CHAPTER IX TN 13: STATISTICAL PROJECTIONS THAT GO BEYOND PROJECTIONS OF PAST TRENDS

By J. Beaman

ABSTRACT

This paper deals with one aspect of making projections: the use of quantitative models that allow one to consider certain types of social change. The general idea behind making projections using an equation like the one following is explained:

$$\text{Probability of participation in an outdoor recreation activity} = \text{General mean} + \text{Income effect} + \text{Age effect} + \text{Family Composition effect} + \text{Urbanization effect} + \text{Error}$$

The author discusses the numerous considerations that relate to using such an equation. These are: (1) the accuracy of the estimates that are made; (2) the problems that arise because of the inadequacy of models due to the existence of interaction effects; (3) the need to consider supply in making estimates by introducing a further term into the equation that reflects amount of supply; (4) the need to consider time budget and purpose of participation factors; (5) the failure of the model to "work" where spontaneous change is involved as opposed to change that is either captured in demographic trends or trends in the model coefficients.

In discussing these points the author refers to a number of research projects which have clarified such matters as what the accuracy of certain projections is, how important interaction effects are, how supply may be considered in using the kind of equation specified.

In concluding portions of the article the author acknowledges the relevance of "futures research" techniques such as the delphi method. There is an attempt to show what the relative role of the quantitative method endorsed should be versus (1) simple trend projection and (2) "qualitative" futures research techniques.

As a general guide the author proposes that trend extrapolation is good for one to three year projection. Even though the regression model forecasting technique (see TN 12, TN 6) can be used for short range (1-3 years), it is not suggested because its use is more complicated than trend extrapolation. The analysis of variance technique with very good models is appropriate to be used for "stable" activities for up to 20 years. When forecasts are of what will happen more than 20 years in the future or when "unstable" activities are considered it is proposed that del-phi forecasting be used in conjunction with analytic methods.

PURPOSE

A purpose of this paper is to discuss the making of projections using models that have some projective efficacy because they tap "demographic" components of trends.

AN INTRODUCTION TO THE PROJECTION STRATEGY ENDORSED

As indicated later, the author acknowledges the limitations of classical quantitative methods. Therefore, In what follows, the reader should not get the idea that the author is only endorsing the use of "quantitative" methods. The relevance of work cited in Linstone and Turoff (1970) and other articles and books are seen as very relevant to one aspect of the problem which is being considered, for example, Mitroff and Turoff (1973), Churchman (1971), Turoff (1970) and Greenall (n.d.). This paper divides "standard quantitative" projection techniques into two large classes: (1) ad hoc, or the use of a trend line extension from existing historical or time series observations, and (2) structural., or the use of explanatory regression equations that take into account age, income, family size, etc. as variables causally related to recreation participation.

For example, quantitative analysis methods associated with the use of structural models that "tap" the influence of socio-economic factors on specific groups allow one to take into account demographic trends such as changing age composition of a population, changing sex composition and changing urbanization.

The values in Figures 1 and 2 are the coefficients necessary for defining equations for predicting participation in hunting and snowmobiling. Figure 1 shows that there is a strongly negative relationship between urbanization and participation in hunting. One sees that age effects (shown above the label AGE on the Y-axis) decreases from about .13 for males around 20 years of age to -.14 for males 65+. In Figure 2, one can note similar age effects on snowmobiling. Information on the data used in deriving the coefficients for Equation 1 that are plotted in Figure 1 is provided in TN 12. Literature on the use of such equations in recreation research is also cited there.

FIGURE 1: RELATIONSHIP BETWEEN PARTICIPATION/NON-PARTICIPATION IN HUNTING AND SELECTED SOCIO-ECONOMIC CHARACTERISTICS (FOR MALES)

