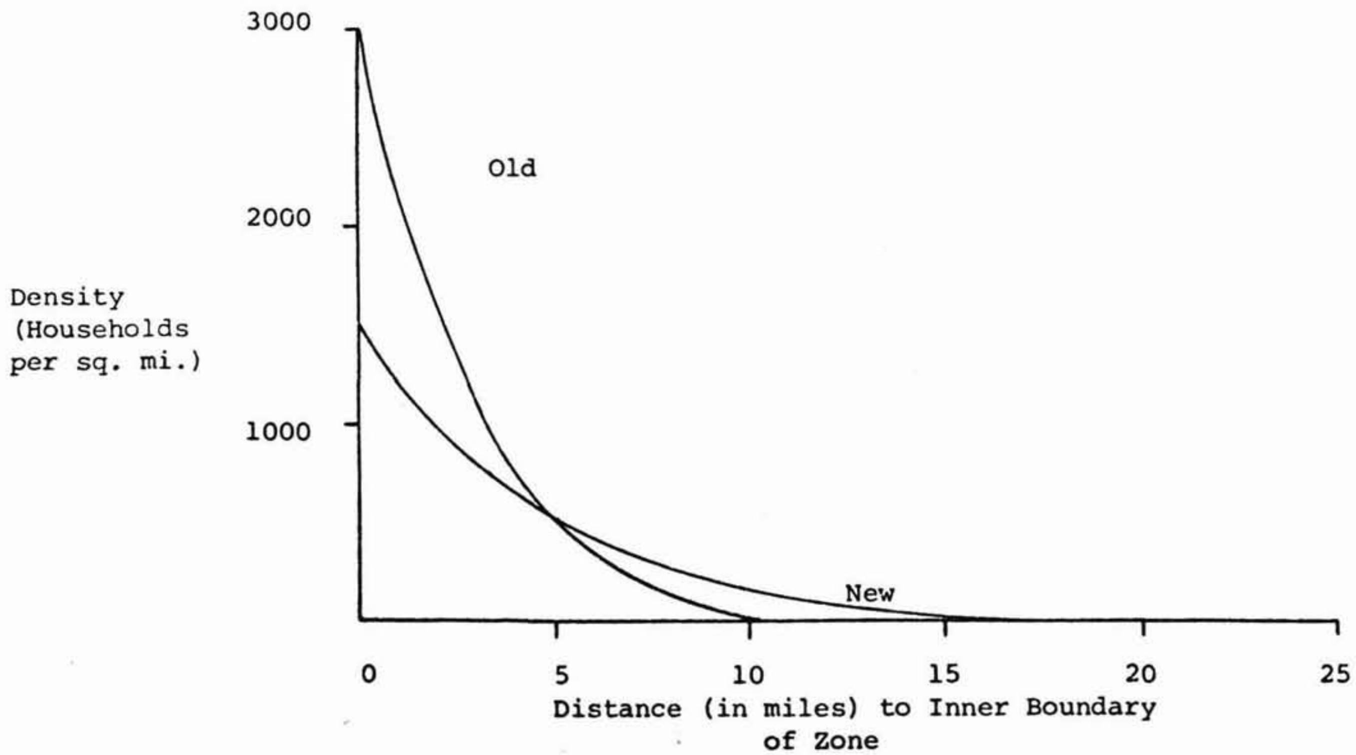
Relaxing assumptions has the anticipated effects. The density curve flattens, for example, when jobs are dispersed. Likewise, if for reasons other than access to jobs, households prefer to look for new units over the outer half of the search radius, the gradient is lowered.

The general lesson to be learned from all this is that one cannot infer causality at the margin over time from cross-sectional distributions at a point in time. The cross section is, in general, the integral over time and space of actions taken during an extended period, and may or may not shed light on the decision rules being followed at any instant.

Figure 5B-2

Comparison of an Older and a Newer City
Assuming a Threshold Search Radius and No Job Mobility, in 1970





As Lee [9] puts it:

> Lacking a coherent body of theory that could explain land-use dynamics, the modelers turned to analogies and descriptive regularities. The untested hypotheses of a variety of social science fields were accepted uncritically and merged without ever establishing the validity of either the individual relationship or the combined structure. One example is the gravity model, which was fitted to aggregate data based on existing land-use patterns and trip behavior, and then employed as a behavioral explanation of future patterns. In between a formal theory was developed to rationalize the transformation, on the postulate that households located so as to minimize some type of trip purpose.
>
> Trip distribution functions (at the heart of any gravity model, which includes most of the land-use models) are fitted to observed trip frequencies for different classes of workers and (sometimes) different classes of households. While valid at the scale of a metropolitan area, the gravity model has no statistical explanatory power at the neighborhood level. This is a near-classic case of imputing individual (or census tract or neighborhood) behavior from aggregate relationships -- the ecological fallacy.

Appendix C

Probability Function Generator

or

PFG

A social scientist frequently wishes to describe how people choose among things. In our own work, the choice is among different locations. But it could just as easily be among different schools, or education levels, or occupations, or any other set of options.

The problem is to estimate either the odds that an individual will choose one of many options, or the proportion of a group that will chose each one. In either case, the desired result is a probability density function (summing to 1.0 over all states) describing choice behavior.

The Probability Function Generator (PFG) was designed to fit such proability density functions to actual choice behavior. Other methods -- particularly regression analysis -- can, in concept, be used to accomplish to same purpose. They have the enomous advantage of having been used a great deal, and, in the process, having evolved a nomenclature and a mathematical rigor that are important in scientific discourse. The advantages of heritage, however, must be balanced against the need for flexibility. We have not found the traditional methods to be flexible enough when it comes to specifying how choices between places are made, and have evolved the PFG to supplement traditional methods when necessary.

Implicit in most existing methods is an underlying model which assumes that people make choices by simultaneously weighing a number of factors, or characteristics, or attributes, of the options to be chosen, and pick the ones with the highest "scores." The best "fit" is the one that minimizes the sum of the squared error between actual choices and the specified function.

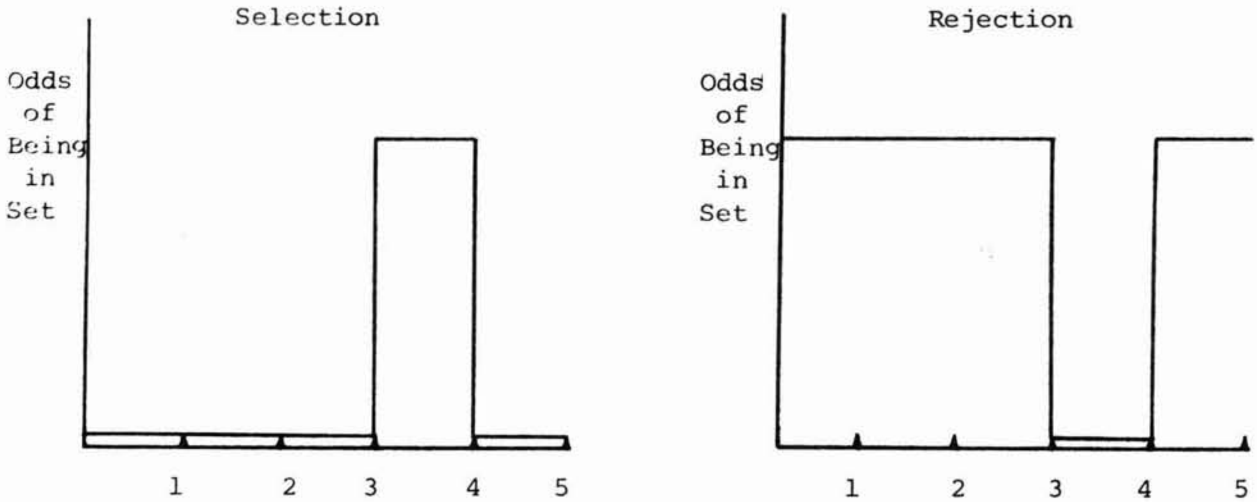
The PFG assumes a somewhat different behavioral model. It assumes that people make choices by sequentially selecting or rejecting certain sets of choices, and that they do this by using a sequence of filters. This is a model that emerges consistently out of our depth interviewing. In residential location, for example, people appear to reject a large number of choices at the start, filtering out areas with which they are not familiar and that are more than a tolerable distance away from their place of work. They then isolate a subset of the remaining areas by selecting the ones that have the right kinds of schools, friends and relatives nearby, and so forth.

The PFG replicates this procedure. It determines which decision rules, or filters, each person type uses in making choices by comparing the choices that would be made using each filter against a set of actual choices, and picking the most "powerful" sequence of filters. The user must, of course, specify the set of potentially useful filters to be evaluated. Since the number of possibilities is enormous, the researcher's first, and most crucial, step is, as usual, one of deduction.

Filters are of two basic kinds: selectors and rejectors. A filter may be used to select a certain set of choices, or it may be used to reject that set.

The sets themselves can be defined in a number of different ways. The simplist way assumes that a person selects (or rejects) those elements that lie in a certain interval within the range of some characteristic, and ignores all others. Continuing with the neighborhood selection process as an example, suppose that we can define a measure of social class for each neighborhood that ranges from 0 to 5, with 5 being high. The "interval" filter specification would then take the form of: "select (or reject) all neighborhoods that are in the class range between 3 and 4." Such a filter might be diagrammed as follows:

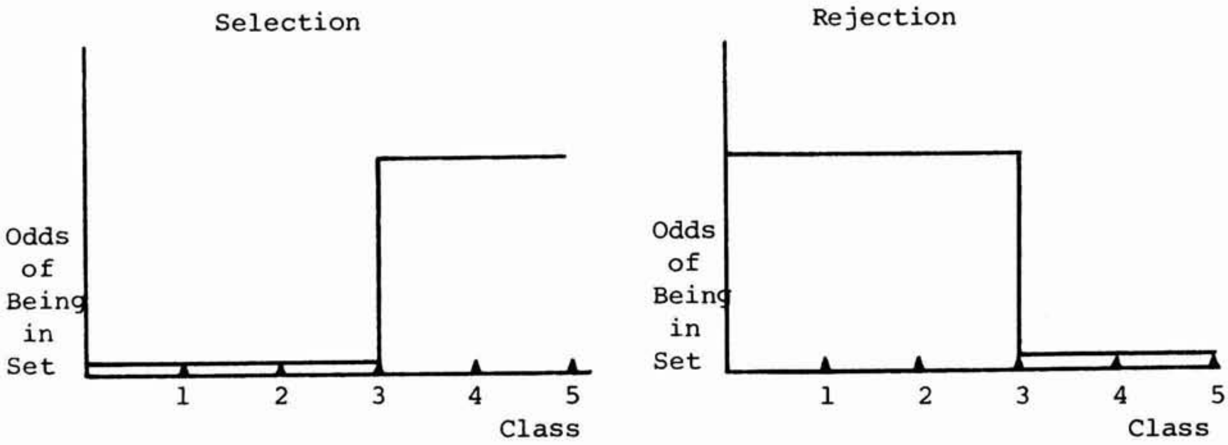
Figure C-1



If used as a selector, only the set of neighborhoods in the range between 3 and 4 would be selected. As a rejector, this same set of neighborhoods would be rejected.

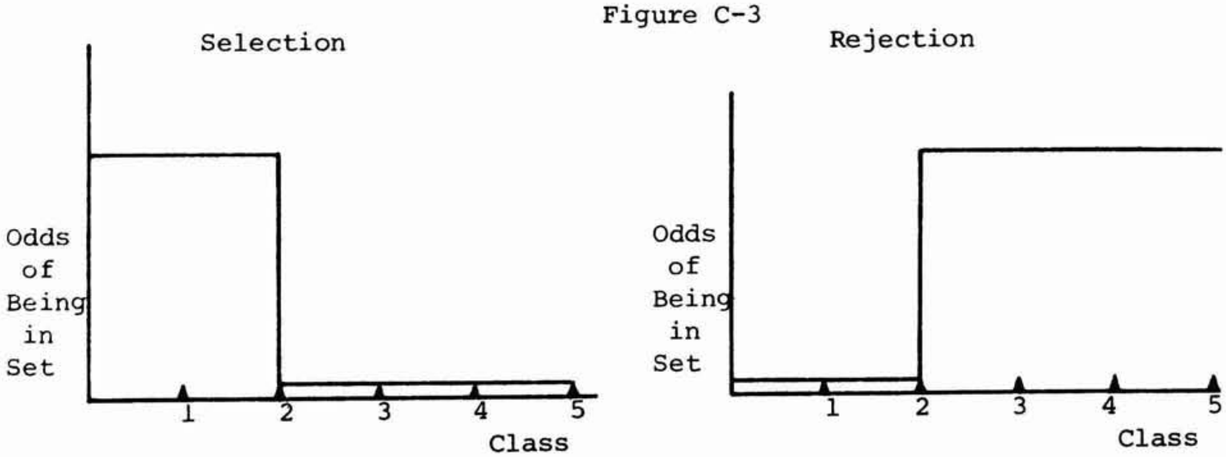
Filters can take on other forms as well. We can say, for example, that a person will select (or reject) all neighborhoods that have a social class rating of 3 or more. Such a filter would look as follows:

Figure C-2



Similarly we might postulate that a person would select (or reject) all neighborhoods with a class rating of 2 or less:

Figure C-3



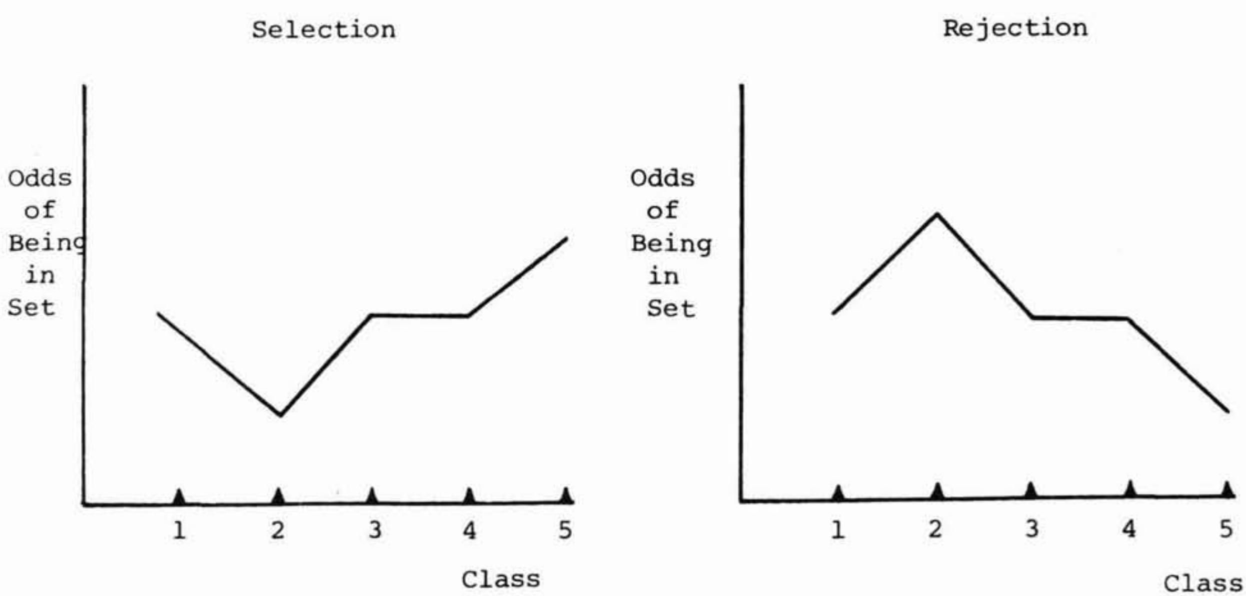
To simplify the user's task somewhat, the PFG offers two options for specifying such "tails." In the first, not being sure of the exact cutoff, the user asks the PFG to scan five different possible cutoffs, either for the cummulative left tail (Figure C-2) or the cumulative right tail (Figure C-3). If the exact cutoff is known, this can be specified as an exact left or right tail.

Similarly, when the user specifies the interval range, five different intervals are tested. There is no option at the moment for specifying a single interval.

The cutoffs that determine where discontinuities will take place are specified by the user. All variables are scaled to fall between 0 and 1 to facilitate comparisons. The user can thus default to fixed intervals (0 .2 .4 .6 .8 1.0) or can specify his or her own breakpoints. Using the social class example, the 1-5 scale would be normalized so that 0 = 0, 1 = .2, 2 = .4, 3 = .6, 4 = .8, and 5 = 1.0. The user can accept this set of intervals, or specify another set which, for example, would isolate only very poor areas (0-.5 on the original scale, 0-.1 on the revised scale) as the first set. The basis for such a choice is, of course, a theoretical one, and flows from the user, not the PFG.

A final filter form comes closest to the continuous functions used in traditional analysis. It asserts that a person chooses (or rejects) a place in proportion to the intensity of some characteristic present at that place:

Figure C-4



The "select" option specifies that selection be in proportion to the intensity of the characteristics; the "reject" option specifies that rejection takes place in proportion to the same intensity.

One thing that emerged from our field interviews was the use of multiple conditions for filtering. For example, a person might select neighborhoods that are above a certain social class and that are no more than a certain distance from work. Violation of either condition led to rejection. Traditionally, such behavior is captured by a multiplicative function -- the Cobb-Douglas being the best known. Such a function might look like:

$$z = a * x^b * y^c$$

Such functions do well in rejecting bad options; whenever x or y go to zero, the function goes to zero. They are less well-behaved in selection, however. If x and y get "big" for the same option, z becomes very large -- a kind of resonance phenomenon.

Such interactive terms are dealt with far more easily in the PFG. Pairs of filters can be used to specify sets destined for selection or rejection, just as single filters serve this purpose. Using the above example, the researcher might specify a filter that simultaneously selected neighborhoods above Class 4 <u>and</u> within 40 minutes driving time, rejecting all others. The function will not "go wild" on the selection side while accomplishing its rejection side efficiently.

Operation

The PFG uses a gradient search procedure to determine which filters should be used in which sequence to best replicate actual behavior. It requires as input:

1. A description of actual behavior ie. a set of choices normalized to take the form of a probability density function.

2. A set of potential filters, specified in terms of characteristics of options and the manner in which they are to be used to isolate a subset of options for selection or rejection.

3. A set of break points for each scaled characteristic indicating where discontinuities may take place. As indicated above, the default is 0, .2, .6, .8, 1.0.

The PFG begins with a "base filter" specified by the user. This is an initial, prior guess regarding the shape of the probability density function. If the user is truly uncertain, he/she can select a "flat prior," ie. assume all options are equally likely at the start.

Working with normalized variables, the PFG tests each potential filter in terms of its ability to improve the base. Filters are tested by creating a trial new base, generated by taking a weighted average of the existing base and the filter, and evaluating it. Trial bases can be evaluated any way the user chooses -- a major source of flexibility. At present, we are using the sum of the absolute value of the difference between the trial base and the known choice behavior.[1] Standard gradient

---

1. We could just as easily have used the more conventional squared error, but find that it tends to overemphasize outliers at the expense of the majority of the options. Since, in our work, outliers are frequently caused by "unusual" events, we prefer to let them remain outliers (and deal with the event) then bend the function inordinantly out of shape to account for them.

search procedures are used to find the optimal weight for that filter vis-a-vis the base, ie. the weight that minimizes the evaluation criterion.

This procedure is followed for all filters. The filter that improves the base the most is selected, and weighted into the base to form a new base. The entire procedure is then repeated, using the new base as "base." The procedure is continued until an exogeneously determined number of filters have been added to the base. At present, we find that most people make effective use of no more than five filters, so 5 is our limit.[1] A diagram describing these steps is presented in Figure C-6.

We can represent the results of this procedure alegebraically.

Define:

P(k)	as the normalized probability vector at Step k.
w(j)	as the weight assigned to the jth filter.
D(j)	as the jth filter, a vector which contains 1 if the observation passes the criterion, and 0 if it fails.
n(j)	as the number of elements in the set defined by Filter D(j). i.e. n(j) = D'(j)·$\underline{1}$, where "'" denotes the transpose of the vector D(j), and $\underline{1}$ denotes a vector of ones.

Then the probability vector after the kth step can be expressed as:

$$P(k) = \sum_{i=0}^{k} \frac{\prod_{j=i+1}^{k} [1-w(j)]}{\prod_{j=i}^{k} S(j)} \cdot w(i) \cdot D(i)$$

---

1. The PFG may select fewer filters if it finds little to be gained by adding a next one.

where:

$$S(k) = \sum_{i=0}^{k} \frac{\prod_{j=i+1}^{k} [1-w(j)]}{\prod_{j=i}^{k-1} S(j)} \cdot w(i) \cdot n(i)$$

and:

$$\prod_{i=j}^{k} f(i) = f(j).f(j+1)\ldots f(k-1).f(k) \qquad \text{if } j \leq k$$

$$= 1 \qquad \text{if } j > k$$

$$w(0) = 1 \text{ (for base)}$$

$$D(0) = \underline{1} \text{ (if assuming flat prior)}$$

$$W(k) \in [0,1] \text{, for all } k$$

$$P(k) \in [0,1]$$

$$P'(k) \cdot \underline{1} = 1.0$$

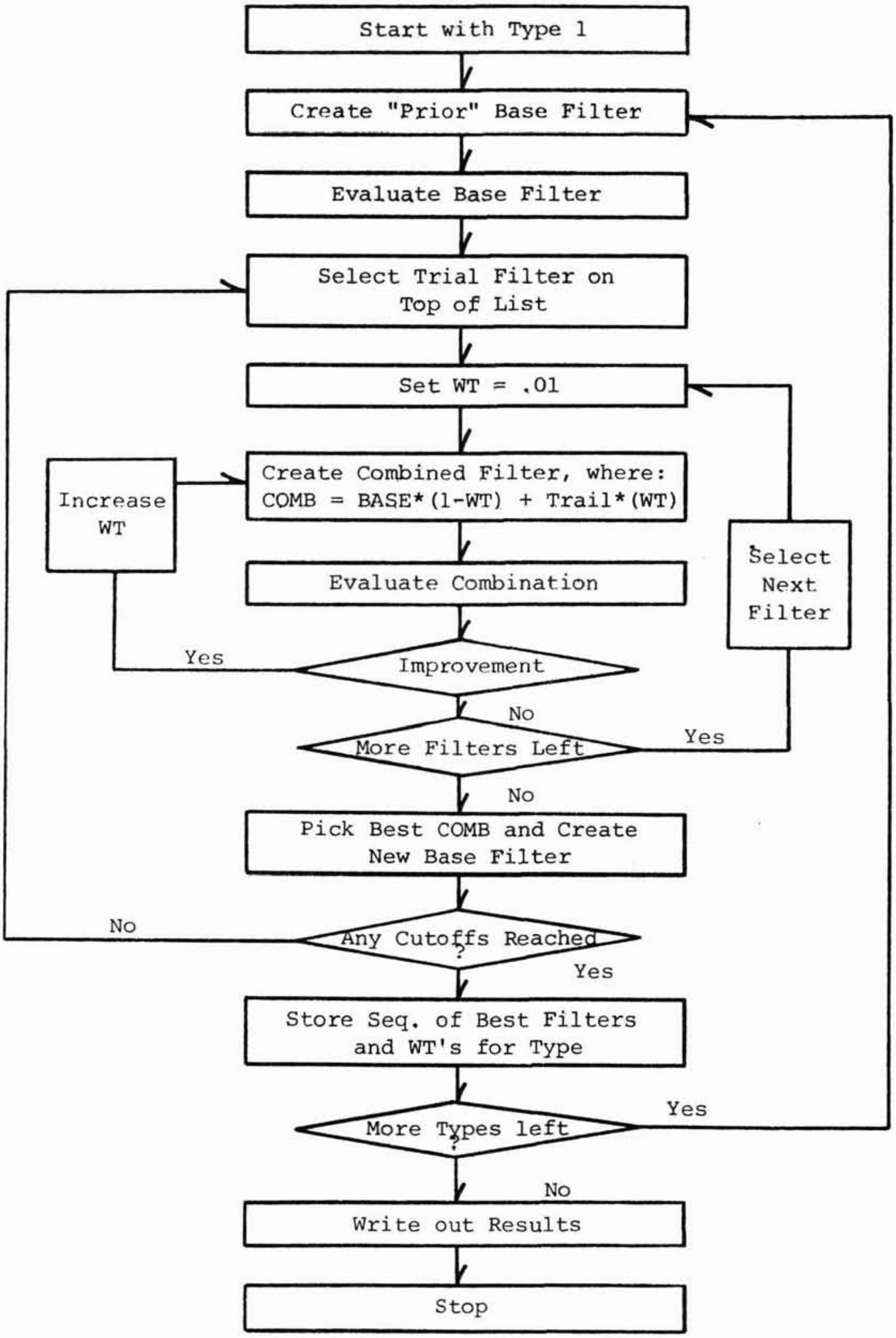
When it is done, the PFG has identified at most 5 filters that, when added to the base sequently at their optimal weights, produce the best fit vis-a-vis known behavior.[1] Since the underlying model is quite diff- erent from the standard linear model, we need a system of notation for describing the results.

---

1. The procedure does not guarantee an overall optimum in much the way stepwise regression fails to produce a global optimum. It only produces a step-wise optimum.

Figure C-6

Schematic Diagram of the Probability Function Generator



Notation

A filter is described in terms of its characteristic (e.g. social class), the way in which subsets are identified (intervals, right or left tails, or continuous variable) and the cut-off points defining where the tails or intervals are defined. To develop the full range of notation let us expand our neighborhood selection example to include another characteristic: race (measured by percent minority). Let us assume further that we have no good intuition regarding the initial base, and adopt a flat prior. We can then describe an individual filter as:

S:Class(GT .8)

where:

S — indicates that the filter is being used to select neighborhoods. "R" would be the reject option.

Class — is the name of the variable

GT — stands for "Greater Than," ie. this filter would select all the neighborhoods in the right tail whose values were greater than .8. Other options are:

LT = Less Than
Int = Interval
Cont = Continous Variable

When Int is used, the range must be given ie. (Int .4-.6)

A jointly defined filter might be specified as follows:

S:[Class(GT .8), Race(LT .2)]

Such a filter would select all neighborhoods whose social class was greater than .8 <u>and</u> whose race did not exceed .2.

A choice function is specified in terms of the desired sequence of filters and the weights associated with their use. We will use the " & " operator to indicate the process of sequential weighting. Assuming two filters were used, a choice function might look like:

P = Base & .85* S:Class(Int, .4-.6) & .13* R:Race(LT .2)

The present measure of "goodness of fit" we use is the sum of the absolute value of the difference between the function and the actual behavior:

$$\text{Measure} = \sum_i \left| F_i - A_i \right|$$

Since both the function (F) and the Actual Behavior (A) sum to 1.0, this measure is itself normalized. Its upper limit is 2.0 We have found that values below .2 represent quite good fits.

It is also possible to compute the traditional R^2, treating the function as an estimate of the actual behavior. On the average the relationship between the difference measure and R^2 is as follows:

Sum of Diff.	R^2
.10	.99
.20	.96
.30	.91
.40	.85
.50	.80
.60	.74

To facilitate comparison with other methods, we will frequently report R^2 as well as our own straight difference measure.

Appendix 5-D

PFG Equations[1] for 27 Household Types in New Haven

```
TYPE:  1 RACE: WH AGE: 20-39 ED: <HS

PROB = BASE  &  0.38 * S:CLAS(INT 0.20-0.40)
             &  0.38 * P:PMIN(GT  0.15)
             &  0.19 * S:PMIN(GT  0.05)
             &  0.13 * S:CLAS(INT 0.40-0.60)
             &  0.54 * R:CLAS(INT 0.60-0.80)

SUMDIF =  0.183
R SQR =   0.880


TYPE:  2 RACE: WH AGE: 20-39 ED: =HS

PROB = BASE  &  0.81 * P:CLAS(INT 0.60-0.80)
             &  0.38 * P:PMIN(GT  0.30)
             &  0.13 * R:CLAS(INT 0.40-0.60)
             &  0.13 * S:PMIN(GT  0.05)
             &  0.19 * R:PMIN(GT  0.15)

SUMDIF =  0.156
R SQR =   0.883


TYPE:  3 RACE: WH AGE: 20-39 ED: >HS

PROB = BASE  &  0.38 * S:PMIN(GT  0.05)
             &  0.38 * P:CLAS(INT 0.60-0.80)
             &  0.19 * P:PMIN(GT  0.15)
             &  0.13 * R:CLAS(INT 0.20-0.40)
             &  0.57 * P:CLAS(INT 0.80-1.50)

SUMDIF =  0.153
R SQR =   0.940
```

1. See Appendix 5-C for an explanation of the notation used in this Appendix.

```
TYPE:  4 RACE: FN AGE: 20-39 ED: <HS

PROB = BASE  &  0.38 * S:CLAS(INT 0.20-0.40)
             &  0.63 * P:PMIN(GT  0.30)
             &  0.81 * P:CLAS(INT 0.0 -0.20)
             &  0.13 * S:PFOR(CONT)
             &  0.19 * S:PMIN(GT  0.05)

SUMDIF =  0.278
R SQR =   0.771


TYPE:  5 RACE: FN AGE: 20-39 ED: =HS

PROB = BASE  &  0.95 * P:CLAS(INT 0.0 -0.20)
             &  0.63 * P:PMIN(GT  0.30)
             &  0.13 * S:CLAS(INT 0.20-0.40)
             &  0.19 * S:PMIN(GT  0.05)
             &  0.13 * P:CLAS(INT 0.40-0.60)

SUMDIF =  0.222
R SQR =   0.892


TYPE:  6 RACE: FN AGE: 20-39 ED: >HS

PROB = BASE  &  0.54 * S:PMIN(GT  0.05)
             &  0.63 * P:PMIN(GT  0.30)
             &  0.19 * S:CLAS(INT 0.60-0.80)
             &  0.28 * S:PMIN(GT  0.50)
             &  0.19 * P:CLAS(INT 0.80-1.50)

SUMDIF =  0.185
R SQR =   0.955
```

```
TYPE:  7 RACE: MN AGE: 20-39 ED: <HS

PROB = BASE   &  0.54 * S:CLAS(INT 0.20-0.40)
              &  0.38 * S:CLAS(INT 0.60-0.80)
              &  0.19 * S:CLAS(INT 0.40-0.60)
              &  0.19 * S:PMIN(GT  0.05)
              &  0.13 * R:PMIN(GT  0.15)

SUMDIF =  0.143
R SQR =   0.975


TYPE:  8 RACE: MN AGE: 20-39 ED: =HS

PROB = BASE   &  0.28 * S:CLAS(INT 0.20-0.40)
              &  0.19 * R:PMIN(GT  0.30)
              &  0.13 * S:PMIN(GT  0.15)
              &  0.13 * S:CLAS(INT 0.60-0.80)
              &  0.19 * R:CLAS(INT 0.80-1.50)

SUMDIF =  0.129
R SQR =   0.978


TYPE:  9 RACE: MN AGE: 20-39 ED: >HS

PROB = BASE   &  0.42 * R:CLAS(INT 0.0 -0.20)
              &  0.19 * S:PMIN(GT  0.05)
              &  0.19 * R:PMIN(GT  0.15)
              &  0.19 * R:CLAS(INT 0.40-0.60)
              &  0.13 * R:PMIN(GT  0.30)

SUMDIF =  0.223
R SQR =   0.378
```

```
TYPE: 10 RACE: WH AGE: 40-64 ED: <HS

PROB = BASE  &  0.95 * R:PMIN(GT  0.15)
             &  0.54 * R:CLAS(INT 0.60-0.80)
             &  0.95 * R:CLAS(INT 0.0 -0.20)
             &  0.95 * R:PMIN(GT  0.50)
             &  0.95 * R:CLAS(INT 0.80-1.50)

SUMDIF =  0.366
R SQR =   0.563


TYPE: 11 RACE: WH AGE: 40-64 ED: =HS

PROB = BASE  &  0.95 * R:PMIN(GT  0.15)
             &  0.28 * S:CTIG(CONT)
             &  0.95 * R:CLAS(INT 0.80-1.50)
             &  0.95 * R:CLAS(INT 0.0 -0.20)
             &  0.19 * S:CLAS(INT 0.40-0.60)

SUMDIF =  0.266
R SQR =   0.809


TYPE: 12 RACE: WH AGE: 40-64 ED: >HS

PROB = BASE  &  0.95 * R:PMIN(GT  0.05)
             &  0.95 * R:CLAS(INT 0.80-1.50)
             &  0.42 * R:PMIN(GT  0.15)
             &  0.95 * S:CLAS(INT 0.0 -0.20)
             &  0.95 * F:PMIN(GT  0.50)

SUMDIF =  0.311
R SQR =   0.838
```

```
TYPE: 13 RACE: FN AGE: 40-64 ED: <HS

PROB = BASE  &  0.95 * R:PMIN(GT  0.15)
             &  0.95 * R:PMIN(GT  0.30)
             &  0.54 * S:CLAS(INT 0.20-0.40)
             &  0.95 * R:CLAS(INT 0.0 -0.20)
             &  0.63 * S:CLAS(INT 0.60-0.80)

SUMDIF =   0.643
R SQR =    0.353


TYPE: 14 RACE: FN AGE: 40-64 ED: =HS

PROB = BASE  &  0.81 * R:PMIN(GT  0.05)
             &  0.95 * R:CLAS(INT 0.0 -0.20)
             &  0.19 * S:CLAS(INT 0.20-0.40)
             &  0.95 * R:PMIN(GT  0.50)
             &  0.95 * R:CLAS(INT 0.80-1.50)

SUMDIF =   0.435
R SQR =    0.558


TYPE: 15 RACE: FN AGE: 40-64 ED: >HS

PROB = BASE  &  0.28 * S:CLAS(INT 0.60-0.80)
             &  0.95 * R:PMIN(GT  0.30)
             &  0.42 * R:CLAS(INT 0.80-1.50)
             &  0.54 * S:CLAS(INT 0.0 -0.20)
             &  0.25 * S:CLAS(INT 0.20-0.40)

SUMDIF =   0.284
R SQR =    0.879
```

```
TYPE: 16 RACE: MN AGE: 40-64 ED: <HS

PROB = BASE  &  0.95 * S:CLAS(INT 0.20-0.40)
             &  0.81 * S:CLAS(INT 0.60-0.80)
             &  0.25 * S:CLAS(INT 0.40-0.60)
             &  0.42 * S:PMIN(GT  0.05)
             &  0.13 * S:PMIN(GT  0.30)

SUMDIF =  0.227
R SQR  =  0.951


TYPE: 17 RACE: MN AGE: 40-64 ED: =HS

PROB = BASE  &  0.38 * S:CLAS(INT 0.20-0.40)
             &  0.19 * S:PMIN(GT  0.05)
             &  0.13 * S:CLAS(INT 0.40-0.60)
             &  0.38 * R:CLAS(INT 0.60-0.80)
             &  0.13 * R:CLAS(INT 0.80-1.50)

SUMDIF =  0.152
R SQR  =  0.980


TYPE: 18 RACE: MN AGE: 40-64 ED: >HS

PROB = BASE  &  0.54 * R:CLAS(INT 0.40-0.60)
             &  0.81 * R:CLAS(INT 0.0 -0.20)
             &  0.63 * R:CLAS(INT 0.80-1.50)
             &  0.13 * P:PMIN(GT  0.0 )

SUMDIF =  0.254
R SQR  =  0.881
```

```
TYPE: 19 RACE: WH AGE: 65+    ED: <HS

PROB = BASE  &  0.63 * R:PMIN(GT  0.15)
             &  0.38 * R:CLAS(INT 0.60-0.80)
             &  0.13 * S:JBAC(CONT)
             &  0.95 * R:PMIN(GT  0.50)
             &  0.13 * S:PMIN(GT  0.05)

SUMDIF =  0.257
R SQR =   0.757


TYPE: 20 RACE: WH AGE: 65+    ED: =HS

PROB = BASE  &  0.28 * S:CTIG(CONT)
             &  0.25 * R:CLAS(INT 0.20-0.40)
             &  0.81 * R:CLAS(INT 0.80-1.50)
             &  0.95 * R:CLAS(INT 0.0 -0.20)
             &  0.13 * R:PMIN(GT  0.05)

SUMDIF =  0.215
R SQR =   0.803


TYPE: 21 RACE: WH AGE: 65+    ED: >HS

PROB = BASE  &  0.95 * R:PMIN(GT  0.15)
             &  0.13 * R:PFOR(CONT)
             &  0.54 * R:CLAS(INT 0.20-0.40)
             &  0.13 * R:CLAS(INT 0.80-1.50)

SUMDIF =  0.294
R SQR =   0.796
```

```
TYPE: 22 RACE: FN AGE: 65+   ED: <HS

PROB = BASE  &  0.95 * R:PMIN(GT  0.15)
             &  0.95 * R:CLAS(INT 0.0 -0.20)
             &  0.85 * R:PMIN(GT  0.30)
             &  0.95 * R:PMIN(GT  0.50)
             &  0.13 * S:PFOR(CONT)

SUMDIF =  0.401
R SQR =   0.738


TYPE: 23 RACE: FN AGE: 65+   ED: =HS

PROB = BASE  &  0.38 * S:PMIN(GT  0.30)
             &  0.25 * R:PMIN(GT  0.05)
             &  0.95 * R:CLAS(INT 0.0 -0.20)
             &  0.81 * R:CLAS(INT 0.80-1.50)
             &  0.95 * R:PMIN(GT  0.50)

SUMDIF =  0.413
R SQR =   0.646


TYPE: 24 RACE: FN AGE: 65+   ED: >HS

PROB = BASE  &  0.63 * R:CLAS(INT 0.40-0.60)
             &  0.13 * R:PFOR(CONT)
             &  0.85 * R:PMIN(GT  0.30)
             &  0.25 * R:CLAS(INT 0.20-0.40)
             &  0.13 * R:PMIN(GT  0.15)

SUMDIF =  0.286
R SQR =   0.885
```

```
TYPE: 25 RACE: MN AGE: 65+    ED: <HS

PROB = BASE  &  0.95 * S:CLAS(INT 0.20-0.40)
             &  0.57 * S:CLAS(INT 0.60-0.80)
             &  0.38 * R:CLAS(INT 0.0 -0.20)
             &  0.13 * R:CLAS(INT 0.80-1.50)

SUMDIF =  0.229
R SQR =   0.944


TYPE: 26 RACE: MN AGE: 65+    ED: =HS

PROB = BASE  &  0.19 * S:CLAS(INT 0.20-0.40)
             &  0.25 * S:PMIN(GT  0.05)
             &  0.13 * R:CLAS(INT 0.80-1.50)

SUMDIF =  0.159
R SQR =   0.978


TYPE: 27 RACE: MN AGE: 65+    ED: >HS

   PROB = BASE  &  0

SUMDIF =  1.000
R SQR =********
```

REFERENCES

1. Alonso, W. Location and Land Use. Cambridge: Harvard University Press, 1964.

2. Birch, Atkinson, Coleman, Parsons, Rosen and Solomon. Models of Neighborhood Change. Cambridge, Mass.: MIT-Harvard Joint Center for Urban Studies. Report to the U.S. Department of Housing and Urban Development, 1974.

3. Birch, Atkinson, Sandstrom, and Stack. The New Haven Laboratory: A Testbed for Planning. Lexington, Mass.: Lexington Books/D.C. Heath, 1973.

4. Butler, Edgar W.F., Chapin, Stuart; et.al.; Moving Behavior and Residential Choice. Highway Research Board, 1969.

5. Friedman, Joseph. Housing Location and the Supply of Local Public Services. Berkeley: Center for Real Estate and Urban Economics, 1975.

6. Ingram, G.K., Kain J.F. and Ginn, J.R. The Detroit Prototype of the NBER Urban Simulation Model. New York: National Bureau of Economic Research, 1972.

