APPLICATIONS OF MICRO-LEVEL DATA SYSTEMS IN URBAN ANALYSIS

Eric G. Moore, Queen's University

1. INTRODUCTION

The emergence of new data systems presents a challenge to produce appropriate methods and models for their effective use. While this statement might offend some logical positivists, it is evident that we have always been influenced by the form in which data are available not only in terms of the analytical methods we employ but also in the ways we think about problems. This has been particularly noticeable in urban analysis where the dominance of ecological studies owes much to the existence of census data by tract and, for similar reasons, discussions of dynamics embrace decades rather than years.

In this paper I discuss the potentialities of a type of data source which recently has been receiving increased attention, namely the annual population enumeration in which basic demographic data are augmented by information on socio-economic and housing characteristics (see, for example, the proceedings of the first five European Symposia on Urban Data Management). Variants on this theme are found in a number of European cities and Wichita, Kansas, the site of our own research, provides the prototype in North America. In evaluating such a system, we need first to consider the types of uses to which such data can be put and then to examine the nature of the benefits which accrue to these applications.

Most uses of urban data systems fall into three broad categories:

i) administrative functions dependent upon routinely gathered data;
ii) planning activities of both an ameliorative (adaptive) and a control nature;
iii) attempts to improve understanding of the ways in which individual cities function with a view to developing better theories of urban processes.

In any case study the applications that can be developed depend to a large degree on the local institutional context. At the heart of the Wichita system lies the population enumeration which is required by Kansas law for assessment purposes. This administrative need drives the data collection process; the resulting limitations on the number and nature of questions asked reflect a sensitivity to potential public reaction which might endanger the quality of the assessment. Under these circumstances we must expect to make compromises and to exercise ingenuity in our planning and theory-building activities for external constraints prevent the design of ideal questions for all needs.

Despite limits on the range of variables, the content of the Wichita system has such strong advantages that we believe it provides a major advance over other operational systems. These advantages stem from three separate characteristics of the data files:

1. access to data on a wide range of variables describing each dwelling unit and its occupants. While confidentiality is maintained, access to the individual records promotes a great deal of flexibility in the construction of detailed cross-classifications.
2. production of a complete enumeration every year. Critics have pointed to the inadequacy of the decennial census in documenting the rapidity and variety of change in our cities, yet we seldom possess alternative sources to indicate the appropriate time frame for monitoring such change. The annual enumeration provides a starting point for measuring the rates of change of different components of the urban system and for determining the most suitable monitoring period for each component.
3. an ability to link individual households and dwelling units over time. Since transactions or changes of state are the basic observations we need to document process, this characteristic is potentially the most important of the three. However, taking advantage of this attribute presents the most challenging methodological problems and relatively little progress has been made to this point.

These three characteristics are hierarchical in that we assume in our discussion that the annual enumerations each contain individual records and that longitudinal files link individual characteristics from one year to the next.

In the remainder of this paper, we use these three data characteristics as a framework for discussing applications of the Wichita system. For a number of reasons, the comments focus on the third category of use, that of improving our understanding of how cities function. Although they form the foundation for development of the system, administrative functions are routinized and of little relevance to our own interests. At the present time, planning applications are of a simple nature mainly comprising an annual 'state-of-the-city' report which serves as useful background information rather than as analytical input for various public agency programs. Failure to utilize the properties of the data files to any greater extent underscores the general lack of experience in using these types of data in the profession as a whole. Methods of handling such data, their presentation and use in appropriate models of urban process are poorly developed and we see it as part of our task to explore the possibilities in these areas. In particular, we need to develop models which relate to the instruments which planners have at their command such as the zoning laws, directed subsidies, mortgage rates, and location of public facilities. In the first instance, however, we need to attain a better understanding of the nature of processes around which we can build future applications.

Much of our initial research attention has focussed on patterns of housing consumption and residential mobility within Wichita and I will use this work to illustrate our arguments. It provides a good case study in that much current planning concern focusses upon the housing market; problems, such as the abandonment of dwellings or the decline in school enrollment, develop quickly, are usually localized within the city, and they must be resolved within a reasonable time frame. Our understanding of the underlying processes is weak; census data are too coarse in both spatial and temporal detail and survey data generally provide too few observations to form a basis for tackling these issues. We therefore need to turn to alternative sources of data.

