CHAPTER 7

AGGREGATION OF DISAGGREGATE MODELS FOR FORECASTING

To be useful in planning and policy analysis, forecasts of individual choice behavior must be aggregated into geographic, socioeconomic, or supply market groups of interest to the planner. Disaggregate models are usually non-linear in explanatory variables, and groups for which forecasts are needed are non-homogeneous in these variables. Hence, the aggregation process is non-linear, and must take into account the distribution of the explanatory variables. Group choice shares are generally not equal to the probability of choice an individual with the average explanatory variables of the group would have. The choice probabilities of individuals in the multinomial logit model, and other choice models used in this study, can be written in the general form

(l) $$P_j = F_j(V_1,...,V_j) \quad ,$$

where P_j = probability of choosing alternative j ;

V_j = "mean" utility of alternative j ;

J = number of alternatives;

and the utility V_j is a function of observed socioeconomic characteristics and level-of-service attributes of alternative j ,

(2) $$V_j = v(LOS_j,SE) \quad .$$

Typically, the function (2) is taken to be linear in specified functions of level-of-service and socioeconomic variables,

$$(3) \qquad V_j = \sum_{k=1}^{K} \beta_k Z^{k}(LOS_j,SE) \quad ,$$

with the β_k interpreted as "importance weights" that are estimated in the demand analysis. Let $z_{jk} = Z^k (LOS_j,SE)$, and define z_{jk} to be the group mean of z_{jk} . Then group average utilities satisfy

$$(4) \qquad \overline{V}_j = \sum_{k=1}^{K} \beta_k \overline{z}_{jk} \quad .$$

A <u>naive aggregate forecast</u> can be obtained by substituting $\overline{V}_j$ in (l) . However,

non-linearity of the function (l) implies that in general the probability evaluated at average utilities of the group will not equal the group choice probabilities (i.e., the average of the individual probabilities). Consequently, the naive procedure will in general yield biased forecasts.

If, for a group, the multivariate distribution of the utilities V_j is known, and denoted by $g(V_1,...,V_j)$, then the aggregate probability for the group satisfies

$$(5) \qquad P_j = \int F_j(V_1,...,V_j)\, g(V_1,...,V_J)\, dV_1...dV_J \quad .$$

This is termed <u>aggregation by statistical integration.</u>[1] For a finite (but usually large) group size, this formula reduces to sum

[1]Koppelman (1975, 1976) referred to this method as integration/summation.

$$(6) \qquad P_j = \frac{1}{N}\sum_{n=1}^{N} F_j(V_{ln},\ldots,V_{Jn}) \quad ,$$

where V_{jn} is the utility of alternative j for individual n. This is known as an enumerated aggregate forecast.

There is an aggregation problem because the enumeration (6) can be expensive, especially for forecasts for groups such as interzonal choice matrices. Nor is the expected value computation (5) practical in most cases. $g(V_1,\ldots,V_j)$ is neither simple nor easy to obtain, and the integral (5) has a known analytical solution for only one special case (McFadden and Reid, 1975).

Other attempts to evaluate (5) have also been impractical. Westin transforms (5) to distributions in probability space. However, his method applies only to binary choice, and requires numerical integration procedures as costly as enumeration (Westin, 1974).

An attempt by Talvitie to approximate the expected value of group shares by a Taylor series has given counter-productive results for typical variances of the explanatory variables because of the poor convergence properties of series expansion (Talvitie, 1973; Koppelman, 1975).

The aggregation error that results from naive aggregate forecasts can be substantial. A study on a Washington, D. C. area sample and model of CBD-oriented work mode-choice has shown this error to range from eight to ten percent RMS, depending on the geographic level of aggregation (Koppelman, 1976).[1] Results on the data and models of the Urban Travel Demand Forecasting Project study show naive forecasting error between 13.8 RMS, for origin zone aggregate trips, and forty-three percent RMS for metropolitan area trips, respectively. Neither of the studies were for interzonal naive aggregation, though it is estimated that for this level of geographic interchange forecasting the error will be about half the origin zone level error above, on the average, and considerably greater for individual interchanges (see Vol. VII, Chapter 7, Table 7.3 discussion).

---

[1]Overall error is measured in terms of the root-mean-square of the errors in the separate alternatives with respect to the equivalent enumerated aggregate forecast. See equation (10) in this chapter.

Naive aggregation can produce larger biases than are produced by estimating and forecasting with strictly aggregate data. The latter procedure has, at least, the ability to replicate baseline shares. However, strictly aggregate models produce biased predictions of the impacts of most policy changes and are insensitive to many policy tests of interest.

The essential advantage of disaggregate models is that they are sensitive to the mix of variables explaining a traveler's choice. Ignoring the distribution of these variables in forecasting produces errors that cancel the advantage of the models. Disaggregate models and an acceptable aggregation method must be used to gain accurate forecasts.