7. Johnston, R.J. "Activity Spaces and Residential Preferences: Some Tests of the Hypothesis of Sectoral Mental Maps." Economic Geography, April, 1972.

8. Kain, John F. "The Journey-to-Work as a Determinant of Residential Location." Papers and Proceedings of the Regional Science Association. 9. 1962.

9. Lee. Douglass B. "Requiem for Large-Scale Models." Journal of the American Institute of Planners. May, 1973.

10. Lowry, I.S. A Model of Metropolis. Santa Monica: The Rand Corporation, 1964.

11. Moore, Eric. Residential Mobility in the City. Washington, D.C.: Association of American Geographers, 1972.

12. Mosteller, Frederick. "Association and Estimation of Contingency Tables." Journal of the American Statistical Association. March, 1968.

13. Muth, R. Cities and Housing. Chicago: University of Chicago Press, 1969.

14. Tiebout, Charles M. "A Pure Theory of Local Expenditures." Journal of Political Economy. 65. October, 1956.

15. U.S. Bureau of the Census. A Public Use Sample of Basic Records from the 1960 Census: Description and Technical Documentation. Washington, D.C.: U.S. Government Printing Office, 1971.

16. U.S. Bureau of the Census. U.S. Census of Housing: 1960. Metropolitan Housing, Final Report HC(2)-124. Washington, D.C.: Government Printing Office, 1962.

Section 6

DYNAMICS OF LOCAL REAL ESTATE MARKETS

This section discusses various activities which occur within neighborhood housing markets once these markets have been cleared. These activities involve homeowners, landlords, lending institutions, and insurers. Homeowners can increase or decrease sale prices of homes and can decide to rent out the homes they own, and landlords can alter rent levels on apartments they own. Homeowners and landlords can alter the configuration of their buildings to increase or decrease the number of units. Finally, lending institutions and insurers make decisions to offer or withhold their services from neighborhoods.

The activities above are discussed with respect to all the housing submarkets within each neighborhood. Housing submarkets are defined by tenure and price or rent level. There are two tenures, owned and rented. Within each tenure there are three price levels: high, medium, and low. Changes in price level can occur in all six submarkets.

Setting Prices on Homes and Apartments

Once housing markets in all neighborhoods have been cleared, the number of vacancies remaining in each submarket indicates how demand and supply forces have been resolved. At this point homeowners and landlords with properties on the market must decide on a sale or rental price to seek. They do this in the context of their neighborhood submarket, which means that prices and rents must reflect the demand and supply forces within the submarket. In general, homeowners and landlords raise prices and rents when demand is greater than supply. A low vacancy rate within a submarket means that households have been moving into the submarket. This permits homeowners and landlords to raise prices and rents.[1] If demand is weak relative to the supply of housing, a high proportion of the available vacant units will remain vacant. Under these circumstances, homeowners and landlords are unlikely to raise prices and rents. In fact, some prices and rents may be decreased as homeowners seek to attract buyers, or landlords compete with each other for tenants.[2] Homeowners attempting to sell their homes incur costs when unable to sell. Landlords unable to rent vacant apartments forego rental income, while certain fixed costs such as taxes continue.

We assume that homeowners will seek the highest selling price they can obtain consistent with a reasonable waiting period. We also assume that most landlords will seek the highest rents they can obtain for their apartments. Apartment management is the sole or main source of income for

---

1. Use of the vacancy rate to explain rent levels on apartments receives empirical support in Smith [13].

2. Instead of lowering the contract rent, some landlords give new tenants one or more rent-free months. Reasons for lowering rent this way instead of reducing the contract rent are varied. Tenants may find a large initial bonus more appealing than a smaller bonus spread over many months. Also, landlords may fear rent control, which typically uses existing rents as base rents.

many landlords, so they may want to charge the highest rent the market will bear. Frictions in adjusting rents (e.g., existing leases specifying the rent level, or lack of information) may mean that rents are not always at the highest possible level. Generally, over the long run, the typical landlord is likely to follow a profit-maximizing course of action. However, according to Krohn and Fleming [7], professional investors and small-scale landlords may take different approaches in setting rent levels. Professional landlords, or those with a purely economic approach to setting rents, are more interested in charging what the market will bear. Small-scale landlords tend to be more casual in their economic relations with tenants. They are often more interested in attracting or keeping a certain kind of tenant than in maximizing their monetary return.

Although only those housing units on the market make price or rent changes manifest, all units are candidates for price or rent changes. If a neighborhood is attractive to movers, as indicated by low vacancy rates, the values of most occupied units rise also. When these units eventually come into the market, their prices or rents will reflect prior changes in their values.

In order to determine the pattern of price and rent changes within each neighborhood submarket, we must measure the relative attractiveness of housing within a submarket. The most desirable measure for this purpose is the vacancy rate in the submarket after clearing. A low vacancy rate results from a tight market. Since some vacancies always exist in a submarket due to the availability of new units or to lags between occupants, we have defined a normal or standard vacancy rate for each market. We assume that homeowners and landlords recognize this and do not change prices or rents when the vacancy rate is at its normal level. When the vacancy rate rises above or falls below its normal level, this creates pressure for price and rent changes.[1] Thus, price and rent changes depend on the relationship between the actual vacancy rate and the normal vacancy rate.

Since the price and rent categories used in the model are rather broad, units can undergo price or rent changes yet remain within the same category. Consequently, in any year, only a small fraction of housing units moves out of one category and into another. Evidence of the upward trend in home prices and rent levels in this country is provided in Table 6-1. Median home values and median gross rents both increased over 50 percent from 1960 to 1970. These increases can be attributed to units rising in price or rent, as well as to new construction of higher priced housing.

For modeling purposes we assume that prices and rents tend to decrease more slowly than they increase. We estimate that in a year no more than 2 or 3 percent of the units in the low and medium price groups can move to a higher group, and no more than 1 percent of the units in the medium and high groups can move to a lower group.[2] These limits, which were empirically determined, reflect the propensity for units to increase in price or rent over time.[3]

---

1. The concept of a natural or normal vacancy rate is used by Smith [13], who defines it as "the rate at which there is no excess demand or supply."

2. Note that housing prices in the model are in constant dollars, while Table 6-1 is in current dollars.

3. Schaaf [11] found, in his study of resale of single-family used homes in Oakland, California, that used-house prices advanced steadily by about 4.5 percent per year over the entire six year period under study.

Table 6-1

Values and Gross Rents of Dwelling Units in 1960 and 1970

Value of Owner-Occupied Units (percentage of units)

	1960	1970
Total units	26,278,000	32,383,000
Under $10,000	38%	17%
$10,000 - 15,000	29	19
$15,000 - 20,000	18	21
$20,000 - 25,000	8	16
$25,000 - 35,000	4	16
$35,000 and above	3	11
Median value	$12,000	$18,200

Gross Rents of Renter-Occupied Units (percentage of units)

	1960	1970
Total units	18,549,000	22,155,000
Under $50	20%	6%
$50 - 99	57	34
$100 - 149	14	32
$150 - 199	2	15
$200 - 299	1	6
$300 and above	0	1
No cash rent	6	6
Median rent	$71	$111

Source: U.S. Department of Commerce, Bureau of the Census, 1970 Census of Housing, Components of Inventory Change, series HC(4)-1, United States and Regions, Table 1.

Conversions and Mergers

Some changes in the housing stock result from activities which alter the tenure or configuration of housing units. The impact of such activities, especially conversions and mergers, may be significant in certain neighborhoods, although the impact on the total metropolitan housing stock is low, as indicated in Table 6-2.

Table 6-2

Net Changes in the U.S. Metropolitan Housing Inventory, 1960 to 1970

All housing units (1970)	46,635,201
All housing units (1960)	38,372,301
Net change: total	8,262,900
Percent change	21.5
Sources of additional units	
Conversions	3%
New construction	95%
Other sources	2%
Total added	12,550,599
Causes of units lost	
Mergers	10%
Demolitions	61%
Other means	29%
Total lost	4,273,596

Source: U.S. Department of Commerce, Bureau of the Census, 1970 Census of Housing, Components of Inventory Change, series HC(4)-1, United States and Regions, Table C.

In some neighborhoods owned housing units are converted to rented units by their owners.[1] If there is a continuing mismatch of demand and supply in the owned housing market, with vacancy rates for owned housing remaining high, some homeowners find it very difficult to sell their homes. Consequently they may decide to rent them out. In other cases, when the owned market is weak and the rental market strong, someone may purchase an owner-occupied unit with the intention of subdividing it and renting out the units. Sometimes the units converted are large houses located in an older section of a city, and sometimes they are simply one-family units in a neighborhood which has become less attractive to households seeking to purchase homes.

In simulating the process of converting units from owned to rented tenure, we use information on the vacancy rates in both owned and rental housing markets to assess the strength of these markets. If the vacancy rate in an owned submarket exceeds the normal vacancy rate by some amount, and if the vacancy rate of the rental housing market in the same price category does not exceed the normal vacancy rate for that submarket, then conversions can occur in the model. The units which are converted in the model are vacant since conversions generally are made only when occupancy of a unit changes. The number of units converted in a given submarket depends on the extent to which the vacancy rate in the owned housing submarket exceeds the normal vacancy rate. However, the number of units converted in any neighborhood is rather low.

Mergers generally refer to housing units created by combining two or more units into one larger unit. Landlords may create large apartments

---

1. Conversion of large apartments into several smaller apartments are not modeled here, as we lack data on individual structure types within neighborhoods.

from several small ones, and homeowners may create a large unit for owner occupancy from several small apartments. However, the process of merging smaller units into larger units is not modeled here, since we lack data on individual structure types within neighborhoods. We assume that the quality and condition of the particular unit are important in determining its suitability for merging.

Denying Loans and Insurance

Refusal of lending and insuring institutions to provide loans or underwrite insurance in a neighborhood can disrupt operation of the housing submarkets in the neighborhood by inhibiting transactions between buyers and sellers. This practice, often termed redlining, can have a variety of effects, some of which are discussed by the National Commission on Urban Problems [9], the National Urban League [10], Searing [12], and the United States Commission on Civil Rights [15]. Homeowners and landlords find it difficult to sell their properties as buyers are unable to obtain mortgage loans through conventional channels. This affects home values and apartment rent levels adversely. Lack of insurance makes it virtually impossible for homeowners and landlords to protect their real estate investments. This reduces their incentive to invest in maintenance, so properties may deteriorate more rapidly.

Lending institutions and insurance companies which systematically deny loans and insurance to people in certain neighborhoods are presumably acting in what they perceive as their own best interests. If lenders think the risk of default is too high, if they think that the housing units are valued too low to provide adequate collateral for a mortgage loan, or if they think a neighborhood is declining in some sense, they are apt to deny loans.

Insurance companies often refuse to issue property insurance in neighborhoods in which they expect insurance payouts to exceed the revenue received from premiums.[1]

If experience in assigning premiums has actually led to losses in an area, redlining is very likely. This is due in part to the inflexible insurance rating system, which usually considers type of residential structure but ignores location of the structure.[2]

Lenders and insurers base their business decisions on the characteristics of applicants, the condition of the housing units, and the nature of the neighborhood. Applications for loans and insurance may be rejected if the applicant has insufficient income or assets, if the condition of the housing unit is not satisfactory, or if the neighborhood has an obsolete housing stock or a diminished locational value.[3] Locational value refers to the accessibility to various amenities, such as shopping areas and government services.

The criteria which prompt bankers and insurers to deny credit and insurance to an individual applicant apply _a fortiori_ to their neighborhood evaluations. They may believe that the risk attached to loans and insurance

---

1. Birch et al. [2] reported that "when underwriters perceive the possibility of underwriting losses in particular areas in which deterioration has begun, and thus the possibility of higher risks than what the present rates allow for, they simply avoid extending insurance coverage to these areas."

2. Syron [14] investigated Boston fire damage data from 1965 to 1967. He found that fire damage per sales value of residences and business establishments was 4.2 times higher in the study core area (the South End, Roxbury, and North Dorchester) than in the remaining neighborhoods in Boston.

3. See Laurenti [8], who found that lending institutions were reluctant to lend if properties were on the downward swing in their life cycles.

increases as the income level of the neighborhood falls. It is widely believed that lenders and insurers use racial composition of the neighborhood as an indication of the income level, and hence, of risk.[1] The presence of minority groups, especially nonwhites, is often linked to low income levels and high risk. For this reason neighborhoods with a substantial number of nonwhites are often redlined. Another basis for denying loans and insurance is the general quality and condition of the housing stock in a neighborhood. If the units are considered obsolete, their value as collateral is apt to be low, even though the condition of the units may be good. If the condition of the units is marginal, their resale value is likely to be low, which would not permit the lending institution to recoup its investment if foreclosure became necessary.

In simulating redlining, we are concerned with cases in which an entire neighborhood faces the possibility of being denied loans and insurance. We assume that the likelihood of redlining is related to two factors: racial composition of the neighborhood and overall condition of the housing stock. Use of racial composition as a variable presumes that lenders and insurers regard this as a surrogate for the general level of financial status of the neighborhood. There may also be elements of discrimination affecting the behavior of lenders and insurers. As the fraction of

---

1. The evidence on discrimination by mortgage lenders is not free of ambiguity. One Baltimore study [15] spoke of discrimination against blacks in general, while another [1] spoke of discrimination against medium-income blacks. A New York study [6] found discrimination against blacks and Puerto Ricans. A study at the Center for Urban Affairs of Northwestern University [3] suggested that lending institutions may be reluctant to lend in largely nonwhite areas. However, one study [5] focusing on heavily black, or transitional, sections of Queens County, New York, found that blacks were able to obtain credit. What may be important to lenders is the racial or ethnic stability of a neighborhood. Lenders may be reluctant to lend in areas undergoing or expected to undergo racial transition out of a fear of declining property values. However, once the racial stability of a neighborhood seems ensured, lenders may be willing to lend to blacks. See also National Urban League [10] and Center for Urban Affairs [4].

minority residents increases, loans and insurance are less likely to be granted. Overall condition of the housing stock reflects its value as collateral, and is probably related to the degree of fire hazard. Marginal stock condition is presumed to discourage bankers and insurers from serving an area.

In conclusion, we find that the processes which characterize housing markets operate in reasonably predictable ways. While parameters may vary across cities, the basic processes should be similar. Hence, we expect changes in price and rent levels, and conversions, to reflect demand and supply forces, and we expect credit and insurance denials to reflect the combined impact of racial composition and housing stock condition.

Mathematical Statement

Setting Prices on Homes and Apartments

Price and rent changes within a neighborhood submarket reflect the balance of demand and supply forces. Every submarket (defined by tenure and price level) within a region has a normal or standard vacancy rate, VSTD, which is determined outside the model on the basis of U.S. Census data on occupied and vacant units for 1960 and 1970. The actual vacancy rate in a neighborhood submarket is compared with the standard vacancy rate, as shown below.[1]

$$VR(T,P) = VAC(T,P) / STOC(T,P) \qquad (6.1)$$

for all tenures (T), all price groups (P) and all neighborhoods (I), where:

VR(T,P) = vacancy rate for Tenure T in Price Group P

VAC(T,P) = number of vacancies of Tenure T and Price Group P

STOC(T,P) = number of housing units of Tenure T and Price Group P.

The pressure for price changes in a submarket depends on the difference between the actual and standard vacancy rates. This difference is determined as follows:

---

1. In this and all subsequent equations in Section 6 the subscript I, which refers to the individual neighborhoods in a region, is suppressed to simplify the exposition.

$$DIFF(T,P) = VSTD(T,P) - VR(T,P) \tag{6.2}$$

for all tenures (T), all price groups (P) and all neighborhoods (I), where:

DIFF(T,P) = the difference between the standard vacancy rate and the actual vacancy rate for Tenure T and Price Group P

VSTD(T,P) = the standard vacancy rate for Tenure T and Price Group P across the region.

Once the difference between the standard and actual vacancy rates is known, the fraction of units changing from one price group to another [PCHANG(T,P)] is determined using the non-linear function below:

$$PCHANG(T,P) = a - a * e^{-\frac{DIFF(T,P)}{a} * sl} \tag{6.3}$$

for all tenures (T), all price groups (P), and all neighborhoods (I), where:

$a = -.01$ for $DIFF(T,P) \leq 0$

$a = .025$ for $DIFF(T,P) > 0$

$.1 \leq sl \leq 5.0$

If the actual vacancy rate is the same as the standard vacancy rate, the value of PCHANG(T,P) is 0.0. A positive value for DIFF(T,P) means the standard vacancy rate exceeds the actual vacancy rate, indicating strong demand. The value for PCHANG(T,P) will be positive and will indicate the fraction of housing units which move from one price level up to the next. A value for DIFF(T,P) less than 0.0 means units will move down one price level.

Two factors limit the extent of price changes. Units in the low price or rent category cannot drop to a lower price group, and units in the high price or rent category cannot rise to a higher price group. Hence, the PCHANG(T,P) values in these situations are 0.0. If the redlining index (REDLIN) exceeds some critical value, the fraction of units permitted to rise in price is further restricted.

Once the fraction of units changing price groups is known, the accounting is straightforward. The model does not adjust the variable for total housing stock (STOC), but instead adjusts the variables for occupied housing stock (TOCC) and vacant units (VAC) separately. For price increases in low and medium price groups:[1]

XTOCC(T,P-1) = PCHANG(T,P) * TOCC(T,P) (6.4)

TOCLOS(T,P) = PCHANG(T,P) * TOCC(T,P) (6.5)

XVAC(T,P-1) = PCHANG(T,P) * VAC(T,P) (6.6)

VACLOS(T,P) = PCHANG(T,P) * VAC(T,P) (6.7)

for all tenures (T) and all neighborhoods (I), where:

XTOCC(T,P-1) = number of occupied units which move up from Price Group P to Price Group P-1

PCHANG(T,P) = fraction of units of Tenure T and Price Group P which move to a higher price group

TOCC(T,P) = number of occupied housing units of Tenure T and Price Group P

TOCLOS(T,P) = number of occupied units which move up from Price Group P to Price Group P-1

XVAC(T,P-1) = number of vacant units which move up from Price Group P to Price Group P-1

VACLOS(T,P) = number of vacant units which move up from Price Group P to Price Group P-1.

---

1. Price groups 1, 2 and 3 refer here to high-, medium- and low-priced housing units, respectively. Thus, a move from price group P to price group P-1 means a move to a higher price level; a move from price group P to P+1 means moving to a lower price level.

For high- and medium-priced units moving to a lower price group:

$$XTOCC(T,P+1) = -PCHANG(T,P) * TOCC(T,P) \quad (6.8)$$

$$TOCLOS(T,P) = -PCHANG(T,P) * TOCC(T,P) \quad (6.9)$$

$$XVAC(T,P+1) = -PCHANG(T,P) * VAC(T,P) \quad (6.10)$$

$$VACLOS(T,P) = -PCHANG(T,P) * VAC(T,P) \quad (6.11)$$

for all tenures (T) and all neighborhoods (I), where:

XTOCC(T,P+1) = number of occupied units which move down from Price Group P to Price Group P+1

XVAC(T,P+1) = number of vacant units which move down from Price Group P to Price Group P+1.

The accounting for price changes is described below. TOCC and VAC are adjusted upward for increases, and downward for decreases, in number of units.

$$TOCC(T,P)_t = TOCC(T,P)_{t-1} + XTOCC(T,P) - TOCLOS(T,P) \quad (6.12)$$

$$VAC(T,P)_t = VAC(T,P)_{t-1} + XVAC(T,P) - VACLOS(T,P) \quad (6.13)$$

for all tenures (T), all price groups (P) and all neighborhoods (I), where the subscript t represents the current year and the subscript t-1 represents the previous year, and where:

XTOCC(T,P) = increase in number of occupied units of Tenure T and Price Group P

XVAC(T,P) = increase in number of vacant units of Tenure T and Price Group P.

Conversions

For units to be converted from owned to rented tenure, the market for owned units must be weak and the market for rented units must be strong. Market strength is determined by comparing the actual vacancy rate with the adjusted standard vacancy rate. That is, in markets for rental units the following criteria are applied:

$$RENTVR(P) = TRTVAC(P) / TRTSTC(P) \quad (6.14)$$

$$RSTDVC(P)' = RSTDVC(P) * k_R \quad (6.15)$$

for all price levels (P) and all neighborhoods (I), where:

RENTVR(P) = vacancy rate for rented units in Price Group P

TRTVAC(P) = vacant units for rent in Price Group P

TRTSTC(P) = total number of rental units in Price Group P

RSTDVC(P)' = adjusted standard vacancy rate for rented units in Price Group P

RSTDVC(P) = standard vacancy rate for rented units in Price Group P

k_R = a parameter for establishing a rental market vacancy rate threshold above which conversions will not occur ($k_R \geq 1.0$).

If the actual vacancy rate is greater than the adjusted standard vacancy rate, the market is weak and no units will change to rented tenure.

If

$$RENTVR(P) \geq RSTDVC(P)', \text{ there will be no conversions.} \quad (6.16)$$

If

$$RENTVR(P) < RSTDVC(P)' \text{ conversions are possible.} \quad (6.17)$$

In markets for owned units these criteria are applied:

$$OWNVR(P) = TOWVAC(P) / TOWSTC(P) \quad (6.18)$$

$$OSTDVC(P)' = OSTDVC(P) * k_W \qquad (6.19)$$

for all price groups (P) and all neighborhoods (I), where:

OWNVR(P) = vacancy rate for owned units in Price Group P

TOWVAC(P) = vacant units for sale in Price Group P

TOWSTC(P) = total number of owned units in Price Group P

OSTDVC(P)' = adjusted standard vacancy rate for owned units in Price Group P

OSTDVC(P) = standard vacancy rate for owned units in Price Group P

k_W = a parameter for establishing an owned market vacancy rate threshold below which conversions will not occur ($k_W \leq 1.0$).

If the actual vacancy rate is less than the adjusted standard vacancy rate, the market is strong, and no units will change to rented tenure.

If

$OWNVR(P) \leq OSTDVC(P)'$, there will be no conversions. (6.20)

If

$OWNVR(P) > OSTDVC(P)'$, conversions are possible. (6.21)

For conversions to occur, the conditions in Equations 6.17 and 6.21 must be met simultaneously. The number of units converted from owned to rented tenure depends on the difference between the actual vacancy rate and the standard vacancy rate, as shown below:

$$CONV(P) = k_C * [OWNVR(P) - OSTDVC(P)] * TOWVAC(P) \qquad (6.22)$$

for all price groups (P) and all neighborhoods (I) where:

CONV(P) = the number of units converted from owned to rented tenure in Price Group P

k_C = a constant used to calculate the number of conversions.

Note that only vacant owned units are converted. The maximum number of vacant units converted in any year is limited to some fraction of the total number of vacant units for sale.

Denying Loans and Insurance

The availability of loans and insurance within individual neighborhoods depends on the overall condition of the housing stock and on the racial composition. The lower the stock condition, or the higher the fraction of minority households, the greater the likelihood of credit and insurance denials (i.e., redlining) within a neighborhood. Nonlinear functions are used to generate an index showing the contribution of each factor to pressure for redlining.

The influence of racial composition on redlining is determined as follows:

$$RACE = a + (b - a) * e^{\frac{(c - PCTMIN)}{(a - b)} * sl} \qquad (6.23)$$

for all neighborhoods (I), where:

RACE = influence of racial composition on redlining

a = 0 for PCTMIN $\leq$ c

a = 1.0 for PCTMIN $>$ c

b = a parameter whose value is typically .50

c = a parameter whose value is typically .05

sl = a parameter whose value is typically 10.

PCTMIN = fraction of minority households.

An increasing fraction of minority households creates increasing index values.

The influence of neighborhood stock condition on redlining is determined as follows:

$$\text{STOCK} = a + (b - a) * e^{\frac{(c-\text{STCOND})}{(a-b)} * s1} \qquad (6.24)$$

for all neighborhoods (I), where:

STOCK = influence of stock condition on redlining

a = 1.0 for STCOND $\leq$ c

a = 0 for STCOND $>$ c

b = a parameter whose value is typically .50

c = a parameter whose value is typically 70

s1 = a parameter whose value is typically -.025

STCOND = index of stock condition.

Low stock condition values create high values for the index.

The index values for stock condition and racial composition are multiplied together to create an index for redlining:

$$\text{REDLIN} = \text{RACE} * \text{STOCK} \qquad (6.25)$$

for all neighborhoods (I), where:

REDLIN = an index showing the extent of pressure for redlining.

Higher redlining index values indicate a greater likelihood of redlining. For each region being studied a threshold value is established.

If the value of the redlining index is below this value, redlining does not occur. If the index value exceeds the threshold, redlining is presumed. This in turn affects price changes, as described earlier in this section, and affects investment in maintenance, as described in Section 8, "Investment in Maintenance of the Housing Stock."

REFERENCES

1. Baltimore Department of Housing and Community Development. Home Ownership and the Baltimore Mortgage Market. Baltimore, Md,: Baltimore Department of Housing and Community Development, 1974.

2. Birch, David, et al. Models of Neighborhood Evolution, Cambridge, Mass.: Joint Center for Urban Studies, 1974.

3. Center for Urban Affairs, Urban-Suburban Investment Study Group, Northwestern University. "An Analysis of the Federal Home Loan Bank of Chicago's Survey of Urban Lending and Savings Patterns by Cook County Insured Associations." Working Paper No. 9, 1975.

4. _____ "Sample of HUD-FHA Single Family Firm Commitments in the Chicago Metropolitan Area." Working Paper No. 12, 1975.

5. City of New York Commission on Human Rights, "Residential Mortgage Financing and Minorities." Unpublished. New York, 1966.

6. Devine, Richard J., Rennie, Winston O., Sims, N. Brenda. Where the Lender Looks First: A Case Study of Mortgage Disinvestment in Bronx County 1960-1970. New York: National Urban League, 1973.

7. Krohn, Roger G. and Fleming, E. Berkeley. "The Other Economy and the Urban Housing Problem: A Study of Older Rental Neighborhoods in Montreal." Cambridge, Mass.: Joint Center for Urban Studies, Working Paper No. 11, August 1972.

8. Laurenti, Luigi. Property Values and Race. Berkeley, Cal.: University of California Press, 1960.

9. National Commission on Urban Problems. Building the American City, Report of the National Commission on Urban Problems to the Congress and to the President of the United States. House Document No. 91-34, Washington, D.C.: Government Printing Office, 1968.

10. National Urban League. The National Survey of Housing Abandonment. New York: National Urban League, 1971.

11. Schaaf, A.H., "Price Behavior in Existing House Markets, 1961-1966." Appraisal Journal, Vol. 37, April 1969.

12. Searing, Daniel A. "Discrimination in Home Finance," Notre Dame Lawyer. Vol. 48, June 1973.

13. Smith, Lawrence B., "A Note on the Price Adjustment Mechanism for Rental Housing." American Economic Review, vol. 64(3), June 1974.

14. Syron, Richard F. An Analysis of the Collapse of the Normal Market for Fire Insurance in Substandard Urban Core Areas. Boston, Mass.: Federal Reserve Bank of Boston, 1972.

15. United States Commission on Civil Rights. Home Mortgage Financing and Racial and Economic Integration. Washington, D.C.: Government Printing Office, 1970.

16. United States Department of Commerce, Bureau of the Census, 1970 Census of Housing, Components of Inventory Change, series HC(4)-1. Washington, D.C.: Government Printing Office. 1973.

Section 7

CONSTRUCTION AND DEMOLITION OF HOUSING UNITS

This section discusses changes in the size of the housing stock caused by construction of privately owned units, construction of publicly subsidized units, and demolitions.

New Construction

Builders respond to a variety of factors in deciding how many new units to construct in a neighborhood submarket.[1] Present demand and supply forces, land costs and land use restrictions, and expectations of future housing demand all influence builders' decisions, as outlined below.

Demand and Supply Forces

Perhaps the most immediate question facing any builder is whether the units built will be salable. In many cases the builder has purchasers for the new units before construction begins, or finds purchasers after construction has started. Sometimes purchasers are not found until the units are completed. If the lag between completion and sale of new units is substantial, a builder's finances can be strained due to the interest payments on construction loans. If the builder anticipates a weak market for new units, the builder is expected to reduce the level of construction. However,

---

1. Submarkets are defined by tenure (ownership status) and price level, as explained in Section 6, "Dynamics of Local Real Estate Markets."

even in a weak market, builders will produce some units as they wish to hold on to their work forces for awhile. A continued weak market forces builders to lay off workers and to idle equipment to avoid high carrying costs. On the other hand, in strong housing markets builders cannot afford to pass up profitable opportunities, so they tend to increase the size of their workforces. The extent to which they can increase construction levels this way is limited, of course, by the availability of suitable labor and equipment.[1] Because of the expense and disruption created by hiring or laying off workers and by adjusting the level of equipment usage, builders have some incentive to maintain operations at a relatively constant level.

The model assumes that builders assess the strength of demand in housing submarkets by examining vacancy rates.[2] We assume that each submarket has a normal or standard vacancy rate which indicates the fraction of units likely to be vacant due to expected lags between occupants.[3] Builders can assess the demand-supply relationship in any given submarket

---

1. The President's Committee on Urban Housing [10] examines the significance of manpower and materials limitations in residential construction.

2. See Maisel [8] for a more detailed discussion of how vacancies in different types of units influence builders.

3. L.B. Smith [12] provides support for the concept of a normal vacancy rate in his study of rental housing in five Canadian cities. He found normal vacancy rates in rental housing markets ranging between 5.0% (in Toronto) and 7.4% (in Montreal). W.F. Smith [13] too has suggested that the normal vacancy rate can vary from market to market. However, he [14] has raised some questions about the significance of the normal vacancy rate. The existence of vacant units is often explained as necessary to the free movement of households. However, vacant units may impose various costs on their owners, who in effect provide external benefits to households. W.F. Smith, however, asserts that it is not clear how the vacancies are produced as a response to the needs of households wishing to move.

by comparing the existing vacancy rate with the normal vacancy rate.[1] When demand is greater than the supply of available units, the vacancy rate will be low. If the actual vacancy rate is lower than the normal rate, builders will increase their construction level. Weak demand relative to the supply of units means a high vacancy rate. If the actual vacancy rate is greater than the normal rate, construction activity will drop.

In making construction decisions at the neighborhood level, builders also take into account the regional demand and supply situation in each submarket. They recognize that housing units in one neighborhood can substitute for similar units in other neighborhoods. As a result, builders assess the submarket demand-supply situation in the region before committing themselves to any new construction. If the regional vacancy rate is higher than the normal vacancy rate for a given tenure and price level of housing, construction levels drop. If the regional market is tight, builders are expected to increase their levels of construction. However, increases or decreases in construction levels by builders are limited by the

---

1. An alternative measure of the balance between demand and supply forces in a submarket is the absorption rate, which is defined as the fraction of available vacant units which are sold or rented in a given period of time. That is,

$$\text{Absorption rate} = \frac{\text{Number of units sold or rented}}{\text{Total available vacant units}} .$$

Hua [5] supports using the absorption rate, as he found that builders respond more to the length of time it took to sell a new unit than to vacancy rates per se, since short waiting times indicate a strong market. Because the absorption rate is a behavioral measure, and the vacancy rate is not, the absorption rate is a more desirable measure for modeling purposes. However, we are not able to generate measures of the absorption rate in the model due to feedback loops. As a result, we have used a measure of housing market activity based on vacancy rates.

factors described above. Consequently, we assume that total construction in the region does not rise more than about 25 percent above the previous year's construction level or fall more than 50 percent below the highest previous construction level. As Table 7-1 indicates, these limits give ample leeway and would at most constrain construction levels in one year (1971), assuming construction patterns in individual regions are similar to those at the national level.

Land Costs and Land Use Restrictions

Most builders, especially those building single-family homes, are sensitive to the price of land.[1] As the supply of available land decreases, land prices are likely to rise. As the National Commission on Urban Problems [9] indicated, a conventional rule of thumb for builders is that the land component of a single-family home should not exceed about 20 percent of the total land-plus-house cost. As land prices rise, builders tend to build fewer homes, especially single-family homes in the low and medium price categories.[2] To simulate this, we consider the fraction of land which is vacant and available for construction. As this fraction becomes smaller, we assume prices for the remaining vacant land parcels rise. Builders react by constructing fewer low-priced units and more high-priced units to balance land and house costs. The President's Committee on Urban

---

1. Goldberg's [3] questionnaire study of builders in the Vancouver, B.C. region found that land price and land availability were important to developers.

2. Winger and Madden [18] found a shift toward increased construction of multi-family structures from 1961 to 1966. This shift appeared to be more connected with rising land prices and builders' efforts to keep overall costs down by building more units per acre, than with shifts in consumer preferences. Behman and Codella [1] found that the most significant factors in explaining rising prices of new homes were rising land prices and increasing size and amenity levels of new homes. See also The President's Committee on Urban Housing [10].

Table 7-1

Private Housing Starts

Year	Number of units	Change from previous year (%)
1959	1,516,800	--
1960	1,252,100	-17%
1961	1,313,000	5
1962	1,462,700	11
1963	1,603,200	10
1964	1,528,800	-5
1965	1,472,800	-4
1966	1,164,900	-21
1967	1,291,600	11
1968	1,507,600	17
1969	1,466,800	-3
1970	1,433,600	-2
1971	2,052,200	43
1972	2,356,600	15
1973	2,045,300	-13
1974	1,337,700	-35
1975	1,161,500*	-13

*Preliminary estimate

Source: United States Savings and Loan League, <u>1976 Savings and Loan Fact Book</u>, Table 7.

Housing [10] indicated that builders of multiple-family structures are not as affected by land costs, since such structures can usually be made taller as needed to keep land and structure costs in the appropriate proportions.

Zoning restrictions pose a significant external constraint on apartment builders.[1] In the few cities without zoning codes, land use patterns may be similar to those in zoned cities.[2] Builders are subject to density restrictions established formally by zoning boards or informally by neighborhood residents.

Density preferences of neighborhood residents can be used to simulate zoning restrictions. This is explained in Section 9, "Land Use and Zoning." Zoning restrictions within the model constrain, but do not totally prevent construction of multiple-family housing units. Some apartment construction is possible in any neighborhood, since the neighborhoods in the model typically encompass a reasonably large geographic area. However, lower density level tolerances, as expressed by the zoning code, will inhibit construction of multiple-family units.

Expected Demand

In an effort to plan their construction activity, builders find it useful to examine demographic and economic trends which are likely to affect patterns of housing demand. Some of these are outlined below.

---

1. In his study of builders in the Vancouver, B.C. area, Goldberg [3] found that many developers felt it was not worth the time and effort to try to have land rezoned for specific projects.

2. Siegan's [11] study of unzoned areas in Houston supports this.

Two demographic trends we have found useful in modeling different metropolitan regions are the shifting preferences for suburban areas over central areas, and the tendency for some sectors of a region to grow faster than others.

The shift of population from central city to suburban areas has long influenced builders. Households have sought to leave areas with high crime levels and unsatisfactory schools in favor of suburban areas where they could have single-family homes and some land.[1] Construction patterns reflect the shift, and construction in inlying and outlying areas responds to current demand and supply forces. Builders we talked to generally feel more optimistic about growth in outlying areas than in central areas. When economic forces cause a slowdown in the level of construction in the region, builders cut back less in outlying areas than in central areas of the region. In part this reflects the pattern of wealth and income distribution in many regions, where lower income households live in the central areas and middle and upper income households live in the suburbs. In modeling this we assume that builders reduce construction levels in suburban areas less than in central areas. When construction levels pick up again, both central and outlying areas respond.

In most metropolitan regions one can distinguish high and low prestige areas. Sometimes these areas form wedges or sectors which radiate from the central core.[2] The boundaries of these sectors can be defined fairly clearly

---

1. Straszheim [16] found there was demand for large lots and suggested that this demand could be satisfied at lower cost in the suburbs than in the central city.

2. See Hoover [4] for a brief discussion of observed sectoral patterns (i.e., why particular activities gradually move outward within the same sector).

by using Census data as well as perceptions of builders. For example, in Houston the highest prestige sector is in the southwest, while the lowest prestige sector is in the northeast. In Charlotte the most prestigious sector is in the south and southeast, while the low prestige sectors are in the north, northwest, and west. Dayton's highest prestige sector is also in the south and southeast, and its lowest prestige sector is in the west. The high prestige sectors tend to grow faster than low prestige sectors, as they attract households which move within or into the region. These arriving households want to be near similar household types, and want their homes to maintain their resale values in case they leave.

The financial status of households in the high prestige sectors also attracts builders and lending institutions.[1] Rising construction costs have made it increasingly difficult for builders to erect housing for lower-income and more recently, middle-income, households. Lending institutions also like to transact business in high prestige areas where default risks are low. In the low prestige sectors, builders slowly cut back their activity over time as the income levels of households in these areas become inadequate for the purchase of newly constructed housing units. In addition, lenders may become reluctant to grant mortgages in low prestige sectors. A self-fulfilling cycle is in effect. Because households, builders, and lending institutions regard an area as prestigious, they favor it as a place to live, build new units, and make loans and the area thrives.

---

1. Kaiser [6] found that the prestige level of a neighborhood, based on occupational, income and educational characteristics of the population and on the median home value, greatly influenced the locational decisions of builders.

For modeling purposes, sectors are defined as groups of contiguous neighborhoods forming a wedge and fanning out to the suburbs. The inclusion of suburban neighborhoods in the wedge depends on how far out development has taken place. Once the sectors have been defined, we measure their prestige levels according to the fraction of household heads with some education beyond high school. The greater the fraction, the higher the prestige level of the sector. For example, in Houston in 1970, the fraction of household heads with more than a high school education in the high prestige sector is 0.52; the fraction in the low prestige sector is 0.14. In Charlotte the corresponding figures are 0.52 and 0.15, and in Dayton, 0.35 and 0.10. In any year the effect of prestige level on a sector is slight. The maximum increase in construction for the highest prestige sector is 4 percent and the maximum decrease for the lowest prestige sector is 2 percent.

In addition to the large-scale, region-wide influences of the urban-suburban shift and the sector prestige levels, there are some smaller-scale trends which reflect the impact of economic pressures. Households tend to choose neighborhoods which are accessible to their workplaces,[1] and builders may develop new areas adjacent to neighborhoods experiencing rapid growth. Strictly speaking, these trends are part of the larger demographic patterns described above. That is, the shift of population to suburban areas is due in part to the increasing employment opportunities found there. The opening of new areas to construction reflects builder's expectations that people will continue to seek housing in the desirable areas of a region.

---

1. Straszheim [16] found that work site and accessibility affected the choice of location and the amount of housing consumed by households.

Builders see opportunities for new construction in areas around new job opportunities or in areas opened up by new highways. They expect, quite reasonably, that an increase in the number of jobs or improved access to a workplace will stimulate demand for housing units within some radius.

Evidence from various sources indicates that the residential location choices of households are constrained by the access afforded to the workplace.[1] Since workers typically limit their commuting time to 40 minutes or less, we assume that the workers employed at any given location will be distributed over a 40 minute travel radius from their workplace, with the majority 25 minutes or less away from work. This is consistent with some recent studies, which have indicated that commuters do not balance travel time and possible residential locations in an effort to economize on location rents or travel time.[2] Instead, commuters seem largely indifferent to the amount of time spent commuting, as long as it is below some threshold.

The process of modeling the impact of job distribution requires information on the existing distribution of employment opportunities and housing units. Builders in the model assume that most workers will choose housing within a 25 minute driving distance from their job sites. They assume that workers will distribute themselves evenly across the neighborhoods surround-

---

1. See Section 5, "Residential Location," for a summary of the evidence and for supporting tables from the Survey of Neighborhood Opinions.