2. THE VALUE OF MICRO-LEVEL DATA

Of itself, the existence of detailed micro-level records for households and dwelling units is not exceptional. Indeed, the Census obtains far more detailed data on individuals each decade than are collected in Wichita. However, it is the ability to readily access these individual records which increases the value of the system for one thereby gains a great deal of flexibility in terms of output format. This attribute, coupled with the size of the enumeration which generates approximately 100,000 records per year, each with data on 25-30 household and dwelling characteristics, permits ideas to be examined in considerable depth.

In considering the value of these data, we make three points: (i) they permit construction of detailed cross-classifications which lead to more policy-relevant descriptions of certain behavior; (ii) with large data files, more complex models of behavior can be constructed, and (iii) the flexibility of aggregation does not tie us to pre-specified classifications. Although it is clear that each of these attributes could equally well be attained from existing census data, it is evident that the latter are not being used in this way.

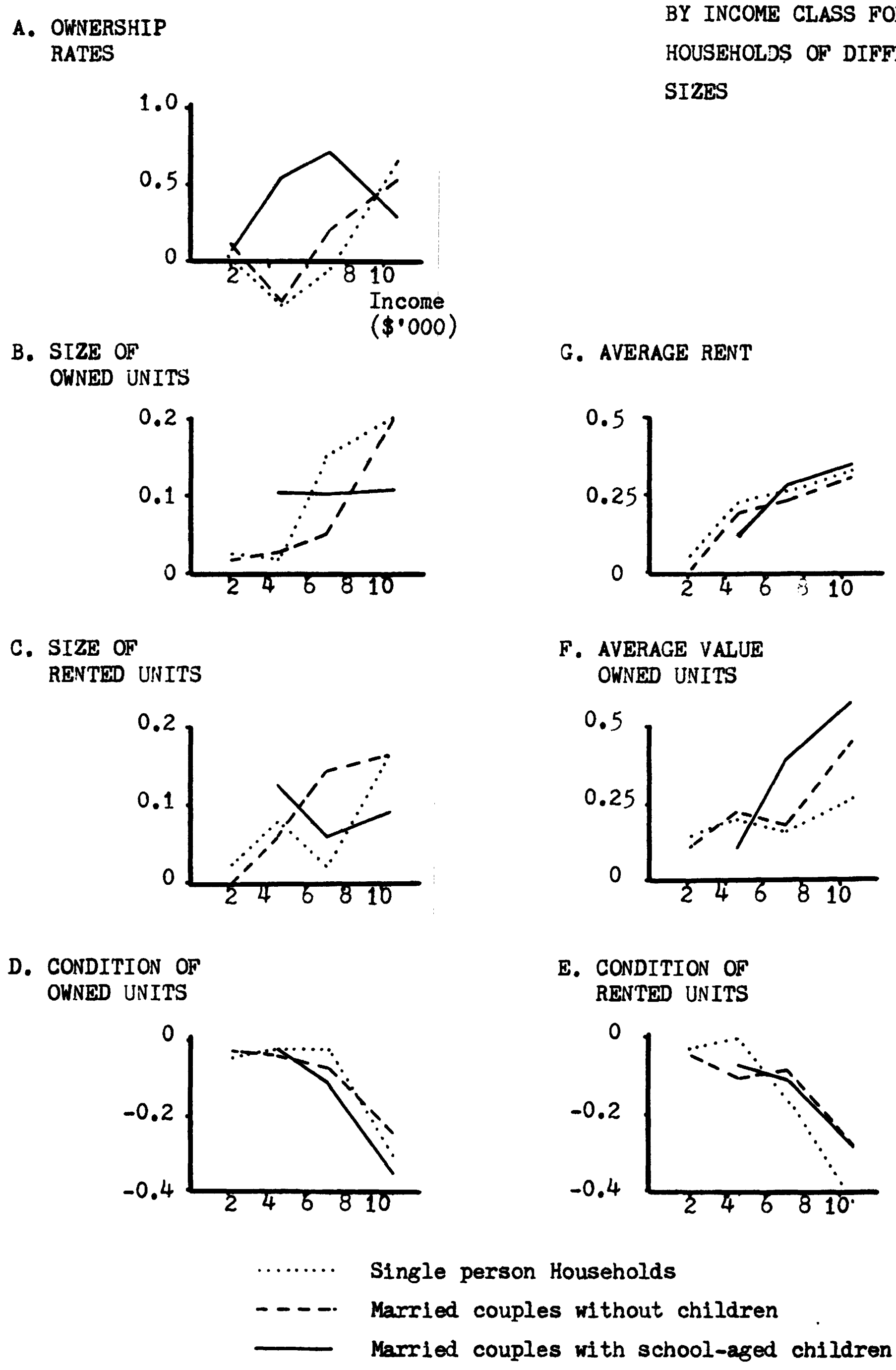
Example 2.1: Existing housing studies have stressed the importance of measuring income elasticities of demand for various dwelling characteristics (tenure, size, condition, etc.). It is argued that such measures contribute both to an understanding of consumer behavior and to the assessment of impacts of public policy on the housing market (de Leeuw, 1971). Although these values have usually been computed in aggregate terms for large populations, it is clear that if policies are to be designed to alleviate existing inequities, they must be directed at specific segments of the market. In these circumstances, we need to know more about income elasticities than can be acquired from aggregate analysis yet few data are available to provide such detailed estimates (notable recent exceptions are found in Struyk and Marshall, 1974 and Straszheim, 1975). Partly, in response to this situation we estimated income elasticities of demand for tenure, size, value, rent, and condition for 31 different household types over four income ranges for both recent movers and the total population of Wichita in 1973 (Moore et al., 1975).