Fortunately, some progress has been made in approximations to enumerated aggregate forecasts, based on classification. These approximations are simple and considerably reduce the error of the naive method. Koppelman (1976) has shown that classification by auto availability and choice set can reduce RMS error (at various levels of aggregation) to three percent versus eight percent by the naive method. Later results on a sample used by this study show that classification based on the values of differences in utilities of choices can reduce error to less than one percent compared to forty percent for the naive method.

The classification methods also have great potential savings in computation efficiency with suitably chosen classes. Their attempts to delineate different choice types approach the minimum amount of information necessary to define the different market segments or traveler types making up the range of possible choices. Examples and recommended practice with these methods will be explained in detail following a review of the state of aggregation theory and a discussion of the planning context, its needs, and resources. Questions of data need and availability, efficiency of predictions, and effects of aggregate size will be discussed in addition to accuracy.

The State of Aggregation Theory

The major thrusts in simplifying the general theory of aggregation have not been as fruitful as the approximate classification approaches. Koppelman (1975) extended Talvitie's Taylor series approximation method for multiple choices to account for covariances between utilities of different alternatives. The aggregate shares are

$$\frac{S_j}{N} = P(j\,|\,\overline{V}) + \frac{1}{2}\sum_{K}^{J}\sum_{\ell}^{J} \left.\frac{\partial^2 P(i\,|\,V)}{\partial V_K \partial V_\ell}\right|_{\overline{V}} \text{cov}\;(V_K, V_\ell) + \ldots \quad , \tag{7}$$

where J is the number of alternatives, V is the vector of utilities of the choices, $\overline{V}$ is its mean in the aggregate of N travelers, and V_K, V_ℓ are the utilities of choices k and ℓ. However, he found that this approximation suffers from larger errors when the series is truncated after the second term than when it is truncated after the first term. That is, the method produces worse results than the naive approach for aggregate sizes down to districts that were about l/l00th of the Washington metropolitan population.

The binary logit curve can be scaled to represent the probit curve within one percent absolute error. Using this, Reid (1974) has shown that the simple form of (5) for normally distributed explanatory variables can be extended from probit to logit models:

$$\frac{S_j}{N} = \left(1 + \exp\left[\frac{\overline{V_i - V_j}}{(1 + \text{Var}(V_i - V_j)\;/2.79)^{1/2}}\right]\right)^{-1} \quad , \tag{8}$$

where V_j and V_i are the utilities of the two choices and $\overline{V_i - V_j}$ is the mean of their difference in the aggregate of N travelers. The assumption of normality need apply only to the scalar utility difference, not to each explanatory variable. Tests have shown this assumption to contribute 4.0% error for a large aggregate forecast with large variance and forty-three percent naive error, suggesting that normality is a reasonable assumption (see Vol. VII, Chapter 7). The method is still limited to binary choice.

Work on the theory of multiple choice by the integration method could be very productive in extending its simplicity and accuracy. Recent research on estimation methods for multinomial probit models has had the side-effect of revealing new efficient approximations to integrals of the form of (5). (Daganzo, *et al.*, 1976; McFadden, 1977).

An aggregate forecast by classification can be defined as a weighted sum of the naive shares in classes identified by some characteristics of the travelers or their alternatives. The aggregate shares are obtained by

$$\frac{S_j}{N} = \sum_{c=1}^{C} P(j \mid \overline{V}_c) \frac{N_c}{N} \quad , \tag{9}$$

where c is the index of the C cells in the classification, $\overline{V}_c$ is the mean utility computed for cell c , and N_c is the number of travelers in cell c . Koppelman (1975) has done the principal study of aggregation by classification using choice subsets and specific variable values as identifiers. He found that where the full alternative set is unavailable to a significant portion of individuals in the aggregate (thirty percent in his tests) that classification based on available choice set is more effective than on any single variable. Additional classification by variables is recommended to concentrate on those with greatest contribution to the utility variance. Choice set classifications reduced RMS error in the Washington sample to five percent. Cross-classification with the variable auto/drivers in the household reduced error to three percent regardless of aggregate size. Naive method error was eight percent. No other methods tested by Koppelsnan gave smaller error. He did not test the statistical integration method.

As stated before, the statistical integration method has been found in this study to be capable of reducing naive error over ten-fold. Koppelman performed Monte Carlo simulations of the aggregation error by each of the above methods to determine the range of error with statistics of the distribution of the utilities of the alternatives. These confirmed that the integration and classification methods usually gave better results than the naive and series approximation methods, agreeing with his empirical results, and with ours. However, the simulations show that for skewed distributions, exceptions may occur to this ranking, especially regarding the desirability of the integration method.

Results of this study have extended the general classification approach in two ways--developing more systematic methods for defining classes based on specific variables and considering classification directly on utility scale values.