2. See Section 5, "Residential Location," for a discussion of these studies.

ing the locus of employment opportunities,[1] and they can assess the impact of this behavior on the housing submarkets affected. If the number of housing units in a neighborhood is smaller than the number of workers likely to seek housing there, builders react to this expected demand by increasing construction in all submarkets of the neighborhood, to an extent determined by the difference between the expected demand and the current supply. At most, we expect the increase to be 10 percent. The minimum increase is set at 1 percent.

Opening Up New Areas

The areas opened to new development by builders usually lie in the path of urban expansion. Typically, most growth occurs in the high prestige sectors. As the land in one area is consumed by new construction, builders search for suitable vacant land nearby to continue the construction. This behavior has been observed in cities like Charlotte and Houston, as shown in Table 7-2.

Before builders open up a new area, they implicitly establish certain criteria which enable them to assess the risk involved. Builders look first at the situation in adjacent neighborhoods currently undergoing heavy construction. If a high percentage of the developable land has been used, further residential construction will be limited, and construction in new areas is likely to be profitable. Builders also consider the rate of construction in the adjacent neightborhoods undergoing development. If it is high enough that all the remaining vacant land is likely to be built upon

---

1. The approach used in modeling this behavior is similar to that used by Lustig and Pack [7] to allocate residences in their effort to develop zoning standards to accommodate households wishing to live near their workplaces.

Table 7-2

Examples of Areas Opened to New Development on a Large Scale in Charlotte and Houston

City	Neighborhood location	Number of Units: New Construction 1960	1970	Stock 1960	1970
Charlotte	Matthews-Pineville	170	358	663	3172
Charlotte	Steele Creek	51	263	863	2073
Houston	Middle West Houston	343	1538	152	6511
Houston	Middle West Houston	218	1214	201	5878
Houston	Far Southeast Houston	207	609	216	5888

within some number of years, builders are likely to seek a nearby tract of land suitable for large-scale development. The tract chosen must contain a large amount of vacant land, since economies of scale in large construction projects permit builders to reduce costs per unit.[1] Moreover, certain fixed costs, like fees to the municipality, or sewer and road installations, may be relatively low on a per unit basis.

Builders also consider competing uses of land in the neighborhood. If more than 15 percent of the land in the neighborhood is used for manufacturing purposes, builders are very unlikely to open the area to new development. For example, in Charlotte, North Carolina, the area north of Interstate 85 is becoming increasingly industrialized and hence less desirable for new residential development, even though there is much vacant land available. In a new area, builders expect to be able to attract households and offer housing units which can compete in price with units in built-up areas. If the neighborhood chosen for new development is already experiencing a high level of construction, it cannot, of course, be opened to new development. However, the mix of construction in the chosen neighborhood may be affected by the types of housing units demanded in the adjacent neighborhood whose growth triggered the search for an area suitable for new development.

Once an area suitable for new development has been located, the builder must decide how many and what types (by tenure and price level) of units to

---

1. The National Commission on Urban Problems [9] discusses a variety of cost savings associated with large-scale housing development. Stevens [15] found that there were economies in single-site construction of multi-family housing. See also The President's Committee on Urban Housing [10].

build.[1] In general, the builder constructs in the new area about the same number of units as are being constructed in the adjacent area already under development. The mix of units built speculatively reflects the types of units being demanded in the rapidly growing neighborhood. However, the developer does not erect low-priced units. High construction costs put most newly constructed units beyond the reach of low-income households.

Exogenous Influences on the Housing Stock

Certain changes in the size of the housing stock are considered exogenous in the model. Some of these changes, such as demolitions, reduce the size of the stock. Other changes, such as construction of public housing, add to the housing stock.

Demolitions typically result from one of two factors. If a more desirable use is found for a parcel of land, it may be advantageous to demolish the existing structure. Such demolitions may occur through the actions of private owners or public authorities. For example, the power of eminent domain may be exercised to seize land for public housing or highways. We do not attempt to model demolitions of this sort, as the results of such politically complex behavior appear to be distributed too irregularly. For modeling purposes we simply acknowledge the existence of these demolitions and reduce the size of the housing stock to reflect their occurrence. The second type of demolition is that which takes place when abandoned units are

---

1. Smith [14] provides a useful outline of the sorts of decisions a residential developer is forced to make when undertaking speculative construction of the sort described here.

torn down.[1] The process of abandonment is simulated, as explained in Section 8, "Investment in Maintenance of the Housing Stock."

The construction of public housing and more generally, units built with direct government assistance, are also treated as exogenous factors. Within the model publicly financed housing is acknowledged by adjusting the size of the housing stock as appropriate to indicate that these units have become available for occupancy. Most publicly subsidized units are rental units, and almost all are in the low and medium rent categories.

A final type of exogenous factor we consider is any special circumstance which somehow inhibits or bypasses normal market forces. For example, quirks of land ownership may discourage or encourage construction in some areas. Local ordinances, like sewer restrictions or growth controls, may influence greatly the patterns of new construction. To take into account these sorts of factors, we utilize information provided by the regions being studied for the affected neighborhoods. Typically, no more than four or five neighborhoods in a region are seriously affected by such special circumstances in our model.

The process of residential construction appears extremely complex, even in our simplified model. Numerous factors influence builders' decisions about where to build, what types of units to build, and how many units to build. Nevertheless, we have found that the process is remarkably similar in all the regions under study.

---

1. Friedman [2] points out that demolitions of abandoned or unfit units tend to be somewhat unsystematic.

Mathematical Statement

New Construction

To simulate the process of new construction, we model the responses of builders to various actors and forces in their environments. Builders assess available information in a complex manner. They are simultaneously influenced by demand and supply forces in the region and in individual neighborhood submarkets, by their expectations of future demand, and by land costs and land use restrictions.

The general functional form of the decision to build is as follows:

$$\text{Construction} = f(\text{demand forces, supply forces, builders' expectations of future demand, land costs, land use restrictions}) \quad (7.1)$$

We model this process by modulating the construction rate of the previous year using a series of multiplicative factors, each of which represents a major influence on builders. Equation 7.2[1] shows construction in the current year to be a function of both construction in the previous year and a series of variables defined below.

$$CON(T,P)_t = CON(T,P)_{t-1} * VADJ * PRSTGE * SFLPRB * PROBRT * LANDAV * REDLIN * WKMULT * AOGCW * RATIO \quad (7.2)$$

for all tenures (T), all price groups (P), and all neighborhoods (I), where:

---

1. In this and most subsequent equations in Section 7 the subscript I, which refers to the individual neighborhoods in a region, is suppressed to simplify the exposition.

$CON(T,P)_t$ = construction by tenure and price in current year

$CON(T,P)_{t-1}$ = Construction in previous year

VADJ = influence of demand and supply conditions

PRSTGE = influence of prestige level of the sector on the neighborhood

SFLPRB = impact of rising land prices on single-family home construction

PROBRT = impact of density restrictions (zoning) on rental unit construction

LANDAV = availability of vacant land suitable for construction

REDLIN = index of pressure for redlining

WKMULT = influence of employment access from the neighborhood

AOGCW = impact of significant exogenous factors

RATIO = impact of region-wide demand and supply conditions

The values of LANDAV and REDLIN in Equation 7.2 are either 0 or 1. That is, if there is no land suitable for construction, or if redlining has occurred, builders refrain from building in the neighborhood.

TOTDEV which represents total construction in an area opened to new development, does not appear in Equation 7.2. In choosing to develop a new area and in deciding how many units to construct, builders follow a different set of guidelines from those in Equation 7.2. The appropriate equation for construction of this sort is the following:

$$TOTDEV = VLTST * TSTMFG * PRTEST * CONUSE * AOGTST * D_J * RESTST_J * CONADJ_J * STCTST_J * CONLMT_J \qquad (7.3)$$

for all qualifying neighborhoods (I), where:

TOTDEV = total construction in a neighborhood opened to new development

VLTST = test for availability of vacant land in neighborhood being considered for new development

TSTMFG = test for level of manufacturing activity in neighborhood being considered for new development

PRTEST = test for prestige value of sector which encompasses neighborhood being considered for new development

CONUSE = test for current construction level in neighborhood being considered for new development

AOGTST = impact of significant exogenous factors

D_J = distance test between neighborhood being considered for new development and adjacent neighborhoods (J)

$RESTST_J$ = test for extent of development in adjacent neighborhoods (J)

$CONADJ_J$ = test for level of construction activity in adjacent neighborhoods (J)

$STCTST_J$ = test for level of construction activity relative to amount of housing stock in adjacent neighborhoods (J)

$CONLMT_J$ = highest level of new construction in an adjacent neighborhood (J).

The value of all the independent variables except CONLMT in Equation 7.3 is either 0 or 1, since they represent criteria which the candidate neighborhood or its adjacent neighborhoods must satisfy in order for development to proceed. If a neighborhood qualifies for new development, the value of TOTDEV indicates the extent of new construction there, and the following holds for such neighborhoods:

$$CON_t = TOTDEV. \tag{7.4}$$

The factors influencing builders are discussed in the sequence in which they are presented in Equations 7.2 and 7.3.

Neighborhood Housing Market Conditions

Within neighborhood submarkets, builders respond to demand and supply conditions. They compare actual vacancy rates to the empirically determined standard vacancy rates,[1] as shown below:

$$VR(T,P) = VAC(T,P) \;/\; STOC(T,P) \tag{7.5}$$

for all tenures (T), all price groups (P), and all neighborhoods (I), where:

VR(T,P) = vacancy rate in submarket of Tenure T and Price Group P

VAC(T,P) = number of vacant units of Tenure T and Price Group P

STOC(T,P) = total number of housing units of Tenure T and Price Goup P.

The vacancy rate comparison variable is VDIFF:

$$VDIFF(T,P) = VR(T,P) - VSTD(T,P) \tag{7.6}$$

for all tenures (T), all price groups (P), and all neighborhoods (I) where:

VDIFF(T,P) = the difference between the actual and standard vacancy rates in the submarket of Tenure T and

---

1. We use the terms standard vacancy rate and normal vacancy rate interchangeably.

Price Group P

VSTD(T,P) = standard vacancy rate in submarket of Tenure T and Price Group P.

If VDIFF(T,P) is positive, the actual vacancy rate exceeds the standard vacancy rate, so builders decrease the construction level in the submarket. If VDIFF(T,P) is negative, they increase construction. The change in construction levels is represented by the function below:

$$VADJ(T,P) = 1 + s1 * VDIFF(T,P) \tag{7.7}$$

for all tenures (T), all price groups (P), and all neighborhoods (I), where:

VADJ(T,P) = a construction modulator reflecting the influence of demand and supply forces on builders within a submarket ($.9 \leq VADJ \leq 1.1$)

s1 = a parameter whose values are typically -8.0 for owned housing and -4.0 for rented housing.

Depending on the value for VADJ(T,P), builders increase or decrease construction levels in neighborhood submarkets. However, builders do not increase construction in the low-priced markets.

Prestige Level of the Sector

Builders' expectations of future demand influence construction decisions. All central city neighborhoods and some of the suburban neighborhoods within each region are assigned to a sector based on Census information. The prestige level of a sector affects builders' expectations of future demand in that sector. The prestige value of a sector is based on the fraction of household heads with more than a high school education.

$$EDUC(II) = SECTCL(II) / SECTHH(II) \tag{7.8}$$

for all sectors (II), where:

EDUC(II) = an initial prestige measure of Sector II

SECTCL(II) = the number of household heads with more than a high school education in Sector II

SECTHH(II) = the total number of household heads in Sector II.

Once the initial prestige values have been calculated, they are scaled to fall around 1.0. At present, the values of the prestige variable are constrained as follows:

$$.98 \leq \text{PRSTGE} \leq 1.04 \qquad (7.9)$$

for all neighborhoods (I), where:

PRSTGE = the prestige value of Neighborhood I in Sector II.

The value of PRSTGE for a particular neighborhood is determined by the sector which includes that neighborhood.

Land Prices

In the model, only single-family home construction is affected by rising land prices. As residential density in a neighborhood increases, land prices also increase, inhibiting new construction. Low-priced housing is affected to the greatest extent by rising land prices, and high-priced housing is least affected. In the face of high land prices, we have found that builders in the model may reduce construction by as much as 20 percent of its previous level. The form of the land price variable in Equation 7.2 is SFLPRB(P), where

$$0.8 \leq \text{SFLPRB(P)} \leq 1.0 \qquad (7.10)$$

for all price groups (P), where:

SFLPRB(P) = the impact of land prices on new single-family construction in Price Group P.

In rental housing submarkets the value of SFLPRB(P) is always 1.

Land Use Restrictions

Land use restrictions in the form of zoning affect construction only of multiple-family dwellings in the model. If rental unit construction in a neighborhood is already above some threshold level, and if there is a large amount of usable vacant land, there is no restriction on rental unit construction. Similarly, if the maximum density permitted by the zoning board is equal to or greater than six units per acre, there is no restriction on rental unit construction. However, if the maximum permitted density is under six units per acre, then

$$PROBRT = MAXDEN / 6, \qquad (7.11)$$

where the value of PROBRT is constrained:

$$PROBRT \geq .09 \qquad (7.12)$$

for all neighborhoods (I), where:

PROBRT = the impact of land use restrictions on rental unit construction

MAXDEN = maximum permitted density.

The value of PROBRT is adjusted according to the relative number of rental units in the neighborhood:

$$PCTRNT = RSTOC / TSTOC \qquad (7.13)$$

$$PROBRT' = (0.1 * PCTRNT + 0.9) * PROBRT \qquad (7.14)$$

for all neighborhoods (I), where:

PCTRNT = fraction of total stock which is rental units

RSTOC = number of rental units

TSTOC = total number of housing units

PROBRT' = revised value of PROBRT

The value of the land use restriction (zoning) variable in Equation 7.2 is constrained as follows:

$$0.81 \leq PROBRT' \leq 1.00. \qquad (7.15)$$

In owner-occupied housing submarkets the value of PROBRT or PROBRT' is always 1.

Availability of Vacant Land

If a neighborhood has no suitable vacant land, builders are unable to construct new units. The form of this constraint in Equation 7.2 is a binary variable, LANDAV. If there is any suitable vacant land, the value of the variable is 1. If there is no vacant land, the value is 0.

Unavailability of Credit and Insurance

If a neighborhood has been systematically denied credit and property insurance by lenders and insurers, builders are quite unlikely to construct new units there. The form of this constraint in Equation 7.2 is a binary variable.[1] If the value of the redlining variable (REDLIN) is above some threshold, the constraint variable has a value of 0. Otherwise its value is 1.

Effect of Employment Accessibility

This section describes the procedure for estimating the impact of employment accessibility on construction levels in a neighborhood. We first determine the capacity of each neighborhood to accommodate housing units (AVAIL). This capacity is based on the expected residential density of the neighborhood (EXPDEN) and the acreage available. We then compare capacity with the expected demand. If we expect more workers (TWORK) to choose a neighborhood than can be accommodated by the available housing stock (TSTOC), builders will increase their level of construction in the neighborhood. The extent of the increase depends on HRATIO, the ratio of TWORK and TSTOC.

---

1. The derivation of the redlining variable is described in Section 6, "Dynamics of Local Real Estate Markets."

Expected residential density is determined as follows:

$$\text{EXPEN(J)} = 1.0 + \text{SLOPE} * \text{JOBS(J)} \tag{7.16}$$

for all neighborhoods (J), where:

$$\text{JOBS(J)} = \sum_{L=1}^{\text{NIND}} \sum_{I=1}^{\text{NTRACT}} \text{RJ(L,I,DT)}, \tag{7.17}$$

$$(\text{DT} \in 0,25)$$

for all neighborhoods (J), and where:

EXPDEN(J) = expected residential density (number of housing units per acre) in Neighborhood J

SLOPE = empirically determined parameter relating job accessibility of a neighborhood to expected residential density in the neighborhood (typical value .00005)

JOBS(J) = number of jobs located within 25 minute driving distance of Neighborhood J[1]

NIND = number of industry types

NTRACT = number of neighborhoods in the region

RJ(L,I,DT) = number of jobs in Industry Type L in Neighborhood I located DT minutes driving time from Neighborhood J.

Using the expected residential density, we calculate the number of housing units which each neighborhood can accommodate as follows:

---

1. We find that builders use this conservative threshold, rather than the actual 40 minute one, in making their decisions.

$$CAPAC(J) = \left[RES(J) + EASYVAC(J)\right] * EXPDEN(J) \qquad (7.18)$$

for all neighborhoods (J), where:

CAPAC(J) = housing unit capacity in Neighborhood J, given the expected residential density of J

RES(J) = amount of land in residential use in Neighborhood J

EASYVAC(J) = amount of vacant easy-to-build land in Neighborhood J.

The number of housing units in Neighborhood J which should be available to people working in Neighborhood I is:

$$AVAIL(I,J) = CAPAC(J) * FRAC \qquad (7.19)$$

for all neighborhoods I and all neighborhoods J, where:

FRAC = 1 if driving time (DT) from I to J $\leq$ 25 minutes

FRAC = .2 if $25 < DT \leq 40$

FRAC = 0 if $DT > 40$.

The neighborhoods (J) which are more than 25 minutes driving time from Neighborhood I are assumed to be less attractive than those less than or equal to 25 minutes driving time from Neighborhood I. Neighborhoods more than 40 minutes driving time from Neighborhood I are not considered at all by workers in choosing a place to live.

The total number of units potentially accessible from I is:

$$TAVAIL(I) = \sum_{J=1}^{NTRACT} AVAIL(I,J) \qquad (7.20)$$

for all neighborhoods (I).

Next, we determine where workers in Neighborhood I are likely to live. We assume workers choose among the available neighborhoods (J) according to the effective capacities of these neighborhoods to accommodate housing; that is:

$$TWORK(J) = \sum_{I=1}^{NTRACT} \sum_{J=1}^{NTRACT} TOTJOB(I) * \left[AVAIL(I,J) \; / \; TAVAIL(I)\right] \tag{7.21}$$

for all neighborhoods (J), where:

TWORK(J) = total number of workers employed in all Neighborhoods I who choose to live in Neighborhood J

TOTJOB(I) = the total number of workers working in Neighborhood I.

Once TWORK is known for all neighborhoods, we can determine whether each neighborhood has sufficient housing units to accommodate all the workers that might be expected to move there. The following ratio is computed:

$$HRATIO(I) = \left[TWORK(I) * CONJOB\right] \; / \; TSTOC(I) \tag{7.22}$$

for all neighborhoods (I), where:

HRATIO(I) = the ratio of the number of workers demanding housing in I to the total housing stock in I

CONJOB = empirically determined ratio of households to workers (CONJOB $<$ 1.0)

TSTOC(I) = total number of housing units in Neighborhood I.

The value of HRATIO(I) indicates the extent to which a neighborhood can satisfy the expected demand for housing. The higher the value of HRATIO(I), the less able the neighborhood is to accommodate this demand. Consequently, high values for HRATIO(I) stimulate construction in Neighborhood I, to a degree based on the following function:

$$WKMULT(I) = 1 + \left[b + sl * (HRATIO(I) - c\right] \tag{7.23}$$

for all neighborhoods (I), where:

WKMULT(I) = the impact on builders in Neighborhood I of employment accessibility for the neighborhood ($1.0 \leq WKMULT \leq 1.10$)

b = a parameter whose value is typically .01

s1 = a parameter whose value is typically .015

c = a parameter whose value is typically 2.0.

The value of WKMULT is constrained in several additional ways. First, if the value of HRATIO is greater than 1.0, WKMULT must be at least 1.01. Second, if the neighborhood under consideration is at the lowest prestige level, WKMULT remains at 1.0. Finally, if total construction in Neighborhood I already exceeds some threshold, WKMULT is set to 1.0. If a neighbhrhood still qualifies for an increase from WKMULT, construction levels in all submarkets of the neighborhood are affected by the increase.

Exogenous Factors

Exogenous factors, such as an unusual stimulus to demand, may cause builders to boost construction levels in a neighborhood. Exogenous factors like sewage restrictions may dampen construction levels in a neighborhood. Where such circumstances could not otherwise be captured by the model, an Act-of-God variable, AOGCW, is employed. In Equation 7.2,

$$0.0 \leq AOGCW \leq 2.0 \qquad (7.24)$$

for all neighborhoods (I), where:

AOGCW = exogenous factor affecting builders' construction decisions.

However, very few neighborhoods in a region are affected by significant exogenous circumstances.

Opening New Areas to Development

If a neighborhood and its adjacent areas meet several criteria, builders will open it to new construction. All of these criteria enter Equation 7.3

as binary variables. For a neighborhood to undergo new development, the value of all of these variables must be 1. If a single criterion is not met, the value of the variable representing that criterion will be 0, and new development will not occur.

The area where development occurs must have sufficient vacant land to permit sizable levels of construction. That is,

$$TLAND = RES + EASYVAC \qquad (7.25)$$

$$VLTST = EASYVAC / TLAND \qquad (7.26)$$

for all neighborhoods (I), where:

TLAND = developed residential land plus available vacant land

RES = land in residential use

EASYVAC = available vacant land

VLTST = vacant land test, indicating fraction of TLAND which is available.

If

$$VLTST < DSSPEC(1) \qquad 7,27)$$

where:

DSSPEC(1) = threshold value for VLTST (typical value .7),

then the area will not undergo new development.

For new development to take place in an area, there must be little or no manufacturing activity.

$$TSTMFG = (LTMFG + HVMFG) / TOTLAND \qquad (7.28)$$

for all neighborhoods (I), where

TSTMFG = test value for level of manufacturing activity

LTMFG = land in light manufacturing use

HVMFG = land in heavy manufacturing use

TOTLAND = total land area.

If

$$TSTMFG > 0.15 \tag{7.29}$$

then new development will not occur.

If a neighborhood is in a low prestige sector, the area is not likely to undergo new development. That is, if

$$PRSTGE < PFACT \tag{7.30}$$

for all neighborhoods (I), where

PRSTGE = the prestige value for Neighborhood I in Sector II

PFACT = the threshold prestige value (typical value 1.0),

then the neighborhood will not be developed.

A neighborhood is excluded from consideration for new development if it is already experiencing sufficient construction to build it out eventually. That is,

$$CONCAP = DSSPEC(2) * EASYVAC \tag{7.31}$$

$$TRIG2 = CONCAP / DSSPEC(4) \tag{7.32}$$

$$CONUSE = \sum_{T=1}^{2} \sum_{P=1}^{3} CON(T,P) \tag{7.33}$$

for all neighborhoods (I), where:

CONCAP = capacity for new construction

DSSPEC(2) = average number of housing units which can be constructed per acre (typical value 2.0)

EASYVAC = available vacant land

TRIG2 = annual construction required to build out the neighborhood

DSSPEC(4) = average number of years required to build out a new area (typical value 20)

CONUSE = total present construction level

CON(T,P) = current construction level in Tenure T and Price Group P.

If

$$CONUSE > TRIG2 \quad (7.34)$$

for all neighborhoods (I), then the present construction level in the neighborhood is high, meaning it is already undergoing development.

Occasionally an exogenous circumstance, such as someone holding land for speculative purposes, prevents an area from being developed. Where this sort of information is available from the region being studied, it is taken into account. In Equation 7.3, AOGTST reflects the value of AOGCW, where

$$0 \geq AOGCW \geq -1.0 \quad (7.35)$$

for all neighborhoods (I), where:

AOGCW = impact of exogenous factors on builders' decisions.

If the value of AOGCW is 0, the value of AOGTST in Equation 7.3 is 1, and new development can occur. If AOGCW is less than 1, AOGTST is 0, and new development does not occur.

Finally, if a futile attempt was made in the past to develop a new area, no further attempt is permitted.

Once a neighborhood meets all the above criteria, it becomes a candidate for new development. Such development must be stimulated by activity in a nearby neighborhood. If one or more of the adjacent neighborhoods (J) meets all the criteria outlined below, then development activity will begin in the candidate neighborhood. The tests below are made for each candidate neighborhood (I).

The center of the adjacent neighborhood must be within a specified distance of the center of the candidate, if it is to influence activity in the candidate neighborhood. That is, if

$$DSSPEC(5) < D(J) \quad (7.36)$$

for all adjacent neighborhoods (J)[1] where:

DSSPEC(5) = threshold distance for Neighborhood J to influence Neighborhood I ($0.5 \lesssim DSSPEC(5) \lesssim 15$, depending on the region)

D(J) = distance between the center of candidate neighborhood I and the adjacent Neighborhood J,

then the adjacent neighborhood is too far away to stimulate development in the candidate neighborhood.

The next criterion is the extent to which the adjacent neighborhood has used up its available land.

$$TLAND(J) = RES(J) + EASYVAC(J) \quad (7.37)$$

$$RESTST(J) = RES(J) \;/\; TLAND(J) \quad (7.38)$$

for all adjacent neighborhoods (J), where:

TLAND(J) = total residential land plus available vacant land in adjacent Neighborhood J

RESTST(J) = fraction of developable land already in residential use in Neighborhood J.

If

$$DSSPEC(6) < RESTST(J) \quad (7.39)$$

---

1. The subscript I refers to the candidate neighborhood, and the subscript J refers to one of the neighborhoods geographically contiguous to the candidate neighborhood. There can be as many as seven neighborhoods contiguous to the candidate neighborhood.

for any adjacent neighborhoods (J), where

DSSPEC(6) = threshold for level of residential use in adjacent Neighborhood J (typical value .45),

then the adjacent Neighborhood J may stimulate construction in the candidate neighborhood, as much of the developable land in Neighborhood J has been used.

If the level of construction in the adjacent Neighborhood J is too low (i.e., below its buildout rate), it will not stimulate construction in the candidate neighborhood. That is,

$$CONADJ(J) = \sum_{T=1}^{2} \sum_{P=1}^{3} CON(T,P,J) \qquad (7.40)$$

$$TRIG4(J) = \left[DSSPEC(2) * EASYVAC(J)\right] / DSSPEC(8) \qquad (7.41)$$

for all adjacent neighborhoods (J), where:

CONADJ(J) = total construction in adjacent Neighborhood J

CON(T,P,J) = total construction in Tenure T and Price Group P in Neighborhood J

TRIG4(J) = annual construction level required to build out adjacent Neighborhood J

DSSPEC(2) = average number of housing units which can be constructed per acre in adjacent Neighborhood J (typical value 2.0)

DSSPEC(8) = average number of years required to build out an area under development (typical value 15).

If

$$CONADJ(J) < TRIG4(J) \qquad (7.42)$$

then the construction level in adjacent Neighborhood J is too low to stimulate construction in the candidate neighborhood.

There may, however, be very little vacant land available in an adjacent neighborhood. A low level of construction in such a neighborhood will cause the neighborhood to be built out entirely, and it could trigger construction in the candidate neighborhood. To prevent this, we establish two additional criteria which the adjacent neighborhood must meet.

First, if

$$\text{EASYVAC(J)} \geq 10 \tag{7.43}$$

for any adjacent neighborhoods (J), the adjacent neighborhood may trigger construction elsewhere.

The second test considers the likely levels of construction in previous years in built-up adjacent neighborhoods. That is,

$$\text{STOCK(J)} = \sum_{T=1}^{2} \sum_{P=1}^{3} \text{STOC(T,P,J)} \tag{7.44}$$

$$\text{STCTST(J)} = \text{STOCK(J)} \;/\; \text{DSSPEC(8)} \tag{7.45}$$

for all adjacent neighborhoods (J), where

STOCK(J) = total housing stock in adjacent Neighborhood J

STCTST(J) = average number of housing units built each previous year in adjacent Neighbrhood J.

The value for STCTST(J) cannot exceed 100. if

$$\text{STCTST(J)} < \text{CONADJ(J)}, \tag{7.46}$$

an adjacent Neighborhood J may stimulate construction in the candidate neighborhood.

If at least one adjacent neighborhood (J) meets the criteria outlined above, new development is likely to be stimulated in the candidate neighborhood. The extent and compsition of the new construction are determined next.

The quantity of new construction in the candidate neighborhood will generally equal the level of construction in the adjacent neighborhood which stimulated the new construction. If two or more adjacent neighborhoods meet the criteria, the highest level of construction is used. However, the construction level in the candidate neighborhood will not exceed some empirically determined limit (ranging from 350 to 500). However, if the construction level in the candidate neighborhood already exceeds that of the qualifying adjacent neighborhoods, construction in the former remains unchanged. That is,

$$\text{TOTDEV (I)} = \text{MIN (CONLMT, SPCLMT)} \tag{7.47}$$

for all qualifying neighborhoods (I), where:

TOTDEV(I) = the level of new construction in the candidate Neighborhood I

CONLMT = the highest level of construction in a qualifying adjacent Neighborhood J

SPCLMT = an empirically determined upper limit to construction in an area undergoing new development.

The composition of construction in the candidate neighborhood is influenced by the types of units demanded in the adjacent neighborhood in the current year's market clearing.

That is,

$$\text{COMPSP(T,P)} = \sum_{T=1}^{2} \sum_{P=1}^{2} \left[\text{TOCC(T,P,J)} - \text{OLDTOC(T,P,J)}\right] \tag{7.48}$$

for all tenures (T), for medium and high priced units, for all qualifying adjacent neighborhoods (J), where:

COMPSP(T,P) = the types of housing units bought and rented, by tenure and price group, in all qualifying adjacent Neighborhoods J

TOCC(T,P,J) = occupied housing units, by tenure and price group, in all qualifying adjacent Neighborhoods J after market clearing

OLDTOC(T,P,J) = occupied housing units, by tenure and price group, in all qualifying adjacent Neighborhoods J before market clearing.

Note that housing units in the low price and rent groups are excluded by COMPSP(T,P), as builders do not find it profitable to construct such housing.

The total change in number of occupied housing units is thus

$$\text{DELOCC(J)} = \sum_{T=1}^{2} \sum_{P=1}^{2} \text{COMPSP(T,P)} \qquad (7.49)$$

for all qualifying adjacent neighborhoods (J), where:

DELOCC(J) = total change in number of occupied housing units in all qualifying adjacent Neighborhoods J.

Using this information, we determine the types of units constructed in the candidate neighborhood as follows:

$$\text{CON(T,P,I)} = [\text{COMPSP(T,P)/DELOCC(J)}] * \text{TOTDEV(I)} \qquad (7.50)$$

for all tenures (T), for medium and high priced units, and for all neighborhoods (I).

Two points must be noted. First, no low priced or low rent units are built in the area undergoing substantial new construction. Second, even if the adjacent neighborhoods do not affect the extent of new construction in the candidate neighborhood, they nevertheless affect its composition.

Region-Wide Demand and Supply Conditions

The demand and supply conditions within a particular regional submarket (defined by tenure and price group) affect builders' decisions at the neighborhood submarket level. If the regional vacancy rate in a submarket is too high (i.e., above the standard vacancy rate), builders curtail construction in this submarket in all neighborhoods. This occurs even if the demand for housing within that submarket in a particular neighborhood is rather strong. On the other hand, if the regional vacancy rate in a submarket is too low (i.e., below the standard vacancy rate), builders increase construction in this submarket in all neighborhoods, even if the demand for housing in that submarket is weak in a particular neighborhood.

We model this behavior by estimating the need for new housing in each submarket for the entire region. After builders have made tentative plans for construction, as outlined in this chapter, they adjust their plans by increasing or decreasing planned construction levels proportionately so that the total construction which actually occurs will approximate the region-wide estimates, by submarket, of construction needed.

Determination of region-wide needs for new construction is based on conditions in each housing submarket at the beginning of the year. This is modeled as follows:

$$SMVACR(T,P) = TOTVAC(T,P) \ / \ TOTSTC(T,P) \qquad (7.51)$$

for all tenures (T) and all price groups (P), where:

SMVACR(T,P) = regional vacancy rate in submarket for Tenure T and Price Group P

TOTVAC(T,P) = total vacancies of Tenure T and Price Group P in the region

TOTSTC(T,P) = total number of housing units of Tenure T and Price Group P in the region.

The regional vacancy rate is compared with the standard vacancy rate for the submarket:

$$DIF(T,P) = SMVACR(T,P) - VSTD(T,P) \quad (7.52)$$

for all tenures (T) and all price groups (P), where:

DIF(T,P) = the difference between the actual and standard vacancy rates within a regional submarket

VSTD(T,P) = the standard vacancy rate for Tenure T and Price Group P within a regional submarket.

If DIF(T,P) is positive, the regional vacancy rate for the submarket exceeds the standard vacancy rate, so builders reduce construction. If DIF(T,P) is negative, builders increase construction. The size of the increase or decrease is derived from the function below:

$$ADJCON(T,P) = 1 + sl * DIF(T,P) \quad (7.53)$$

for all tenures (T) and all price groups (P), where:

ADJCON(T,P) = the multiplier for Tenure T and Price Group P to increase or decrease construction levels ($.5 \lesssim ADJCON(T,P) \lesssim 1.25$)

sl = a parameter whose value is typically -100 for owned housing and -15 for rented housing.

Once the multiplier ADJCON(T,P) is known for a regional submarket, the amount of new housing needed for this submarket in the current year is calculated as follows:

$$CONNEW(T,P) = OLDCON(T,P) * ADJCON(T,P) \qquad (7.54)$$

for all tenures (T) and all price groups (P), where:

CONNEW(T,P) = construction needed in the current year for submarket of Tenure T and Price Group P

OLDCON(T,P) = total construction in the previous year in submarket of Tenure T and Price Group P.

If the value of CONNEW(T,P) falls below some predetermined level, it is adjusted as follows:

$$TOTNEW = \sum_{T=1}^{2} \sum_{P=1}^{3} CONNEW(T,P) \qquad (7.55)$$

where

TOTNEW = total construction needed for the region.

If

$$TOTNEW < (CONMAX * 0.5), \qquad (7.56)$$

where

CONMAX = the highest level of construction within the region in a previous year of the model run,

then each value of CONNEW(T,P) is multiplied by

(0.5 * CONMAX)/TOTNEW,

where

$$[(0.5 * CONMAX) / TOTNEW] > 1.0 \qquad (7.57)$$

Builders make their final construction plans within neighborhoods according to the values of CONNEW(T,P), the amount of new construction needed within a regional submarket.

To prevent excessive increases in construction in any neighborhood submarket, the construction algorithm constrains total construction to a specified limit. That is,

$$\text{TENCON(T)} = \sum_{P=1}^{3} \text{CON(T,P,I)} \tag{7.58}$$

for all tenures (T) and all neighborhoods (I), where:

TENCON(T) = total construction by tenure within Neighborhood I

CON(T,P) = total construction in Tenure T, Price Group P, and Neighborhood I.

That is, if

$$\text{TENCON(T,P)} > \text{CONLIM} \tag{7.59}$$

for all tenures (T) and all neighborhoods (I), where

CONLIM = limit to amount of new construction possible in either tenure in any neighborhood (CONLIM $\leq$ 1000),

then

$$\text{REDUC} = \text{CONLIM} \ / \ \text{TENCON(T)} \tag{7.60}$$

$$0 < \text{REDUC} < 1.0 \tag{7.61}$$

where:

REDUC = multiplier for adjusting construction levels.

Finally, builders compare their tentative construction plans with their explicit regional construction goals.

$$\text{SUMCON(T,P)} = \sum_{I=1}^{\text{NTRACT}} \text{CON(T,P)} \tag{7.62}$$

for all tenures (T) and all price groups (P), where:

SUMCON(T,P) = total planned construction within each submarket in the region

NTRACT = total number of neighborhoods in the region.

Builders compare SUMCON(T,P) with the target levels indicated by CONNEW(T,P). That is

$$RATIO(T,P) = CONNEW(T,P) / SUMCON(T,P) \qquad (7.63)$$

for all tenures (T) and all price groups (P), where

RATIO(T,P) = the ratio of target construction levels to planned construction levels.

If the value of RATIO(T,P) is less than 1.0, it is increased for neighborhoods lying outside the central city, since builders do not reduce construction in outlying areas as much as in inlying areas. In such cases, the value of RATIO(T,P) is increased by

$$[1.0 - RATIO(T,P)] * SUBRAT, \qquad (7.64)$$

for all tenures (T) and all price groups (P), where

SUBRAT = empirically determined ratio adjustment factor $(0.1 \leq SUBRAT \leq 0.25)$.

Once the values for RATIO(T,P) have been determined, builders decide on their final construction levels [CON(T,P)]:

$$CON(T,P)' = CON(T,P) * RATIO(T,P) \qquad (7.65)$$

for all tenures (T) and all price groups (P).

Exogenous Changes in the Housing Stock

Neither public housing starts nor demolitions are simulated in the model. For those years in which the simulated level of abandonment exceeds the reported level of demolitions in a neighborhood submarket, we use the figure for abandonment as the estimate of demolitions. Adjustments to the size of the housing stock to reflect public housing starts and demolitions are made to the variable representing the number of available vacant units, as follows:

$$AVAC(T,P)' = AVAC(T,P) + AOGPH(T,P) - DEMO(T,P) \quad (7.66)$$

for all tenures (T), all price groups (P), and all neighborhods (I), where:

AVAC(T,P)' = number of available vacant units after adjusting for public housing starts and demolitions

AVAC(T,P) = number of available vacant units before adjusting for public housing starts and demolitions

AOGPH(T,P) = number of public housing starts

DEMO(T,P) = number of demolitions.

REFERENCES

1. Behman, S. and Codella, D. "Wage Rates and Housing Prices." Industrial Relations, vol. 10. February, 1971

2. Friedman, Lawrence M. Government and Slum Housing. Chicago: Rand McNally & Company, 1968.

3. Goldberg, Michael A., "Residential Developer Behavior: Some Empirical Findings," Land Economics, February, 1974.

4. Hoover, Edgar M., "The Evolving Form and Organization of the Metropolis," in Harvey S. Perloff and L. Wingo, editors, Issues in Urban Economics. Baltimore: Johns Hopkins Press, 1968.

5. Hua, Chang-i. Housing Vacancy and Housing Market. Pittsburgh, Pa.: School of Urban and Public Affairs, Carnegie-Mellon University, 1974.

6. Kaiser, E.J., A Producer Model for Residential Growth. Center for Urban and Regional Studies, University of North Carolina, Chapel Hill, N.C.: 1968.

7. Lustig, Morton and Janet R. Pack, "A Standard for Residential Zoning Based Upon the Location of Jobs." Journal of the American Institute of Planners, Vol. 40 (5), September, 1974.

8. Maisel, Sherman J., "A Theory of Fluctuations in Residential Construction Starts." American Economic Review, June, 1963.

9. National Commission on Urban Problems. Building the American City, Report of the National Commission on Urban Problems to the Congress and to the President of the United States. House Document No. 91-34, Washington, D.C.: Government Printing Office, 1968.

10. President's Committee on Urban Housing. A Decent Home Washington, D.C.: Government Printing Office, 1968.

11. Siegan, Bernard H. Land Use without Zoning. Lexington, Mass.: Lexington Books, D.C. Heath and Co., 1972.

12. Smith, Lawrence B., "A Note on the Price Adjustment Mechanism for Rental Housing." American Economic Review, 1974.

13. Smith, Wallace F. Aspects of Housing Demand - Absorption, Demolition and Differentiation. Research Report No. 29, Center for Real Estate and Urban Economics. Berkeley: University of California, 1966.

14. Smith, Wallace F. Housing: The Social and Economic Elements. Berkeley, Cal.: University of California Press, 1971.

15. Stevens, Barbara, "Single-Site Economies in the Construction of Multi-Family Housing." Land Economics, Vol. 51 (1), February, 1975.

16. Straszheim, Mahlon R., "Estimation of the Demand for Urban Housing Services from Household Interview Data," Review of Economics and Statistics, February, 1973.

17. United States Savings and Loan League. 1976 Savings and Loan Fact Book. Chicago: United States Savings and Loan League, 1976.

18. Winger, A.R. and Madden, J. "The Application of the Theory of Products: The Case of Residential Construction." Quarterly Review of Economics and Business. Summer, 1970.