For the purposes of this paper, it is perhaps most useful to look at a number of conclusions which arise from this work:

1. elasticities vary not only from one population subgroup to another but also vary for different income ranges for the same household type (violating the oft-adopted assumption of the Engel curve). This situation is clearly illustrated for different-sized households in Figure 1.
2. the functional form of the relationship between income and elasticity is different for different household types (for example, in the case of tenure it increases with increasing income for households with several pre-school children whereas it decreases for elderly couples whose children no longer live at home).
3. from a policy standpoint, the consistently low values of demand elasticities for low-income families with children suggests any subsidy program designed to improve housing conditions for this group should be specifically tied to housing expenditure rather than be in the form of a general income subisdy (although it cannot, of course, address the more general policy issue of whether such a goal is desirable (de Leeuw, 1974)).

Of course, many questions are left unanswered by a single study such as this. Perhaps the most important concerns the effect of local supply conditions on elasticity measures. Some progress could be made in this regard by replicating the study in successive years but it would seem that a more valuable strategy would

FIGURE 1: VARIATION IN ELASTICITIES BY INCOME CLASS FOR HOUSEHOLDS OF DIFFERENT SIZES



be to employ existing micro-level census data in a wide range of urban environments. The outcome would have particular value in suggesting differences between cities in their response to alternative housing programs.

Example 2.2: Closely related to the previous example are those instances in which we need to impose controls on a study design to test specific hypotheses. A typical case concerns the detection of racial discrimination in the housing market. Do blacks have lower propensities to own, do they pay higher prices than whites for similar housing, and so forth? How do we compare the two groups such that we do not confound income, age, and family structure differentials in our evaluation? Two studies (Altman and Tam, 1974 and Tam, 1975) have examined differences in rental payments and ownership rates respectively. Both studies concluded that the degree of racial difference was noticeably smaller than had generally been assumed in the literature once the population differences had been controlled. An example of these differences is provided in Figure 2 which identifies racial variations in ownership rates for different socio-economic and family types.

Example 2.3.: Once we address problems at this level we realize that other traditional procedures do not sit quite so comfortably. Can we continue to utilize simple linear models which ignore interactions between independent variables? Can we assume that the same functional forms are appropriate for modelling a given behavior over different household types? When we recognize these as important issues, what methods are appropriate to answering these questions? Since much of the data contained in the enumeration is either nominal (sex, migration experience, structure type) or ordered categorical (number of rooms, income class, rent category), we have been able to make extensive use of recent developments in multidimensional contingency table analysis (particularly the material contained in Fienberg et al., 1975). In examining the structure of ownership rates, Tam (1975) applied these procedures in logit form, from which he was able to identify a number of important interactions, particularly that between income and household size, which contribute both statistically and interpretively to explanation of rate differentials. The more difficult issue is whether the contribution they make is worth the additional effort in data collection and analysis. Certainly the differences between estimates which include and exclude the interactions are statistically significant; in absolute terms the differences are small (often less than ten percent) and whether they are important depends on the particular context in which they are being computed.