A systematic picture of priorities for class identifier variables can be gained by looking at the covariance matrices of the within-aggregate variation of the explanatory variables at the desired level of aggregation. The binary choice aggregation theory has established the direct relationship between this matrix and aggregate shares for probit models and for logit models with limited assumptions. Though no corresponding theoretical relationship exists in the multiple choice case, the covariance matrices are still better indicators of which variables are appropriate classifiers than *a priori* knowledge, especially if these vary with aggregate level. Table 68 shows a normalized covariance matrix of the intra-city aggregate values of the major explanatory variables in two alternatives of a prediction model. The model used is shown in Part II, Chapter 2, Table 12 of this volume. The covariance elements are normalized by dividing by the largest of their values (in this case it is the variance of the utility of CARS/DRIVERS in the traveler's household) in order to give a picture of the ranking of the contributions of each variable to the overall utility variance. The underlying model is a logit function, linear in its explanatory variables.

TABLE 68 Covariance Matrix of Intra-City Values of Explanatory Variables

	Auto-Alone Alternative Variables					Bus-with-Walk-AccessAlternative				
	Drivers	AUTO-ONVEH	Drivers	Employm. Dens.	Cost	Buswalk	Bus-ONVEH	Hdway	Xfer-wait	Sum Alt.2= 1.73
C/DR	1.0	.06	<	<	.07	<	<	<	<	
AONV	.06	.51	<	<		-.07	.23	<	.06	
DR's	<	<	.27	<			<	<		
EMP	<	.10	<	.11		-.06	<			
$COST_1$	.07	.10			.07		<			
BWLK	<	-.07	<	-.06		.84	.33	.05	<	.84
BONV	<	.23	<	<		.33	.21	<	.07	.54
HDW	<	<	<			.05	<	.09	<	.14
XFT	<	.06				<	.07	<	.14	.21
SUM Alt.1= 2.62	1.26	.91	.27	.11	.07					4.35 3.96[3]

NOTES:

1. Covariance values are normalized by multiplying by their respective model coefficients and dividing by the maximum such "utility" covariance element: $\beta_k S^2_{k\ell}\beta_\ell / \beta_1 S^2_{11}\beta_1$ (in this case the maximum utility component covariance is the variance of the first variable CARS/DRIVERS) .
2 < = element is less than .05 .
3. Total variance of differences in utility between alternatives 1 and 2.

Hence, each element in the matrix is the product of the covariances of the individual variables and their correponding model coefficients $S_{k\ell} = \beta_k S^2_{k\ell}\beta_\ell / \beta_1 S^2_{11}\beta_1$; k and ℓ are the matrix indices of the K variables in the model, $1 \leq k,\ \ell \leq K$; $\beta_k\beta_\ell$ are the linear utility coefficients for the k-th and ℓ-th variables; $S^2_{k\ell}$ is the sample covariance between the k and ℓ-th variables; and $\beta_1 S^2_{11}\beta_1$ is the corresponding utility variance for the largest matrix element. These elements are called utility component covariances because each expresses the part of the total sample variance of the linear utility for one mode contributed by a single variable (diagonal variance elements) or pairs of variables (covariance elements). The sum of the matrix elements for any mode equals the total variance of the utility for that mode.

For binary logit choice models the relationship between these covariances and aggregation error is understood under the assumption of normally-distributed explanatory variables (equation (8)). Thus Table 68 provides a convenient way to rank the individual variables and their combinations, by their importance for reducing aggregation error for binary choice. The larger the variance of the difference of utilities between the two modes the larger the aggregation error.[1]

The variables that have individual values that are most important for reducing aggregation are the ones with the largest values in Table 68, such as BUSWALK, CARS/DRIVERS, and AONVEH (time riding in autos). The inter-modal covariances show the relative importance of these pairs of variables in aggregation. However, because inter-modal elements contribute to the variance of binary utility difference with negative sign it is the negative values such as that between CARS/ DRIVERS and BUSWALK that increase the aggregation correction. Positive values indicate a compensating reduction in aggregation error. The amount of aggregation error that will result in a forecast using the individual values of only selected variables can be estimated by comparing equation (8) with and without the unnormalized values of the matrix elements to

[1]Because most of the variables in this model are unique to a mode, variances of intermodal utility differences are equal to the individual modal elements. Where the same variable appears in the utility expression for two modes, the differences of utilities can be obtained from the table by the expression for the variance of such a difference: Var (X - Y) = Var X + Var Y - 2 Cov (X,Y).

be considered.[1]

Inspection of Table 68 shows that a large part of the variance of the utilities, and hence aggregation correction, can be recovered by considering only a few of the variables in the model. For example, three of the eleven variables in the model used here--CARS/DRIVER, BUSWALK, and BONVEH (riding on bus)--constitute fifty-nine percent of the utility variance. When only their individual values are used for prediction, they correct for sixty-nine percent of the aggregation error produced by the naive method. Auto (AONVEH) time, in contrast, contributes nothing to the correction of aggregation error by its inclusion with this set since its covariance with BUSWALK cancels out its variance contribution alone.

The covariance matrix alone gives incomplete information on the advantage gained in classification because of the correlations between the classifier variable(s) and the others. This can also be evaluated by adding the information from a corresponding correlation matrix of the variables within the same prediction aggregate. The additional power of a classifier due to this effect is given by the sum of the product of covariances and correlation coefficients of all variables correlated to the classifier. With this procedure, one can minimize aggregation error, given limits on the number of classification variables, or minimize the number of classifiers, given tolerable limits on error.