Section 8

INVESTMENT IN MAINTENANCE OF THE HOUSING STOCK

The condition of the housing stock in a neighborhood reflects numerous past decisions by homeowners and landlords regarding investment in maintenance. Changes over time in the stock condition of a neighborhood affect decisions of households moving within, into, and out of the neighborhood. Such changes also affect the maintenance behavior of current homeowners and landlords, as well as the business decisions of lending institutions and property insurers. If the stock condition remains good, builders may construct new units in the neighborhood. If the overall condition deteriorates significantly, property owners tend to reduce their maintenance, and some units may become unfit for habitation. Such units may be abandoned and subsequently demolished.

Deterioration of the Stock

Through normal usage and the effects of the elements, the physical condition of all housing units deteriorates over time. Condition refers to the state of repair or habitability of a dwelling unit. Changing tastes, as well as improving technology, can create technological obsolescence. Numerous features of a house are subject to obsolescence, including architectural style, quality of insulation, types of built-in appliances, and the number of bathrooms. These sorts of features determine the quality of a housing unit. The quality of a unit refers to its intrinsic characteristics, as constructed or renovated. The quality of a unit may be high even though the condition is poor. Conversely, the condition may be good but the quality may be low.

The rate at which housing units deteriorate in value is difficult to assess. As an approximation we assume that the value of all units deteriorates at a rate of 6.5 percent per year. We assume that most housing units will lose their value if not maintained at all for 10-15 years. However, the rate at which particular units deteriorate depends on numerour factors, such as climate, the intrinsic quality of the unit (e.g., the durability of its roofing or siding), the condition of the unit, and the intensity with which the unit is used.[1] The exact figure used is not too important, as our main concern is with changes in the relative condition of the housing stock across neighborhoods.

Investment in Maintenance

Homeowners and landlords invest in maintenance because they wish to preserve the condition of the housing they own. However, the investment behavior of homeowners differs from that of landlords. Homeowners invest in maintenance because they wish, in general, to consume

---

1. Eisenstadt [3], in her study of apartment buildings in New York City, found that the expenditures required to maintain and operate private rental housing to the standards set by the City regulations varied with a building's physical characteristics.

high levels of housing services.[1] Our analysis of maintenance behavior by different household types[2] indicates that the age, ethnicity, and educational characteristics of homeowners may be related to their maintenance efforts[3], as shown in Table 8-1. For example, foreign-born households may be unable to maintain their units as well as they might like due to low income levels. External factors also affect maintenance behavior. In general, home and apartment owners are reluctant to maintain their units considerably above or below the average for the neighborhood. Although maintenance raises the value of housing units, it is difficult for a unit to rise to a value considerably above the average in the neighborhood. This tends to impose an upper limit on maintenance investments. Homeowners and landlords may have an incentive to maintain their units below the neighborhood average. Units treated this way are protected somewhat against declines in value by the higher-value units surrounding them. However, homeowners and landlords who

---

1. Grigsby [4] and others have noted the tendency for homeowners to maintain their units better than landlords as a result of their desire for more pleasurable occupancy. Sweeney [10] has developed a model to explain such observations. Dildine and Massey [2], however, question whether owner-occupied homes are better maintained than renter-occupied dwellings.

2. See Volume 4 of this report, The Behavioral Foundations of Neighborhood Change, for a discussion of investment in maintenance by homeowners.

3. Winger [14] studied a sample of urban homeowners and found household income, age, and education to be significantly associated with maintenance expenditures.

Table 8-1

Maintenance Expenditures Reported by Homeowners

Age of Respondent:	20-39	40-64	
	n = 234	n = 350	
	.026	.022	
Education of Respondent:	<HS	HS	>HS
	n = 108	n = 184	n = 290
	.020	.024	.025
Ethnicity of Respondent:	Native	Foreign-born	Minority
	n = 476	n = 29	n = 81
	.024	.018	.021

Reported maintenance expenditures are standardized by dividing expenditure by home value. Values reported are annual averages over the three year period preceding the survey.

Source: Birch et al., The Behavioral Foundations of Neighborhood Change, Section 4.

under-maintain may face social or legal pressure to increase the level of maintenance. If a neighborhood becomes less attractive to households moving into or within the region, homeowners and landlords are presumed to invest less in maintenance, as they become concerned about recouping their investments when selling or renting their housing units. If lenders and insurers reduce or deny access to credit and insurance within a neighborhood, maintenance levels are likely to fall. Unavailability of credit inhibits sale and purchase of housing units. Lack of property insurance makes it difficult for owners to protect the values of their homes. Finally, owners of units in deteriorating neighborhoods are unlikely to maintain their units as well as they might if the units were in neighborhoods with stock in better condition.[1] Owners in such deteriorating areas may expect further deterioration and hence not wish to invest in maintenance which is unlikely to increase the sale value of their units.

Landlords invest in maintenance in their apartment buildings only if they expect to recoup maintenance expenditures from rent payments.[2] This means the financial status of the tenants is a key determinant of investment in maintenance in the apartments they occupy.

---

1. Public Affairs Counseling [7] describes the maintenance behavior likely to occur in sound, deteriorating, and dilapidated neighborhoods.

2. For detailed examinations of landlord investment behavior in inner city areas, see Sternlieb [9] and Stegman [8].

Abandonment

Changes in the condition of a neighborhoods's housing stock over time reflect the balance between deterioration of the housing units and the maintenance efforts of homeowners and landlords. If units receive little maintenance, they deteriorate and may eventually become unfit for habitation.[1] Units which reach this point are typically abandoned. Although some owner-occupied units are abandoned, most abandoned units are rental units which the landlord no longer wishes to maintain. Abandonment in the model is viewed as the logical and likely outcome of various complex processes, including shifts in demand patterns, changes in maintenance behavior, and impacts of property tax, all of which affect housing units long before they are abandoned.[2] Once abandonment begins in a neighborhood more is likely to occur, as the neighborhood becomes less attractive to many households. This is due in part to the host of problems created by abandoned buildings, such as vandalism, arson, and squatting, and to the likely withdrawal of lending and insurance services from the neighborhood.[3] Thus, abandonment can result in a rapid deterioration of neighborhood property values, as well as a collapse in the neighborhood housing market. Land freed by the demolition of abandoned buildings is apt to have little

---

1. For a listing of neighborhood stages which precede abandonment, see Public Affairs Counseling [7].

2. Numerous books and articles have explored the process of abandonment. For some reviews of the research in this area, see U.S. Department of Housing and Urban Development [12] and Linton, Mields, and Coston, Inc. [6].

3. See Vandell et al. [13] for a discussion of how disinvestment by financial institutions can be both a cause and an effect of neighborhood decline.

or no economic value for a long period of time. Eventually neighborhoods with many abandoned structures may become candidates for urban renewal programs.

Simulation of Deterioration, Maintenance, and Abandonment

The simulation of the processes described above requires several steps. A scale to measure housing stock condition is needed so changes in condition can be taken into account. The units in a neighborhood must be allowed to deteriorate over time, and homeowners and landlords must be permitted to invest in maintenance. Finally, units which have deteriorated significantly must be abandoned.

The first problem is to develop a scale which reflects the condition of the housing stock in a neighborhood. Any single number which purports to capture housing condition obviously represents considerable simplification. However, the need for a simple measure of housing condition within a neighborhood stimulated development of such a measure based on Census Bureau data on housing unit condition from the 1960 census of housing. Using these data, we assign to each neighborhood an index which reflects the average housing unit condition in 1960. In 1970, the Census Bureau did not collect data on housing condition, as such data were considered unreliable. We recognize this problem of questionable reliability, but we have no superior alternatives. To overcome the lack of data for 1970, we use the 1960 measures of housing stock condition and adjust them to reflect new construction as well as deterioration of the stock between the 1960 and 1970 censuses.

The next step in the simulation is to adjust the stock condition index to reflect use and weather.[1] The index is also adjusted to take into account new private and public construction.

We then determine the extent of maintenance in each neighborhood, which permits us to revise the value for the housing condition index. Using data from various sources, especially our Survey of Neighborhood Opinions,[2] we have developed an index reflecting the propensities of homeowners of different types to invest in housing unit maintenance. In the case of tenants, the propensities to invest indicate the extent to which the tenants are able to stimulate investment in maintenance by their landlords. Since homeowners and tenants are stratified according to age, ethnicity, and education in the model, the measures of propensity to invest utilize these variables.

Other factors can influence the level of investment by homeowners and landlords as well. As indicated above, individuals are assumed not to invest significantly above or below the general neighborhood level. As the demand patterns for housing units in the neighborhood shift, investment behavior is adjusted. If the neighborhood has been redlined by lenders and insurers, maintenance expenditures are reduced. If the condition of the housing stock is very low, or if abandonment has begun, maintenance expenditures are also reduced.

---

1. This differs from the approach taken by Ingram et al. [5]. In their filtering submodel, changes in the condition of a housing unit occur only if the unit has been vacated. They estimate that only one-quarter of the standing stock undergoes a change in condition each period.

2. See Volume 4 of this report, The Behavioral Foundations of Neighborhood Change, for an explanation of the survey data.

The final step in the simulation is to estimate the number of housing units that will be abandoned in any year. Abandonment is not begun until the stock condition has fallen to a relatively low level. The number of units abandoned depends on the value of the stock condition index. The lower the index, the more units are abandoned. However, the number of units abandoned cannot exceed some limit, approximately 2 percent of the housing stock in the neighborhood, each year. Most of the abandoned units are rental units, and most are in the low rent or price category.

Mathematical Statement

For each neighborhood, the maintenance submodel keeps track of changes in the condition of the housing stock which are caused by deterioration and maintenance. It also estimates abandonment levels in each neighborhood.

As explained above, the index of neighborhood housing stock condition is based on census data on the number of sound, deteriorating, and dilapidated units. The following scale is used to represent housing stock condition at the neighborhood level:

100	Newly constructed units
90-99	Very good condition
70-89	Good condition
50-69	Fair to poor condition (corresponds to census category "deteriorating")
0-49	Poor condition (corresponds to census category "dilapidated")

At the beginning of a model run, index values for stock condition in each neighborhood are created. These values are revised each year to reflect construction of publicly financed housing, private construction, and abandonment during the previous year. Thus,

$$TEMPST = TSTOC - TCON - TOTPH + STABAN \quad (8.1)$$

for all neighborhoods (I), where:

TEMPST	=	housing stock prior to adjustment for new construction and abandonment

TSTOC	=	housing stock after adjustment
TCON	=	private construction during the previous year
TOTPH	=	publicly financed construction during the previous year
STABAN	=	number of units abandoned during the previous year.

Once TEMPST is determined, we can adjust the stock condition index:

$$SUM1 = TEMPST * STCOND \quad (8.2)$$

$$SUM2 = (TCON + TOTPH) * 100 \quad (8.3)$$

$$SUM3 = STABAN * 30. \quad (8.4)$$

$$STCOND = (SUM1 + SUM2 - SUM3) / TSTOC \quad (8.5)$$

for all neighborhoods (I), where:

SUM1	=	contribution to stock condition index attributable to housing stock before adjustment
SUM2	=	contribution to stock condition index attributable to new construction (public and private)
SUM3	=	contribution to stock condition index attributable to abandoned units
STCOND	=	revised stock condition index reflecting new construction and abandonment.

Next we determine the deterioration in the condition of the housing stock due to normal depreciation. The stock condition index, STCOND, is multiplied by (1.0 - DETER); DETER is the annual rate of depreciation in the housing stock.

The condition of the housing stock can be improved through maintenance. Since different household types have different propensities to maintain if if they are owner-occupiers, or different capacities to stimulate maintenance if they are renters, the total effect of maintenance decisions is based on a weighted average of the maintenance efforts of all households in a neighborhood. First we find the expected average maintenance level:

$$TMAINT = \sum_{K=1}^{27} PROPIN(K) * N(K,I) \qquad (8.6)$$

$$AVGMNT = TMAINT / TOTN \qquad (8.7)$$

for all neighborhoods (I), where:

TMAINT	=	weighted sum of maintenance efforts by all households
PROPIN(K)	=	maintenance effort expected from Household Type K
N(K,I)	=	number of households of Type K
AVGMNT	=	average maintenance effort expected from all households
TOTN	=	total number of households.

Since homeowners and landlords do not invest significantly above or below the average level of maintenance in the neighborhoods, no value for PROPIN(K) can lie more than one standard deviation from the average, AVGMNT. A new average maintenance value, TOTINV, is calculated as in Equations 8.6 and 8.7; this new value is based on the revised values of PROPIN(K).

The value of TOTINV is revised for each Neighborhood I to reflect the influences of demand for housing, the pressure for redlining, and the effects of stock condition.

The level of demand for housing in Neighborhood I affects the extent of maintenance activity. High levels of demand stimulate maintenance, while low levels of demand inhibit maintenance activity. That is,

$$DDMULT = 1 + (s1 * TEXD) \qquad (8.8)$$

for all neighborhoods (I), where:

DDMULT	=	multiplier based on the strength of housing demand ($.9 \leq DDMULT \leq 1.1$)
s1	=	a parameter whose value is typically .001
TEXD	=	total excess (i.e., unsatisfied) housing demand.

To reflect demand influences, TOTINV is multiplied by DDMULT for all neighborhoods (I).

The influence of pressure for redlining is expressed in the non-linear function:

$$RDMULT = a + (b-a) * e^{-\frac{(c-REDLIN)}{(a-b)} * s1} \qquad (8.9)$$

for all neighborhoods (I), where:

RDMULT	=	multiplier based on the pressure for redlining ($0 \leq RDMULT \leq 1$)
a	=	1.0 for REDLIN $\leq$ c
a	=	0 for REDLIN $>$ c
b	=	a parameter whose value is typically .50
c	=	a parameter whose value is typically .75
REDLIN	=	an index of pressure for redlining
s1	=	a parameter whose value is typically 3.

To reflect the pressure for redlining, TOTINV is multiplied by REDMLT.

The condition of the housing stock in the neighborhood affects the maintenance expenditures, as shown below:

$$STMULT = a + (1.0-a) * e^{\frac{(c-STCOND)}{(a-1.0)} * sl} \quad (8.10)$$

for all neighborhoods (I), where:

STMULT	=	multiplier based on the stock condition ($.9 \leq STMULT \leq 1.03$)
a	=	.90 for STCOND $\leq$ c
a	=	1.03 for STCOND $>$ c
c	=	a parameter whose value is typically 65
STCOND	=	stock condition index
sl	=	a parameter whose value is typically .005.

To reflect the condition of the housing stock, TOTINV is multiplied by STMULT.

Once the final value of TOTINV is determined, the final value of stock condition can be calculated:

$$STCOND' = STCOND * TOTINV \quad (8.11)$$

for all neighborhoods (I), where:

STCOND'	=	the current index for stock condition, reflecting the influence of deterioration and maintenance .

The extent of abandonment in each neighborhood can then be estimated as follows: For STCOND $\leq$ c,

$$ABNNOT = b + sl\ (STCOND - c). \quad (8.12)$$

For STCOND $>$ c,

$$ABNNOT = a + (b-a) * e^{\frac{(c-STCOND)}{(a-b)} * sl} \quad (8.13)$$

for all neighborhoods (I), where:

ABNNOT	=	the fraction of housing stock not abandoned $(0 \leq \text{ABNNOT} \leq 1)$
a	=	0 for STCOND $\leq$ c
a	=	1.0 for STCOND $>$ c
b	=	a parameter whose value is typically .98
c	=	a parameter whose value is typically 60
sl	=	a parameter whose value is typically .003.

Once ABNNOT is determined, we can calculate the fraction of units abandoned (ABNPCT).

$$\text{ABNPCT} = 1.0 - \text{ABNNOT} \qquad (8.14)$$

for all neighborhoods (I).

The number of units abandoned is constrained as follows:

$$\text{ABNLMT} = \text{PRMINV}(4) * \text{TSTOC}(I) \qquad (8.15)$$

for all neighborhoods (I), where:

ABNLMT	=	maximum number of units which can be abandoned
PRMINV(4)	=	fraction of units which can be abandoned in one year
TSTOC	=	total number of housing units.

The total number of units abandoned (ABANSM) is

$$\text{ABANSM} = \text{ABNPCT} * \text{TSTOC} \qquad (8.16)$$

for all neighborhoods (I).

If

$$\text{ABANSM} > \text{ABNLMT}, \tag{8.17}$$

ABANSM is set equal to ABNLMT. Another constraint on the extent of abandonment is the stock condition. If the stock is in good condition, as indicated by a stock condition index value greater than 60, no abandonment occurs.

The abandoned units are taken first from the low-priced stock in proportion to the tenure balance. For example, if 35 percent of the low-priced stock is owned units, 35 percent of the units abandoned are owned units. If the number of units to be abandoned exceeds the number of low-priced units, units in the medium-priced group are abandoned. However, no more than 20 percent of the medium-priced stock can be abandoned in any single year. Thus

$$\text{ABAN(T,3)} = \text{STOC(T,3)} \ / \ \text{TOTLOW} * \text{ABANSM} \tag{8.18}$$

for all tenures (T), all low-priced units, and all neighborhoods (I), where:

ABAN(T,3)	=	number of abandoned low-priced units of Tenure T
STOC(T,3)	=	number of low-priced housing units of Tenure T
TOTLOW	=	total number of low-priced units.

If

$$\text{ABAN(T,3)} > \text{STOC(T,3)} \tag{8.19}$$

for all tenures (T) and all neighborhoods (I), then some medium-priced units are abandoned. That is,

$$\text{EXCESS(T)} = \text{ABAN(T,3)} - \text{STOC(T,3)} \tag{8.20}$$

for all tenures (T) and all neighborhoods (I), where:

EXCESS(T) = number of abandoned units not in the low-priced group.

The number of abandoned low-priced units is set equal to STOC(T,3) for all tenures (T) and all neighborhoods (I). The number of medium-priced units abandoned is constrained as follows: If

$$\text{EXCESS(T)} > \text{STOC(T,2)} * .2 \tag{8.21}$$

then

$$\text{EXCESS(T)} = \text{STOC(T,2)} * .2 \tag{8.22}$$

for all tenures (T) and all neighborhoods (I), where:

STOC(T,2) = number of medium-priced housing units of Tenure T.

The number of medium-priced units abandoned is:

$$\text{ABAN(T,2)} = \text{EXCESS(T)} \tag{8.23}$$

for all tenures (T) and all neighborhoods (I).

REFERENCES

1. Birch, David L., et al. The Behavioral Foundations of Neighborhood Change. Cambridge: The Joint Center for Urban Studies, 1977.

2. Dildine, Larry L. and Fred A. Massey, "Dynamic Model of Private Incentives to Housing Maintenance." Southern Economic Journal, Vol. 40 (4), April 1974.

3. Eisenstadt, Karen M. Factors Affecting Maintenance and Operating Costs in Private Rental Housing. New York: Rand Institute, Paper No. R-1055-NYC, August 1972.

4. Grigsby, William G. Housing Markets and Public Policy. Philadelphia: University of Pennsylvania Press, 1963.

5. Ingram, Gregory, John F. Kain, and Royce Ginn. The Detroit Prototype of the NBER Simulation Model. New York: Columbia University Press, 1972.

6. Linton, Mields, and Coston, Incorporated. A Study of the Problems of Abandoned Housing. Washington, D.C., November, 1971.

7. Public Affairs Counseling, "The Dynamics of Neighborhood Change." U.S. Department of Housing and Urban Development, Office of Policy Development and Research, May 1975.

8. Stegman, Michael. Housing Investment in the Inner City. Cambridge, Mass.: MIT Press, 1972.

9. Sternlieb, George. The Tenement Landlord. New Brunswick, N.J.: Rutgers, The State University, 1966.

10. Sweeney, James L., "Housing Unit Maintenance and the Mode of Tenure." Journal of Economic Theory, vol. 8(2), June 1974.

11. Sweeney, James L., "Modeling Housing Markets: Commodity Hierarchies in a Spatial Setting." Stanford Journal of International Studies, Vol 9, Spring 1974.

12. United States Department of Housing and Urban Development. Abandoned Housing Research: A Compendium. Washington, D.C.: Government Printing Office, 1973.

13. Vandell, Kerry, Barbara Hodas, and Rachel Bratt. Financial Institutions and Neighborhood Decline. Cambridge, Mass.: Joint Center for Urban Studies, November 1974.

14. Winger, Alan R., "Some Internal Determinants of Upkeep Spending by Urban Homeowners." Land Economics, Vol. 49, November 1973.

Section 9

LAND USE AND ZONING

Land use restrictions and land use changes are both considered in the model. Restrictions on land use reflect economic and political pressures generated by households and business firms. Land use changes result from decisions by builders to construct residences and business structures.

Land Use Restrictions

Zoning codes embody the most significant restrictions on land use within cities. However, zoning restrictions may be more important as a reflection of people's land use preferences than as a factor determining land use patterns.[1] For example, Houston is an unzoned city, yet land use patterns in Houston are similar to those found in other large cities as Siegan [7] has indicated.

Zoning codes generally establish density limits within specified areas by stating what types of buildings may be constructed. For example, the zoning code may prohibit buildings intended for multiple-family occupancy or commercial and manufacturing use. However, shifting residential

---

1. Support for this is found in a report issued by the National Commission on Urban Problems [1]: "... experience has shown that it is always difficult and often impossible to preserve land for the purpose designated on a land-use plan in competition with free market forces." See also Stull [10] and Tarlock [12].

patterns, as well as pressures generated by movement of businesses, result in changing land use patterns. Over time this creates forces which may alter land use restrictions. In many areas the passage of time, along with deterioration of the housing stock, results in down-zoning. As the average income level of households falls, so does the average space consumed per person. These rising residential densities reflect a tolerance for higher densities among certain types of households. Households which tolerate high-density residential living presumably do so because of preferences (e.g., having lived in high-density areas) or because of income constraints.

The model treats zoning as primarily a political process, a view supported by Babcock [2], Mandelker [3], and Neutze [6]. The density preferences or tolerances of the residents in a neighborhood are presumed to have considerable impact on the decisions of the zoning board.[1] When an elected zoning board makes land use decisions which reflect the preferences of neighborhood residents, there is no reason for residents to seek to change the board. However a disparity between residents' preferences and zoning board decisions results in pressure on board members to act differently.[2] Failure of the board to heed such pressure may result in replacement of board members through the political process, as Souter [9] has indicated. New members of the zoning board would presumably be more sympa-

---

1. See National Commission on Urban Problems [5].

2. One common safety value which zoning boards have is the power to grant variances to existing zoning ordinances. See Merrifield [4] for a discussion of justifications and effects of zoning variances.

thetic to the wishes of neighborhood residents.

In order to model the zoning board decisions, we simulate "elections" in which neighborhood residents vote for preserving or increasing the current density level.[1] The problem is to assess the density preferences of the residents. Most households are presumed to prefer the existing density levels, this being one of the reasons for which they selected the neighborhood.

Households which move from a more to a less densely populated neighborhood may be willing to tolerate higher density levels than those that exist. As Smith [8] has indicated, residential space may have low value to such families relative to other consumption items. These families live at higher densities which enables them to spend less on housing. When these households move, they tend to bring their approval, or at least tolerance, of higher densities. Such households support (or do not oppose) decisions by zoning boards to raise the permitted density level.

The effect of commercial activity on household density tolerances and on zoning board decisions must also be considered. Neighborhood residents are assumed to tolerate higher density residential development in areas of commercial activity. Hence, if a significant fraction of the land in use in a neighborhood is in commercial use, zoning boards generally permit high-

---

1. Stull [11], in an econometric study of Boston suburbs in 1960, found that homeowners "in general preferred living in communities where a high proportion of local land was occupied by single-family homes to those where this proportion was low, <u>ceteris paribus</u>. They revealed these tastes by paying a premium for homes in the first type of community. Second, these same homeowners discriminated among different categories of nonsingle-family use. Commercial uses, industrial uses, multiple-family uses, and vacant land were not perceived as equally undesirable. Again, these tastes were revealed by the prices which were paid for homes in communities which contained these uses in different proportions."

density apartment construction, as well as additional commercial construction. Rental housing construction often serves as a buffer or transition zone between commercial and single-family residential areas, as Mandelker [3] has indicated. In general, zoning boards do not permit rental units to be distributed equally throughout a metropolitan region.[1]

In order to simulate density restrictions at the neighborhood level, three pieces of information are required. These are the present density level, the level of commercial activity, and the density preferences of the households in the neighborhood. Most households prefer to maintain the existing residential density level, but some households are willing to have it doubled or quadrupled. As the number of households willing to tolerate higher density living increases, the zoning board is expected to respond by raising the permitted density level.

If at least 10 percent of total neighborhood land in use is in commercial use, then the zoning board is presumed to place no restrictions on residential density.

Land Use Changes

Information on the distribution of land in use and on vacant or unavailable land is recorded in the model under the following categories:

Residential use

Light manufacturing use

Heavy manufacturing use

---

1. The National Commission on Urban Problems [5] points this out and suggests that zoning boards have many reasons for restricting multiple-family housing construction.

Trade and service use

Vacant, easy-to-build

Vacant, hard-to-build

Unavailable (e.g., roads, schools, parks).

As builders, businesspeople, and government officials make decisions which affect land use, the information on land use for each neighborhood is updated.

Construction of single- and multiple-family housing units uses up land which is vacant and usable (i.e., land which has sewage facilities, good drainage, etc.). The amount of land required for a housing unit depends on the location of the unit, whether the unit is in a multiple-family structure, and the price or rent level of the unit. Units constructed in neighborhoods where land is expensive, such as in the central city, typically are built on smaller plots than are similar units constructed elsewhere. As a consequence, neighborhoods with expensive land are apt to have high residential densities, as is the case with high-rise apartments. Housing units designed for single families, built on separate parcels of land, require more land per unit than do units built in multiple-family structures. Land use per unit by miltiple-family structures is fairly consistent across submarkets.

Although demolitions of housing units are not simulated in the model, demolitions are accounted for, as they free land which may or may not become available for use by builders. The extent to which land freed by demolitions becomes available for other uses must be determined empirically for each neighborhood.

In simulating land use changes due to residential construction, we take into account location, tenure, and price level of the units. For each metropolitan region we establish a set of standard lot sizes for each tenure and price level combination.[1] The lot size value is increased or decreased according to the residential density of the neighborhood relative to the average residential density of the metropolitan region.

Changes in industrial and commercial land use are also modeled. The amount of land required for each worker depends on the nature and location of his or her business. Once the increase in number of jobs has been estimated, as explained in Section 10, "Employment Location," the amount of land required for industrial and commercial purposes can be determined. For this we require information on the average land per job by industry. Land used for industry and commerce may not need to be of the same quality as that used for residences.

Simulating changes in industrial and commercial land use involves accounting for increases in the number of jobs, by job type, in each neighborhood.

---

1. Information on standard lot sizes is supplied by each region modeled.

Mathematical Statement

Land Use Restrictions

This section of the model determines the maximum residential density permitted in a neighborhood, based on the density tolerance of the residents. This determination requires information on the density tolerance of each household type, the distribution of household types, and the extent of commercial activity in an area.

Most household types prefer the existing density level, so they "vote" for no change. For these household types,

$$V(K) = 1$$

where:

V(K) = the vote by Household Type K showing the change in density level this household type would tolerate.

Certain household types[1] are willing to tolerate increased density levels. The value of the votes of Household Types 2, 10, 14, and 19 is

$$V(K) = 2,$$

and the value of the votes of Household Types 1, 13, and 22 is

$$V(K) = 4.$$

1. Household types are identified according to the characteristics of their heads, as follows:

		Native			Foreign-born			Minority		
Education:		< HS	HS	> HS	< HS	HS	> HS	< HS	HS	> HS
	20-39	1	2	3	4	5	6	7	8	9
Age	40-64	10	11	12	13	14	15	16	17	18
	65+	19	20	21	22	23	24	25	26	27

The density level permitted by the zoning board will reflect the tolerances of the different household types, weighted by the number of households of a type in the neighborhood:[1]

$$\text{MAXDEN} = \sum_{K=1}^{27} \frac{N(K)}{TOT} * V(K) * \text{RDEN} \qquad (9.1)$$

for all neighborhoods (I), where:

MAXDEN	=	maximum density permitted
N(K)	=	number of households of type K
TOT	=	total number of households
RDEN	=	present residential density (total number of housing units ÷ total number of acres in residential use).

If at least 10 percent of the land in use is devoted to trade and services, residents will tolerate high residential density, as explained above. That is,

$$\text{SUM} = \text{RES} + \text{LTMFG} + \text{HVMFG} + \text{COMM} \qquad (9.2)$$

$$\text{PTS} = \text{COMM} / \text{SUM} \qquad (9.3)$$

for all neighborhoods (I), where:

SUM	=	total land in use[2]
RES	=	land in residential use
LTMFG	=	land in light manufacturing use

---

1. In this and all subsequent equations in Section 9, the subscript I, which refers to the individual neighborhoods in a region, is suppressed to simplify the exposition.

2. All land measurements in the model are in acres.

HVMFG	=	land in heavy manufacturing use
COMM	=	land in trade and services use
PTS	=	fraction of total land in use devoted to trade and service activity.

If, for any neighborhood,

$$PTS > .10 \quad (9.4)$$

then MAXDEN is increased by 5. This increase signifies that density restrictions have been eliminated from the neighborhood.

Land Use Changes

Changes in land use resulting from new residential, commercial, and industrial construction are simulated. Information on current land use patterns at the neighborhood level is provided by the region being modeled for each of the following categories:

residential use

light manufacturing use

heavy manufacturing use

trade and service use

vacant-- easy-to-build

vacant-- hard-to-build

unavailable (e.g., roads, schools, parks).

For each year simulated, the data on land use by category are revised to reflect changes due to new construction and demolitions. Since total acreage in a neighborhood is constant, any increase in one category (e.g. new residential use) must be offset by a decrease in another (e.g., vacant land).

To simulate land use due to new residential construction, we require information on average lot sizes by tenure and price category, average residential density in the region, and residential density in each neighborhood.

Average lot sizes by tenure and price category are provided by the region being modeled.[1] We calculate average residential density within the region outside the model so that we can use data from two periods (1960 and 1970). Average residential density within a neighborhood is found as follows:

$$RDEN = TOTSTC/RES \qquad (9.5)$$

for all neighborhods I, where:

RDEN	=	residential density
TOTSTC	=	total number of housing units
RES	=	land in residential use.

To determine the lot size values applicable to a given Neighborhood I, we compare the residential density in the neighborhood to the residential density for the region. Neighborhoods with density lower than the region average have larger than average lot sizes. Neighborhoods with higher than average density have smaller than average lot sizes. To determine the adjustment to the average lot size for a neighborhood, the following equation is used:

$$ADJLOT = a + (1-a) * e^{-\frac{(SMDEN-RDEN)}{(a-1)} * s1} \qquad (9.6)$$

---

1. Typical lot sizes for high-, medium-, and low-priced single family housing units might be 0.5, 0.3, and 0.2 acres, respectively. A typical unit-size value for units in a multiple-family dwelling might be 0.1 acre per unit.

for all neighborhoods (I), where:

ADJLOT	=	lot size adjustment factor
SMDEN	=	residential density in the region
RDEN	=	residential density in Neighborhood I
a	=	2.0 for RDEN $\leq$ SMDEN
a	=	0.67 for RDEN $>$ SMDEN
s1	=	a parameter, whose value is typically 0.5.

For rental units, the value of ADJLOT cannot exceed 1.25.

Once the appropriate lot size values are known for a neighborhood, the change in land use due to new construction can be estimated as follows:

$$USE = \sum_{T=1}^{2} \sum_{P=1}^{3} LOTSZE(T,P) * ADJLOT * CON(T,P) \tag{9.7}$$

for all neighborhoods (I), where:

USE	=	land required for residential construction
LOTSZE(T,P)	=	standard lot size values for Tenure T and Price Group P
ADJLOT	=	adjustment to lot size values to reflect residential density
CON(T,P)	=	construction of new units of Tenure T and Price Group P.

The land use accounting is adjusted as shown below:

$$RES_t = RES_{t-1} + USE \tag{9.8}$$

$$EASYVAC_t = EASYVAC_{t-1} - USE \tag{9.9}$$

for all neighborhoods (I), where:

RES	=	land in residential use in the present year (t) or the previous year (t-1)

EASYVAC = vacant land suitable for residential construction in the present year (t) or the previous year (t-1).

If the amount of land required for new residential construction exceeds the amount of vacant easy-to-build land within any neighborhood, new construction in each submarket is reduced by an equal fraction so that the amount of land required equals the vacant land available.

The amount of land freed by demolitions that may become available for other uses depends on the number of units demolished in a neighborhood and the lot size of each of these units. That is,

$$\text{DEMUSE} = \sum_{T=1}^{2} \sum_{P=1}^{3} \text{LOTSZE(T,P)} * \text{ADJLOT} * \text{AOGDEMO(T,P)} * \text{DEMLND} \tag{9.10}$$

for all neighborhoods (I), where:

DEMUSE = acres of land freed by demolitions which become available for use

AOGDEMO(T,P) = total demolitions of units of Tenure T and Price Group P

DEMLND = fraction of land freed by demolitions which becomes available for further use (empirically determined).

In general, the quality of the demolition data from the regions under study is low. If the number of reported demolitions in a given tenure and price group for a neighborhood is less than the simulated number of units abandoned, the value for demolitions in Equation 9.10 is replaced by the value for abandoned units.

Once DEMUSE is determined in Equation 9.10, the land use accounting is adjusted. Land in residential use is decreased, and vacant land suitable for construction is increased, as below:

$$RES_t = RES_{t-1} - DEMUSE(I). \quad (9.11)$$

$$EASYVAC(I)_t = EASYVAC(I)_{t-1} + DEMUSE(I). \quad (9.12)$$

The final adjustments in land use reflect changes in industrial and commercial activity. The first step is to calculate the average land needed for an industrial or commercial job for one worker. The average land used to support one industrial worker is

$$APJ = (LTMFG + HVMFG) / RJIND \quad (9.13)$$

for all neighborhoods (I), where:

APJ	=	average land used (in acres) per job
LTMFG	=	land used in light manufacturing activity
HVMFG	=	land used in heavy manufacturing activity
RJIND	=	total number of workers employed in light and heavy manufacturing.

However, if the number of manufacturing workers in a neighborhood is below 250, Equation 9.13 is not used. Instead, the average land used per manufacturing worker for the entire region (AVEDEL) is used. At the neighborhood level the value of APJ is constrained as follows:

$$AVEDEL * .05 \leq APJ \leq AVEDEL * 10.0 \quad (9.14)$$

This eliminates extreme values for the average land use per job in a given neighborhood.

Once the average land used per job is known, we can estimate the number of acres required for increases in heavy and light manufacturing activity. We assume that the proportions of land devoted to light and heavy manufacturing activity which prevailed in the past will remain unchanged in the future in each neighborhood. These proportions are calculated as follows:

$$TOTMFG = LTMFG + HVMFG \quad (9.15)$$

$$SHAREL = LTMFG / TOTMFG \quad (9.16)$$

$$SHAREH = HVMFG / TOTMFG \quad (9.17)$$

for all neighborhoods (I), where:

TOTMFG	=	total land used for light and heavy manufacturing activity
SHAREL	=	fraction of total devoted to light manufacturing activity
SHAREH	=	fraction of total devoted to heavy manufacturing activity.

The total change in number of manufacturing jobs is split between heavy and light manufacturing according to SHAREL and SHAREH.[1] The concomitant changes in land use are as follows:

$$ACREH = \sum_{L=1}^{6} SHAREH * APJ * EMPDEL(L) \quad (9.18)$$

$$ACREL = \sum_{L=1}^{6} SHAREL * APJ * EMPDEL(L) \quad (9.19)$$

for all neighborhoods (I), where:

ACREH	=	total acres required for new heavy manufacturing jobs
EMPDEL(L)	=	total number of new jobs in Industry L
ACREL	=	total acres required for new light manufacturing jobs.

1. There are ten different job categories. Categories 1 through 6 are industrial, and Categories 7 through 10 are commercial.

The total number of acres required for new jobs in commerce is derived similarly. Average land use per commercial job is

$$APJ = COMM/ RJCOM \tag{9.20}$$

for all neighborhoods (I), where:

COMM	=	land used for commerce
RJCOM	=	total number of workers employed in commerce.

If the number of commercial workers in a neighborhood is below 250, Equation 9.20 is not used. Instead, the average land use per commercial job for the entire region is used. At the neighborhood level the value of APJ is constrained as in Equation 9.14, using average land use per commercial job for the entire region, to eliminate extreme values of average land use per job in a given neighborhood.

Once the average land use per commercial job is known, the number of acres needed for new commercial activity (ACRETS) can be calculated as follows:

$$ACRETS = \sum_{L=7}^{10} APJ * EMPDEL(L) \tag{9.21}$$

for all neighborhoods (I).

The total number of acres required for new manufacturing and new commercial jobs (TACRES) is thus

$$TACRES = ACRETS + ACREH + ACREL \tag{9.22}$$

for all neighborhoods (I).

The acreage required for these new jobs is taken from available vacant land in both the easy-to-build and hard-to-build categories according to the relative availability of such land. That is,

$$TOTVAC = VACEASY + VACHARD \qquad (9.23)$$

for all neighborhoods (I), where:

TOTVAC	=	total vacant land available
VACEASY	=	easy-to-build vacant land available
VACHARD	=	hard-to-build vacant land available.

A variable is designated for each type of vacant land showing the portion of the whole it represents:

$$SETB = VACEASY / TOTVAC \qquad (9.24)$$

$$SHTB = VACHARD / TOTVAC \qquad (9.25)$$

for all neighborhoods (I), where

SETB	=	fraction of vacant land which is easy-to-build
SHTB	=	fraction of vacant land which is hard-to-build.

Land for new jobs is taken first from easy-to-build vacant acreage (ETB), as follows:

$$ETB = SETB * TACRES \qquad (9.26)$$

for all neighborhoods (I).

If there is not enough easy-to-build vacant land to meet the demand specified in Equation 9.26, the shortfall (SHFALL) is

$$SHFALL = ETB - VACEASY. \qquad (9.27)$$

Land for the remaining new jobs is taken from hard-to-build acerage (HTB) as follows:

$$HTB = SHFALL + SHTB * TACRES \tag{9.28}$$

for all neighborhoods (I).

If there is insufficient hard-to-build vacant land to accommodate the new jobs, the estimates of land needed for manufacturing and commercial jobs are reduced by an equal fraction so that only vacant hard-to-build land is used.

The final step is to adjust the land use data to reflect the land consumed by changes in manufacturing and commercial activity.

$$LTMFG' = LTMFG + ACREL \tag{9.29}$$

$$HVMFG' = HVMFG + ACREH \tag{9.30}$$

$$COMM' = COMM + ACRETS \tag{9.31}$$

$$VACEASY' = VACEASY - ETB \tag{9.32}$$

$$VACHARD' = VACHARD - HTB \tag{9.33}$$

for all neighborhoods (I), where:

LTMFG'	=	revised land use for light manufacturing activity
HVMFG'	=	revised land use for heavy manufacturing activity
COMM'	=	revised land use for commerce
VACEASY'	=	revised easy-to-build vacant land available
VACHARD'	=	revised hard-to-build vacant land available.

REFERENCES

1. American Society of Planning Officials, Problems of Zoning and Land-Use Regulation. Washington, D.C.: National Commission on Urban Problems, Research Report No. 2, 1968.

2. Babcock, Richard F. The Zoning Game: Municipal Practices and Policies. Madison, Wis.: University of Wisconsin Press, 1966.

3. Mandelker, Daniel R. The Zoning Dilemma. New York: Bobbs-Merrill Co., Inc., 1971.

4. Merrifield, Lewis B. "The General Welfare, Welfare Economics and Zoning Variances," Southern California Law Review Vol. 38, Summer 1965.