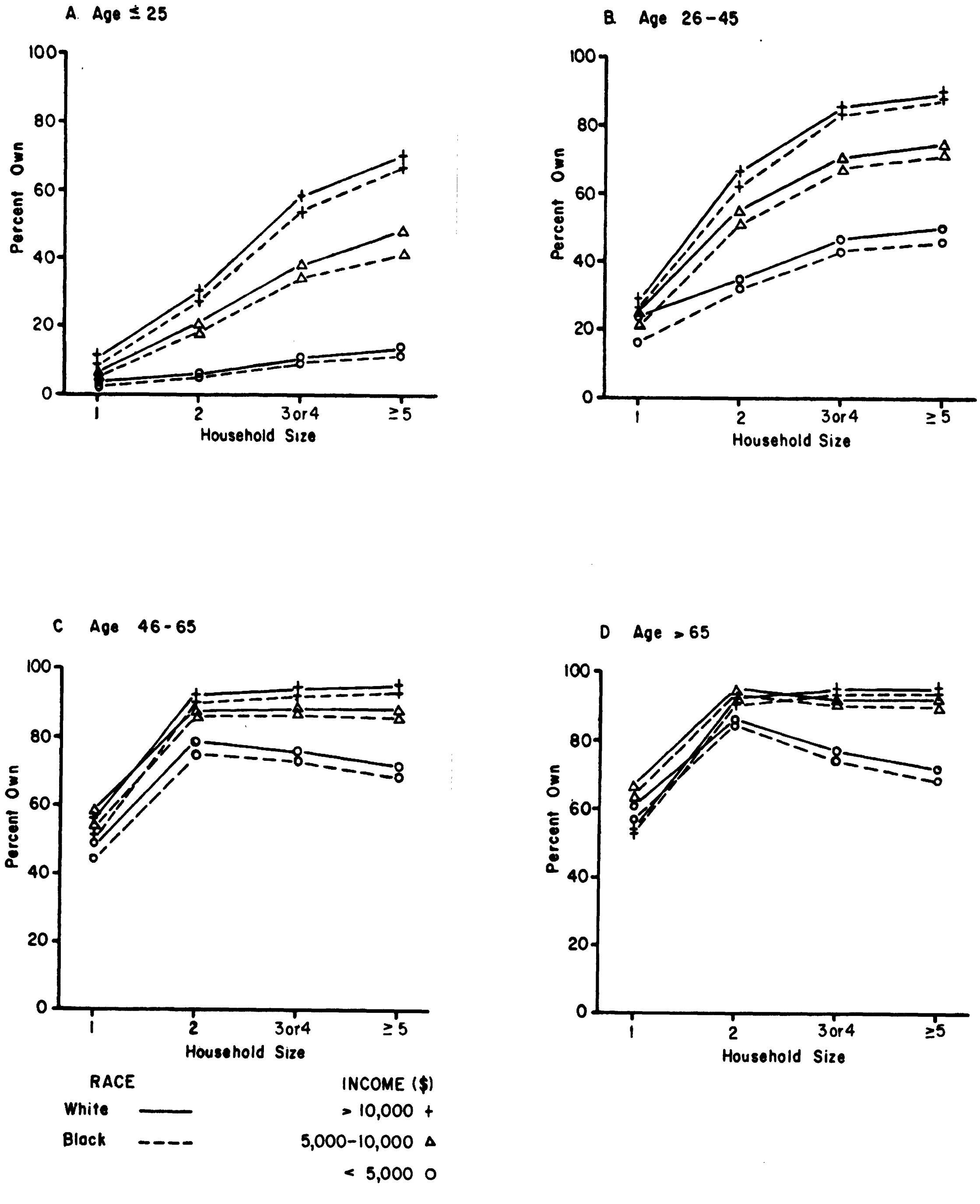
Example 2.4: Implicit through all of the above comments is that we are able to maintain considerable flexibility in the construction of classifications. Access to the individual records in the estimation of elasticities, for example, permits us to combine individual data on age of household head, age and number of children, and number of other household members to produce complex life cycle classes. In addition, we are not tied to any one *a priori* grouping, a fact which is particularly important in the spatial treatment of certain problems. We are not bound by specific tract or block group assignations which makes the data much more amenable to direct manipulation for questions such as those concerning access to facilities, school redistricting or selection of bus routes. It is well known that analytical results are sensitive to the classes we choose and therefore we need to retain this flexibility to ensure that our inferences are not subject to avoidable bias.