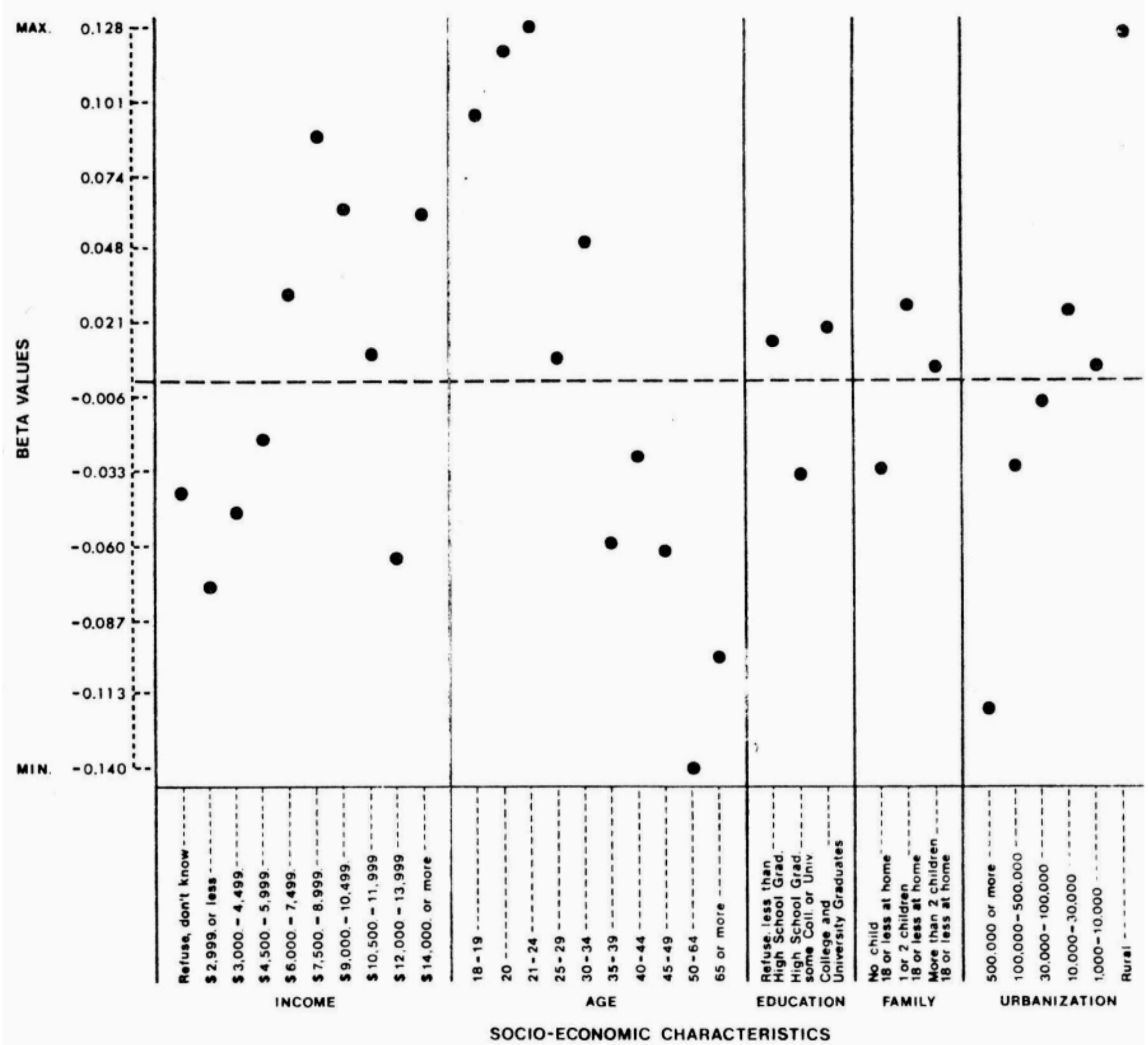
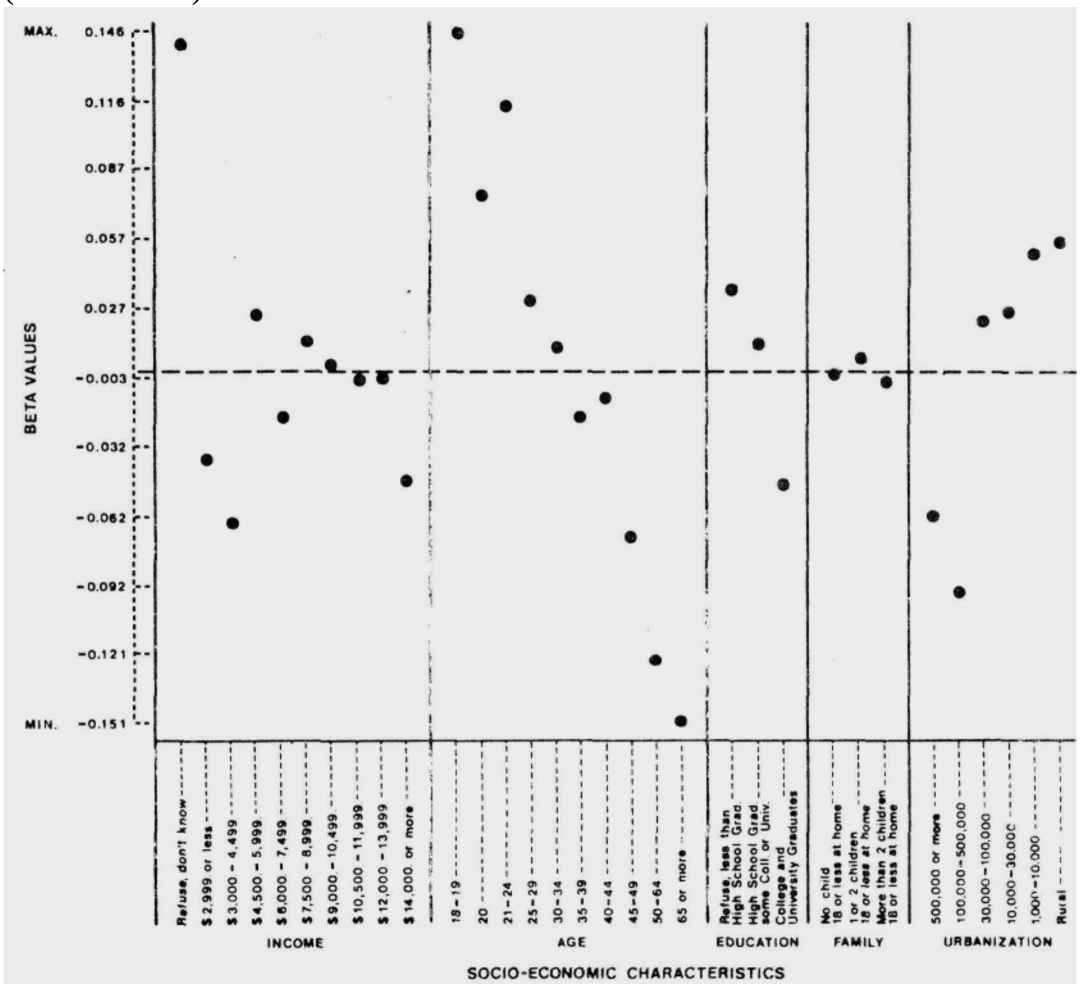