5. National Commission on Urban Problems. Building the American City. Report of the National Commission on Urban Problems to the Congress and to the President of the United States. House Document No. 91-34, Washington, D.C.: Government Printing Office, 1968.

6. Neutze, Max. The Suburban Apartment Boom. Washington, D.C.: Resources for the Future, Inc., 1968.

7. Siegan, Bernard H. Land Use Without Zoning. Lexington, Mass.: D.C. Heath and Co., 1972.

8. Smith, Wallace F. The Low-Rise Speculative Apartment. Research Report No. 25. Berkeley, Cal.: Center for Real Estate and Urban Economics, University of California, Berkeley,1964.

9. Souter, Don V. "How a Board of Zoning Appeals Functions," Michigan State Bar Journal Vol. 40, No.5, May 1961.

10. Stull, William J. "Land Use and Zoning in an Urban Community," American Economic Review Vol. 64, June 1974.

11. Stull, William J. "Community Environment, Zoning, and the Market Value of Single-Family Homes," Journal of Law and Economics Vol. 18, No. 2, October 1975.

12. Tarlock, A. Dan. "Toward a Revised Theory of Zoning," in American Society of Planning Officials, Land Use Controls Annual-1972.

Section 10

EMPLOYMENT LOCATION

The location decisions of employers[1] are major determinants of the structure and growth potential of neighborhoods in the region. Decisions by employers to expand or contract affect employment opportunities within the metropolitan area, affecting the structure of the regional labor market and levels of regional unemployment. For the purposes of this study, the decisions of firms have been grouped in the following way:

1. location choices of new firms
2. expansion decisions
3. decisions to move into a neighborhood
4. decisions to close operations
5. contraction decisions
6. decisions to move out of a neighborhood.

The first three can be categorized as decisions that will have a positive employment impact on neighborhoods, while the latter three have a negative impact. By defining decisions in this way, we will examine not only new employment change within the metropolis, but also the components of gross employment change within metropolitan neighborhoods.

Three basic structural issues must be dealt with in any theory of employment location: (1) the classification of firms, (2) the mapping of jobs

---

1. The discussion in this section often uses the terms business or firm in place of the term employer. This section is in fact concerned with employment, be it in a public or private enterprise. In particular, government jobs are considered in our analysis, although not as completely as private establishments because of the limitations of our chief source of data.

to firms and vice versa, and (3) the level of geographic aggregation used for analysis. Once we have resolved these three issues, the exposition of theory will outline the decision processes related to the six components of business demographic change and will describe the links between job location theory and the other processes of neighborhood change with which we are concerned in this report.

The principal function of the submodels described in this section is to estimate changes in the number of jobs by industry type and neighborhood. Through the mechanisms of the unemployment rates, which are determined in the submodel described in this chapter, the labor force participation rates, and the access of each neighborhood in the metropolitan area to every other neighborhood, the location of jobs is linked to household location and the rest of the posited theory of neighborhood change.

Employment location data sources which were frequently used in past studies contained only aggregate information and restricted the investigator to the anlaysis of net employment change. The use of net employment change as the dependent variable of analysis obscures a large amount of locational activity. In addition, changes in the number of jobs provided by an establishment are important components of net change and may occur for reasons other than locational decisions. Struyk and James [7] point out that net employment change reveals nothing about the actual decisions that are being made and can, therefore be of little help in informing policy makers on the impact of their decisions. Finally, it cannot be assumed that the processes that are components of net employment gain and net employment loss are mirror images of each other. It is necessary to examine the components of change in order to determine whether or not employment in communities grows in the same way it declines.

The primary data source used in this study is a file maintained by the Dun and Bradstreet Corporation, called the Dun's Market Identifiers (DMI) file, used in the company's credit rating function.[1] The DMI file contains a substantial amount of information on each establishment listed, including the number of employees, the net worth, the street and mailing addresses, and the relevant Standard Industrial Classification (SIC) codes. Each establishment is assigned a unique code (the DUNS identifier) which remains the same as long as the firm remains in the file and does not change its legal status. By comparing the DMI file with other sources of data, we have been able to determine the gaps in coverage in the DMI file. On the whole, the Dun and Bradstreet files are quite complete, but must be regarded as an extensive sample, rather than as a census, of firms. The following types of establishments tend to be underrepresented in the files:

1. newly formed firms
2. government agencies
3. service industry establishments
4. small firms.

We have obtained the files for the years 1969, 1972, and 1975. By using the DUNS identifier, we can distinguish the six components of employment change and therefore, bypass the conceptual difficulties of dealing with net employment change. The six components of employment "demography" are determined from the file as follows:

Birth	The creation of a new firm which did not exist (anywhere in the U.S.) at the beginning of the period.

1. For a detailed description of how the Dun and Bradstreet files have been abstracted and reduced into a compact research data base, see Allaman [1].

Death	The demise of a firm that existed in a neighborhood at the beginning of the period, but was not to be found (anywhere in the U.S.) at the end of the period.
Expansion	An increase in the number of employees of a firm that does not move outside the neighborhood being studied.
Contraction	A decrease in the number of employees of a firm that does not move outside of the neighborhood being studied.
Inmigration	The arrival in the neighborhood of a firm that was previously outside it.
Outmigration	The departure from the neighborhood of a firm that was previously located there.

In our data reduction, a firm that moves and changes employment size (expands or contracts) is counted as a mover. The neighborhood from which the establishment has moved will lose the number of jobs the firm possessed at the beginning of the period and the receiving neighborhood will gain the jobs the firms possesses at the end of the period.

The Classification of Firms

The classification of firms lies at the heart of any theory of intra-metropolitan employment location. The problem in classifying firms is to maximize variation in location behavior between groups of firms with a minimum of variation within groups. In addition, the number of groups created must be small in order to simplify the analysis. A proliferation of firm classification categories would inordinately complicate any explanation of firm behavior. Finally, any classification of firms should be intuitively plausible, as well as analytically sound. The dimension (or dimensions) along which firms are classified is usually a firm-specific variable (such as size of firm, age of firm, or the product or service produced by it.

After several experiments, described in Volume 4, Section 5, of this report we decided to compare the patterns of location of the major industry classifications described in Table 10-1. The industries were divided among three categories, plant, office, and store, based more or less on the types of physical plant and labor force an establishment might be expected to have. Industries are shown according to the Plant-Office-Store (POS) classification in Table 10-2. It was assumed that enterprises needing similar physical plants and similar work forces search roughly in the same places for sites and within broad limits are distributed similarly across metropolitan areas.

The POS categorization is simple and has intuitive, as well as empirical, merit. The classification was created by charting the distribution of the ten industrial categories over a metropolitan area and noting which categories were similarly distributed (see Volume 4, Section 5). As the ten categories fell into place, it was possible to discern other patterns in physical plant, workforce, and the type of market. The plant category consists of industry groups that can be expected to demand industrial, loft, and warehouse space. Railyards, docks, and truck trans-shipment areas are closely associated with warehousing districts where wholesalers might be expected to operate. In addition, heavy industrial users tend to be located near transportation and warehousing facilities. It cannot be demonstrated that all manufacturing industries share locational propensities; in fact, the opposite is often the case. Nevertheless, the constraints on the number of categories made it necessary to lump all manu-

Table 10-1

Industry Categories derived from Standard Industrial Classification

Category	Industry	Two-digit SIC Code (1967)
1	Non-durable manufacturing; other	01-14,20-23,26-31,99
2	Durable manufacturing	19,24,25,32-39
3	Construction	15-17
4	Transportation	40-47
5	Communications and utilities	48,49
6	Wholesale trade	50
7	Retail trade	53-59
8	Finance, insurance, and real estate	60-67
9	Services	70-89
10	Government	91-94

Table 10-2

Plant-Office-Store (POS) Classification

PLANT

1. Nondurable manufacturing
2. Durable manufacturing
3. Construction
4. Transportation
6. Wholesale trade

OFFICE

5. Communications and utilities
8. Finance, insurance, and real estate

STORE

7. Retail trade
9. Services
10. Government

facturing industries into the same class.

The decision to include wholesale trade in the plant class, instead of in the store class with retail trade, is an innovative one. Wholesalers employ more white-collar workers than does the typical manufacturing concern. However, because of the necessity of storing goods, the physical plant needs of wholesalers are more nearly akin to plants (and possibly offices) than to stores. The construction industry was difficult to place, because of the unstable and footloose nature of such work. Construction companies more than likely do not maintain their offices in prime downtown or suburban office space, but rather may locate in areas where they can store their equipment. In terms of the number of workers, the construction industry is small; therefore, its correct placement in a POS category should not be crucial.

The industries that are included in the office category have highly distinctive work forces and locational patterns. Finance, insurance, and real estate, and communications and utilities form a natural grouping since both employ very large proportions of clerical workers and both industries are strikingly concentrated in the CBD. Across all the metropolitan areas studied in this project, the office category includes many fewer employees than the plant or store subdivisions. However, the two industries are so exceptional that they must form a category apart. The office category occupies much of the prime office space in CBD, along suburban commercial strips and in suburban office parks. This category is the most homogeneous of the three when the work force and building types of the constituent industries are considered.

The third category, Stores, is quite heterogeneous, with industries heavily concentrated in office and retail commercial floor space. Retail trade, services, and government were included here because much employment in these industries is population-oriented. Retailers seek locations that are accessible to potential customers. The same is true for personal or consumer services. Business services most likely belong in the office class, but were placed in stores in order to maintain the integrity of SIC categories at the one-digit level. With the exception of facilities designed to serve the entire region, employment in government services such as in schools, post offices, and social service agencies is distributed according to the population. In addition, many of the government's population-oriented services are located in storefronts. The work forces of these industries include relatively large proportions of managerial and professional personnel in comparison to plants or offices, as well as the expected numbers of clerical and sales workers.

The number of categories that can be managed in our model is small, restricting our flexibility in recombining major SIC categories. The POS division contains innovative aspects and will aid us later in interpreting the model and theory. It is important to remember when reading the balance of this section that each job classified in plant, office, or store may not, in fact, be located in a plant, office, or store, respectively. Rather, the descriptive types are meant to be a general indication of the kinds of jobs in a given area.

Location of Jobs or Location of Firms

The second basic issue in developing a theory of employment location concerns the difference between modeling the location of jobs and the location of firms. Economic activity is clearly not spread evenly over a geographic area, but is lumped into discrete units called firms. While a large number of firms have a small number of employees, a large proportion of the jobs in a region are in fact situated in a relatively small number of large facilities.[1] It is certainly true that in building a behavioral model of decision-making, we should consider firms as discrete units. However, it is the number of jobs in a geographic area, not the number of firms, which is required in the structure of our model. Putman articulates the consequences of modeling the location of jobs:

> ...some continuous function allocation models [i.e., models which consider jobs] are capable of producing reasonably good forecasts of non-market-sensitive employment location in some circumstances. However, they are unable to deal with discrete facility location. This inability is likely to be a significant problem in dealing with non-market-sensitive employment types. When these models attempt to forecast significant changes in the location of non-market-sensitive industry types, they are open to this criticism. Thus, although these models are capable of producing reasonable forecasts for market-sensitive employment types,... their capability to deal with non-market-sensitive employment is restricted to situations where new or relocating facility location is unlikely to take place and where growth or decline (or both) are more or less regularly distributed.[2]

---

1. In 1959, Birch et al. [4] found for manufacturing firms in the New Haven SMSA that there were 396 plants with 1 to 9 employees, 380 with 10 to 249 employees, and 34 with over 250 employees.

2. See Putman [6], pp. 219-220.

Putman's criticisms are most applicable to the multiple regression models that have typically sought to predict the location of employment. While our theory shares features with the criticized continuous function allocations, we acknowledge that it may be impossible to predict large facility changes in location or changes in employment because the causes are so idiosyncratic that these changes must be considered on a case-by-case basis. Even if a change in employment in a large establishment can be predicted by the theory, the chances for error are large because the averaging and leveling effects of the law of large numbers will not obtain.

In trying to account for the locational propensities of firms, one must attempt to predict the behavior of a small number of large firms. In addition, it is necessary to specify the employment size distribution of each type of firm for each metropolitan area under consideration. If the level of geographic aggregation is small enough, the relocation of a large firm can mean large discontinuities in employment change. In view of the fact that we will be predicting changes in employment rather than changes in numbers of firms, the level of geographic aggregation is quite important and will be discussed below. Our decision to model employment may obfuscate the "lumpiness" of business location decisions and in fact implies that the appeal of an area to a business is proportional to the number of jobs in that area.

The most important reasons for modeling the location of jobs rather than the location of firms are, first, the nature of the links between employment change and our theory of neighborhood evolution and, second, the difficulty of mapping back and forth between firms and jobs. It is the relationship between the labor force and the number of available jobs that links employment location to a general theory of neighborhood change. Communities often seek to be the locus of large firms in order that the wages provided to local workers will, through the multiplier effect, expand the local economy far in excess of the firm's payroll. In addition, the existence of surplus jobs will attract labor from outside the region. Firms are sought for declining areas, not in order that that neighborhood might be the locus of production, but for the jobs the firm will make available to the area's unemployed residents. Localities endure a wide range of negative externalities from industries in order to preserve jobs for their residents. Our theory of labor supply is explicated below.

Level of Geographic Aggregation

Our general system of neighborhood characterization and boundary delineation is discussed in Section 3. If the geographic unit is highly disaggregated, trends in behavior are not apparent and problems of discontunuity are encountered when large firms relocate. If, for example, a large manufacturing firm relocates in an area with few manufacturing jobs, the percentage increase is huge. On the other hand, high levels of aggregation on the basis of spatial proximity may hide significant differences in even a considerable amount of locational activity over the region.

Commonly used geographic areas such as census tracts or zip code areas often use natural features (e.g., rivers) or man-made barriers (main streets, railroad tracks) as boundaries. In most cases such barriers do, indeed, define neighborhoods and for most uses, such boundaries are useful and efficient. For the purpose of charting employment location change, however, the use of main streets as boundaries can be confounding. Although a strip highway or a town's main commercial street does divide two neighborhoods, the businesses located on that street (regardless of which side they are on) form an integral business district. In addition, if a large firm is located on one side of a boundary street, the neigborhoods on either side of the boundary will reflect disproportionate employment totals and change. It is important to recognize these difficulties, although it remains necessary to use accepted units of geographical aggregation in order to ensure comparability with other sources of data.

In order to create a theoretically and empirically sound geographical basis for the model, we have used two levels of aggregation, neighborhoods and regional job areas (RJA's), which are groups of zip code areas[1] which consist, in turn, of neighborhoods. The RJA represents a non-contiguous area; a zip code is assigned to an RJA on the basis of its industry mix. We have postulated eight different types of RJA's as listed in Table 10-3. The jobs in each POS category are totaled over the entire region and the proportion of each POS class across the region is found for each zip code area. Cutoff proportions are determined for each metropolitan area and zip code areas are described as having a negligible, low (though significant), or high number of jobs as compared with the rest of the region in each POS class. The central business district (CBD) is usually designated as one zip code area which is always assigned to RJA 1. In certain cities, additional zip code areas adjacent to the CBD may also be assigned to RJA 1.

---

1. The Dun and Bradstreet data are most easily analyzed using zip code areas as geographical units.

Table 10-3

Regional Job Areas (RJA's)

RJA Number and Name	Zip Code Area Job Proportion Plant	 Office	 Store
1. Central Business District	L	H	H
2. Central Industrial District	H	H	H
3. Plants Only	H	[O / L]	[L / O]
4. Plants Plus	H	H	[L / O]
	H	[L / O]	H
5. Other Only	[L / O]	H	[H / L / O]
	[L / O]	[H / L / O]	H
6. Moderate Plants	L	[L / O]	[L / O]
7. Moderate Other	O	L	O
	O	O	L
	O	O	L
8. Residential	O	O	O

1. H O L, for example, means a high proportion of plants, a negligible proportion of offices, and a low, but significant, proportion of stores. Brackets [ ] indicate "or." For example, H H [L / O]. means that zip code areas with POS proportions H H L or H H O should be included in RJA 4.

Volume 4 explains the RJA's in greater detail

In certain cities, additional zip code areas adjacent to the CBD may also be assigned to RJA 1. This is the only RJA that must be contiguous because of the definition of the central business district as the center of the region.

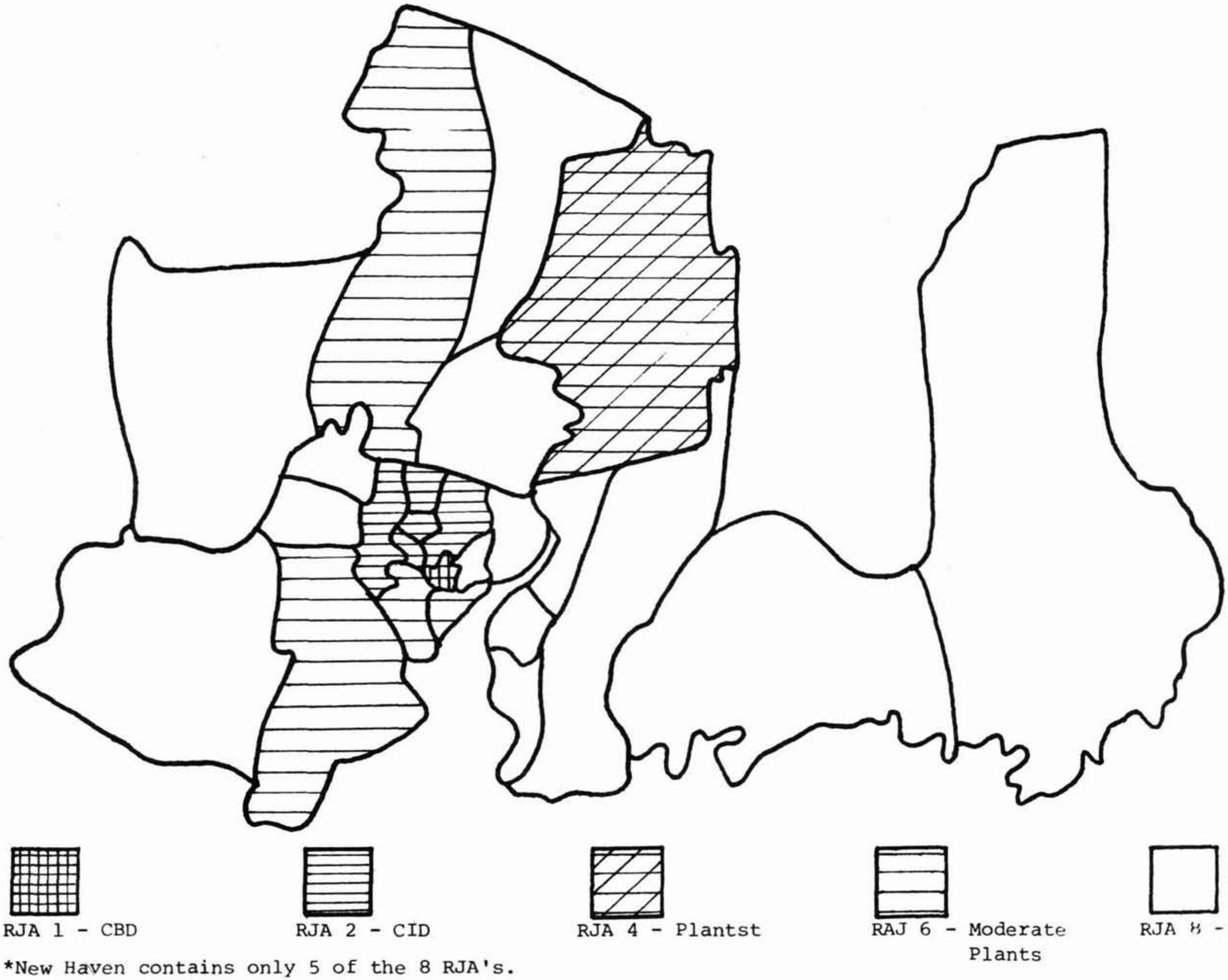
Zip code areas are assigned to RJA's using the industrial mix during an arbitrarily chosen initial year. The zip code areas and constituent neighborhoods will remain in that RJA for the time period being modeled, even though the changing distribution of employment has, according to the assignment rules, switched that zip code area to another RJA by the end of the period. The permanent assignment of a zip code area to an RJA, regardless of changes in employment location, results from the assumption that the locational decision makers' perceptions of the character of a neighborhood will lag somewhat behind the actual condition of the area and that these perceptions are relatively stable over the short forecasting period of our model of neighborhood evolution.

RJA's have been defined for the six metropolitan areas under consideration in this report. Figure 10-1 is our sample of the New Haven region.

Neighborhoods are not distributed uniformly across the regional job areas. RJA 8, the residential regional job area exhibiting negligible amounts of employment compared to the rest of the metropolitan area, usually contains more than half of the neighborhoods in the region. This is reasonable, as industrial and commercial land use tend to be concentrated and separated from residential land use. In addition, the neighborhoods in RJA 8 are overwhelmingly suburban; conversely, most suburban neighborhoods are in RJA 8. Both Rochester and Worcester lack a well-defined inner city industrial-commercial district (RJA 2) in which there are high concentra-

Figure 10-1

Regional Job Areas - New Haven*



*New Haven contains only 5 of the 8 RJA's.

tions of all three types of jobs. Larger metropolitan areas appear to have neighborhoods represented in all RJA's, while smaller areas lack the full range of regional job areas.

Descriptive typologies and planning tools may occasionally use such discontinuous areas, but rarely has a non-contiguous geographical basis been used for analytical purposes. Generally, investigators have used contiguous, compact geographical areas which have been assembled so that they will have equal populations, equal areas, or be homogeneous along some neighborhood attribute. We think that in making locational decisions, businesspeople exclude many neighborhoods from consideration at the outset because the amount and mix of enterprises in those RJA's indicate that the neighborhood is not likely to offer suitable buildings, environment, public services or infrastructure.

Two levels of aggregation, the RJA's and the neighborhoods, are used in order to create theoretically and empirically sound geographical divisions. First, suitable territory in the metropolitan area (the specific RJA) is chosen because of a favorable industry mix, and second, a specific neighborhood with desired characteristics is chosen. This two-stage process will be described more fully below.

Location of Jobs Over Time

As outlined above, the net employment change for a given geographical area consists of the gross employment changes produced by the six demographic processes -- births, expansions, in-moves, deaths, contractions, and out-moves. The first three processes sum to the gains and the last three to the losses in employment for a given geographical area. How do the causes of these six different processes differ from each other and how does our theory take these differences into account?

The birth of a new firm and the movement of an established firm to a new site are obviously locational decisions for which the businessperson considers the characteristics of alternative neighborhoods. In addition, a firm which has expanded considerably may make an explicit decision to remain at a site rather than move to accommodate the expansion. In general, however, changes in the size of the firm and its death have little to do with the characteristics of the neighborhood in which the change has taken place. Such changes, for example, depend upon the state of a particular industry in the region or the nation and the high mortality rate of new businesses.

Table 10-4, which summarizes the relative importance of the components of employment change for two of the metropolitan areas studied, shows that deaths and contractions combine to dominate the losses aggregated across each region. Similarly, births and expansions dominate employment gains. In other words, movement of firms is a relatively unimportant cause of neighborhood change.[1]

Based on our findings to date,[2] we assume that all areas within a

---

1. Struyk and James [7] found that expansions and contractions of manufacturing firms remaining in one location accounted for most of the net change within zip code areas in a three-year period. They also discovered that the natural increase (the difference between births and deaths) of manufacturing establishments was more significant to the decentralization of manufacturing jobs than was net migration.

2. See Volume 4, Section 5 of this report for a fuller explanation.

Table 10-4

Components of Employment Change as a Percentage of Losses/Gains, 1969 through 1972

CHARLOTTE

	Gains						
	Plant	Office	Store		Plant	Office	Store
Births	.24	.46	.44	Deaths	.51	.30	.60
Expanders	.56	.32	.43	Contractors	.34	.28	.25
Inmovers	.19	.22	.13	Outmovers	.15	.42	.15
	1.00	1.00	1.00		1.00	1.00	1.00
Births + Expanders	.81	.79	.87	Deaths + Contractors	.85	.58	.85
Total Gains	24876	1133	9314	Total Losses	23733	681	7723

NEW HAVEN

	Plant	Office	Store		Plant	Office	Store
Births	.19	.18	.35	Deaths	.42	.07	.58
Expanders	.55	.19	.46	Contractors	.46	.83	.25
Inmovers	.26	.63	.19	Outmovers	.12	.10	.17
	1.00	1.00	1.00		1.00	1.00	1.00
Births + Expanders	.74	.37	.82	Deaths + Contractors	.88	.92	.83
Total Gains	6748	609	4995	Total Losses	14122	2225	5402

metropolitan region are losing jobs at approximately the same rate (depending on the industry mix). That is, we posit a constant loss rate by POS category for each metropolitan area. These findings on loss rates are consistent with other studies. For example, Struyk and James [7] report that geographic location had no effect on the death rate of businesses in the metropolitan areas studied. The Inter-Area Migration Project, a companion study to this one, investigated nationwide inter-regional employment location and found that "changes in employment ... take place through a relatively constant loss rate (averaging about 10 percent of all jobs in each area in each year) and a differential rate of gain."[1]

As mentioned above, we assume a two-stage decision-making process (first, choice of RJA and second, choice of neighborhood). The two-stage process complements the two levels of geographic aggregation. A businessperson accumulates general information about relatively large areas of the region (RJA's) and selects the neighborhoods in an RJA which have been chosen by similar establishments in the past. For example, financial institutions have tended to locate in the CBD in the past and we assume that this will have a strong effect on future behavior in the short term. The DMI files were used to determine the probabilities that the gains for each POS category would be made in a given RJA. In other words, the gains by POS classes in each RJA are distributed in the same proportion in which such gains have been apportioned in the past. The firm's choice of RJA is based purely on inertia or past business locaton decisions.

---

1. David Birch, et al. [2], p.4.

At the level of neighborhood choice, there are two different kinds of characteristics which affect choice of neighborhood within an RJA. The first are inertial characteristics, those which summarize past locational decisions. Any other characteristic is of the second type, called attractiveness variables. These describe the neighborhood in terms of attributes like the tenure and value of the housing stock.

It seems reasonable to assume that inertia exerts a powerful, stabilizing force on the system of employment location since the stock of physical plants changes very slowly. Further, the past choices of similar firms undoubtedly affect business locational decisions. There are four neighborhood properties which sum up the past locational practices of businesses in the neighborhoods. First, there is floor space made available by firms which have undergone employment losses. Neighborhoods with large employment bases lose many jobs because of constant loss rates, and therefore have more slots available to be filled by replacement employment. Because of the ease of acquiring recently vacated space, we believe that neighborhoods with many slots will attract new jobs.

The second through fourth inertial factors, one for each POS category, indicate the industry mix in a given neighborhood. The proportion of jobs in each neighborhood POS category to total employment in the neighborhood shows the extent to which neighborhoods which in the past have been attractive to offices plants, or stores will continue to be so. Firms tend to be attracted to areas with high concentrations of similar firms. For example, we have found that firms in the office category are overwhelmingly attracted to neighborhoods with high office job concentrations. Another finding, not quite as significant, is that firms in one category may not locate in areas with heavy concentrations of firms from other categories.

This is true of plants and stores.

The second type of characteristic affecting choice of neighborhood is the attractiveness of neighborhoods to business decision-makers. This attractiveness may be in terms of the site's suitability for the firm or its proximity to desirable residential areas. Investigators of business location have suggested many factors which businesspeople consider in determining suitability or attractiveness of an area. It is common for investigators to consider industry locational decisions as either market-sensitive (mostly trade) or non-market-sensitive (mostly manufacturing) or to consider the location factors for each separately, if both categories are included in the study. Hoover and Vernon [5] mention factors such as the availability of space suitable for modern manufacturing techniques; ease of access to transportation, including the disadvantages of congestion in the central city; and the supply of the right kind of labor force, as being important in the locational choices of manufacturers.

This study also relates that smaller plants need the external economies of the central city (this is known as the incubator theory). Struyk and James [7] have found the incubator theory to be incorrect, although they note the importance of external economies, as indicated by the geographic concentration of industry, to manufacturing location. Putman [6] mentions such factors as access to inputs, tax expense, land cost, the availability of an appropriate labor force, the availability of suitable space, and access to markets, all as significant considerations in the locational choice process.

The attractiveness variables which we used in developing our final model include characteristics which indicate the social organization of a neighborhood and those which measure the availability of land for economic development. Social organization includes the status of a neighborhood and the dominant tenure (owned or rented) of the housing stock. Preliminary observation of urban land use patterns suggests that commercial and office development often coincides spatially with apartment dwellings. We also thought that commercial employers in an effort to serve an affluent clientele, and office firms to be in a pleasant environment, would seek those neighborhoods with high ratios of high value to low value housing.

Our measure of available land for economic development is based not only on the amount of vacant land in a neighborhood, but also on the accessibility of the neighborhood to those sections of the metropolitan region with a large amount of economic activity. All other things held equal, one might expect that a neighborhood with a substantial amount of vacant, easy-to-build-upon land which is adjacent to a highly developed area would attract more jobs than one which was relatively remote from such development. In addition, a neighborhood with a good road might be expected to attract more employment than would one with poorer roads. A measure called "attractive available land" has been created to combine the effects of the amount of land available for building and the amount of economic activity that is accessible to the neighborhood.

Finally, the percentage of commercial and industrial land in a neighborhood indicates how receptive a community has been to non-residential land use and whether the community would have the infrastructure necessary to support additional jobs.

Mathematical Statement

The number of jobs in each neighborhood and major job category (as listed in Table 10-1) is determined in five steps successively for the region, for regional job areas (RJA), and for the neighborhood. First, the number of jobs in the region by each major job class is determined. The multiplicative factors that govern regional growth are exogeneous to the model. By definition, net employment change is composed of gross gains and gross losses. In the second and third steps, respectively, regional job losses by POS category can be found using regional loss rates, and gross gains are found using the net change and the losses.

Regional job gains are of course not clumped in the center of the region, but must be spread across the metropolitan area. Each new job in each POS category has a fixed probability of locating in any RJA over the period of the model run. The probabilities, by RJA and POS type, and the loss rates are computed from the Dun and Bradstreet files. Finally, the new jobs for each RJA are placed in neighborhoods within the RJA on the basis of the attractiveness of that area to the particular type of job, and the number of jobs in each neighborhood is totalled for the next time period using the job gains and regional loss rates.

Regional net employment change

The total number of jobs for the next time period is found through Equation 10-1.

$$TOTJOB = RJ(11,NTRONE) * (1. + SPEC(IND)) \quad (10\text{-}1)$$

where:

TOTJOB	=	total number of jobs in the region for the next time period
RJ (11,NTRONE)	=	total number of jobs in the region for the current time period
SPEC(IND)	=	the compound rate of job growth of the region for each year, IND.

Job losses by RJA and POS. It is assumed, as explained above, that the net change in employment and the loss rates for each POS class are determined exogeneously and regionally. Job losses for each Industry K and Neighborhood I are given by the loss rates and the current number of jobs in each neighborhood and POS category.

$$LOSS(K,I) = RJ(K,I) * LRATE(K) \quad (10\text{-}2)$$

where:

LOSS(K,I)	=	the number of jobs lost in each Industry Type K and Neighborhood I
RJ(K,I)	=	the employment for each K,I
LRATE(K)	=	the job loss rate for each Industry K.

The values of LRATE are presented in Table 10-5.

Regional job gains

Job gains for each POS type LI are calculated from the job losses and the net changes which are summed from industry types (K) to POS classes. The losses and net changes are found from the following equations:

Table 10-5

Job Loss Rates by POS and City, 1969-1972

	Plant	Office	Store
Charlotte	.142	.074	.139
Dayton	.165	.086	.132
Houston	.177	.082	.149
New Haven	.197	.355	.173
Rochester	.195	.088	.175
Worcester	.207	.203	.159

$$BSUM\ (LI) = \sum_{K \in LI} SUML(K) \qquad (10\text{-}3)$$

$$BSUMD\ (LI) = \sum_{K \in LI} RDEM(K) \qquad (10\text{-}4)$$

where:

BSUM(LI)	=	the sum of losses for each POS (plant, office, store) Type LI.
SUML(K)	=	the sum of losses over the region for each Industry Type K
BSUMD(LI)	=	the sum of net changes for each POS Type LI.
RDEM(K)	=	the net change over the region for each Industry Type K.

The sum, BSUM (LI) + BSUMD(LI), represents the regional employment gains for each POS category. It is now necessary to determine how these gains distribute geographically among the RJA's. Until this step, the gains and losses have been determined outside the system boundaries of this model. The aggregate net change of employment in each industry, RDEM(K), is determined by a model closely linked to this one.[1] The two models can be considered as one grand model that describes the interregional flows of people and jobs on one level and the intraregional distribution of people and jobs on another.

RJA job gains

The job gains for each POS group are divided among the

1. See Birch, Allaman and Martin [3] for a description of that model.

RJA's according to the probabilities determined from the Dun and Bradstreet files.

BGAIN (LI,LB) = PROB(LI,LB) * (BSUM(LI) + BSUMD(LI)) (10-5)

where:

BGAIN(LI,LB) = the employment gains for each POS Type LI, and RJA LB.

PROB(LI,LB) = the probability that a new job in POS Class LI will locate in RJA LB.

This equation corresponds to the first level of locational decision-making described earlier in this section. At this point, establishment decision-makers determine which RJA is best suited to their industries. The next stage of the decision-making process is to find a particular neighborhood in which to locate.

Determination of neighborhood employment levels for each POS category.

Neighborhood employment levels are calculated in two steps. First, the probabilities that gains in an RJA will go to a particular neighborhood are determined by measuring the "attractiveness" of the neighborhood, and second, these probabilities are used to determine the level of employment in the neighborhood for the next time period. The probabilities are specified using the PFG package described in Volume 3, Section 5, Appendix C. These probabilities are determined on the basis of characteristics of each neighborhood and the extent to which these characteristics explain the actual distribution of gains given for a base year. The neighborhood characteristics we have used are described above. The variables that are used to measure these characteristics are defined below. The first variable, SLOT, is calculated for each POS type and is entered into the calculations for only that type.

SLOT	=	the number of jobs lost in the neighborhood during the previous time period. The variable is found by multiplying the loss rate of the POS type by the total number of jobs in that neighborhood.
HRNT	=	the proportion of high-value rental housing units to the total number of rental units in a neighborhood, expressed as a percentage.
CILD	=	the proportion of land in commercial and industrial use to the total land in a neighborhood, expressed as a percentage
ROAD	=	the road of highest quality (determined by number of lanes and type of intersections with other roads) crossing or touching the neighborhood.
HLRT	=	the ratio of high-value housing units to low value housing units in that neighborhood.
AALD	=	the product of the percentage of regional jobs within easy commuting distance of the neighborhood and the percentage of easy-to-build vacant land.
%PLT	=	the ratio of plant jobs to the total number of jobs in that neighborhood, expressed as a percentage.
%OFC	=	the ratio of office jobs to the total number of jobs in that neighborhood, expressed as a percentage.
%STO	=	the ratio of store jobs to the total number of jobs in that neighborhood, expressed as a percentage. The sum % PLT, %STO, and %OFC should be 1.0 for each neighborhood.

We assumed that jobs will be attracted or repelled by the characteristics represented by all variables other than AALD and ROAD in proportion to the values of these variables (i.e., the continuous function option of the PFG is used). AALD and ROAD are used in a very different way. We have structured the PFG so that firms will be attracted to a given neighborhood if the percentage of available attractive land (AALD) is above a given threshold and if the highest class of road measured in our model is present in that neighborhood.

The gains are distributed among neighborhoods according to the equation:

GAIN(LI,I) = BGAIN(LI,LB) * TRPROB(LI,I) (10-6)

where:

GAIN(LI,I)	=	the number of new jobs of POS Type LI going to each Neighborhood I.
TRPROB(LI,I)	=	the probability computed by the PFG that one new job of Type LI will be located in Neighborhood I.

The matrix TRPROB(LI,I) is computed using a set of 24 equations (one set for each city) generated by the PFG. It is assumed that there are no *a priori* reasons for an employer to choose one neighborhood in an RJA over another. We are more successful at predicting the location of jobs in certain RJA's (4 and 7, in particular) than in others, (most notably RJA 8).

Table 10-6 shows those variables that were selected by the PFG as having the greatest effect on the location of gross gains. The most important feature of this table is the prominence of the SLOT variable as an attractor of jobs. Neighborhoods with large numbers of Plant, Office, or Store jobs (and, therefore, large numbers of slots available due to constant loss rates) attract large numbers of Plant, Office, and Store jobs, respectively. %PLT, %OFC, and %STO are useful, especially in explaining the location of Office and Store jobs. Office jobs locate in those neighbor-

Table 10-6

Total Number of Times PFG[1] Chose Variables as Selector (+) or Rejector (-)

Variable	Plant +	Plant -	Office +	Office -	Store +	Store -
INERTIA						
SLOT (cont)	31	0	20	0	27	0
%PLT (cont)	6	1	3	1	1	1
%OFC (cont)	2	3	15	0	7	0
%STO (cont)	0	1	1	0	7	0
ATTRACTIVENESS						
HRNT (cont)	4	0	6	0	8	0
CILD (cont)	8	0	7	0	10	0
AALD (GT .40)-ROAD(GT .60)	9	4	3	4	9	2
AALD (GT .60)-ROAD(GT .60)	5	9	3	4	7	8
ALLD (GT .80)-ROAD(6T .60)	3	5	4	6	4	5

1. The PFG and the variable notation used here is described in Volume 3, Section 5, Appendix C.

hoods that already have Office jobs. Store jobs are attracted to neighborhoods with high proportions of Store or Office jobs, because Store jobs include commercial, service, and government jobs, many of which are, in fact, located in offices.

The other variables are most important in the location of Store jobs and least important in the location of Plants. The AALD-ROAD variable is mixed in its effects. It acts as both in attractor and a rejector in all three POS categories. The AALD-ROAD variable is most used as a selector characteristic by Store employers.

Finally, within each neighborhood for each POS class, it is assumed that the gains are distributed in the same way as the number of jobs in the past time period. The updated number of jobs is simply the current number of jobs, plus employment gains, minus losses.

$$RJ(K,I)_t = RJ(K,I)_{t-1} + TPROP(K,I) * GAIN(LI,I) - LOSS(K,I) \qquad (10\text{-}7)$$

where:

$RJ(K,I)_t$	=	the number of jobs in each Neighborhood I and Major Job Type K for time period t.
TPROP(K,I)	=	the proportion of jobs in an Industry Type K to the number of jobs in POS Category LI, Industry Type K represented in each neighborhood.

Labor Market

The labor market section of the model consists mainly of a series of mappings -- from jobs by industry to jobs by occupation (the demand side) and from persons by type of person to labor force by occupation (the supply side) -- in order to determine unemployment rates by occupation (see Figure 10-2). We relate the demand and supply of labor in a labor market that is occupation-oriented.

To account for net commuting from outside the model area, the population is first adjusted by multiplying each element in the population vector by an exogenously predicted factor. The population POP(K) is then mapped into the labor force by type of person, LFP(K), using labor force participation rates by type of person, LFPR(K):

$$LFP(K) = LFPR(K) * POP(K) \qquad K = 1,36. \qquad (10\text{-}8)$$

We assume that the labor force participation rates change slowly to reflect the increased propensity for women to enter the labor force and the general rise nationwide in participation rates.