When applying this covariance analysis to multiple choice, simpler approximate guides to the important aggregation variables can be gained by observing the covariance matrices of major subsets of the alternatives. These should be alternatives which have shares different from 1/J and are considerably different from one another.

The procedure also does not account for skewed utility distributions. Skewedness will weaken the analysis. This does not appear be a problem in the tests of this study (see Vol. VII).

The covariance analysis on a subsample of the population can also guide larger sample data collection and processing priorities for planning and policy analysis.

---

[1]Only the individual values of some variables or their variance may be available in a forecasting situation. In some cases, variances are possible to obtain or estimate where covariances are not Talvitie and Dehghani, 1976).

Utility Scale Classification

Though the covariance analysis procedure above can systematically guide selection of classification variables, the process is inefficient. If more than a few variables contribute significantly to aggregation variance, the number of cross-classifications becomes large for achievement of small error. This is especially true of predictions for large geographic aggregates.

A more efficient method of classification is possible for the class of simply scalable models of the form of equation (1). These models express choice among J alternatives as a function of J utility scales of the individual decision-maker's characteristics and choice attributes. As long as each scale is only a function of the attributes of one choice alternative, regardless of complexity, classification directly on these utilities is much more efficient than on the individual variables.

Cross-classifications between the individual variables include much information that does not enter into a simply scalable choice model. The essential information to predict each individual's choice in these models is contained in the total utility of each alternative. Cross-alternative classification between these utilities describes the full distribution of individual choice factors in an aggregate prediction sample. Because this procedure bypasses individual variable cross-class information, which does not enter into the choices, it requires fewer classes. Classification directly on the total utility in choice picks up the total variance including the minor variables, not just the variance of the limited number of interactions feasible in classification, by specific variables. This further increases its efficiency. Defining relatively homogeneous classes of utility combinations across modes is getting at the essence of the classification approach--the grouping of individuals with uniform choice situations. Because the procedure operates on the utility scales, it is termed the utility classification method of aggregation.

This method would not capture the full information entering into non-simply scalable models such as probit types that incorporate inter-relationships of cross-alternative attributes with separate model coefficients (Hausman and Wise, 1976; McFadden, 1977). Aggregate predictions with these models require individual variable interactions. This is accomplished by the classification by variable values and enumeration methods previously discussed.

The utility class sizes and boundaries cannot be defined in the same way as for individual variables, because utilities are not discrete and intuition gives no guide on utility thresholds in choice such as it does, for example, for BUSWALK time. However, utilities are directly related to choice ranges. For the binary logit case the optimum divisions would be differentiating utility differences near the maximum non-linearity of the choice function (near choice probabilities of .25--or 1.1 on the utility scale). For multiple choice logit model classification criteria should concentrate on the differences of utilities of pairs of alternatives with the same relative probabilities as in the binary case, thus distributions of utilities of pairs of alternatives falling, for example, in the range $\pm$ 1.1 to $\pm$ 0.5 should be differentiated most closely.

Various cluster analysis techniques could be employed to achieve isolation of classes (Green and Wind, 1973; Friedman and Rubin, 1967). In practice, tests have shown that *ad hoc* class divisions are quite adequte and more feasible in the press of obtaining predictions.

One such procedure is the successive division of an aggregate sample about the median (or mean) values of the differences of the utilities of pairs of alternatives, starting with the pair with the largest variance of utility difference. The procedure cycles through all pairs of utilities, further subdividing the first pairs if the variance of utility differences in the resulting classes are still large. The variance criteria for desired accuracy can be estimated with the covariance analysis procedures previously discussed, the ranges for which divisions are fruitful from the discussion above. This procedure resulted in a smaller error than any of the individual variable classifications considered in tests of this study for regional aggregation with only two utility classes for a four-alternative model. With eight classes it resulted in insignificant error. In contrast, the class counts defined on explanatory variables or on geography ranged from four to 136, with only the larger divisions equalling the accuracy of the two segment utility classification.

The focus of this aggregation method on the total utility value need not preclude the retention and association of the values of specific explanatory variables with the classes. This would be needed for interpretation, predicting sub-segment choice or the analysis of the effects of policy changes on choices. These can all be accommodated by characterizing each class cell by, in addition to the average value of the J total utilities of its members, the average value of any desired model variable or socioeconomic descriptor. Thus the characteristics of the homogeneous choice groups may be interpreted and sub-segment predictions may be obtained. Together with the values of the underlying model scale coefficients, the mean utilities may be adjusted for policy changes on specific variables to simply analyze their effects on aggregate choices. A complete dscription and examples of the use of the utility classification method in regional travel prediction and policy analysis is given in Volume IX of the Urban Travel Demand Forecasting Project's Final Report Series.