FIGURE 2: RELATIONSHIP BETWEEN PARTICIPATION/NON-PARTICIPATION IN SNOW MOBILING AND SELECTED SOCIO-ECONOMIC CHARACTERISTICS (FOR MALES)



In words, the equation developed in TN 12 is based on the idea that:

Probability of participation in an outdoor recreation activity	=	General mean	+	Income effect	+	Age effect	+	Family Composition effect	+	Urban-ization effect	+	Error

As described in TN 12, it is possible to go from such an equation for an individual to an equation for predicting participation for a population:

(1) $NP = UT + \Sigma_J B(1,J)n(1,J) + \ldots + \Sigma_Q B(S,Q)n(5,Q)$

WHERE NP = the number of participants in an outdoor recreation activity in the population ,

U = the general mean for an activity,

T = the number of persons in the population considered,

B(1,J)...B(5,Q) = effect of being in a category J...Q of socio-economic variables 1...S,

n(1,J)...n(5,Q) = the number of persons in the population considered who are in a category J...Q of socio-economic variable 1...5.

It is important to note that the "urbanization" effects shown in Figure 1, when used in Equation 2, suggest that an increase in urbanization will result in a decreasing participation in hunting even though the effects (coefficients) shown graphically in Figure 1 may remain constant over time. Specifically, Equation 2 indicates that the number of people participating in hunting in a given year in Canada depends on: (1) the population of Canada in the given year; and, (2) the number of people in various socio-economic groups.

Other shifts, beside urbanization, in the population include aging of the population and increasing education levels. In Figures 1 and 2 the effects shown above the headings Age, Education and Urbanization suggest that being old, highly educated and/or being in an urbanized area are associated with negative effects on hunting participation. Thus three demographic shifts result in an increase in the number of people having negative differentials for participation in hunting. These demographic shifts in the population might "explain" an observed decrease in the number of hunters.