The next task is to convert the labor force by type of person, LFP(K), into the labor force by occupation, LFO(OC):

$$LFO(OC) = \sum_{K} LFP(K)*POCC(K,OC) \qquad (10\text{-}9)$$

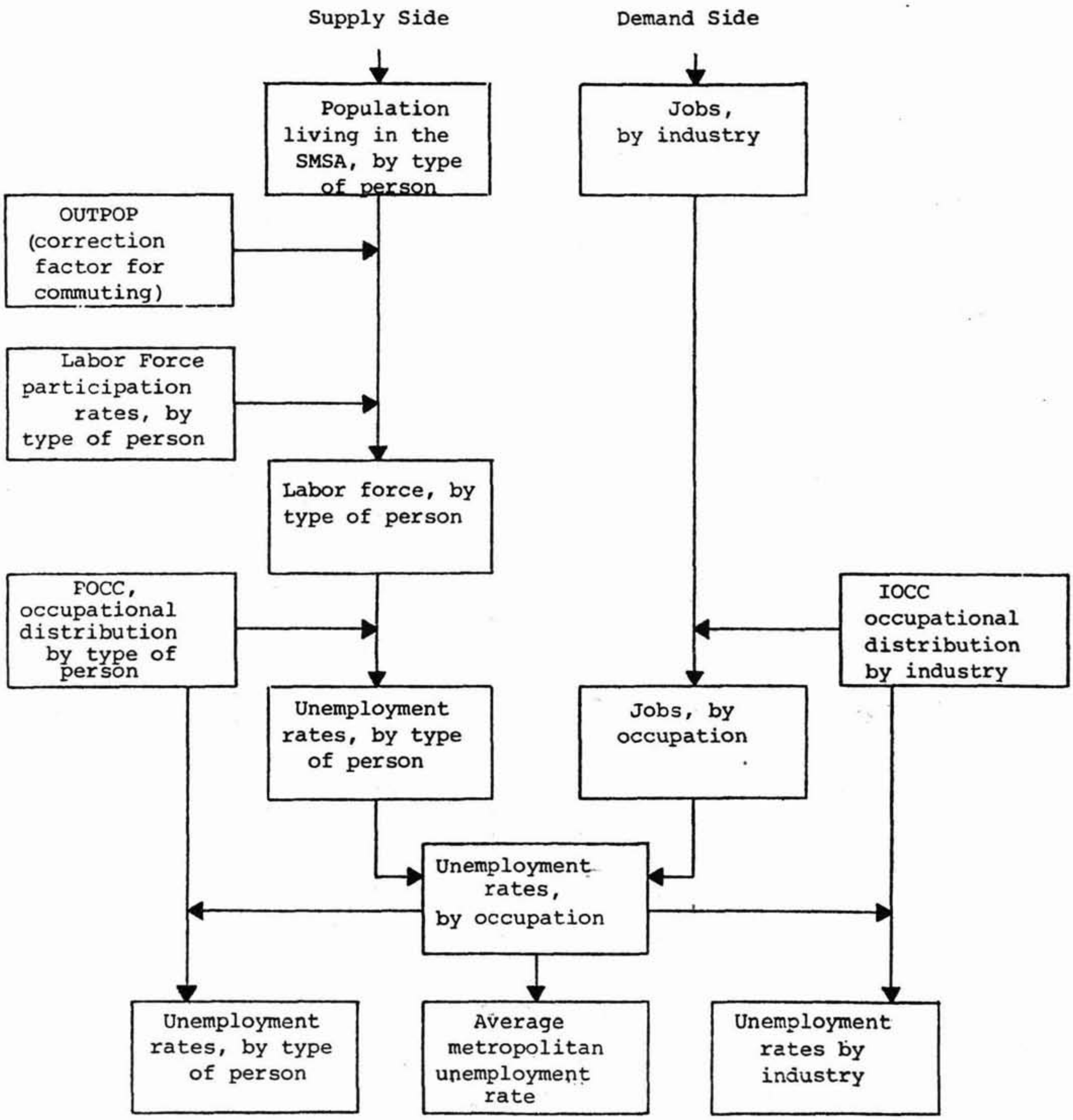
where:

POCC(K,OC) = the proportion in the labor force of Person Type K with Occupation OC.

POCC(K,OC) the distribution of the labor force over occupations, does not remain constant over time, but changes to reflect the changes in workers' job skills in response to the shifting importance of various industry types in the region. POCC(K,OC) is adjusted to maintain a constant ratio among the unemployment rates for each occupation type. The calculations on the supply side of Figure 10-2 are now completed and we

Figure 10-2

The Labor Market Submodel



shall turn our attention to the demand side.

Only one mapping is involved here, from jobs by industry to jobs by occupation, OCC(OC):

$$OCC(OC) = \sum_{L} REG(L)*IOCC(L,OC) \qquad (10\text{-}10)$$

where:

REG(L) = the number of jobs in Industry L.

IOCC(L,OC) = the proportion of the jobs in Industry L that fall in Occupation OC.

We assume that the relative distribution of jobs remains constant among occupations within an industry.

Employment Rates

The supply and demand sides are integrated through the computation of unemployment rates by occupation, UNEMO(OC)

$$UNEMO(OC) = (LFO(OC) - OCC(OC))/LFO(OC)$$

$$OC = 1,8 \qquad (10\text{-}11)$$

OCC(OC) in Equation 10-11 stands for available jobs in Occupation OC, which is equal to the sum of filled and unfilled jobs. We do not make any distinction between filled and unfilled jobs in the rest of the model. When we talk about jobs we mean available positions. This is reasonable in that one can expect physical facilities to be designed for available jobs and to have a corresponding effect on land use, for instance. The approach is less satisfactory when jobs are viewed as sources of demand for retail trade and services. Only workers who actually exist create such demand at their places of work. However, the errors introduced are certainly negligible when compared to other expected errors in the model and the data.

The final step in the labor market section involves mapping from unemployment rates by occupation to unemployment rates by type of person. The average unemployment rate for the SMSA is also determined. The unemployment rates by type of person, UNEMP(K), are:

$$UNEMP(K) = \sum_{OC} UNEMO(OC)*POCC(K,OC) \qquad (10\text{-}12)$$

Equation 10-12 simply says that the unemployment rate for person type K is the weighted sum of the unemployment rates by occupations, the weights given by POCC(K,OC). Finally, the average unemployment rate, UNEMME, is computed as:

$$UNEMME = \frac{\sum_{OC} UNEMO(OC)*LFO(OC)}{\sum_{OC} LFO(OC)} \qquad (10\text{-}13)$$

REFERENCES

1. Allaman, Peter. "Using the Dun and Bradstreet Data for Research on Regional Economic Development." Cambridge, Mass.: MIT-Harvard Joint Center for Urban Studies. Inter-Area Migration Project Working Paper No. 4, 1975.

2. Birch, David L., Peter M. Allaman, and Elizabeth A. Martin. Inter-Area Migration Project: Progress Report. Cambridge, Mass.: Joint Center for Urban Studies and MIT and Harvard University. March, 1976.

3. Birch, David L., Peter A. Allaman, and Elizabeth A. Martin. A Model of Population and Employment Change for Metropolitan and Rural Areas in the U.S. Cambridge, Mass.: Joint Center for Urban Studies of MIT and Harvard University. March, 1977.

4. Birch, David, Reilly Atkinson, Sven Sandstrom, and Linda Stack. The New Haven Laboratory: A Test-bed for Planning. Lexington, Mass.: Lexington Books, 1974.

5. Hoover, Edgar and Raymond Vernon. Anatomy of a Metropolis: The Changing Distribution of People and Jobs Within the New York Metropolitan Area. Cambridge, Mass.: Harvard University Press, 1959.

6. Putman, Stephen H. "Intraurban Employment Forecasting Models: A Review and a Suggested New Model Construct," Journal of the American Institute of Planners. July, 1972.

7. Struyk, Raymond J., and Franklin J. James, Jr. Intrametropolitan Industrial Location: The Patterns and Process of Change. Lexington, Books, 1975.

Section 11

SCHOOLS

Schools are a factor both in the search for new housing and in residents' satisfaction with their present housing. People, particularly parents, are often concerned about the kinds of services the school provides, the school's reputation, and other factors affecting the public image of the school. The attractiveness of a given neighborhood to residents or potential residents can be affected by their perception of the schools in that neighborhood.

But schools do not operate challenge-free. Residents -- new and old, parents and non-parents -- have some control over schools. They vote for school board members, they approve or disapprove school bond issues, and in some areas, they vote directly on school budgets.[1] In addition, they can also utilize personal contact, either individually or through group associations, to communicate their discontent with the schools to school administrators.[2] School administrators must react to resident response

---

1. Even when the school board is appointed, residents vote for the persons responsible for the appointment, so control is present though indirect.

2. Throughout this section unless otherwise noted, school administrators or administration will refer to both the school board and the school superintendent. For modeling purposes the complex interaction between the board and the superintendent is suppressed. Justification for compressing the two is based on the findings of the recent National School Boards Association survey [27]. Considerable confusion in the public mind existed regarding the locus of the legal responsibility for running the schools; 58 percent of the people surveyed thought it was the school board, 30 percent thought it was the school administration, and 40 percent thought it was the PTA.

if they are to remain administrators and if they are to receive continued support for their educational programs. In turn, their reactions will affect residents' views of and satisfaction with the schools. Consequently, the interaction of residents and educational administrators is important because it affects both the schools and residents' perceptions of the schools.

This interaction is intensified when one or both of the following conditions exists: schools are modified as the prevailing notion of educational criteria changes (for example, the educational revolution following Sputnik), or schools are altered as the residential composition of neighborhoods evolves and produces new needs, demands, and standards. As schools change, under either or both of these conditions, residents' perceptions of schools will be affected. As perceptions shift, satisfaction with residential location may also be affected. Extreme dissatisfaction can result in residential relocation. We therefore decided to examine resident/school administration interaction, and to simulate this interaction and the changes that result from it.

The school submodel simulates two forms of behavior:

1. residents' perceptions of and responses to changes in schools
2. school administrators' reactions to residents' behavior.

In the simulation, school policy is represented by selected school characteristics. Residents indicate their dissent with school policy by voting in an annual election. They vote to raise or lower the value of a school characteristic if its current value differs from that desired by the residents. School administrators react to residents' expressed preferences (the outcome of the election) by adjusting the value of the school characteristic.

The simulation is based on the theory that the demographic and socioeconomic composition of the community are indicators of the kind of education the community prefers. Implicit in the theory is the notion that different people perceive, behave, and respond in different ways to schools and changes in schools. As the composition of the neighborhood changes, the community's educational preferences also change. School administrators will tend to respond selectively to shifts in residents' preferences, whether they are a result of shifts in the composition of the population or shifts in the current notion of desirable educational policy.

Residents were grouped along four dimensions so we could cluster them by the traits that affect their response to schools. Three dimensions already existed in the model: age, education, and race/ethnicity. The fourth dimension is social class. These dimensions have been subdivided in the following ways:

Age:	20-39, 40-64, 65+
Social Class:	lower (LC), working (WC), middle (MC), and upper status (UC)
Education:	less than high school (<HS), high school (HS), more than high school (>HS)
Race/ ethnicity:	native white (W), foreign-born white (F), non-white (MIN).

Each resident type has his/her own perception of schools and his/her own way of acting on these perceptions.

<u>Perceptions of Schools</u>. Perception of schools is defined here as the intersection of educational attitudes and expectations. These represent a person's preferences regarding the qualitative and quantitative aspects of education. For example, educational expectation may refer to the amount of education each person is to receive, or to the pupil-teacher ratio. Educa-

tional attitude refers to the conditions under which people want that education to be delivered -- for instance, in a highly structured setting as opposed to an open school. Educational attitudes and educational expectations are both influenced by social class, prior educational experience, and, to a certain extent, ethnicity. Attitude toward education is also influenced by religious affiliation.[1]

The literature indicates that lower class people stress formal education less than members of the upper classes. Higher status members, especially younger people, tend to see education as having intrinsic value. Lower classes consider education worthwhile when it is job related and leads to occupational success and job security.[2] This class differentiation is compounded by racial factors. Researchers found that blacks generally regarded formal education as a necessary, if not sufficient, prerequisite to success in occupational mobility.[3]

Prior educational experience also influences educational attitudes and expectations. Three general conclusions can be made about adult educational attitudes. First, people with more education tend to be more positive about their educational experience and consequently, more likely to continue their education or to have high educational expectations for their children. Second, younger, better-educated people differ in educational attitudes from

---

1. See Berelson, Lazarsfeld, and McPhee, Voting [2].

2. See Johnstone and Rivera, Volunteers for Learning [19]; Hyman, "The Value Systems of Different Classes" [18]; and Cloward and Jones, "Social Class: Educational Attitudes and Participation" [7].

3. See Tollett, "Higher Education and the Public Sector" [33], and The Second Newman Report. National Policy and High Education [30].

older or less-educated people in that the former emphasize discipline less and individual development more. Third, people who are secure economically and socially are more likely to consider formal education intrinsically valuable than are people for whom education represents an opportunity for occupational and social mobility.[1]

Given these conclusions, three broad categories of educational attitudes can be identified:

Attitude Type	Attitude
A	The purpose of education is primarily functional; a rigid structure is preferred.
B	Education has innate value; a rigid structure is preferred.
C	Education has innate value; a more open structure is preferred.

"Functional" denotes that education is worthy as a means to an end. "Innate value" means that education itself is important. "Rigid structure" is defined as an emphasis on basics within a highly disciplined setting. More "open structure" is defined as an emphasis on individual exploration and development in a flexible structure. Each resident type falls into one of these categories. These attitude categories are then used to estimate the educational expectations for each school characteristic of each resident type.

---

1. These findings come from the research of many scholars who have attempted to assess adult attitudes toward continuing education as a function of their prior educational experience. For a more detailed discussion, see: Darrell Anderson and John A. Niemi, Adult Education and the Disadvantaged Adult [1]. Jack London, "The Influence of Social Class on Adult Education Participation" [23]; Harry L. Miller, Participation of Adults in Education [25]; Coolie Verner, Adult Education [35]; Jane Zahn, "Differences Between Adults and Youth Affecting Learning" [36].

The educational expectation is a numerical value, the educational ideal of each resident type. The degree of difference between the ideal value and the actual value of a school characteristic has an important effect on whether and to what degree residents will express dissent.

Expression of Dissent

Dissent can be expressed in many ways, most of which can be considered political activities. People can express their disapproval of school policy at the polls, in person, or indirectly through personal or political pressure. The method of expressing dissent is largely a function of social class.

In considering school-related political activity, we must remember that normally only a small percentage of residents, regardless of class membership, will be active on issues affecting the schools. Schools tend to be submerged into the background of daily life, to emerge into the forefront only when problems or issues arise.

The following discussion of school-related political activity centers around social class, however three additional sets of distinctions have also been taken into consideration: normal vs. crisis behavior, behavior based on differences in school district size, and behavior based on dissimilarities within each social class described.

Lower-class members have been deemed likely to be more apathetic and generally to be withdrawn from political activity as it applies to the school system.[1] Working-class members are alienated and also tend to be apathetic. However, under crisis conditions they will express dissent

---

1. See Boskoff, Voting Patterns in a Local Election [3]; Carter, Voters and Their Schools [5]; Minar, "The Community Bases of Conflict in School System Politics" [26]; Verba and Nie, Participation in America [34]; Milbraith, Political Participation [24].

through formal channels (e.g., voting) or they may use organized protest as a means of expressing their views.[1] Middle class and upper status members will vote, run for election as school board members, or they may move. Upper status members may also exercise direct or indirect pressure on school administrators, or they may send their children to private schools.

In the simulation, all forms of dissent are expressed as votes for or against a given school characteristic. As mentioned above, elections are held annually and residents vote to change a school characteristic depending on their perceptions of schools. The number of people who vote will depend on two things: (1) the degree of difference between each resident's ideal numerical value for a school characteristic and its actual value (the greater the difference the more likely is he/she to vote), and (2) the degree of political activity in which each resident ordinarily engages. Five measures of political activity have been identified. Each resident will fall into one of these five categories (see Table 11-1).

Individual votes are aggregated to determine the results of the election. However, the final outcome is affected by the distribution of resident types in various communities. An important attribute of resident types which affects the whole voting process is each resident's educational level. It is important to distinguish between those residents with four or more years of college and those with some education beyond high school. A device, the Power Ratio, has been developed to compensate for the lack of discrimination in educational levels of residents.

---

1. See Crain and Rosenthal, "Community Status as a Dimension of Local Decision-Making" [9], and Lipsky, "Protest as a Political Resource" [22].

Table 11-1

Resident Political Activity

Response Expected	Characteristics of Response	Conditions Generating Response
(AP) Apathy	Little likelihood of activity although change in characteristic is viewed negatively	Alienation: No sense of political efficacy
(CA) Conditional Apathy	Will act when aroused by issue; when issue subsides, activity ceases	Issue has emotional overtones -- pressure from peers or political leaders
(LA) Limited Activity	Partial or spasmodic activity -- not likely to be sustained	Concern over change in characteristic is strong and sense of political efficacy reinforces decision favoring increased involvement
(COA) Considerable Activity	Increase in activity of people who are likely to be in positions of power or influence	Concern over change in characteristic is strong, as is sense of political efficacy

The Power Ratio refers to the proportion of high-cost housing to low-cost housing in a given school district. The assumption is that, for the most part, higher-income and better-educated people will live in high-cost housing. The greater the Power Ratio, the greater the proportion of higher-income, well-educated residents.

When the Power Ratio exceeds 5.0, the educational expectations of residents between the ages of 20 and 64 with an educational level beyond high school are adjusted. This adjustment increases the difference between the actual value of the school characteristics and the ideal value of the school characteristic, thus increasing the intensity of response.

School Characteristics

Initially, six characteristics were selected for inclusion in the sub-model. These were teacher qualification, curriculum, class size, the school's public image, discipline, and desegregation/integration.

Preliminary analysis resulted in the elimination of three of the school characteristics. According to the data supplied by the school districts, the percentage of high school dropouts, the surrogate for public image, showed no significant change over the years and varied only slightly from one district to another. A second characteristic dropped was Teacher Qualification -- the percentage of teachers (in elementary schools) with education at the master's level or higher. This characteristic showed little difference across communities. The trend was consistently upward, a factor largely accounted for by the decrease in available jobs, the contracts negotiated by teachers' unions, and the increase in the supply of younger, better-educated teachers. Discipline was dropped because of measurement problems.

Desegregation/integration (the percentage of minority students in the school district) was kept but is treated differently from class size and curriculum. The percentage of minority students increases under normal migration, natural increase, and voluntary or court-mandated desegregation efforts. Change in this characteristic is a concern of most parents, regardless of ethnicity, but the causes for change are too complex to replicate in a simulated election. Therefore, changes in percentage of minority students are computed annually, independent of the voting process.

Class size and curriculum varied over time and from one district to another (see Table 11-2). They are defined as follows.

Curriculum: The percentage of high school graduates going on to four-year colleges was selected as a surrogate for curriculum. No distinction was made between public and private colleges.

Class size: The measure used for class size was the mean class size in the elementary schools in a given school district.

These characteristics were intended to represent some of the policy issues that might become a source of contention in the community. The Gallup surveys on the public's attitude toward education were the primary sources used for guidance in the selection of these characteristics.[1] The surveys were particularly helpful because they were done over a period of several years and used a number of categories, thus demonstrating the effect of parental status on perception and the changes in concern about different school characteristics over time.

---

1. See Elam, The Gallup Polls of Attitudes Toward Education, 1969-1973 [11]; Gallup, "Sixth Annual Gallup Poll of Public Attitudes Toward Education" [13]; Gallup, "Seventh Annual Gallup Poll of Public Attitudes Toward Education" [12].

Table 11-2

Changes in School Characteristics

Mean Class Size

School District	1960	1970	1975
New Haven	26.2	21.4	19.5
Branford	29.2	25.4	23.1
East Haven	20.2	25.0	22.8
North Haven	29.3	23.3	21.9
Hamden	26.9	21.4	20.2
Orange/Woodbridge	25.5	22.5	20.5
West Haven	29.4	26.5	23.0

Percentage of High School Graduates to Four-Year Colleges

School District	1960	1970	1975
New Haven	40.2	33.5	28.1
Branford	32.0	40.5	43.1
East Haven	20.0	22.0	23.0
North Haven	41.0	43.0	45.0
Hamden	51.0	74.0	55.0
Orange/Woodbridge	64.0	65.0	68.0
West Haven	38.0	35.0	31.0

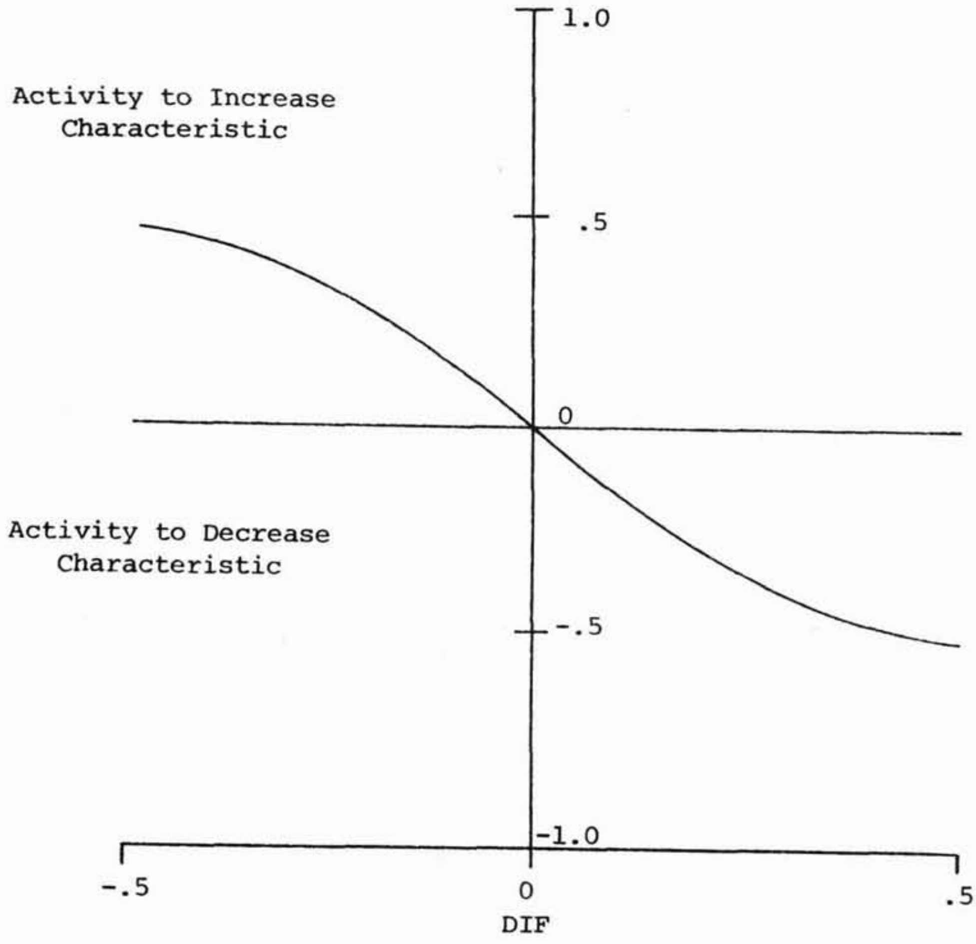
Much of the data for 1960 was extrapolated from data for years between 1960 and 1970 because of the lack of available records for 1960.

School Administrators

School administrators react to residents' expressed preferences by modifying school characteristics. Figure 11-1 graphs a typical response function. The vertical axis represents the limits within which the school administrators can make adjustments to the characteristics. The horizontal axis represents the vote as a percentage of the total vote, and gives the direction of the vote. Positive values mean that most voters want to increase the value of the characteristic; negative values mean that they want the value of the characteristic decreased. Their reactions are based on the results of the election. There is a separate administration reaction curve for each characteristic. However, the educational philosophy espoused by administrators often differs from one district to another. More importantly, the constituency responded to may also differ from one community to another so that there may be more than one kind of administrator reaction to voter response. The fact that, over the years, some communities have maintained the tradition of sending large percentages of high school graduates to four-year colleges even though the population mix has changed has been considered in this submodel as an example of differences in administrator reaction. To account for this, the Administrator Reaction Function is automatically adjusted in the submodel whenever the net vote to lower the percentage of students going to four-year colleges is negative. After any adjustments have been made to the Administrator Reaction Function, the reaction to the resident response is computed and numerical values for the school characteristics are produced for the next year. The cycle then repeats itself.

Figure 11-1

Sample Voter Response Curve



Many factors govern the reactions of school administrators. They can be divided into two groups, internal and external factors. Among the internal factors are the relationship of the school board and the school superintendent, bureaucratic tradition, the power exerted by teacher and other school employee unions, and faculty, staff, and board expectations of the superintendent.[1] External factors include the sources of power within the community, the degree and mode of resident dissent, the administration's identification of its constituency, and the degree to which the superintendent and the school board share this identification.[2]

The school superintendent is faced with very real limitations in the degree to which he can respond to net resident preference. He is constrained on the one hand by a multiplicity of resident views and often no clear consensus regarding direction. He is also constrained by systemic obligations and requirements.[3] As a result, the amount of change that he can implement is relatively small. This is attended to in the model by

---

1. See Rogers, 110 Livingston Street [29]; Corwin, Education in Crisis [8]; Cheng, Altering the Collective Bargaining Structure in Public Education [6]; Gross, Ward, and McEachern, Exploration in Role Analysis: Studies of the School Superintendency Role [16]; Orsi, School-Board-Superintendent Relationships - A Case Study [28].

2. See Kirst, The Politics of Education at the Local, State and Federal Levels [20]; Havighurst and Neugarten, Society and Education [17]; Zeigler, Jennings, and Peak, Governing American Schools [37]; Gans, The Urban Villagers [14]; Gittell, Participants and Participation [15]; DeGood, "Can Superintendents Perceive Community Viewpoints?" [10]; Koerner, Who Controls American Education? [21].

3. See Campbell, Cunningham, and McPhee, The Organization and Control of American Schools [4]; Selakovich, The Schools and American Society [31]; Sexton, The American School [32].

limiting his capacity to alter the characteristic to a maximum of 5 percent in either direction.

This section has described how the model simulates the interaction of residents and school administrators by using the voting process as the vehicle for expressing and meeting challenges to change. The factors which shape residents' response and the constraints which limit administrators' reactions have been described. The process is seen as ongoing with peaks of activity occurring under situations of stress but with constant challenge being presented and met as times and residents change.

Mathematical Statement

In the School Submodel, residents vote to raise or lower the numerical value of school characteristics. School administrators react to the vote by raising or lowering the value of the characteristic in accordance with the net result of the vote. Votes are tallied by school district. The first step is to determine the percentage of high and medium cost owned and high cost rental housing to low cost owned and rental housing in order to calculate the ratio of the percentage of high/medium to low for the school district. This is called Power Ratio (PR).

$$PR(ID) = PH/PL$$

where

PR(ID) = the power in the school district as determined by PH/PL

PH = the percentage of high-and medium-cost owned and high-cost rental housing

PL = the percentage of low-cost owned and rental housing

PR is used to adjust the school preferences for percent-to-college in districts with a high power ratio. The school characteristics are class size (J=1) and percent to college (J=2).

For each school characteristic, residents are assumed to have a preference or educational expectation which is determined by their age, ethnicity, class and education. Table 11-3 illustrates the 108 resident types. Table 11-4 lists the educational expectations for each type for the school characteristic: Percentage of high school graduates going to four-year colleges. This value is expressed in the submodel as:

VALUE (J,K): The numerical preference assigned to each characteristic (J) for each population type (K).

Table 11-5 lists the preferences (educational expectations) for each of the 27 population types used in the submodel. The school characteristic, percent to college, is adjusted as follows:

VALUE (2,K) = VALUE (2,K) * ADJ

where:

VALUE (2,K) = the preferred value of School Characteristic 2 by population Type K

ADJ = the amount by which the value is raised if the power ratio (PR) is high.

Next, the difference between the actual value of the school characteristic and the value preferred by residents is calculated:

DIF (J,K) = SCHOOL (J,ID) - VALUE (J,K)

where:

DIF (J,K) = The difference between the actual value and the preferred value

SCHOOL(J,ID)= The actual value of the School Characteristic J in School District ID

VALUE (J,K) = The preferred value of the school characteristic for Population Type K.

Table 11-3

Classification of Residents

Age		Less than HS			HS			Greater than HS		
		White Native	Foreign Born	Minority	White Native	Foreign Born	Minority	White Native	Foreign Born	Minority
20-39	LC	1	2	3	4	5	6	7	8	9
	WC	10	11	12	13	14	15	16	17	18
	MC	19	20	21	22	23	24	25	26	27
	UC	28	29	30	31	32	33	34	35	36
40-64	LC	37	38	39	40	41	42	43	44	45
	WM	46	47	48	49	50	51	52	53	54
	MC	55	56	57	58	59	60	61	62	63
	UC	64	65	66	67	68	69	70	71	72
65+	LC	73	74	75	76	77	78	79	80	81
	WC	82	83	84	85	86	87	88	89	90
	MC	91	92	93	94	95	96	97	98	99
	UC	100	101	102	103	104	105	106	107	108

Table 11-4

Characteristics of Resident Types

Resident Type	Age	Class	Education	Race/ Ethnicity	Educational Expectation
1	20-39	LC	< HS	W	.05
2	20-39	LC	< HS	F	.05
3	20-39	LC	< HS	MIN	.10
4	20-39	LC	HS	W	.10
5	20-39	LC	HS	F	.10
6	20-39	LC	HS	MIN	.15
7	20-39	LC	> HS	W	.15
8	20-39	LC	> HS	F	.15
9	20-39	LC	> HS	MIN	.20
10	20-39	WC	< HS	W	.10
11	20-39	WC	< HS	F	.10
12	20-39	WC	< HS	MIN	.15
13	20-39	WC	HS	W	.15
14	20-39	WC	HS	F	.15
15	20-39	WC	HS	MIN	.30
16	20-39	WC	> HS	W	.15
17	20-39	WC	> HS	F	.15
18	20-39	WC	> HS	MIN	.30
19	20-39	MC	< HS	W	.20
20	20-39	MC	< HS	F	.20
21	20-39	MC	< HS	MIN	.35
22	20-39	MC	HS	W	.30

Table 11-4 (continued)

Resident Type	Age	Class	Education	Race/ Ethnicity	Educational Expectation
23	20-39	MC	HS	F	.30
24	20-39	MC	HS	MIN	.40
25	20-39	MC	>HS	W	.60
26	20-39	MC	>HS	F	.65
27	20-39	MC	>HS	MIN	.65
28	20-39	UC	<HS	W	.30
29	20-39	UC	<HS	F	.30
30	20-39	UC	<HS	MIN	.40
31	20-39	UC	HS	W	.35
32	20-39	UC	HS	F	.35
33	20-39	UC	HS	MIN	.45
34	20-39	UC	>HS	W	.65
35	20-39	UC	>HS	F	.65
36	20-39	UC	>HS	MIN	.65
37	40-64	LC	<HS	W	.15
38	40-64	LC	<HS	F	.15
39	40-64	LC	<HS	MIN	.15
40	40-64	LC	HS	W	.20
41	40-64	LC	HS	F	.20
42	40-64	LC	HS	MIN	.25
43	40-64	LC	>HS	W	.25
44	40-64	LC	>HS	F	.25
45	40-64	LC	>HS	MIN	.30

Table 11-4 (continued)

Resident Type	Age	Class	Education	Race/ Ethnicity	Educational Expectation
46	40-64	WC	< HS	W	.20
47	40-64	WC	< HS	F	.20
48	40-64	WC	< HS	MIN	.25
49	40-64	WC	HS	W	.25
50	40-64	WC	HS	F	.25
51	40-64	WC	HS	MIN	.30
52	40-64	WC	> HS	W	.20
53	40-64	WC	> HS	F	.20
54	49-64	WC	> HS	MIN	.30
55	49-64	MC	< HS	W	.25
56	40-64	MC	< HS	F	.25
57	40-64	MC	< HS	MIN	.30
58	40-64	MC	HS	W	.40
59	40-64	MC	HS	F	.40
60	40-64	MC	HS	MIN	.40
61	40-64	MC	> HS	W	.75
62	40-64	MC	> HS	F	.75
63	40-64	MC	> HS	MIN	.75
64	40-64	UC	< HS	W	.35
65	40-64	UC	< HS	F	.35
66	40-64	UC	< HS	MIN	.45
67	4o-64	UC	HS	W	.40
68	40-64	UC	HS	F	.40
69	40-64	UC	HS	MIN	.50

Table 11-4 (continued)

Resident Type	Age	Class	Education	Race/ Ethnicity	Educational Expectation
70	40-64	UC	>HS	W	.75
71	40-64	UC	>HS	F	.75
72	40-64	UC	>HS	MIN	.75
73	≥65	LC	< HS	W	.05
74	≥65	LC	< HS	F	.05
75	≥65	LC	< HS	MIN	.05
76	≥65	LC	HS	W	.15
77	≥65	LC	HS	F	.15
78	≥65	LC	HS	MIN	.20
79	≥65	LC	>HS	W	.20
80	≥65	LC	>HS	F	.20
81	≥65	LC	>HS	MIN	.25
82	≥65	WC	< HS	W	.10
83	≥65	WC	< HS	F	.10
84	≥65	WC	< HS	MIN	.15
85	≥65	WC	HS	W	.20
86	≥65	WC	HS	F	.20
87	≥65	WC	HS	MIN	.30
88	≥65	WC	>HS	W	.10
89	≥65	WC	>HS	F	.10
90	≥65	WC	>HS	MIN	.15
91	≥65	MC	< HS	W	.15
92	≥65	MC	< HS	F	.15

Table 11-4 (continued)

Resident Type	Age	Class	Education	Race/ Ethnicity	Educational Expectation
93	≥65	MC	<HS	MIN	.20
94	≥65	MC	HS	W	.25
95	≥65	MC	HS	F	.25
96	≥65	MC	HS	MIN	.30
97	≥65	MC	>HS	W	.50
98	≥65	MC	>HS	F	.50
99	≥65	MC	>HS	MIN	.50
100	≥65	UC	<HS	W	.20
101	≥65	UC	<HS	F	.25
102	≥65	UC	<HS	MIN	.30
103	≥65	UC	HS	W	.25
104	≥65	UC	HS	F	.25
105	≥65	UC	HS	MIN	.35
106	≥65	UC	>HS	W	.50
107	≥65	UC	>HS	F	.50
108	≥65	UC	>HS	MIN	.50

Table 11-5

Resident/Population Types

Population Type	Resident Type	Educational Expectations
10	1	.05
11	13	.15
12	34	.65
13	2	.05
14	14	.15
15	26	.65
16	3	.10
17	15	.30
18	27	.65
19	46	.20
20	58	.40
21	70	.75
22	38	.15
23	50	.25
24	63	.75
25	39	.15
26	51	.30
27	72	.75
28	82	.10
29	94	.25
30	106	.50
31	83	.10
32	95	.25
33	98	.50
34	84	.15
35	96	.30
36	99	.50

Residents' responses are depicted mathematically by a voter response function. The shape of the function is based on the likelihood of school-related political participation of various social classes. People's probable attitudes and their estimated educational expectation for a particular school characteristic are also considered. Figure 11-2 graphs a typical voter response function. The horizontal axis represents the difference (DIF) between the ideal value and the actual value for a given characteristic. The vertical axis, an activity scale, represents the expected level of activity. Zero level represents consonance between actual and ideal value, with no change desired. As the positive value on the vertical scale approaches 1.0, the intensity of the activity to raise the level of the characteristic rises. As the curve moves toward -1.0, the intensity of activity to lower the value of the characteristic rises. The degree or strength of activity is indicated by the slope and depth of the curve. The equation of the function graphed in Figure 11-2 is:

$$\text{Activity} = a - a * e^{-\frac{\text{Dif}}{a} * sl}$$

where:

$$a = .5 \text{ for } x < 0$$

$$a = -.5 \text{ for } x \geq 0$$

$$sl = -2$$

Residents then vote, and the result of the vote is:

VOTE (J,ISIGN,ID) = POP(K,ID) * Activity (K)

where:

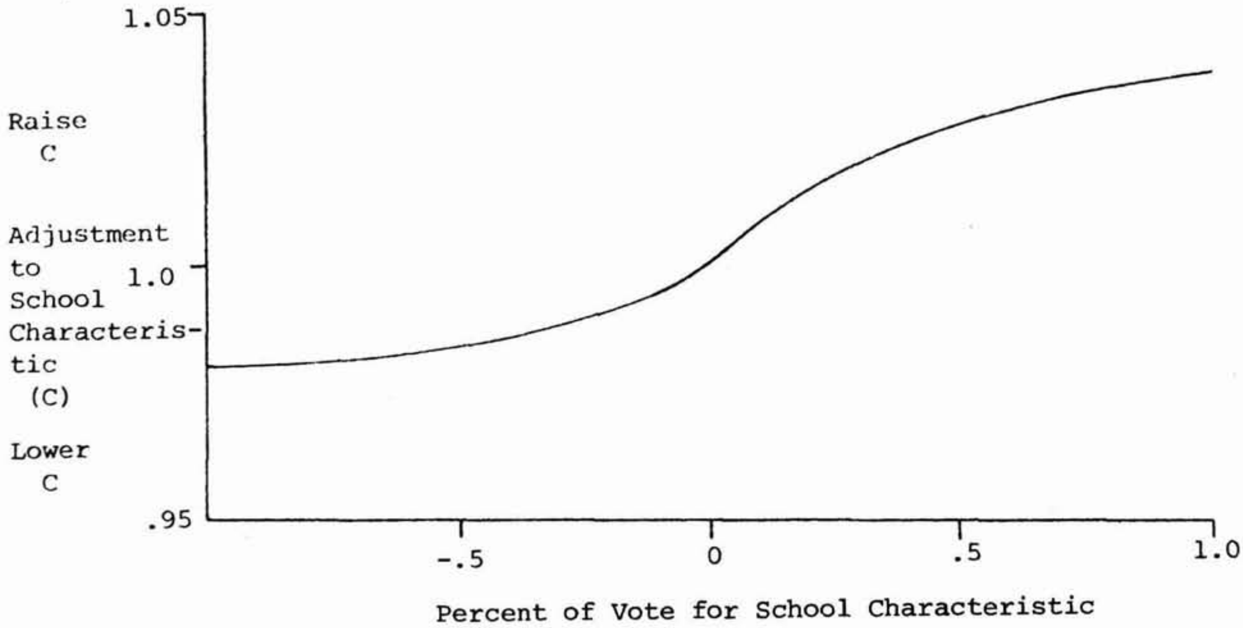
VOTE = The cumulative vote

J = The characteristic voted on

ISIGN = The direction of the vote, i.e., to raise or lower the value of the characteristic

Figure 11-2

Administration Reaction Function



ID = The school district

POP(K,ID) = The population by type, by district

Activity(K) = The odds that residents of Type K will take action to increase or decrease Characteristic J. Activity is derived from the response functions mentioned earlier.

The school administration reacts to the results of residents' votes.

PVT(J) = (VOTE (J,I,ID) - VOTE (J,2,ID)) / SDPOP(ID)

where:

PVT(J) = The net vote as a percentage of the adult population for Characteristic (J)

VOTE(J,1,ID) = The positive vote by characteristic (J) for School District (ID)

VOTE(J,2,ID) = The negative vote by characteristic (J) for School District (ID)

SDPOP(ID) = The total adult population in the school district.

The amount of adjustment to Characteristic J is then derived from the administrator's response function:

SCHOOL (J,I) = SCHOOL (J,I)*Y

where:

SCHOOL (J,I) = School Characteristic J by neighborhood I and

Y = the amount by which SCHOOL (J,I) will be modified [$Y \in .95, 1.05$].