Forecasting in Practice: Planning Requirements Versus Resources

The planning and policy analysis environment in public agencies is characterized by different needs for analysis complexity and by available resources. Forecasts can be needed for aggregate sizes from small geographic zones to entire regions, for demand levels by mode or their financial and environmental impacts or for impacts across socioeconomic, political, and geographical groups. Issues of interest can range from the effects of pricing on work mode-choice to the effects of transport and zoning policies on congestion and urban form. The implied models range from simple binary choice to complex joint-choice structures and supporting data. Some analyses may allow long-scheduled turn-around time; others need results immediately. Available data may be only for aggregates. Institutionalized procedures and skills, and budgets, may not allow major additions to analytic libraries.

The aggregation procedure complicates this situation because it adds a new step to the traditional forecasting process and may not have similar requirements or errors across these ranges of requirements. An assumption of level-of-service homogeneity will make more error for large aggregate forecasting than it will for interzonal predictions. Defining homogeneous classes for one dimension of a joint choice for example, between modes and destinations, may be quite different than for the other. Tolerable errors are different for sketch planning and specific project analysis. It does not seem possible to give general guidelines on aggregation, especially ones that will suffice across wide geographic forecasting scales or differing requirements for simplicity.[1]

Some general guidelines will prevail. A version of the classification approach will probably be best for most aggregated forecasts. The questions will be as follows: Which variables will be considered? How many classes are necessary?

[1]Planning simplicity requirements are more than hopes of something for nothing in tightly-budgeted agencies. Budgets are tighter, but the credibility of large complex models is also low. Decision-makers need simpler, more conceptual models that they can understand. Disaggregate models are a paradox in that they basically have this conceptual simplicity, yet, in applications, seem to hunger for even more data and forecasting complexity to achieve their promised accuracy. The aggregation methods must be simple if the models are to be understood and be financially feasible. They must be simple and inexpensive if they hope to be extended into other realms of analysis, such as joint-choice among the many transport and land-use dimensions.

The differences needed in aggregation guidelines can be clarified by listing at least a few existing and emerging forecasting environments faced in planning. Table 69 shows categories of planning systems. Different ranges of output aggregates, levels of complexity, and resource budgets that affect aggregation procedures are also shown. Typical data resources associated with these categories are discussed below.

The traditional four-step interzonal transportation forecasting systems are complex even before the aggregation issue is added. They do embody a high degree of geographical classification and only require forecasts of small aggregates. The naive method is tempting in its simplicity to apply in these cases, but it still gives significant error. Accurate results require additional data, yet these systems have little tolerance for increases in complexity even if their error is high. Their behavioral generality is now only moderate due to their sequential choice assumptions. Improvements to this are challenging their complexity (Ruiter and Ben-Akiva, 1977).

The data available in these systems are usually only zonally aggregated values, sometimes supplemented by small individual traveler data or the supporting survey data from which intrazonal classes may be reconstructed at a cost. Four-step systems are established in most larger U.S. urban areas.

Policy analysis systems have highly aggregated outputs--usually the whole region plus some socioeconomic segments. They can include fairly general behavioral interactions and are defined to be low-effort. They are moderately expensive but accurate when employing aggregation by enumeration, and inexpensive and fairly accurate with classification. The coarse aggregates of these models require that they consider more variables and class cells when using the latter procedure than, for example, would be necessary for interzonal (four-step) forecasting systems.

To the extent that these systems need improvements in simplicity system generality, or faster turn-around they require more efficient aggregation methods. For example, if they are to reach the level of simplicity allowing hand computation or their inclusion as small components in much larger context analyses (such as category 4) they require simpler aggregation methods. Their small, but in-depth, disaggregate data sets support these needs. Such data sets are available in only a few urban areas.

TABLE 69 Categories of Planning and Policy Analysis Systems

Planning Category	Physical Scope	Output[a] Detail	Generality[b]	Resources[c]	Example System	(ref.)[d]
1 Four-step	large	fine	med.[e]	high	UTPS	UMTA (1976)
2 Policy Analysis	any	coarse	med.	low	SRGP MKT. SEGM.	Atherton, *et al.* (1976) Dunbar (1977)
3 Special Area	corridor	coarse	low	very low	Chicago Parking	Lisco & Tahir (1974)
4 General Policy	any	coarse	high	moderate	RAND	Bigelow, *et al.* (1973)

Notes:

l. Output detail refers to choices, geography, or market segements.

m. Generality refers to issues, system, and behavioral scope.

c. Resources include data, computation, and skill requirements and usually in the inverse of turn-around time of analysis.

d. Examples are referred to by common names as well as the references to text.

e. UTPS is of medium generality because its sequential behavioral assumptions restrict its role.

Part IV, Chapters 3 and 4 discussed methods of overcoming the problem of traditionally collecting such data from scratch in new areas. Policy Analysis Systems data can also be derived from the individual survey files supporting four-step systems at some cost. They cannot utilize the highly aggregated or idealized data typically used by sketch-planning models, which are seen to have a similar planning role but are not usually derived from behavioral models (Scheifler, *et al.*, 1975; Mann, 1975).