Total Population is the only variable in which there is a change which raises the total number of hunters. However, since Canada's population growth rate is approaching zero, there is good reason to believe that in ten years, with another ten per cent urbanization in Canada, there will be a decrease in hunting participation larger than the -3.59% and -5.5% "decline" rates for hunting obtained from the Canadian Outdoor Recreation Demand (CORD) Study data (see TN 22). Percentage growth rates are the average annual change in the per cent participating per year. For example, percentage growth rate 1967 to 1969 = 100 ((14% participating in 1967 - 13% participating in 1969)/ (2 years from 1967 to 1969))/14% participation in 1967 = 3.5%.

In the context of the kind of change in hunting participation described above, one can "understand" what is happening when trends in certain activities are observed. Without either going to a Delphi panel or having the results of several surveys providing data for a trend line, it is possible that reasonable and reliable predictions can be made. In fact, an important point is that the trends in rates obtained from surveys may indicate a certain rate of increase or decrease in participation in an activity that is actually reflecting the way that changes in the age distribution, urbanization and income, etc. of a population are interacting to produce rates that are observed in the years considered.

When a projection is made in which socioeconomic characteristics are independent variables, one need not rely on drawing trend lines through points that are determined by complex interactions of demographic factors. Rather, one can take into account the changes in socioeconomic variables that may be expected to occur in the future. And, in this regard, it is important to note that socio-economic changes may be projected by accepted "sound" methods or projections are sometimes even available from central statistical organizations (e.g. Statistics Canada) so it is possible, with some confidence and in some cases little effort, to see what future participation in recreation activities might be.

SOME CONSIDERATIONS CONCERNING THE USE OF THE PROPOSED PROJECTION STRATEGY

Accuracy of Estimates

One obvious evaluation criterion to employ when considering the use of the kind of projection approach endorsed here is whether the coefficients that must be used in the model can be determined accurately enough that they should be used in making projections. CORD TN No. 6 has been developed to give one perspective on this topic so there is little need to discuss the issues that are covered there in detail. However, there are a few points which can be noted without entering into lengthy discussion. Firstly, an important determinant of the accuracy of coefficients is the size of the sample on which a model is based. A second important factor to consider is whether the socio-economic variables that are used in a model are strong determinants of participation in a given activity. This is particularly important since some activities that have a low level of participation may also only have very weak relationships to socio-economic characteristics. Thus, extremely large sample sizes in terms of a random sample from the population of Canada or a particular area may be required to develop certain models.

One might suspect that the size of the area for which projections are made would be important in relation to the accuracy of estimates that can be achieved. However, inaccuracy in estimates of participation in relation to the size of an area for which estimates are made reflect only inherent variability in levels of participation in relatively small populations. The fact that very highly variable levels of participation for small populations may be noted in surveys and expected when projections are made need not, and usually would not, reflect a deficiency of the modelling approach. It would reflect a reality of dealing with small populations.

The preceding comments have been based on using dummy variable regression models like the one introduced earlier. The CORDS work has already established the effects of certain interaction effect variables that were not considered in the simple models presented in TN 12. Technical Notes have been prepared which show the structural problems of models by pointing out (1) the importance of interaction effects between socio-economic variables in explaining participation in outdoor recreation activities (see TN 20) and (2) the effect of the quantity and quality of supply of facilities on individuals' frequency of participation (see TN 2, 3, 9 and particularly 29; also see the concluding remarks in the review of Chapter VII of this volume).

There is reason to suspect that model coefficients are changing. Over time, changes in the time budgets of people, for example the rise in the number of three-day weekends, may affect coefficients of many variables. Also, exogenous influences not directly related to socio-economic characteristics, such as the energy crisis and innovations in outdoor recreation equipment, certainly result in a changing environment in which people recreate.

Finally, research on National Survey data collected in Canada has shown that there is a clustering of people according to the activities in which they participate (see Romsa 1973 and the Annex to TN 10). This clustering reflects the situation that exists because there are large numbers of people who participate in very few similar activities, and a few people who participate in very many activities. The existence of the clustering phenomena suggests that, ultimately, analysis of variance models (and other models that pose a basically linear structure as explaining participation in an activity based on characteristics of individuals in a population rather than considering equations for sub-groups of the populations) will have very serious limitations in terms of making accurate predictions. Limitations relate to problems with the structure of the linear model and it thus not correctly reflected changes occurring.