3. THE BENEFITS OF AN ANNUAL ENUMERATION

Possession of a sequence of enumerations stimulates questions concerning the nature and rate of change from a comparative static perspective. The simplest procedures, identifying changes in selected population and housing characteristics from one year to the next, in fact constitute the most valuable output from the system as far as local agencies are concerned. The school board can obtain up-to-date tabulations of the distribution of pre-school children, the urban renewal agency can identify any marked shifts in vacancy rates in the housing stock and the advance plans department can monitor its own short-run population projections. Most of the computations are straightforward and much use is made of standard graphic packages such as SYMAP. Although these outputs are not fed into any formal models of change they provide invaluable information on both the current state and recent changes in the system which can serve as background material for many local decisions.

A second set of questions concerns the short-run stability of relationships between variables. A wide variety of models in the areas of housing, transportation, and delivery of social services assume either implicitly or explicitly that calibrated parameters are stationary; perhaps even more fundamentally they assume that the structure of the relationships between the variables remains the same. Although the identification of instability in these attributes of our models perhaps tells us more about their misspecification than about the variability

FIGURE 2: RACIAL DIFFERENTIALS IN OWNERSHIP RATES: MALE-HEADED HOUSEHOLDS, 1971



of the real world, we must make use of all available information in an area in which theory is so weak.

Example 3.1: Our continuing interest in residential mobility stems from a belief that this phenomenon lies at the root of many changes in neighborhood characteristics. In attempting to build a model of the propensity to move which is relevant to discussions of neighborhood change we have placed a strong emphasis on occupancy characteristics of dwellings. We thereby recognize a need to consider the type of housing that is being consumed by a household as well as life cycle attributes which, together with tenure, form a dominant focus of previous research (Rossi, 1955; Speare, 1970).

We begin with a simple model based on the 1971 and 1972 enumerations: a four-way table was constructed using the following variables: [1] - movement propensity (1 = move; 2 = not move); [2] - tenure; [3] - size of household; [4] - size of dwelling (number of rooms). Program ECTA, which applies the procedures developed by Goodman (1971) for fitting models to multi-way tables, was used to identify that a significant interaction exists between size of household and size of dwelling and movement propensity (a [124] interaction in the model terms). The structure of the element-by-element effects in this interaction suggests a marked pattern of increased movement propensities in both undercrowded and overcrowded situations. In the present context, the issue is whether this structure is maintained over time. Applying a similar analysis to data from the 1972-3 enumerations we are able to show that, although the specific parameter values (the effects) exhibit some variability, both the sequence in which different models are fit to the data and the pattern of the interaction effects are the same in both years.

Identification of such stability suggests that we have a valid basis for estimating mobility rates in this context (although certainly not the only one). The question still arises as to whether we need to add further variables in the specification of the model to account for the variation in parameter values but this seems to be more a pragmatic than a theoretical issue.

4. THE LINKAGE BETWEEN SUCCESSIVE FILES

Documentation of successive states of a system is not sufficient to provide the analytic base for policy decisions. If we are to identify the leverage points at which policies are most likely to be effective, we need to know more about the components of change, who participates and under what circumstances. In order to achieve this level of understanding we need to be able to follow specific individuals through the system, to identify the transactions in which he partakes; in other words, we need a longitudinal file.