Special Area studies employing behavioral models are usually small scale, simple forms of the category 2 above. They have coarse or total area, yet small aggregates. They have very low budgets and are feasible only if the data and aggregation requirements are very simple. They require the simplest and most general rules for aggregation such as methods to estimate utility variance from aggregate data or classification segment definition without individual data.

The General Policy category is seen as an extension of the (medium generality) policy models of category 2 into a greater individual and system behavioral scope. Examples are the consideration of equilibration for supply-demand and joint transportation and locational decisions. The message of this category is that they are not feasible, within the generality and accuracy potential of joint-choice behavioral models, without very simple aggregation procedures. Enumeration would become very expensive; class cells counts could grow greatly.

The use of behavioral models with the various planning categories thus requires different data efforts and aggregation guidelines. The common threads are: (1) additional disaggregated data is needed in varying degrees for their beneficial use, and (2) some form of the classification method is likely to be required. Questions of the amount and type of these efforts are addressed in the conclusion of the next section on aggregation practice.

Methods of Aggregation in Practice

This section gives some examples of aggregate forecasting done with disaggregate models, shows results of controlled aggregation error tests on actual data, and concludes with recommendations on the practice for the four planning categories of Table 69.

Liou, *et al.* (1975) attempted aggregate forecasts with disaggregate models in the four-step interzonal forecasting context. Recognizing that naive aggregation gave biases, even for small geographic zones, they used the method of statistical differentials, developed by Talvitie (1973), to obtain the aggregate shares in each zone. No actual data was available to compute the within-aggregate variances required for this method. It was assumed that the desired variances were equal to the variances of the between-aggregate variable values in each of two classes of trips--urban and suburban. Only level-of-service variables were involved in the model. Because no true aggregate predictions were possible, these results were evaluated by comparison with on-the-ground observations and against a traditional method. The aggregation method was considered a success because it more than equaled the accuracy of the traditional method at a lower effort. The aggregated results showed improvements over those using the naive method.

This comparison is confounded by the additional sources of error due to the model and data when comparing actual with predicted aggregate choices. The assumption of equal intra- and inter-aggregate variance will be highly dependent on the level of aggregation and variables used. The statistical differentials method has since been shown to be counter-productive relative to the naive method. It is expected that the improvement in errors over the naive method in this study was due to the better convergence properties of the statistical differentials method on low variance data. More recent efforts on interzonal forecasting have used income-level classification within the zones to reduce naive method error (Ruiter and Ben-Akiva, 1977).

Policy analysis studies have employed both the methods of enumeration and of classification-by-variable-values in their predictions. The study by Atherton, *et al.* (1976) for the Federal Energy Administration accepted the burden of enumerated computations for each traveler for large alternative sets and joint-choice, thus producing accurate results. Forecasts were for the total geographic sample but by three socioeconomic segments and trip purposes. Cost was about $100 per policy test at the metropolitan scale. This is a good demonstration of the feasibility of the enumeration method where data is available. Its retention of the full file of individual characteristics also offers flexibility in selecting outputs for any sub-aggregates. Its limitation is that on sample sizes feasible for enumeration computations, the statistical accuracy of predictions is poor for sub-segments smaller than ten percent of the sample.

Dunbar (1977) applied cross-classification by values of specific model variables in his policy analysis model. Forecasts were also for whole urban areas. He chose three ranges of transit access, and two each of autos per worker and length of trip, for a total of twelve class cells. The classification criteria were largely observations of natural ranges of the data, medians, bi-modal distributions, and the use of variables expected to contribute most of the utility variance. He mentioned that the process involved much judgment, and that his classification of the transit access variable on *a priori* grounds was inefficient.

The aggregation error was 0.9 RMS for the three mode shares predicted. This indicates that a moderate number of roughly determined classes produces good aggregation results. The model used had three level-of-service variables and one socioeconomic variable--autos per worker. The data set was the 1973 Nationwide Personal Transportation Survey.

Lisco and Tahir (1974) used a modified form of geographic classification in a special area study of the effect of parking taxes in the Chicago CBD. They only had data for the zonal averages of each model variables, so their effort was on the best way of using this limited data to efficiently obtain total area predictions. This was done by combining zones into rings which they judged would have similar mode-splits and proceeding by classification for area predictions on the weighted zonal averages of the variables in the rings. There was no basis for evaluation of the accuracy of these results. This study was one of the first to recognize that more than geographic homogeneity was necessary for aggregation. It also reduced the procedure to sufficiently simple classes and counts that it could be applied with very low effort.

The messages from these experiences are:

1. Regional planning agencies may not have the data necessary to accomplish aggregation.
2. Enumeration methods are viable for non-segmented forecasts with moderate model structure complexity.
3. Cross-classification by variable values, though inefficient in selection criteria and computation, gives accurate results.
4. Extremely simple classification procedures in special area studies may be accurate enough for low budget forecasts.