Model structure and the resultant problems that arise when a model is not properly structured has been a recurring theme in the preceding comments. Some specific discussion sections may help the reader see the nature of special "accuracy" problems caused by improperly structured models. These sections may also help the reader see why certain lines of research were pursued in the CORDS to try and overcome structural problems with models.

Interaction Effects

Interaction effects have received some limited attention in the analysis of recreation behaviour. In the Meuller and Gurin report of the ORRC Study (1961) analysis of variance runs for males and females were carried out separately because it was recognized that there was a male-female interaction that would have invalidated the results of an analysis including both males and females. There are a few articles that incorporate interaction effects into the development of models to explain participation in outdoor recreation (VanDoren and Lentnek 1969 and TN 20). On the whole, however, there has not been a concerted effort to discover which interaction effects usually occur in relation to participation in certain activities, or how to deal with these effects.

The essence of the problem is that if, for example, an age-education interaction is important in explaining participation in hunting (as AID runs with Canadian National Survey data have indicated), then part of the variance that could have been explained by an appropriately structured analysis of variance model is not explained if the model derived does not contain interaction effects. Such a structurally deficient model not only does not explain all the variance that could be explained, it obviously cannot provide clear insights into some aspects of behavior in a system in which interactions are occurring. Thus problems may arise when the model is applied using an equation similar to Equation 3. Using socioeconomic projections of a population for some later time, the effects on participation resulting from changes in the distribution of people in various expected age-education categories will not be accounted for properly.

If the structure of a model does not reflect the way people behave, but is simply molded to fit a set of data, it should not be expected to provide a good representation of some future reality. Specifically, (1) models' efficacy in predicting future patterns is lowered and (2) questionable trends may be imputed from the data. The magnitude of structural error depends on the degree of interaction in the portion of the population which is being modeled. Generally, what is the magnitude of interaction? Preliminary results obtained from Canadian National Survey data indicate that the interaction effects in models developed involve a sum of squares (a measure of explanative power) as large as the sum of squares associated with the first order effects, even when analyses are run separately for males and females (see TN 20).

Supply Considerations

Researchers have known for a long time that the supply of facilities for participation in an outdoor recreation activity affects the amount of participation that occurs. However, the exact nature of this relationship has not been established. The work on demand functions for outdoor recreation at reservoirs and similar work on demand functions for parks has cast some light on the relationship between supply and participation. However, most demand function models do not treat supply in a way that allows one to generate figures for the total amount of participation that will arise in a city that is surrounded by a variety of units of supply at various distances (for example, see the literature on demand function models cited in Coomber and Biswas 1972).

Cicchetti, et al. (1969; Cicchetti 1973) have developed an equation that shows the relationship between supply and participation in activities in a given origin area. Because of the unavailability of data for specific locations where people participated, Cicchetti's equation applies only to aggregate supply. Thus it was not possible to include any locational or distance effects on the supply factor. It follows that if one attempts to develop a model similar to Equation 3, it should contain a supply component in addition to various socioeconomic characteristics. This supply factor should reflect the relative level of supply of facilities for a given activity that is available to the relevant population.

If a model is developed without a supply factor as an independent variable then the same problems arise as when a model is developed without a measurement of the interaction effects. As noted in the last section, part of the variance that could be explained by the model is not. This variance either (1) turns up being counted as error when it is, in fact, systematic, or (2) causes distortion in the coefficients of the model. The result will be invalid projections when the model is applied to both (a) participation estimation for specific areas that deviate from some norm such as the average national supply of National Parks and (b) projections of participation for five or ten years for a given area without considering how quantity or quality of supply change.

In recognition of the considerations noted above, some CORDS analysis work has been on the development of the supply factor model (see TN 29). Resources have also been devoted to work on modelling considerations that generalize the Cesario origin destination flow model (see TN 33). Promising results have been achieved in the supply factor modelling effort. The work related to the Cesario model is of a much more general nature and offers a promise of allowing an understanding of the structure of relations involved when a new site is "plugged in" among a group of alternative sites and when people modify their behaviour in relation to alternative sites which are available.