The function for estimating Y has the general form:

$$Y = a + (1-a) * e^{-\frac{x}{a-1} * sl}$$

where:

a = .95 for $x < 0$

a = 1.05 for $x \geq 0$, and

sl = .06 (for the graph depicted).

The adjustment is aggregated by neighborhood for the school district

$$SD\ (J,ID) = \sum_{I} SCHOOL\ (J,I)$$

where:

SD(J,ID) = the adjusted characteristic by school district

The percentage of minority children 0-19 in the school district is calculated in order to keep track of changes in minority student population in the various school districts over time.

PMNST = POPMIN(ID)/POPSCH(ID)

where:

PMNST = the percentage of minority youth 0-19 in the school district

POPMIN(ID) = the number of minority youth 0-19 by school district

POPSCH(ID) = the total number of youth 0-19 by school district

The percentage of minority students in the public schools is estimated as a function of the percentage of minority youth in the school district:

PMIN = Z*PMNST

where:

Z = .20 + ABS (PMNST - .20).

The next year begins with the modified values for school characteristics and percentage of minority students (PMIN), and the entire process is repeated.

References

1. Anderson, Darrell and Niemi, John A. Adult Education and the Disadvantaged Adult. Syracuse: Syracuse University, 1970.

2. Berelson, Bernard R., Lazarsfeld, Paul F., and McPhee, William N. Voting. Chicago: University of Chicago Press, 1954.

3. Boskoff, Alvin. Voting Patterns In a Local Election. Philadelphia: Lippincott, 1964.

4. Campbell, Roald F., Cunningham, Luvern L. and McPhee, Roderick F. The Organization and Control of American Schools. Columbus, Ohio: Merrill Books, 1965.

5. Carter, Richard F. Voters and Their Schools. Stanford: Stanford University, 1960.

6. Cheng, Charles William. Altering the Collective Bargaining Structure in Public Education: Citizen Participation in Educational Decision Making, New York: Praeger Publishers, 1976.

7. Cloward, Richard A. and Jones, James A. "Social Class: Educational Attitudes and Participation" in Passow, A. Harry (Ed.) Education in Depressed Areas. New York: Columbia University, 1963.

8. Corwin, Ronald G. Education in Crisis. New York: Wiley, 1974.

9. Crain, Robert L. and Rosenthal, Donald B., "Community Status as a Dimension of Local Decision-Making," American Sociological Review, Dec. 1967, Vol. 32, No. 6, pp. 970-984.

10. De Good, Kenneth C. "Can Superintendents Perceive Community Viewpoints?" Administrator's Notebook, Vol. VIII, Nov. 1959, No. 3.

11. Elam, Stanley (Ed.). The Gallup Polls of Attitudes Towards Education, 1969-1973. Bloomington, Indiana: Phi Delta Kappa, Inc., 1973.

12. Gallup, George H. "Seventh Annual Gallup Poll of Public Attitudes Toward Education," Phi Delta Kappan, Vol. 57, No. 4, December, 1975, pp. 227-241.

13. Gallup, George H. "Sixth Annual Gallup Poll of Public Attitudes Toward Education," Phi Delta Kappan, Vol. 66, No. 1, September, 1974, pp. 20-32.

14. Gans, Herbert J. The Urban Villagers. New York: The Free Press, 1962.

15. Gittell, Marilyn. Participants and Participation. New York: Praeger, 1967.

16. Gross, Neal, Mason, Ward S. & McEachern, Alexander W. Explorations in Role Analysis: Studies of the School Superintendency Role. New York: John Wiley & Sons, 1958.

17. Havighurst, Robert J. and Neugarten, Bernice L. Society and Education. Allyn and Bacon, 1967, 3rd ed.

18. Hyman, Herbert H. "The Value Systems of Different Classes," in R. Bendix and S. M. Lipset (Eds.) Class, Status and Power. Glencoe: The Free Press, 1953.

19. Johnstone, John and Rivera, Ramon J. Volunteers for Learning. Chicago, Aldine Publishing Co., 1965.

20. Kirst, Michael W., Ed. The Politics of Education at the Local, State and Federal Levels. Berkeley: McCutchan, 1970.

21. Koerner, James D. Who Controls American Education? Boston: Beacon Press, 1968.

22. Lipsky, Michael. "Protest as a Political Resource," The American Political Science Review, Vol. 62, No. 4, December 1968, pp. 1144-1158.

23. London, Jack. "The Influence of Social Class Behavior on Adult Education Participation." Adult Education, Vol. 20, No. 3, 1970, pp. 140-153.

24. Milbraith, Lester W. Political Participation. Chicago: Rand McNally, 1965.

25. Miller, Harry L. Participation of Adults in Education. Brookline: Center for the Study of Liberal Education for Adults, 1967.

26. Minar, David W. "The Community Basis of Conflict in School System Politics," American Sociological Review, Vol. 31, No. 5, October 1966, pp. 822-835.

27. National School Boards Association. Public Knowledge and Attitudes Concerning Local School Boards. Princeton: The Gallup Organization, Inc., 1975.

28. Orsi, D. Raymond. School Board - Superintendent Relationships - A Case Study. Dissertation. Temple University, 1969.

29. Rogers, David. 110 Livingston Street. New York: Random House, 1968.

30. The Second Newman Report: National Policy and Higher Education. Cambridge: MIT Press, 1973.

31. Selakovich, Daniel. The Schools and American Society. Lexington, Massachusetts: Xerox College Publishers, 1973, 2nd ed.

32. Sexton, Patricia Cayo. The American School. Englewood Cliffs: Prentice-Hall, 1967.

33. Tollett, Kenneth S. "Higher Education and the Public Sector," in B. Lawrence, George Weathersby and V. W. Patterson (Eds.), Outputs of Higher Education. Boulder: Western Interstate Commission for Higher Education, 1970.

34. Verba, Sidney and Nie, Norman H. Participation in America. New York: Harper and Row, 1972.

35. Verner, Coolie. Adult Education. Washington, D.C.: The Center for Applied Research in Education, 1964.

36. Zahn, Jane C. "Differences Between Adults and Youth Affecting Learning," Adult Education, Vol. 17, No. 2, 1967, pp. 67-77.

37. Zeigler, L. H., Jennings, M. K. and Peak, G. W. Governing American Schools. North Scituate, Massachusetts: Duxbury Press, 1974.

Section 12

VALIDATION

A major question that must be asked of any model is: Does it work? Does it behave the way the world behaves? Certainly no one will, or should rely on a model that cannot replicate the processes it is trying to simulate with acceptable accuracy. A model must do more than "look reasonable." It must cause neighborhoods to evolve on the computer the way they actually evolve in the world.

To see if our model is valid in these terms, we invested heavily in data collection for the 1960's and for the early 1970's. With these data, we are able to determine if, by following the rules we have established for them, our actors cause changes in neighborhoods on the computer the same way they caused changes during the 1960's and the early 1970's in neighborhoods in the six regions.

Measurement Accuracy

A common misconception among many planners and administrators is that the number of people in households or jobs are known for particular points in time. It is assumed that census figures or directories offer a complete and perfect count for the item of interest. One planner, for example, called us recently to complain that our estimate of the population of his city of several hundred thousand people was off by 83, and why couldn't we do something about it.

The reality, of course, is that there is a considerable amount of uncertainty in any estimate of any measure of people, or housing units, or jobs, or land use. The greater the level of detail, the greater the uncertainty. Census estimates of the total population of a region, for example, might be accurate to within two or three percent, while neighborhood estimates might vary, on the average, by seven or eight percent. As we move away from census years and data, the errors get much worse. Table 12-1 presents our estimates of the accuracy with which the "real world" is measured during a census year (1960 or 1970). In non-census years, using local data sources not designed for enumeration purposes, a general rule of thumb is that these accuracy figures should be multiplied by a factor of two or three, depending on the reliability of the local source.

No model can be more accurate than the accuracy of the data against which its performance is being compared. Any accuracy greater than this is spurious -- there is no way of knowing whether the model or the data are right. Thus, for a census year, we can not expect to estimate neighborhood population better than about plus or minus five percent, and in an off-census year, 10 percent to 15 percent represents the lower limit. Any errors less than this are more than likely an accident, and are certainly not a credit to the model. For all we know, within this range the model could be right and the world (as measured) wrong. Our friend's complaint about the error of 83 was thus absurd; anything within about two or three thousand would have been as good as perfection.

The growth rate of an area also affects the accuracy with which it can be measured (and modeled). Faster growing areas tend to have higher turnover rates and, of course more rapidly expanding neighborhoods. In

Table 12-1

Average Percent Error in Measurements of Population, Households, Housing Units and Employment for Census Years

Population, Households, and Housing Units

	Region	County	Neighborhood
Totals	± 1%	± 3%	± 5%
Marginals[1]	± 2%	± 5%	± 9%

Employment

	Region	District
Totals	± 1%	± 7%
Marginals	± 5%	± 15%

---

1. Marginals are the one-way breakdowns that might appear at the margins of a table of, say, age by race by education. The age distribution is thus a "marginal."

Houston, for example, there are several neighborhoods that have grown from a few hundred people to 15 or 20 thousand in three or four years. Under these circumstances, it is much more difficult to say how many people are in any particular place at a chosen point in time, simply because the number has changed significantly during the sampling interval. This is particularly true for locally generated estimates that may be put together over a period of several months. Needless to say such growth is also more difficult to model accurately. Being off in timing by only one year can lead to an error of 20 or 25 percent. Errors for places like Houston and Charlotte will thus tend to be higher than errors for New Haven and Worcester.

In order to compare the accuracy of our model with the accuracy of the data we have adopted as our measure:

$$\text{Average Error} = \frac{\sum |S - A|}{\sum A}$$

That is, the average error is the sum of the differences between actual and simulated values (for whatever item is being considered) without regard to the sign of that difference, divided by the sum of the actual values .

Experience with the 1960's

Tables 12-2 to 12-5 show the accuracy with which we have been able to simulate the experience of the 1960's for population, households, housing units and employment for various levels of substantive and geographic detail. As anticipated, the errors increase as the level of detail increases. Also errors are greater for the more rapidly growing areas -- especially Houston.

Table 12-2

Average Percent in Projection of 1970 Population
from a 1960 Base, by Area

Area	Level of Detail	Geographic Disaggregation: Region	County	Neighborhood
New Haven	Totals	.032	.032	.101
	Marginals	.033	.033	.156
Worcester	Totals	.022	.022	.075
	Marginals	.050	.050	.123
Dayton	Totals	.000	.022	.107
	Marginals	.010	.037	.152
Rochester	Totals	.012	.015	.075
	Marginals	.021	.029	.123
Charlotte	Totals	.015	.021	.070
	Marginals	.020	.051	.130
Houston	Totals	.009	.050	.162
	Marginals	.017	.063	.228

Table 12-3

Average Percent Error in Projections of 1970 Households
from a 1960 Base, by Area

		Geographic Disaggregation:		
Area	Level of Detail	Region	County	Neighborhood
New Haven	Totals	.025	.025	.071
	Marginals	.025	.025	.143
Worcester	Totals	.039	.039	.051
	Marginals	.064	.064	.126
Dayton	Totals	.003	.012	.081
	Marginals	.014	.042	.144
Rochester	Totals	.002	.017	.059
	Marginals	.007	.024	.122
Charlotte	Totals	.001	.006	.054
	Marginals	.026	.058	.139
Houston	Totals	.005	.037	.101
	Marginals	.006	.051	.198

Table 12-4

Average Percent Error in Projections of 1970 Housing
from a 1960 Base, by Area

Area	Level of Detail	Geographic Disaggregation: Region	County	Neighborhood
New Haven	Totals	.026	.026	.071
	Marginals	.065	.065	.148
Worcester	Totals	.033	.033	.047
	Marginals	.057	.057	.134
Dayton	Totals	.009	.016	.081
	Marginals	.053	.085	.178
Rochester	Totals	.011	.021	.060
	Marginals	.062	.064	.158
Charlotte	Totals	.003	.010	.054
	Marginals	.051	.082	.161
Houston	Totals	.002	.036	.102
	Marginals	.038	.073	.183

Table 12-5

Average Percent Error in Projections of 1970 Employment from a 1960 Base, By Area

Area	Level of Detail	Geographic Disaggregation: Region	County	Neighborhood
New Haven	Totals	.003	.003	.112
	Marginals	.003	.003	.190
Worcester	Totals	.000	.000	.095
	Marginals	.000	.000	.199
Dayton	Totals	.000	.019	.328
	Marginals	.000	.087	.531
Rochester	Totals	.000	.031	.179
	Marginals	.000	.061	.302
Charlotte	Totals	.000	.027	.184
	Marginals	.000	.144	.399
Houston	Totals	.000	.035	.225
	Marginals	.000	.069	.438

In general, the errors in all six regions are now approaching the measurement accuracy of the data. There is still room for improvement, but we have achieved our overall goal of replicating neighborhood change (building from a behavioral base) with reasonable accuracy. The results for Rochester are particularly satisfying to us. Rochester was tested last, and it only had to be run once for the 1960-1970 period. It required no recalibration.

As anticipated, Houston has posed the most problems from a modeling standpoint because of its rapid growth. New Haven and Dayton could also use some further work, particularly at the neighborhood level.

Of particular concern is the difficulty in modeling employment location. While less accurate measurement in the first place is a partial cause of the greater average errors, measurement accuracy is not the whole story. Employment location decisions are "lumpy", especially at the neighborhood level. One plant moving or expanding on contracting can change a neighborhood's employment level quickly and significantly. Variations of this sort tend to average out for higher levels of aggregation, and, fortunately, our model is not too sensitive to variations across neighborhoods so long as they compensate each other for slightly larger areas. Much work needs to be done, however, if we are to understand and explain how the business executive makes fine-grained location decisions.

The pattern of errors underlying the aggregates is fairly consistent across regions. Most simulated neighborhood values are very close to actual values but a few diverge significantly. In our initial runs of the model, the large errors were attributable to coding errors and clerical errors in the data. As these more mechanical sources of distortion were removed, what remained was a group of anomalies -- special land holdings, sewer restrictions, new flood plain restrictions, restrictive real estate practices, a large, publically assisted housing project, a massive, localized program of demolition, a large plant lay-off and so forth. None of these special events or circumstances were known to the model and hence they could not be taken into account. As we move into the next phase of model development, one of our first tasks will be to identify and account for each special event or circumstance so that the rest of the model will not be unduly influenced by them.

Our efforts to simulate changes in schools began well after the rest of the submodels were developed, and are just now begining to pay off. Table 12-6 summarizes our results for New Haven -- the first region for which we were able to obtain the necessary information. As can be seen from the table, we are now replicating change in New Haven's schools with considerable accuracy for both class size and percent of high school graduates who go on to college. Data collection is practically completed in Charlotte and Dayton, and model results will be available for those areas soon. Data from the remaining three areas have not, as yet, been made available to us.

Table 12-6

Errors in Projections of 1970 School Characteristics, starting from a 1960 Base for Cities and Towns in the New Haven Area

Class Size

School District	City/Town	Actual 1970	Sim. 1970	Diff.	% Diff.
1	New Haven	21.4	22.88	1.48	6.9%
2	Branford	25.38	25.07	0.31	1.2%
3	East Haven	25.0	25.15	0.15	0.6%
4	North Haven	23.25	23.67	0.42	1.8%
5	Hamden	21.39	23.23	1.84	8.6%
6	Orange/Woodbridge	22.5	21.96	0.54	2.4%
7	West Haven	26.5	25.72	0.78	2.9%

Percent to College

School District	City/Town	Actual 1970	Sim. 1970	Diff.	% Diff.
1	New Haven	33.5%	34.8%	1.3	3.9%
2	Branford	40.5%	40.2%	0.3	0.7%
3	East Haven	22.0%	22.5%	0.5	2.3%
4	North Haven	43.0%	26.9%	3.9	9.1%
5	Hamden	53.0%	52.2%	0.8	1.5%
6	Orange/Woodbridge	65.0%	68.2%	3.2	4.0%
7	West Haven	35.0%	34.5%	0.5	1.4%

Validation in the 1970's

As might be expected, validation beyond the last decennial census year (1970) becomes far more difficult than checking at a census year. We have nevertheless attempted to piece together whatever sources were available to us to see how well we were doing.

For those areas made up of whole counties (all SMSA's outside of New England) we were able to compare our county estimates of total population with those made by the Census Bureau for 1975. The results are presented below:

Area	Average County Population Error in 1975
Dayton	.06
Rochester	.05
Charlotte	.05
Houston	.09

The standard error of the census estimates is about four percent; we can expect to do no better than that. On the average, our 1975 county estimates approach the standard error of the estimate. In general, the smaller counties tended to have the largest percentage errors, but the absolute errors for these counties were only a few thousand people.

Below the county level, we must rely on locally generated data sources to provide checks on our accuracy. As mentioned above, those local sources tend to be patched together from scraps of data that can be obtained from schools, utilities, local school censuses, building permits, and so forth, and are far less accurate than census figures. In general, their error rates are two to three times higher than census errors. Referring back to Table 12-1, we would expect post-1970 errors for total population (the only item generally available) to be about 10-15 percent at the neighborhood level. Again any model accuracy greater than this is spurious.

Three cities provided us with population estimates at the neighborhood, district or town level: Worcester, Charlotte, and New Haven. Our average population errors for these three are presented in Table 12-7. By and large, our average errors are close to the measurement accuracy of the data. In other words we cannot do much better on the average. Detailed comparisons on a neighborhood-by-neighborhood basis reveal patterns similar to those for 1970. Most neighborhoods are well within the standard error of the data source; a few lie significantly outside the range, and are candidates for further investigation. Of particular concern is our over-estimate for the City of New Haven, which we are now looking into.

A potentially powerful source of validation data on a universal basis is the Landsat satellite. Landsat is capable of producing estimates of land use on an acre-by-acre basis for each region in the country every 17 days in theory, and once a year practically speaking. The satellite records information on each acre in four different frequencies. The information is transmitted to ground stations, where it is compared against known examples of actual land use and thereby calibrated. Once calibrated, the Landsat data can be used to make acre-by-acre estimates

Table 12-7

Average Percentage Error for Population and Housing Units for Various Subareas in Worcester, Charlotte and New Haven for 1975 Predictions Starting From a 1970 Base.

Areas	Item	City of Worcester	Districts in Worcester	Neighborhoods in Worcester
Worcester	Total Population	.02	.08	.12
	Total Housing Units	.03	.08	.09
Charlotte		Mecklenberg County	Districts in Mecklenberg	Neighborhoods in Mecklenberg
	Total Population	.05	.11	.12
	Total Housing Units	.01	.10	.11
New Haven		New Haven SMSA	City of New Haven	Other Towns in SMSA
	Total Population	.04	.11	.06

of land use and change in land use over time. Also, thanks to much careful work by the Jet Propulsion Laboratory (JPL) in developing much of this technology for urban areas, it is possible to aggregate the data to census tracts (or any other set of geographic boundaries) and prepare land use tabulations by type of land use by neighborhood over time.[1]

By careful selection of categories, we are able to compare our estimates of what is on the ground with those made by Landsat. We are just beginning to make such comparisons, and find that on our first pass, the satellite did a relatively good job in some instances, and made horrendous mistakes in others. The problem of course, comes in calibration, not in the accuracy with which the satellite itself measures. Our initial set of calibration sites were not always well chosen and led to ambiguity in resolving distinctions between, say, commercial and industrial usage. Figure 12-1 summarizes some of these difficulties. JPL is now going through another round of calibration based on our first experience and feels that with more careful selection of test sites, we can improve the accuracy considerably.

At present the average errors by land use type run about 30% to 40% over all neighborhoods in a region, with occasional accuracy as good as 4% or 5% and with those horrendous, sometimes comical mistakes. The central business district of Houston, for example, now shows up as a large lake (perhaps it rained before the image was taken, exposing a

---

1. The procedure used by JPL to develop the land use maps and tabulations is described, with examples, in Appendix A to this section.

Figure 12-1

Some Reasons for Variations between City-supplied Land Use Data and Satellite-Derived Classification Data

1. Difference in date - many times the city data is for 1970 and the satelite overflight is in 1972 (1975 for Rochester).

2. Training site as delineated includes too much information (e.g., a residential site includes an elementary school - this would perhaps cause many commercial cover types to be classified residential because of the spectral similarity between schools and commercial buildings.)

3. Training site as delineated includes too little information. If only the grassy areas of an industrial park are given as an industrial classification, for example, there will be a confusion with other grassy classes such as parks, golf courses, etc. Also the buildings left out of the site may be classified as commercial, due to spectral similarity.

4. Too few sites to represent a class are given as input to the classifier If all of the residential sites are marked in highly vegetated areas, the brand new and relatively low-in-vegetation subdivisions will be classified as perhaps commercial (due to low amount of vegetation) or as unknown.

5. The sites were not accurate for the date of the imagery (e.g., a rural site in early 1972 became graded for development at the time of the imagery in late 1972. This will yield a class of rural that actually represents all "developing" classes.)

6. A land use map was not provided. Since several classes have similar spectral signatures (occupy the same ranges in all four bands), such as commercial and industrial, confusion results between them when input together into a classifier. The resulting output is either mostly one class or mostly the other class. Without at least a land use map with general delineations, a good guess cannot be made as to what the output is.

7. Some classes are made up of a variety of land cover types that become confused with other classes. Industrial parks, for example, are made up of highly reflective buildings, grassy areas and asphalt parking lots. Many times these cover types cannot be separated from other cover types in the rest of image. Commercial classes are confused with the buildings, parks confused with grassy areas, and roads confused with the parking lots. The industrial area may end up being classified within three entirely different classes.

8. Atmospheric conditions such as haze, clouds, and smog affect the spectral responses received. This causes a residential site response in a clear area, for example, to be different from an otherwise identical residential site in a hazy area.

sea of wet roof tops to Landsat). Bugs of this sort will be out of the system in a few months.

Validation in the Future

We have learned a great deal from our validation procedures. Many cherished theories with which we began the modeling effort have fallen by the wayside. Others have emerged as we tried to explain variations between simulated and actual outcomes. The present accuracy of the model is attributable in no small measure to our efforts to check our results against the best estimates of real-world phenomena we could obtain. We intend to expand our validation work in the future.

The 1980 census offers an obvious opportunity to check model results against another entire decade of neighborhood change. In the meantime, however, we must find ways to get accurate measures for intra-censal years. One step is to improve the realibility of Landsat by doing a better job of selecting ground truth sites and of calibrating the Landsat data against them. The process is now underway.

A second source which we have underutilized for lack of time, not interest, is the file on neighborhood condition developed by the R.L. Polk Company. As an offshot of its directory business for those areas in which it canvases, Polk has been able to estimate population and households and housing unit occupancy by neighborhood every year or two. This source is available for many of the regions in which we are now working. We plan to experiment with its use in Charlotte in the months ahead.

The real key in the long run, however, will be locally generated data files. A few cities like Wichita, Kansas keep detailed track of who is living in what kind of housing units in each neighborhood each year through its own census taking procedures. Others, like Charlotte and Detroit, maintain machine-readable files on each parcel of property in their cities, including information on condition, value and occupancy. Through the use of automated address coding guides[1] it should be possible to aggregate local records such as the parcel files, or assessors records or school enrollment cards, to neighborhoods and begin to observe and measure neighborhood change with respectable accuracy. Until we reach this point, managing cities will be largely an intuitive process and modeling them will still depend more heavily than it should on measurements that are five or ten years apart.

---

1. The Census Bureau has initiated the creation of one such set of files, called DIME files, for most cities in the United States. All DIME files are not of the same caliber, however.

Appendix A

THE CREATION OF LAND USE MAPS AND TABULATIONS FROM LANDSAT IMAGERY[1]

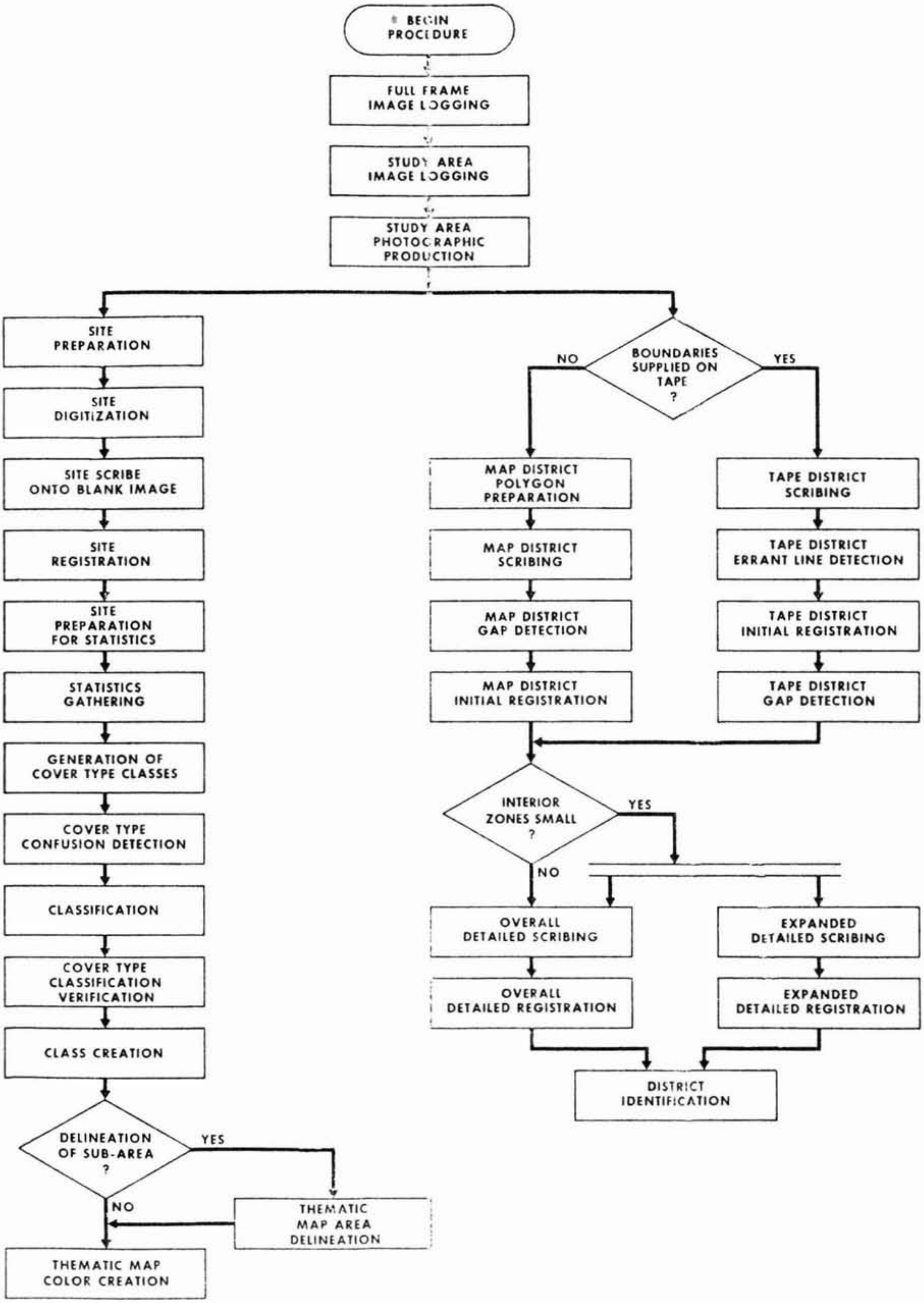
A plausible solution to the verification process rests in the use of thematic maps and tabulations of land use derived from recent Landsat digital imagery. In this paper, the procedures offered by the Multiple Input Land Use System (MILUS) data management system will be described as they have been used to create Landsat-derived color thematic maps and statistical tabulations of land use for the Dayton, Ohio Standard Metropolitan Statistical Area (SMSA).

The steps of the MILUS procedure are outlined in Figure 1. Each box listed contains a sequence of both computer program applications and editing decisions made by the analyst. In the interests of clarity and brevity, only those steps vital to the understanding of the MILUS system will be described using the Dayton case as an example. The upper three boxes in Figure 1 outline the steps used to prepare the Landsat image for future processing. The Goddard-formatted tapes are first logged and then made compatible with the JPL/VICAR image processing system. The program simultaneously generates the full frame at corrected orthophoto projection with any desired aspect ratio. Using generated histograms of reflectance over the entire frame, the desired contrast can be applied to produce an image for viewing.

1. Prepared by Nevin A. Bryant and Gary L. Angelici of NASA's Jet Propulsion Laboratory.

Figure 1

MILUS PROCESSING STEPS



From the full frame image, a rectangle including the study area is digitally extracted and stored on one tape for the sake of future processing economy. At the same time, data transmission errors (such as sixth band striping) are corrected. A set of images is then generated which displays the range of spectral reflectance available in each of the four spectral bands. These images (Figure 2) are referred to fequently by the analyst in subsequent steps to assess their relative spectral heterogeneity.

The steps used by JPL in the creation of land cover maps is outlined along the left column of Figure 1. Training sites that delineate the variety of classes of land use desired were provided by the Dayton regional council of government. Figure 3 shows one of the USGS 7.5 minute topographic maps displaying circumscribed polygons denoting particular land uses of interest in the Dayton study area.

The sites are manually transferred to a mylar oversheet on a 20x24 inch photographic enlargement of the Landsat MSS band exhibiting the highest contrast. In order to eliminate later spectral confusion between classes, the sites are subdivided into cover types. Thus agricultural land use may be subdivided into a variety of spectrally diverse cover types. The cover type sites and two control points for rigid rotation and scale approximation are then digitized via a coordinate digitizer. The control points on the Landsat image provide the transformation for a computer routine to produce output with the digitized training site polygons in Landsat row and column format.

After removing obvious digitizing errors, training site data are then transformed by software into a numerical data set in image form. In order to inspect whether the digitizing is geographically accurate, the image of training site contours is digitally added to the Landsat image. Any

Figure 2

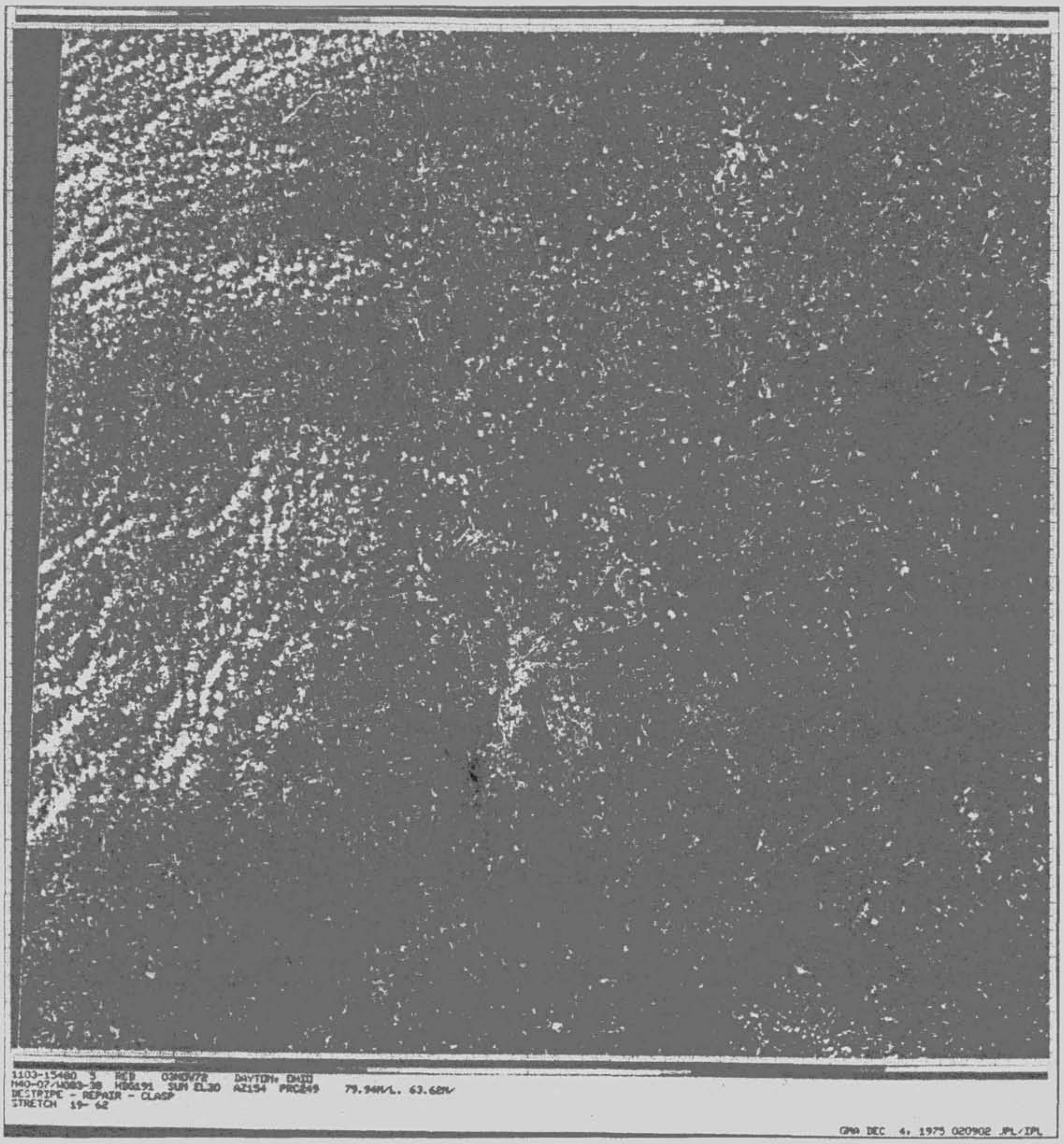
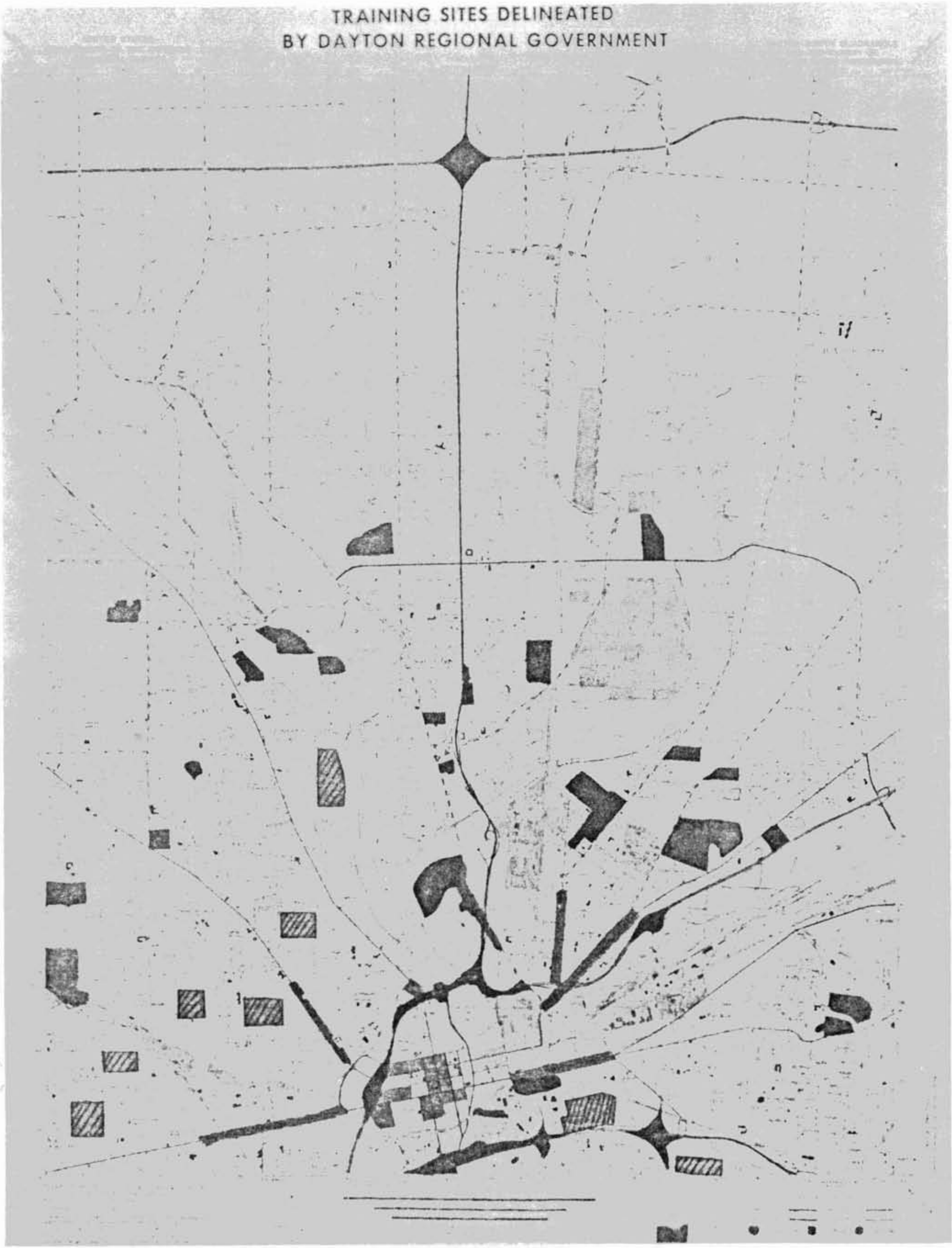

1103-15480 5 RED 03NOV72 DAYTON, OHIO
N40-07/W083-38 MBG191 SUN EL30 AZ154 PRC249 79.94M/L. 63.62M/
DESTRIPE - REPAIR - CLASP
STRETCH 19- 62
GMA DEC 4, 1975 020902 JPL/IPL

Figure 3



misregistration of sites is alleviated by close inspection followed by readjustment of the contours.

Once adequately registered, the sites are individually inventoried for their spectral reflectance. A computer routine gathers the spectral range of each training site in the four raw MSS bands and displays the results in two forms as illustrated in Figure 4.