Aggregation Error Tests

This study has provided controlled evaluations of aggregation error by different methods in different situations.[1] Similar isolation of aggregation error on empirical data has been done only by Koppelman (1975).

The tests of this study were done on a sample of 771 workers drawn from about half of the San Francisco Bay Area. Their choice of mode of commuting to work was described by the four alternative models shown in Part II, Chapter 2, Table 12 of this volume. The overall exact mode shares were 55.6% auto alone, 17.4% bus-with-walk-access, 3.9% bus-with-drive-access, and 23.1% shared ride. The data were from household surveys and transportation network minimum-path simulations with modifications to produce trip attributes temporally and spatially disaggregated to the individual schedules and locations of the trips modeled.

The measure of aggregation error was the percent root mean square of the choice shares as defined by Koppelman:

$$(10) \qquad \left(\sum_{j=1}^{J} \left(\frac{\hat{P}_j - P_j}{P_j} \right)^2 P_j \right)^{1/2} ,$$

where J is the number of choice alternatives,

$\hat{P}_j$ is the aggregate share of alternative j estimated by the tested method, and

P_j is the aggregate share by enumeration.

[1]Controlled error refers to the isolation of aggregation from model and data error. It is expressed as the difference between the results from the test method and those by enumeration (with a sufficiently large sample to make sampling error in the minor shares negligible).

Table 70 shows the error from the naive aggregation method applied at three levels of geographic classification on our sample. Predictions are all for the total region. The impact variables were aggregated to the classes shown to represent results for either of two forecasting situations: geographic classification for <u>regional forecasts</u> or the average absolute errors when making separate predictions for all the cells at a geographic level. These errors are equivalent.

TABLE 70 <u>Percent Aggregation Error for the Naive Method for Three Scales of Geographic Classification</u>

Error Measure	Region (1)	Cities (17)	Traffic Analysis Zones (200)
RMS	40.0	17.9	13.8

Multinomial Choices (4 alternatives)

Notes:

1. Errors all based on predictions made for the full sample; inputs are the average values of the explanatory variables in classes as defined by the residential origin of the trips at the indicated geographic scale.
2. The numbers in the parentheses at each geographic classification are the number of cells at that scale.

The errors are large. It is obvious that geographic classification alone is not an adequate aggregation method for this region and model. Error does decrease with geographic scale. Smaller samples apparently do have less variability. Because classification/aggregation was done on the basis of residential origin only these results do not represent what would be expected from interzonal aggregate forecasts. However, it is expected that such errors for this sample would be about half those shown. Individual inter-district errors would be worse.

Koppelman's results were lower in magnitude and insensitive to geographic scale. He showed an 8.5 percent RMS error for a superdistrict scale similar to our cities. Several reasons are speculated for this difference: he had a CBD trip oriented sample only; his average shares were more nearly equal;[1] the choice model he used was simpler; and, the level of service data was less disaggregate.

Table 71 shows the error on our sample by five available methods of aggregate prediction of regional choice shares. The naive method entry is the same as that for the regional level on the previous table.

TABLE 71 Percent Aggregation Error for Regional Choice Shares Prediction by Five Methods

Error Measure	Naive	Statistical Differentials	Classification by City (17)	Classification by Auto Ownership	Classification by Utility Scale[3] (4)
RMS	40.0	121.0	17.9	21.7	3.1

Multinomial Choice (4 alternatives)

Notes:

1. Errors are all based on predictions made for the full sample.
2. The statistical differentials method followed the original Talvitie method (1973), which ignores the covariances of the utilities between alternatives. These were unavailable for this report. It also shows errors larger than for the naive case (Koppelman, 1975) .
3. An approximation was used for the error in the alternative with the smallest choice share.

[1]It is expected that the average choice shares close to zero bias point (equal shares), caused by Koppelman's results to be insensitive to geographic level. The larger the aggregates the less individual cells varied from the low bias equal shares case.

All of the methods except that of statistical differentials reduce error below that of the naive procedure. These results confirm earlier tests that the Taylor series approximation produces counterproductive results due to its poor convergence properties on large variance data. Unfortunately, such data is where the correction is important. Geographic classification, as shown above, reduces error but insufficiently.

The predictions by classes of auto ownership also reduce the majority of the naive error but are not very accurate at the regional scale of forecasting. Koppelman showed auto availability classification to reduce naive error by one-third, with a result of three percent. Our classifier halved naive error, but from a larger base. Hence while specific variable classification may be acceptable for one data set and geographic scale, it may not be for another.

The method of utility classification gives a much greater reduction relative to naive error than the other methods. Only four class cells were used. They were defined by successive divisions of the sample about the means of the distribution of the difference between pairs of utilities (see the utility classification section above describing this method). Obviously, this procedure is more efficient than the others. This is to be expected at the regional level where the total utility variance of the sample is present. It would also be true within small aggregates unless a much smaller number of variables dominated the variability of the sample. This is not the case for within-city aggregates, as seen from Table 68.