But when one refers to alternatives being available, two issues arise. if there exists a system of parks and a new, similar park is created, then an alternative to the existing parks has been created. However, it is recognized that there are alternatives to visiting parks in terms of outdoor recreation behaviour. Burton (1971) and Gillespie (1973) have discussed a methodology for the determination of the substitutability of activities by the use of a factor analysis procedure. Hendee and Burdge (1974) have reviewed a large number of considerations related to the substitutability of one activity for another (also TN 37 for Beaman's comments on the Hendee and Burdge article). In discussing apparent trade-offs between activities, Beaman and Lindsay (TN 37) suggest that one should be careful to recognize that there are substitutable and complementary activities, activities such that if people participate, they tend to participate in both of the activities or neither. One need only review the works already cited here or by Bishop and Witt (1972), which are devoted to operationalizing the Ontario TORPS (Travel and Outdoor Recreation Planning Study, Ellis and Wolfe 1970) to see how many academic challenges and practical problems relate to understanding substitutability.

Regarding supply considerations, the cluster analysis work (Romsa 1973 and TN 10) has implications for participation in, and trade-offs between, activities. Some of the implications of clustering analysis and problems in using factor analysis in determining trade-offs or complementarity between activities are noted in TN 32. The main point here, however, is that a cluster analysis model is a drastically different model than a linear additive model. The way in which people respond to supply when there is an intricate interrelationship between supply and a collection of activities in which people have participated is simply not "explained" by a group of linear models. Obviously the supply for any number of activities interacts with other participation variables to produce observed behaviour and when a number of supply units for a number of facilities relate to how decisions are made, there is a complex non-linear situation (see the "trade-off" matrix specified in Reference 1.)

Though important work is proceeding on the effect of supply on participation, the results are not yet definitive. It must therefore be recognized that models are still deficient in terms of the way that supply is introduced into the modelling framework. The result is that it is still necessary to "live with" some structural error related to supply and behaviour (see the concluding remarks of the review of Chapter VII).

Time Budget and Purpose of Participation Considerations

Implicit in much of what has been suggested about changing patterns of behaviour and the effect of socioeconomic variables is a concern with time budgets. One component of the supply base for any activity is the availability of time to participate in that activity. The potentially drastic effects on participation in any number of activities if there is a shift to a three- or four-day week is obvious. It is very likely that activities participated in during the week involving little travel may be ignored in favour of activities involving more travel. But very little demand research has rigorously considered expenditures of time and changing supplies of time as these affect recreation consumption. Thus, it would be pretentious to claim that any recreation research effort has provided a really usable way of entering time budget information into a supply-demand or supply-participation model (for general time budget research, see Reference 15). The amounts of discretionary time, the kind of packages in which it is available and the availability of resources to use this time have fundamental implications for the amount of participation in outdoor recreation activities that will occur (for a classical discussion, see Reference 8). Until better theoretical work is developed and better data sets are available for analysis, time budget considerations will enter into the participation models in a very inadequate way. Thus, the incorporation of time budget information for tapping important trends in time use cannot be expected to be realized in the near future.

The only way it now seems possible to structure models to explain some of the trends to the change in time budgets is to include gross or aggregate time use information in models. In Canada, in 1975 two time budget data sets existed that could be used in model development (Halifax time budget data collected by members of the Dalhousie Institute of Public Affairs and Vancouver data collected by members of the Department of Sociology, University of British Columbia).

"Spontaneous" Change vs Trends In Model Coefficients

Understanding trends in the coefficients that define structural models is much more complicated than recognized in time budgets as showing sources of variation. The discussion above has suggested that models do not adequately consider changes in time budgets and supplies of facilities. Furthermore, these models have been shown to ignore interaction effects and how these change model coefficients over time. Implicit in the discussion, however, is the notion that in theory all these sources of change in the coefficients can be understood and controlled.

Though this paper endorses seeking an understanding of the structure of relations that "explain" behaviour, several practical non-structure oriented procedures can be suggested for making necessary projections. First, when several surveys are available that are large enough that model coefficients have small standard deviations compared to their magnitude, then trends in regression coefficients over time should be considered so that, as well as projecting demographic characteristics, relevant regression coefficients may be projected. The author considers this a case where demographically induced trends, and trends in the related coefficients, may be used in a rigorous way for making projections. This may be done by using both projected coefficients and projected population characteristics in Equation 3 or a generalization of this equation to include supply factors and interaction effects.