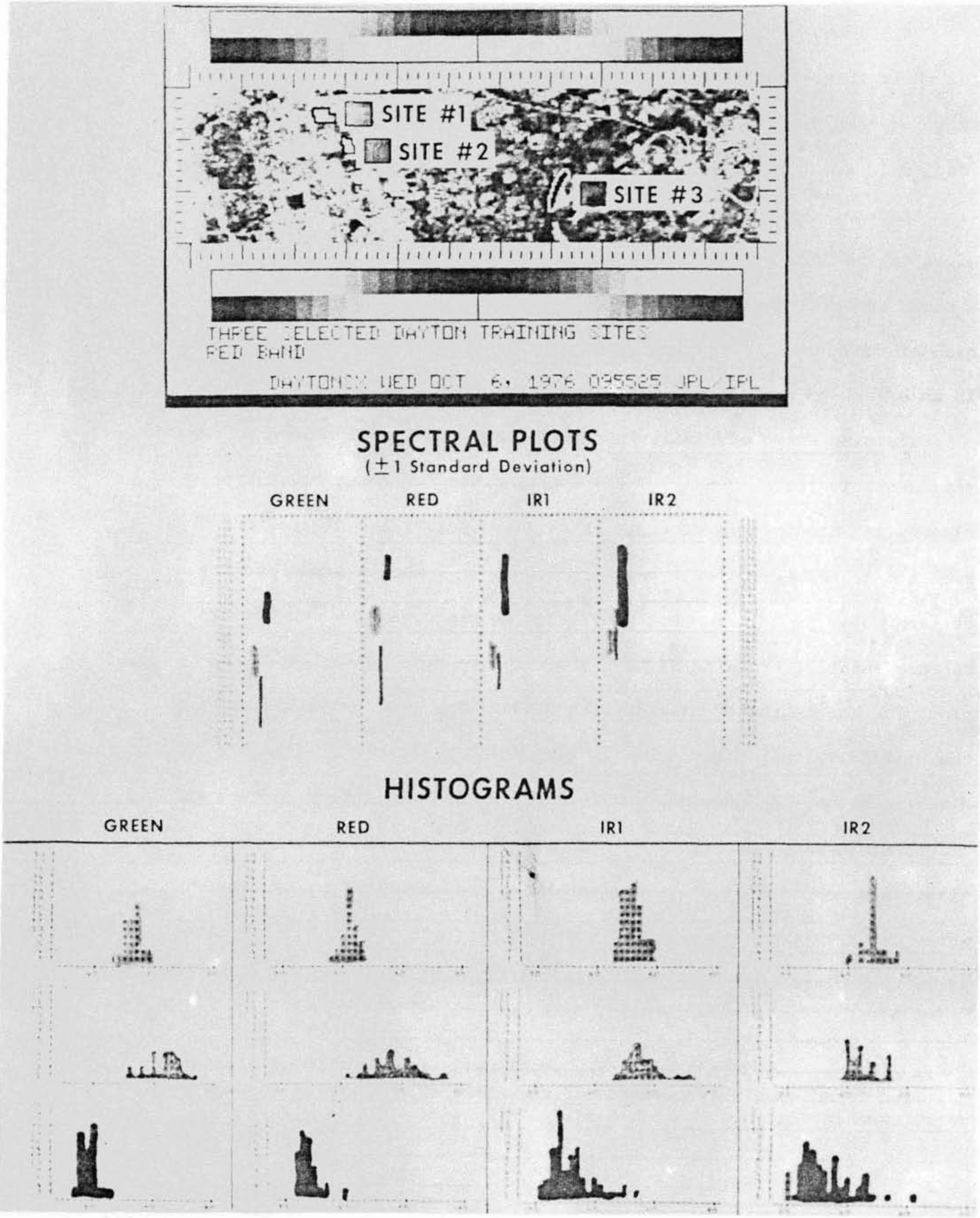
Spectral representation of three selected Dayton training sites is given in the form of spectral plots and histograms. A spectral plot with a broad spectral range requires the inspection of the histogram for bimodal tendencies. Training sites are further subdivided or discarded in an effort to avoid confusion between spectrally unique cover types.

Training sites of similar spectral range are then combined to create representative cover types. The spectral plots for these cover type classes are produced in order to check for separability of spectral reflectance between cover types (see Figure 5). The internal quality of each cover type is inspected by creating histograms, which detect unrepresentative training sites. The relative size of each spectral plot plus the degree of overlap with dissimilar cover types indicates the difficulty that might occur in separating cover types during classification. A Bayesian-Parallelpiped hybrid classifier is used to aid in minimizing the amount of amiguity of class assignment.[1] The classification is initially verified by a misclassification routine which computes the percentage of the pixels of the training sites in a particular class that have been actually classified as that class.

---

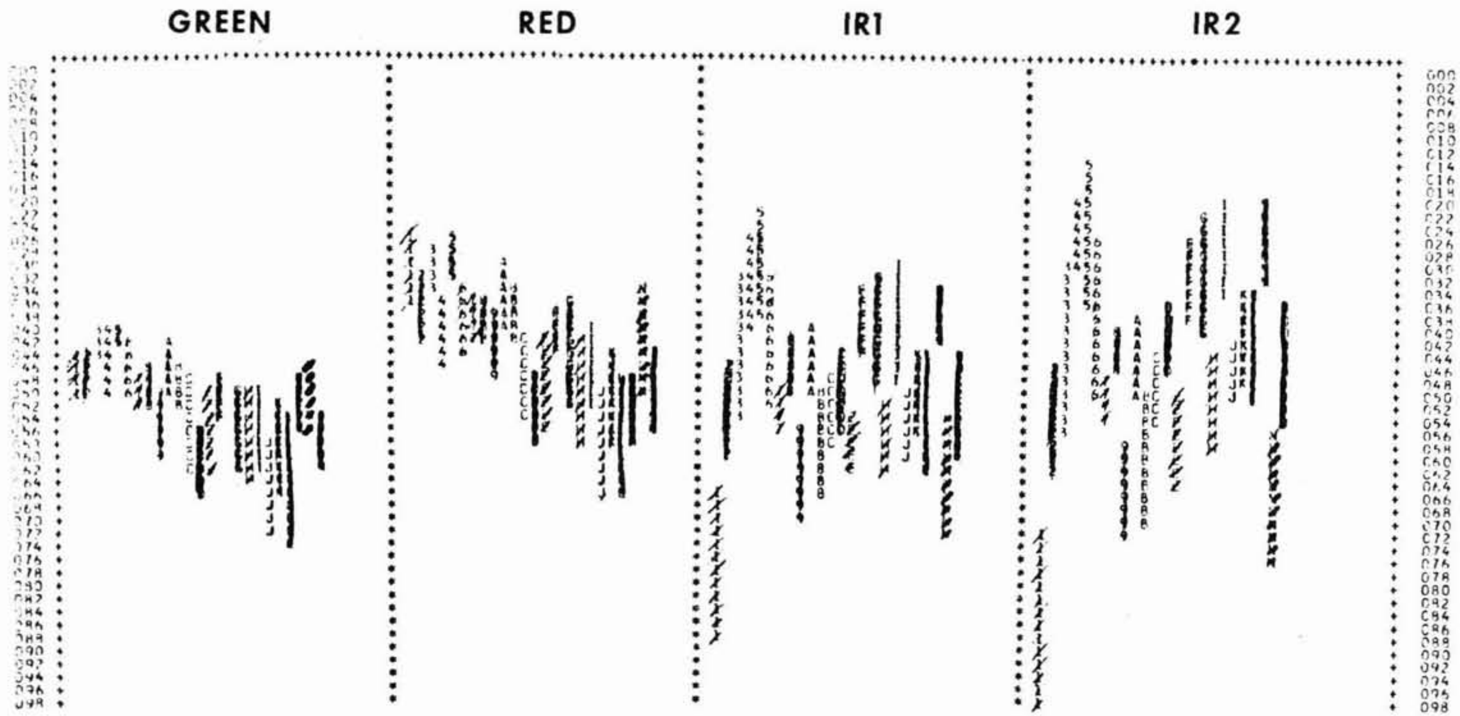
1. Addington, J.D., "A Hybrid Classifier Using the Parallelpiped and Bayesian Techniques", Proceedings of the American Society of Photogrammetry, Washington, D.C., March 9-14, 1975

Figure 4


SITE #1
SITE #2
SITE #3
THREE SELECTED DAYTON TRAINING SITES
RED BAND
SPECTRAL PLOTS
(± 1 Standard Deviation)
GREEN
RED
IR1
IR2
HISTOGRAMS
GREEN
RED
IR1
IR2

SPECTRAL PLOTS OF DAYTON COVER TYPES

(± 1 Standard Deviation)



COVER TYPES

- VACANT (Light)
- VACANT (Dark)
- VACANT-FOREST
- WATER
- AGRICULTURE (Dark)
- AGRICULTURE (Light)
- HIGH DENSITY RESIDENTIAL (Light)
- HIGH DENSITY RESIDENTIAL (Dark)
- LOW DENSITY RESIDENTIAL
- MULTIPLE FAMILY
- MOBILE HOMES (Light)
- MOBILE HOMES (Dark)
- LARGE COMMERCIAL (Dark)
- LARGE COMMERCIAL (Light)
- SMALL COMMERCIAL
- HIGH DENSITY OFFICE BUILDINGS
- LOW DENSITY OFFICE BUILDINGS
- HEAVY INDUSTRY (Dark)
- HEAVY INDUSTRY (Light)
- LIGHT INDUSTRY
- EXTRACTIVE
- GOVERNMENT AND INSTITUTIONAL (Dark)
- GOVERNMENT AND INSTITUTIONAL (Light)
- ROADS AND INTERCHANGES

Figure 5

Re-execution of the site selection process and classification is necessary if the error statistics are unacceptable. A second verification technique used is the creation of a color map of cover types. Viewing this thematic map and the misclassification routine output may reveal new classes not before stipulated or different classes than previously intended. Once an adequate classification is attained, each cover type is collapsed into land use classes resembling as closely as possible those classes initially desired. Since the class, "water", is often misclassified in the classification routine, a binary mask of the raw IR2 band multiplied by the classification map is utilized to create a more accurate representation.[1] The final stages of collapsing cover types into classes produces the color thematic map. A color-coded legend is digitally placed on the image and the positives output for color reconstruction. A black and white version of the final product is shown in Figure 6.

Viewing such an image is visually stimulating and applicable to planning in the broadest sense but useful data is obtained primarily by aggregating data by some administrative district and printing a table of values for each class.[2]

---

1. Blackwell, Richard J., "The Trophic Classification of Lakes Using ERTS Multispectral Scanner Data", Proceedings of the American Society of Photogrammetry, Washington, D.C., March 9-14, 1975.

2. Bryant, N.A. and Zobrist, A.L., "IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Datatype", Proceedings of the LARS Symposium on Machine Processing of Remotely Sensed Data, Lafayette, Indiana, June 29, 1976.

DAYTON, OHIO SMSA
LAND USE
NOVEMBER 1972
VACANT
FOREST
AGRICULTURE
EXTRACTIVE
WATER
MOBILE HOMES
CLOUDS
COMMERCIAL
OFFICE BUILDINGS
INDUSTRIAL
INSTITUTIONAL
UNKNOWN
CENSUS TRACTS

Figure 6

A thematic land use map can be tabulated only after a number of preparatory steps are performed (see right half of Figure 1). The first step requires the selection of the administrative districts by which the land use will be aggregated and, more important procedurally, the form in which the boundaries will be provided. If the boundary lines are supplied only on a map, for example, additional steps must be performed. Control points must be located, the boundary lines digitized onto tape, and finally output as Landsat row and column positions. Boundaries supplied on tape (such as the Census Bureau Urban Atlas File tapes of census tracts) are already digitized, thereby allowing immediate execution of subsequent steps after an initial logging procedure.

Following tape copying or digitizing steps, the lines are digitally output (scribed) onto a blank image. This task is performed, in the case of tape input, by converting the unit of measure in which the tape was digitized (latitude and longitude for Urban Atlas File tapes) to the line and sample coordinates of the Landsat image. Inputing control points from a map featuring the unit of measure used for the tape file and identical ones on the image into a conversion routine produces an accurate transformation. This image is then inspected closely for digitizing errors, such as missing lines, gaps left between lines, or errant lines that do not represent actual boundaries. Photographic enlargements aid the analyst in finding the line and sample coordinates of these errors. Alterations to the digitized data or in software parameters used for reading the input tape finally produce an image of boundary lines ready for overlaying onto the Landsat image. Upon digitally summing the boundary line image and the Landsat image, the technician can view the overall registration quality. Registration within 50 (fifty) pixels of the accurate location is usually sufficient at this point in the process. Cross misregistration requires rechecking control point numbers or shifting the starting line and sample of the boundary line image.

The precise registration of the boundaries requires the use of additional control points and application of a continuous surface fitting algorithm. With an enlargement of the image of the boundaries overlaid onto the Landsat frame, corresponding points of the district lines and on the Landsat frame can be located and marked. When a collection of such points is found throughout the image, their row and column coordinates are listed as well as the differences between the present and the correct coordinates. After placing the points in a proper format, a computer routine readjusts the boundaries to fit as precisely as the digitizing accuracy allows. Occasionally the digitizing is inexact, or areas of digitizing are skewed in relation to the entire file. In this case, another set of points must be determined and the routine executed again in order to bring the lines into proper registration.

If any administrative districts are so small that the ratio of border to interior pixels is unacceptable, the area on the Landsat frame that includes such areas is digitally enlarged by repeating pixels in both the row and column direction three or more times. The district boundaries are again overlaid on the image, but now the area within each district has increased while the district boundary lines remain only one pixel wide. The Dayton case did not require such an expansion. The boundary lines, once adequately registered, are then processed by a routine that fills holes left due largely to double-digitizing. Each district is then encoded with a unique digital value, and the border lines are randomly assigned to the adjoining districts. An image displaying the different regions in four or five shades of grey is created (Figure 7).

At this point the classification map image and the uniquely encoded geo-reference image, both registered to the Landsat image, can be interfaced.

A computer routine digitally overlays both images row by row and outputs histograms of data representing the number of pixels of each class in each administrative district. The columns are then sorted, aggregated, and transferred in accordance with the user's specifications and new columns listing area (in acres, etc.) and calculating percentages of each land use are created. After the uniquely encoded areas are properly labeled by their district names, a tabulation of land use by administrative district is produced. A page from the Dayton tabulation is offered in Figure 8.

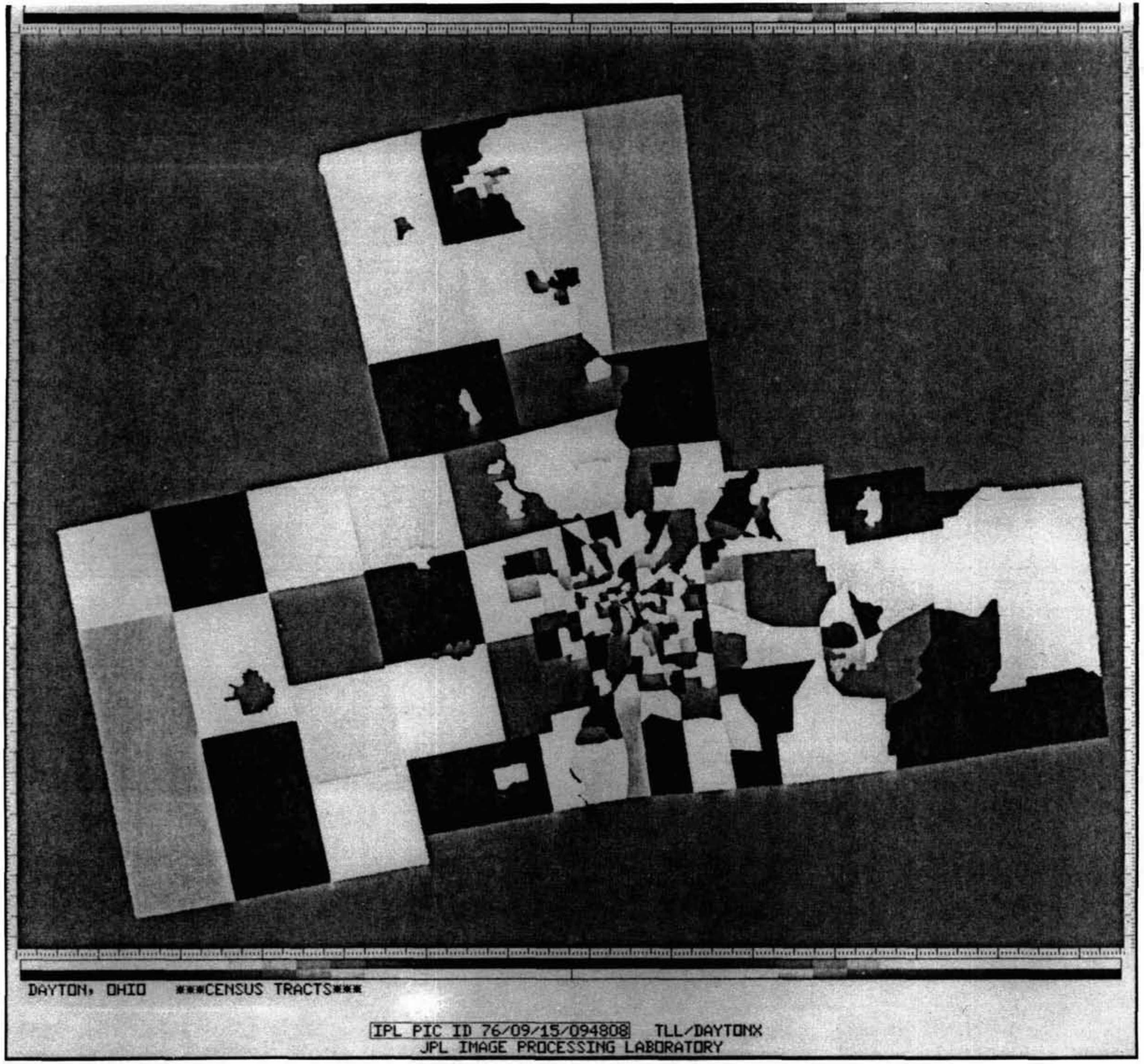

DAYTON, OHIO ***CENSUS TRACTS***
IPL PIC ID 76/09/15/094808 TLL/DAYTONX
JPL IMAGE PROCESSING LABORATORY

Figure 7

Figure 8

DAYTON, OHIO SMSA
LAND USE BY CENSUS TRACTS
NOVEMBER 1972
PRELIMINARY RESULTS

TRACT	AREA	VACANT AREA	VACANT PCT	FOREST AREA	FOREST PCT	AGRICULTURE AREA	AGRICULTURE PCT	EXTRACTIVE AREA	EXTRACTIVE PCT	WATER AREA	WATER PCT	MOBILE HOMES AREA	MOBILE HOMES PCT
1.00	127	6	5.0	6	5.0	5	3.7	0	0.0	25	20.0	13	10.0
2.00	76	6	8.3	0	0.0	2	2.1	3	4.2	5	6.3	13	16.7
3.00	131	5	3.6	1	0.6	11	9.4	0	0.0	6	4.8	3	2.4
4.00	217	3	1.5	2	0.7	21	9.5	0	0.0	0	0.0	3	1.5
5.00	223	3	1.4	2	0.7	21	9.2	0	0.0	5	2.1	6	2.8
6.00	323	35	10.8	16	4.9	44	13.7	0	0.0	0	0.0	6	2.0
7.00	312	36	11.7	6	2.0	74	23.9	0	0.0	0	0.0	11	3.6
8.00	234	8	3.4	2	0.7	16	6.8	0	0.0	0	0.0	14	6.1
9.00	331	21	6.2	8	2.4	27	8.1	0	0.0	0	0.0	30	9.1
10.00	359	9	2.6	8	2.2	21	5.7	0	0.0	0	0.0	30	8.4
11.00	201	5	2.4	0	0.0	9	4.7	0	0.0	3	1.6	6	3.1
12.00	253	28	11.3	2	0.6	22	8.7	0	0.0	0	0.0	36	14.4
13.00	286	30	10.5	11	3.9	49	17.1	0	0.0	0	0.0	27	9.4
14.00	753	93	12.4	46	6.0	82	10.9	0	0.0	3	0.4	76	10.1
15.00	390	28	[illegible]	40	10.2	62	15.9	2	0.4	21	5.3	9	2.4
16.00	149	2	1.1	2	1.1	13	8.5	0	0.0	0	0.0	3	2.1
17.00	131	8	6.0	0	0.0	2	1.7	0	0.0	6	4.8	0	0.0
18.00	55	0	0.0	0	0.0	6	11.4	0	0.0	0	0.0	0	0.0
19.00	111	0	0.0	0	0.0	17	15.7	0	0.0	2	1.4	3	2.9
20.00	188	0	0.0	0	0.8	14	7.6	0	0.0	8	4.2	3	1.7
21.00	370	15	4.4	11	3.0	46	12.4	0	0.0	0	0.0	22	6.0
22.00	348	5	1.4	6	1.8	47	13.6	0	0.0	13	3.6	8	2.3
23.00	353	71	20.2	9	2.7	40	11.2	0	0.0	0	0.0	21	5.8
24.00	547	286	52.3	5	0.9	79	14.5	0	0.0	0	0.0	8	1.4
25.00	283	25	8.9	2	0.6	17	6.1	0	0.0	6	2.2	33	11.7
26.00	419	43	10.4	8	1.8	27	6.4	2	0.4	0	0.0	36	8.7
27.00	1724	378	21.9	495	28.7	338	19.6	2	0.1	149	8.6	24	1.4
28.00	506	127	25.0	19	3.8	43	8.4	0	0.0	8	1.6	30	5.9
29.00	250	2	0.6	4	0.6	30	12.0	2	0.6	16	6.3	21	8.2
30.00	145	0	0.0	4	3.3	8	5.4	2	1.1	5	3.3	2	1.1
31.00	212	16	7.5	8	3.7	24	11.2	0	0.0	27	12.7	6	3.0
32.00	294	24	8.1	11	3.8	62	21.0	0	0.0	24	8.1	[illegible]	[illegible]
33.00	327	21	6.3	3	1.0	32	9.7	14	4.3	49	15.0	12	4.3
34.00	443	40	9.9	6	1.4	17	3.9	11	2.5	38	8.6	62	13.9
35.00	229	13	5.7	0	0.0	3	1.4	8	3.5	17	7.6	8	3.5
36.00	92	[illegible]	8.8	3	3.8	0	0.0	2	1.9	16	19.2	2	2.4
37.00	129	0	0.0	2	1.2	0	0.0	2	1.2	22	17.3	9	3.8
38.00	319	11	8.6	6	2.2	0	0.5	5	1.5	127	39.8	9	7.2
39.00	594	68	11.7	41	7.0	81	13.9	5	0.9	30	5.1	32	5.5
40.00	145	2	1.1	2	1.1	0	0.0	0	0.0	13	8.7	24	16.3
41.00	158	11	7.0	0	0.0	3	2.0	0	0.0	0	0.0	6	4.0
42.00	444	133	25.5	41	9.5	47	10.7	0	0.0	1	0.7	13	2.8
43.00	190	3	1.8	0	0.9	9	5.3	0	0.0	0	0.0	23	12.3
44.00	220	15	6.6	0	0.0	3	1.4	3	1.4	0	0.0	35	14.4
45.00	65	3	5.9	0	0.0	0	0.0	0	0.0	5	7.3	9	14.6
46.00	373	6	1.7	8	2.1	5	1.3	6	1.7	90	24.2	3	0.8

DAYTON, OHIO SMSA
LAND USE BY CENSUS TRACTS
NOVEMBER 1972
PRELIMINARY RESULTS

TRACT	AREA	RESIDENTIAL AREA	RESIDENTIAL PCT	COMMERCIAL AREA	COMMERCIAL PCT	OFFICE BUILDINGS AREA	OFFICE BUILDINGS PCT	INDUSTRIAL AREA	INDUSTRIAL PCT	INSTITUTIONAL AREA	INSTITUTIONAL PCT	OBSCURED BY CLOUDS AREA	OBSCURED BY CLOUDS PCT	UNKNOWN AREA	UNKNOWN PCT
1.00	127	15	[illegible]	22	17.3	13	10.0	14	11.3	3	2.5	0	0.0	0	0.0
2.00	76	[illegible]	12.5	24	31.3	8	10.4	6	8.3	0	0.0	0	0.0	0	0.0
3.00	131	32	24.1	62	47.0	5	3.6	2	1.2	2	1.2	0	0.0	0	0.0
4.00	217	144	66.4	38	17.5	6	2.9	0	0.0	0	0.0	0	0.0	0	0.0
5.00	223	90	40.4	62	27.7	24	10.6	9	4.3	0	0.0	0	0.0	2	0.7
6.00	323	210	65.2	11	3.4	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
7.00	312	177	56.5	6	2.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
8.00	234	153	65.5	38	16.2	3	1.4	0	0.0	0	0.0	0	0.0	0	0.0
9.00	331	218	66.0	25	7.7	2	0.5	0	0.0	0	0.0	0	0.0	0	0.0
10.00	359	267	74.4	22	6.2	2	0.4	0	0.0	0	0.0	0	0.0	0	0.0
11.00	201	[illegible]	[illegible]	25	12.6	9	4.7	3	1.6	0	0.0	0	0.0	0	0.0
12.00	254	108	42.5	46	18.1	3	1.3	6	2.5	2	0.6	0	0.0	0	0.0
13.00	286	144	50.3	28	[illegible]	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
14.00	753	361	47.5	75	10.5	11	1.5	5	0.6	2	0.2	0	0.0	3	0.4
15.00	390	158	40.7	55	14.2	11	2.8	6	1.6	0	0.0	0	0.0	0	0.0
16.00	149	101	68.1	25	17.0	3	2.1	0	0.0	0	0.0	0	0.0	0	0.0
17.00	131	55	44.6	43	32.5	5	3.6	6	4.8	2	1.2	0	0.0	2	1.2
18.00	55	27	48.6	17	31.4	2	2.9	3	5.7	0	0.0	0	0.0	0	0.0
19.00	111	51	45.7	35	31.4	0	0.0	2	1.4	2	1.4	0	0.0	0	0.0
20.00	199	157	80.7	9	5.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
21.00	370	196	53.0	38	10.3	11	3.0	9	2.6	0	0.0	0	0.0	2	0.4
22.00	344	190	54.8	62	17.7	17	5.0	0	0.0	0	0.0	0	0.0	0	0.0
23.00	353	148	41.9	28	8.1	11	3.1	5	1.3	0	0.0	0	0.0	0	0.0
24.00	547	108	19.7	9	1.7	2	0.3	0	0.0	0	0.0	0	0.0	3	0.6
25.00	283	118	40.8	58	20.7	16	5.6	8	2.8	0	0.0	0	0.0	2	0.6
26.00	419	237	56.6	44	10.5	6	1.5	11	2.6	0	0.0	0	0.0	5	1.1
27.00	1724	235	13.5	70	4.0	16	0.9	13	0.7	0	0.0	0	0.0	2	0.1
28.00	506	152	30.0	70	13.8	36	7.2	3	0.6	0	0.0	0	0.0	19	3.8
29.00	250	71	[illegible]	52	20.7	11	4.4	5	1.9	0	0.0	0	0.0	0	0.0
30.00	145	57	35.1	60	41.3	3	2.2	3	2.2	2	1.1	0	0.0	0	0.0
31.00	212	24	11.2	90	42.5	6	3.0	6	3.0	5	2.2	0	0.0	0	0.0
32.00	294	57	15.4	65	30.1	16	5.4	0	0.0	0	0.0	0	0.0	0	0.0
33.00	327	52	15.5	70	[illegible]	30	9.2	22	6.8	5	1.4	0	0.0	6	1.9
34.00	443	68	15.4	117	26.4	27	6.1	47	10.7	3	0.7	0	0.0	6	1.4
35.00	229	6	2.8	65	28.9	22	9.7	51	22.2	2	0.7	0	0.0	0	0.0
36.00	82	2	1.5	22	26.5	9	11.5	13	15.4	6	7.7	0	0.0	2	1.9
37.00	129	6	4.9	30	23.5	25	19.8	17	13.6	2	1.2	0	0.0	2	1.2
38.00	319	11	3.4	43	13.4	43	13.4	60	18.5	25	8.0	0	0.0	0	0.0
39.00	594	77	[illegible]	164	28.2	32	5.4	32	5.4	14	2.4	0	0.0	6	1.1
40.00	145	3	[illegible]	51	35.1	22	15.2	14	9.8	5	3.3	0	0.0	2	1.1
41.00	158	24	15.0	96	61.0	6	4.0	9	6.0	2	1.0	0	0.0	0	0.0
42.00	444	133	30.0	49	11.0	14	3.2	8	1.8	3	0.7	0	0.0	0	0.0
43.00	190	63	33.1	74	41.2	6	3.5	0	0.0	0	0.0	0	0.0	0	0.0
44.00	220	22	10.1	74	33.8	38	17.3	6	2.9	5	2.2	0	0.0	17	7.5
45.00	65	0	0.0	38	58.5	5	7.3	2	2.4	0	0.0	0	0.0	0	0.0
46.00	373	5	[illegible]	150	40.3	54	14.4	19	5.1	21	5.5	0	0.0	2	0.4
47.00	196	25	12.5	112	57.3	24	12.1	13	6.5	0	0.0	0	0.0	0	0.0

Change Detection

In monitoring the change in the uses of land over time, a few steps are required in addition to those performed for the creation of one-point-in-time land use maps and tabulations. These steps can be categorized under the topics of registration, differencing, classification and tabulation.

The span of time over which the change of land use is to be monitored must first be decided. A LANDSAT overflight including the geographic area of interest must then be obtained, attempting to choose a date as close as possible to an anniversary date of the initial classified imagery. This is done to minimize the spectral difference of the two images resulting only from seasonal changes and sun angle difference. The LANDSAT tapes are received, and the immediate study area is extracted to include at least as much area as that of the initial date. An enhancing contrast stretch is executed to reveal common features in the imagery for both dates. Line and sample coordinates of identical locations in the two images are determined interactively using a correlation routine which searches for spectrally similar features. The coordinate numbers are then entered into a rubber-sheet geometric rectification routine which will draw the second image into pixel-by-pixel registration with the early date. This pixel alignment is then tested by differencing the original early date imagery from the rectified late date imagery. Any misregistration is revealed as sharp edges that emerge from an otherwise gray picture.

Once the two images are adequately registered, the procedures of determining the geographic areas of change are initiated. It has been found best to compute first the ratio of band 5 to band 7 of the LANDSAT MSS for each date. The early date ratio is then subtracted from the last date ratio and a constant value added. The difference is enhanced by allowing it to conform to a Gaussian distribution. Finally, from accompanying statistics, points on either side of the mean are computed that lie two standard deviations from the mean. All areas outside of these points on the distribution are the areas in which change has occurred between the two dates. These areas show up as black on Figure 9 an August 1972 to September 1975 change detection image for the Houston, Texas SMSA.

The creation of a late date classification is simplified by using the change detection procedures just described. Only those areas that show change are digitally extracted from the second image and classified. Since the remaining areas by definition did not change, the early date classified data for unchanged areas can be added to the late date classified data in changed areas to obtain a complete second date classification.

From the revised classification, a tabulation, such as performed for the early date, can likewise display land use statistics by administrative district for the second date. To obtain a comparison between early and late date statistics, a composite tabulation will reveal general land use changes. But if the specific nature of the change (e.g., how many acres changed from rural to urban) is important, another series or steps must be instituted. First, the land use classes are aggregated into rural/urban dichotomies and then entered into a routine which permits comparison of individual pixel values. Using polygon

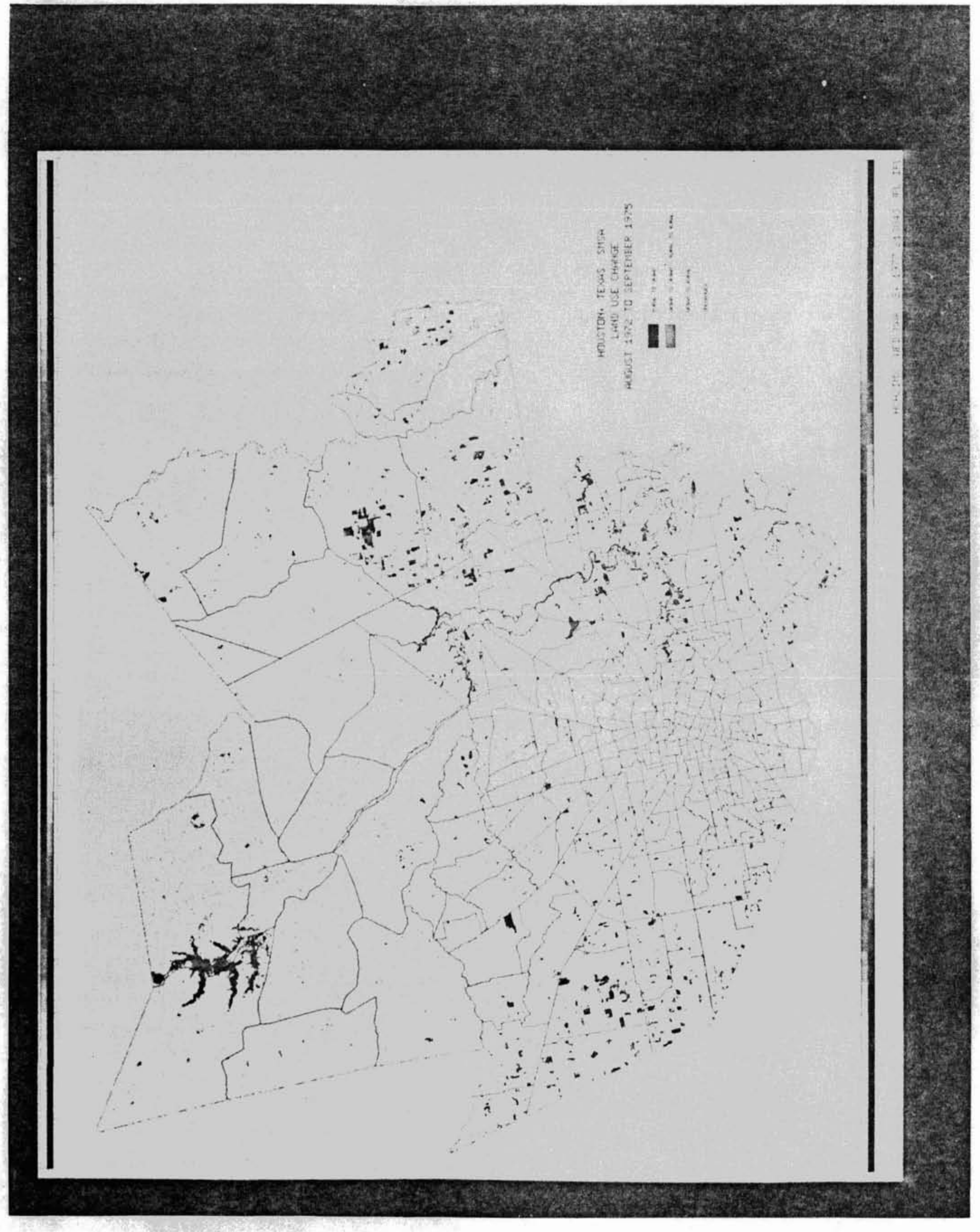
Figure 9


2249-16121 5 RED 28SEP75 HOUSTON, TEXAS
H30-18/W095-13 HDG190 SUN EL47 AZ133 PRC121 SCALE 79.98M/PXL
GEOMA - F - F
SEP75 (RED/IR2) - AUG72 (RED/IR2)
F - GAUSSIAN STRETCH - STRETCH
IPL PIC ID 77/02/02/210508 /HOUPOHYX
JPL IMAGE PROCESSING LABORATORY

overlay routines on the now classified change detection image and the geo-reference image of neighborhoods, a tabulation and image notes the dynamics of the changes that occurred between the two dates in question Figure 10).

In general, the MILUS batch processing procedure is time consuming. Work is progressing on the conversion of nearly all of the steps, presently done in batch mode, to interactive processing. This improvement would reduce the overall time involved in editing and registering of data sets from several months to a matter of weeks. With improved satellite resolution projected for the 1980's and interactive data handling capabilities based upon the procedures presented here, timely cost-effective updates of land use files can become a reality.

Figure 10


HOUSTON, TEXAS SMSA
LAND USE CHANGE
AUGUST 1972 TO SEPTEMBER 1975

Section 13

CONCLUSION

One of our initial uncertainties hinged on the question of generality. Could we build a single model that would work equally well in all six cities, or would we be forced to develop six different models? The significance of the question goes far beyond modeling technique. At issue is the cause of variation from city to city.

We know that cities differ.[1] Do they differ because the people in them behave differently or do they differ because of differing circumstances from one place to the next? It is an important question from a scientific standpoint. If we cannot assume constant behavior -- if people follow different rules in different places -- then our task of understanding how cities change will indeed be a frustrating one. The answer has practical implications as well. Uniform behavior would mean uniform response to federal and state programs, after controlling for variations in circumstances.

Based on the information we have gathered thus far, we feel that behavior is remarkably similar across cities and regions. Our interview results show little variation among Houston, Dayton, and Rochester after controlling for differences in climate, topography, and so forth. Our single model now runs with similar accuracy in all six cities. Variations from city to city can be handled through straight-forward adjustments in parameters.

---

1. For ample evidence on this point, see Coleman and Parsons, City Selection for the Joint Center Study of Neighborhood Evolution and Decline (Cambridge, Mass.: MIT-Harvard Joint Center for Urban Studies, 1975).

By implication, we have come to the conclusion that the considerable variation in urban form from place to place is largely a function of the historical development of the city, its climate, and its topography. The kinds of structures, people, and jobs, and their spatial relationships to one another, depend to a great extent on when in the past the city experienced its most rapid period of growth. Was the city built during the industrial revolution and filled by immigrants from Ellis Island? Was it built on an expanding railraod or canal and filled by western-bound migrants? Did much of its growth take place after the invention of the automobile and the highway? Was its growth influenced by the strong flow from south to north before, during, and after World War II? These things, rather than differences in the behavior of groups of people today, explain much of the variation from place to place. Today's rules are more or less the same; variations in outcome are caused mainly by variations in the environment in which the rules are being followed.

In retrospect, it would be surprising if this were not the case. Most people absorb the same information through the same media, drive the same kinds of automobiles on the same kinds of highways, borrow money from the same capital market at the same rate, consume the same goods at the same prices, receive an education from the same basic textbooks, and hire a labor force from a highly mobile labor pool. Great variations in decision rules should not be expected under these circumstances.

As the design of our model has progressed, we have capitalized on the generalizability of behavior and the uniqueness of circumstances by placing all the circumstances in datasets outside of the structure of the model and basing the model's internal workings on the more general behavioral decision rules. By so doing, we can move from city to city without changing the model's structure. All that a new user must do is fill in the "circumstantial boxes" on the outside to specify the configuration of his or her city. The internal structure remains unchanged.

As indicated in Section 12, our first major job in the months ahead is to identify more of the unique circumstances that have a strong influence on outcomes in 4 or 5 percent of the neighborhoods in most regions. The model must have information about special land holdings, flood plains, sewage restrictions, large rivers and mountains, and so forth before it can simulate neighborhoods where such things exist.

Another thrust in the future will be in the area of current validation. We plan to work extensively with satellite data, Polk data, and data from the Dun and Bradstreet file as well as locally generated sources, to validate the model through the present on all fronts, and to develop the capability to move beyond the 1980 census after it is compiled.

As the model takes even greater account of local anomalies and demonstrates its robustness through extensive validation tests, we will be putting it to work evaluating policies and translating results into decision-making terms for local planners and administrators. In the area of program testing, we will be examining strategies of rehabilitation, demolition, job creation, and restricted land development.

Plans are already underway to translate model results into terms of direct relevance to public officials. For example, we will estimate the line items in a city budget based on neighborhood change, changes in federal and state revenue sources, and actions taken by the mayor to control costs. We will estimate the likely incidence of fire, and evaluate steps that might be taken to prevent it. We will predict school enrollments by grade and school system. We will anticipate fare revenues in a public transit system. We will estimate the likely demand for health services, by type of service and neighborhood. We have already developed and tested such "translators," as we call them, for estimating energy consumption by fuel source and use, and for predicting bank deposits and demand for mortgage credit. Once developed for one city, a translator may be used (after suitable recalibration) in any other because the inputs (the model's outputs) are the same everywhere. Only the local circumstances vary.

During all of this application and testing, we will constantly be evaluating and gradually be modifying the Community Analysis Model's internal workings. While the present version runs accurately enough for many purposes, it is still far from perfect. As we clean up the remaining local anomalies, we will be able to see more clearly where and why the behavioral decision rules are not defined sharply enough. If the future is anything like the past, the insights gained through this process should lead to an improved set of rules, a better theory, and eventually another generation of CAM. Model building in the CAM tradition is a learning process, not the production of a "canned" item. The model never sits still. It evolves with our understanding of the decision rules and the information flow that drives them. The basic framework is now in place and we are in a good position to build on it in the future.

☆ U.S. GOVERNMENT PRINTING OFFICE : 1979 O—285-193

U.S. DEPARTMENT OF HOUSING
AND URBAN DEVELOPMENT
WASHINGTON, D.C. 20410

OFFICIAL BUSINESS

PENALTY FOR PRIVATE USE, $300

FILE COPY

DEPARTMENT OF HOUSING AND
URBAN DEVELOPMENT
POSTAGE AND FEES PAID
HUD-401

U.S.MAIL

January 1979
HUD-PDR-363-2