Since we are concerned with the components of change in small residential areas, a logical place to start using the linkage properties of the data is in the context of demographic accounting models (see, for example, Stone (1971), Rees and Wilson (1973)). The basic accounting model used in our studies (Gale and Moore, 1972; Moore and Gale, 1974; Moore, 1975) is specified as follows:

$$\underline{a}(t+1) = (\underline{a}(t) - \underline{d}(t))[(I-R(t))F(t) + R(t)G(t)] + \underline{n}(t)$$

where

- $\underline{a}(t)$ is the s-element vector of frequencies at time t in each occupancy class (these classes being defined by the assignment of a household with given characteristics to a dwelling with specific characteristics)
- $\underline{d}(t)$ is the vector of decrements to these occupancy frequencies due to demolitions in the stock
- $R(t)$ is the s x s diagonal matrix of movement rates for each occupancy class for the period t - t+1
- $F(t)$ is the s x s matrix of conditional transitions rates between occupancy classes for dwellings not experiencing household relocations
- $G(t)$ is the corresponding transition matrix for dwellings receiving household relocations
- $\underline{n}(t)$ is the vector of occupancy increments due to new construction.

R(t), F(t) and G(t) contain the parameters to be estimated for the accounting model. We might note at the outset that, given the structure of the files, all estimates of these values are prospective since we can measure the rates of occurrence of these events between t and t+1, given the characteristics at time t; this is an attribute that cannot be achieved with retrospective histories.

Although we have ideas as to the appropriate structure of R(t) in some circumstances, we have few a priori notions as to the nature of F(t) and G(t) and must proceed in an empirical vein. In experimenting with this structure we have focussed on the problem of changes in overcrowding in the housing stock at the neighborhood level, using as an example, the individual changes in occupancy for two areas in Wichita from 1971-1973 (Moore, 1975).

Three separate issues are involved

in using this structure: (i) estimation of the parameters R(t), F(t), G(t); (ii) determination of the degree of stability in the parameters over time, and (iii) an evaluation of the consequences of any observed non-stationarities in the parameters. Table 1 provides estimates of the parameters for one neighborhood near the center of the city in 1972-3. We can see that the dwellings not experiencing household relocations play as important a part in change in this area as the movers. Furthermore, although overcrowded units are more likely to experience a household move, they are also more likely than non-overcrowded units to be overcrowded after a new family moves in; this latter property tends to reduce the overall effect of mobility in decreasing overcrowding. Such behavior strongly suggests that supply constraints are serving to restrict the housing choice for large (and probably low-income) families.

Attempts to test for parameter stationarity reveal a complex situation with some elements (notably F(t) appearing stationary and the remainder fluctuating from one year to the next. Unfortunately, with large numbers of observations even small differences are statistically significant which led us to look at other attributes of the parameter shifts. The matrix $[(I - R(t))F(t) + R(t)G(t)]$ was replaced by the single matrix P(t). The latter was estimated separately for 1971-2 and 1972-3, the resulting matrices being labelled P(1) and P(2). In two separate simulations, P(1) and P(2) were repeatedly applied to the occupancy vector for 1971 to generate two time paths for the neighborhood. Although there is certainly a change in the long-run frequency of overcrowded rental units resulting from parameter shifts between the two years, the numerical difference is small and the overall structure of local housing conditions remains the same.

The outcome of this preliminary analysis is to suggest further lines of research. We need to clarify our ideas with respect to the nature of parameter shifts which are significant to particular local situations. One obvious strategy is to experiment with different hypothetical forms for the parameter matrices to determine the degree of variation in different elements necessary to produce overall structural change. Once this has been accomplished we can then address ourselves to the question of what types of action are likely to produce that level of change in the parameters.

5. CONCLUSIONS

We still have a great deal to learn about the use of micro-level data in general and longitudinal files in particular, even in the restricted field of housing analysis. In using files such as these, the gaps in our knowledge of micro-level behavior are rapidly exposed. Even the simple act of documenting events contributing to change, as in our overcrowding example, provides insights as to the relative importance of different processes which we have not been able to obtain before.

We therefore see the gap between development of process-oriented models and applications as still being considerable. For some time to come, the main applications of the files will be two-fold: (i) the provision of up-to-date information on single variable distributions as background to decision making, and (ii) the use of the files as a detailed sampling base for in-depth surveys. More sophisticated use depends first on the identification of the nature of specific processes and second on the integration of policy instruments into models of these processes.

REFERENCES

Altman, N.D.E. and Y.K. Tam (1974), An Analysis of Price Discrimination in the Rental Housing Market: Wichita, Kansas - 1971, Working Paper No. 5, PRMCCR (Program of Research on Metropolitan Change and Conflict Resolution, Peace Science Department, University of Pennsylvania).