Another consideration relates to what might be called "spontaneous change". Innovation in a given area may result in the development of the snowmobile or a device that has not yet been marketed. Obviously trends in participation in presently marginal activities or activities of the future cannot be extrapolated to suggest future participation in these activities. It is in these cases that demographically induced trends are irrelevant to predicting future levels of participation unless participation in one activity is traded-off with another. In this case, valid projections may be made for aggregates of activities but not for specific ones in a collection of activities among which trade-offs are occurring.

In the context of specific established activities, looking at demographically induced trends may be relatively useful while this approach has little to offer in the context of new or "unthought of" activities. For these activities, the use of the Delphi technique, or other futuristic projection techniques, are alternatives. The boom in snowmobiling was not predictable in 1930 in the sense that 45 years ago one could not have looked at certain coefficients, age distributions, etc., and have predicted the present level of snowmobiling. Further, it is almost certain that even with data on snowmobiling in 1969 and 1972, one cannot come up with realistic projections of what snowmobiling will be in 1990 based on current trends and trends in regression coefficients. Snowmobiling in 1969 was in a transition period and it is questionable whether some equilibrium had been reached only three years later.

CONCLUSION

In this paper, two kinds of participation projection models have been identified. The one which is of major concern is the structural model which explicitly considers change in participation as being directly linked to socioeconomic changes in the populations under consideration. This type of change may be described as resulting in "demographically induced trends" in participation rates.

When several surveys on participation rates and demographic trends are available, appropriate structural parameters can be estimated and trends in these parameters can be established. Given the expected future values of these parameters, it becomes possible to project future participation levels. This is a logical extension of the assumption that coefficients observed at a given time are relatively stable, and thus are valid and reliable enough to apply five, ten or twenty years in the future.

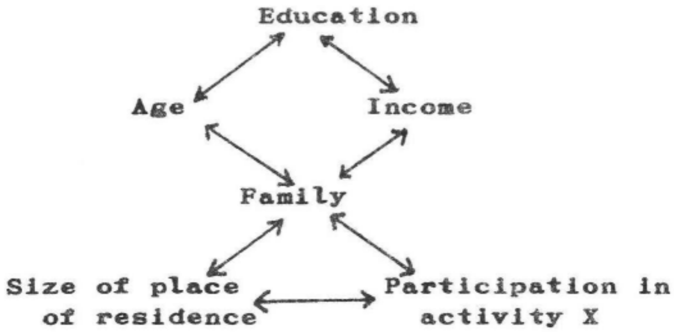
The author believes that when demographic changes are the predominant reason for changing participation levels, there is good reason to argue that accurate projections of participation in the future will only be achieved by the procedures endorsed here, rather than by simply projecting an observed trend in a dependent variable. However, if a planner Is interested in future levels of new activities (e.g. sky-diving or snowmobiling, or the future importance of currently non-existent activities which can be expected to grow out of future technologies and fads) then the "demographically induced trends" projection approach is either too risky or impossible.

Problems in operationalizing the demographic trends approach are due to:

1. accuracy of model coefficients;
2. problems with model structure; and,
3. changing model structure over time.

Therefore, it means that it is necessary to evaluate (a) all projections made on the basis of common sense, (b) "check" projections made by comparison with simple trend projections, (c) make re-evaluation of models as new data, allowing new parameter estimates, become available, and (d) evaluate and revise models as new research results become available allowing improvement of models with existing data.

Finally, many readers may find it surprising that in this paper there has been no discussion of causal modelling. Obviously, one problem with the kind of model discussed earlier is that the variables do not have the reciprocal non-causal relation implied by considering all the independent variables as exogenous. It is really only a matter of priorities and time that has prevented the use of path analysis on the econometric method of multiple-stage least-squares to further clarify how models should be structured. Certainly the author is interested in seeing the value of path coefficients for a model like that below:



Incidentally, in the above the rather strange "role" of age reflects the fact that age with cross sectional data for a given year at a given point in time reflects the generation and values of a generation by its relation to year of birth.

In other words, many research steps remain to be taken before the general projection procedure discussed here becomes a tested tool for use by a technician. Projection is still partially "art" as well as "science". A researcher or planner should be willing to abandon any "accepted" projection methods which have more problems than those that were described for the modelling approach. It is necessary to continually update and revise the processional's repertoire of modelling techniques as improvements are made.

Unlike many engineering fields, recreation's "accepted techniques" are generally far from totally satisfactory for even "simple" problems.