Table 72 shows the variation in aggregation error by cell count for the utility classification method. These results are shown for a three-alternative subset of the four-mode choice model above (auto alone, bus-with-walk-access and shared ride). The cell definition criteria were the same. The error can be reduced below all of the other methods tested with only two utility classes. With eight it becomes negligible. The instability in error reduction with cell count is partly due to the *ad hoc* cell definition criteria used. It is also due to the non-linearity of the RMS measure; an error measure composed of a linear sum of absolute errors in each alternative decreased monotonically with cell count.

TABLE 72 Percent Aggregation Error by Cell Count for the Utility Classification Method for Regional Share Predictions

Error Measure	Total Class Cell Count			
	2	4	6	8
RMS	6.4	2.3	2.6	0.5

Multinomial Choice (3 alternatives)

Some perspective on the importance of these errors is gained by noting examples of model and data error--the two other contributions to total forecasting error. As with aggregation, these errors are influenced by many factors. A model, forecasting on the same data and sample on which it was estimated, predicts exactly by definition, as long as alternative-specific constants are allowed. Errors arise as it is transferred to new samples, with sample size, with changes in the environment, due to unobserved factors, not modeled over time and with different or erroneous data collection.

Two empirical results of model and data error are illustrative. Koppelman (1975) found a 19.7 percent RMS error from these sources in predicting super-district shares with a model calibrated on a sample of the Washington area data. This study found a twenty-three percent RMS error in forecasting the redistribution of choice shares among four modes using a new data set (eighty percent of which was a new sample) and with the passage of three years. These errors are less than those for naive aggregation but more than most of the remaining methods.

Conclusions

Some general conclusions and guidelines can be made, followed by more specific recommendations for each of the planning categories defined earlier.

Aggregation error cannot be ignored, that is, the naive method cannot be used. Even where aggregates are small and other errors are large, such as in four-step interzonal forecasting, it would usually be better to use models calibrated, as well as forecasting, with aggregate data than accept naive error.

A classification procedure, especially one based on the utility scales, will be superior if efficient, rather than enumerated forecasts are desired. When classification is by specific variables, many more cross-classifications will be needed than on the basis of utility scale values. The utility classification procedure is useless for non-simply scalable models such as those of Part IV, Chapter 2. Classification by variables must be used with that category of models. The statistical differentials method is counter-productive.

For the traditional regional forecasting environment with UTPS (four-step) type systems, it is necessary to obtain some individual zonal data, most profitably socioeconomic variable values, in order to construct at least a few classes in each interchange. This will be more important if the disaggregate models being used to modify aggregate forms of these procedures have high ambitions for response to issues, thus including many variables, or if they employ joint-choice.

Policy analysis forecasting on single groups (non-segmented) can take advantage of the accuracy of the enumeration method unless their objective for sub-aggregate outputs, quick response or low cost use is important. In either case, they can best use the utility classification method unless non-simply scalable models are being used. In this case, the intra-aggregate utility component covariance matrix analysis discussed in the State of the Theory section above may be used to most efficiently define the classes.

The Special Area studies can profitably employ the utility classification procedure if disaggregate data is available. Even if it is not, that principle of attempting to classify the sample along lines of homogeneous choice groups should be followed in collecting and grouping aggregate or approximate data for forecasting.

Generally, policy analysis systems employing simultaneous choice or iterations for equilibration are the ones that require the simplest form of classification to reduce each component of the computation effort. The message for data requirements in all of these categories is that some degree of disaggregate data information within the forecasting aggregates must be available or recoverable. Otherwise strictly aggregate methods are superior. UTPS frameworks can get by on a small additional amount of individual data. The policy analysis categories require full disaggregate household data and, at least, interzonal level-of-service data.[1]

Chapters 3 and 4 of Part IV have dealt with the problem of deriving disaggregate data from more conventional sources when it cannot be collected directly.

Classification methods of forecasting, especially those based on selecting homogeneous utility or choice groups, realize the full potential for efficiency of disaggregate models in transportation planning that was first seen when it was found that these models could be calibrated on a fraction of the observations necessary for aggregate models. Now forecasting can be done by identifying only a limited number of utility classes, that is, decision-maker types. Forecasts for any aggregates or sub-aggregates are simply a process of determining the relative numbers of the population in each of these classes. Thus, each class will have associated with it the number of travelers in each forecasting--aggregate, socioeconomic, so forth--that has utilities falling within the class. The total aggregate outputs are just the sums of the shares weighted by the aggregate populations. The forecasting problem has been reduced from one of computing choice shares for great numbers of aggregate outputs to locating the proportions of these aggregate cases in a small number of behavioral market segments. More in-depth data is necessary, but forecasting computations are greatly reduced.

Volume VII of the Urban Travel Demand Forecasting Project Final Report Series <u>Aggregation Methods and Tests</u>, gives more details on the subjects of this chapter.

[1]Omission of intrazonal transit access data in these forecasts has been found in this study to give a 4.2 percent RMS error in average zonal shares.