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland (1975), Discrete Multivariate Analysis: Theory and Practice, M.I.T. Press, Cambridge.

Gale, S. and E.G. Moore (1973), A Research Program for the Description and Examination of Occupancy Shifts and Neighborhood Change, Working Paper No. 1, PRMCCR (Program of Research on Metropolitan Change and Conflict Resolution, Peace Science Department, University of Pennsylvania).

Goodman, L.A. (1971), "The Analysis of Multi-dimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications," Technometrics, 13, 33-61.

de Leeuw, F. (1971), "The Demand for Housing: A Review of Cross-Sectional Evidence," The Review of Economics and Statistics, 53, 1, 1-10.

________ (1974), What Should U.S. Housing Policies Be? Working Paper 729-1, The Urban Institute, Washington, D.C.

Moore, E.G. (1975), The Dynamics of Overcrowding, Working Paper No. 8, PRMCCR (Program of Research on Metropolitan Change and Conflict Resolution, Peace Science Dept., Univ. of Pennsylvania).

Moore, E.G. and S. Gale (1974), "Some Dynamic Indicators of Neighborhood Change: A Case Study," Proceedings of the American Statistical Association: Social Science Section, St. Louis 1974, 381-6.

________, N.D.E. Altman, S. Clatworthy, and Y. Tam (1975), The Effect of Income on Housing Consumption by Household Type, Technical Paper No. 4, PRMCCR (Program of Research on Metropolitan Change and Conflict Resolution, Peace Science Department, University of Pennsylvania).

Rees, P.H. and A.G. Wilson (1973), "Accounts and Models for Spatial Demographic Analysis 1: Aggregate Population," Environment and Planning, 5, 1, 61-90.

Rossi, P.H. (1955), Why Families Move, The Free Press, Glencoe.

Speare, A. (1970), "Home Ownership, Life Cycle Stage and Residential Mobility," Demography, 7, 4, 449-58.

Stone, R. (1971), Demographic Accounting and Model Building, OECD Technical Reports in Education and Development, Paris.

Straszheim, M.H. (1975), An Econometric Analysis of the Urban Housing Market, Urban and Regional Studies No. 2, National Bureau of Economic Research, New York.

Struyk, R.S. and S. Marshall (1974), "The Determinants of Household Home Ownership," Urban Studies, 11, 289-299.

Tam, Y.K. (1975), "Racial Differentials in Homeownership Rates," Working Paper No. 11, PRMCCR (Program of Research on Metropolitan Change and Conflict Resolution, Peace Science Department, University of Pennsylvania.

ACKNOWLEDGEMENT

The support of the National Science Foundation Grant GS-39387 and the assistance of the Wichita-Sedgwick County Metropolitan Area Planning Department are gratefully acknowledged.

TABLE 1. PARAMETER ESTIMATES FOR BASIC ACCOUNTING MODEL, 1972-3.

	State	1	2	3	4	5	6
(i)	M(1972)						
1.	Owned - Overcrowded	.103	0	0	0	0	0
2.	- Not Overcrowded	0	.099	0	0	0	0
3.	Vacant	0	0	.591	0	0	0
4.	Rented - Overcrowded	0	0	0	.488	0	0
5.	- Not Overcrowded	0	0	0	0	.484	0
6.	Vacant	0	0	0	0	0	.591
(ii)	F(1972)						
1.	Owned - Overcrowded	.765	.193	0	.035	.007	0
2.	- Not Overcrowded	.014	.948	0	.001	.038	0
3.	Vacant	0	0	1	0	0	0
4.	Rented - Overcrowded	.138	.025	0	.750	.088	0
5.	- Not Overcrowded	.003	.098	0	.028	.871	0
6.	Vacant	0	0	0	0	0	1
(iii)	G(1972)						
1.	Owned - Overcrowded	.091	.394	.273	.030	.212	0
2.	- Not Overcrowded	.051	.394	.263	.051	.240	0
3.	Vacant	.190	.810	0	0	0	0
4.	Rented - Overcrowded	.018	.064	0	.170	.556	.193
5.	- Not Overcrowded	.006	.055	0	.088	.671	.179
6.	Vacant	0	0	0	.173	.827	0

All zeros are logical zeros.