

AGGREGATE TRAVEL DEMAND FORECASTING FROM DISAGGREGATED BEHAVIORAL MODELS

Daniel McFadden and Fred Reid, University of California, Berkeley

This paper derives an expression for obtaining aggregate (interzonal) travel forecasts given a probit disaggregate demand model and zonal averages and intrazonal variances of the independent variables. It also derives expressions for the biases in aggregate model calibrations resulting from zonal homogeneity assumptions in the variables. The conditions under which these biases are important are discussed. Expressions are also determined for obtaining consistent, unbiased estimates for both aggregate and disaggregate models that take into account nonhomogeneous zones and practical data limitations.

•FORECASTS of aggregate travel demand between zones, determined as a function of policy instruments, are a basic input to the transportation planning process. Traditionally, these forecasts have been based on models of aggregate interzonal flows, calibrated by using zonal average trip attributes and socioeconomic characteristics. Interzonal travel demand is, tautologically, the result of aggregating individual travel decisions for the zonal population. Traditional aggregate models do not attempt to make this connection. By contrast, the recent travel demand literature has emphasized the individual decision-making unit and the behavioral foundations of travel demand (1, 2, 4, 8, 9).

Aggregate and disaggregate travel demand models are sometimes viewed by transportation planners as mutually exclusive or competitive approaches to the forecasting problem. We shall argue that, to the contrary, they are complementary. [A good analogy can be made between the problem of modeling travel demand and the physical theory of a perfect gas. The observable macroparameters of a gas, such as pressure and temperature, are linked by empirical laws such as Boyle's law; however, the behavior of individual gas molecules is described by the kinetic theory of gases. If the gas is in equilibrium, then Boyle's law predicts macrobehavior accurately, and it is unnecessary to consider the molecular structure of the gas. However, the empirical laws fail for gases in disequilibrium, and it is necessary to turn to the molecular theory. This requires the further step of describing the distribution of molecular behavior to obtain the desired macroaverages; this is provided by Maxwell statistics. Equilibrium in the analogy to travel demand is identified with the concept of an identical choice environment for all individuals in a zone. Aggregate demand models, like the empirical gas laws, should forecast accurately under this homogeneity condition but cannot be expected to succeed when zones are heterogeneous. Disaggregated behavioral models, therefore, should have the properties of implying the aggregate models under the homogeneity conditions and of forecasting correctly under heterogeneity conditions. The specification of the distribution of individual behavior in the population (as with Maxwell statistics applied to gases in disequilibrium) will play a critical role in the disaggregate travel demand model.] The disaggregate models provide a theoretical foundation for the aggregate models and provide conditions under which the aggregate models will give valid forecasts. The aggregate models may provide the most convenient means of forecasting when zonal homogeneity conditions are met.

This paper demonstrates that the aggregate and disaggregate models have a common foundation and that it may be possible to use a synthesis of the models to facilitate calibration and improve forecasting accuracy. The following questions are answered:

1. Given a correctly calibrated disaggregated behavioral model, how can aggregate interzonal flows be forecast?
2. What biases are introduced in the calibration of aggregate models when zones are not homogeneous? How do these biases affect demand forecasts? What corrections can be applied?
3. What biases are introduced in the calibration of disaggregated behavioral models when some independent variables are approximated by zonal averages? What corrections can be applied?

DISAGGREGATED BEHAVIORAL MODELS

The axioms of disaggregated behavioral demand modeling are that individuals represent the basic decision-making unit and that each individual will choose one alternative among those available that he or she finds most desirable or useful. This depends on the attributes of the alternative and the socioeconomic characteristics of the individual. For simplicity, we will develop this model and the results only for the classical problem of modal split for a work trip. The substance of our conclusions continues to hold for more complex aspects of travel behavior. The full power and generality of the disaggregated approach become apparent primarily in forecasting complex demand systems.

Suppose an urban area is partitioned into zones, indexed $1, \dots, J$. Let $k = 1, \dots, K$ index the individuals in the population, and let N_{ij} denote the set of individuals commuting from zone i to zone j . Each individual is assumed to face a binary modal choice, and a and b are the modes. Now, consider a specific individual k having a vector of socioeconomic characteristics s^k and facing vectors of attributes x^{ak} and x^{bk} for the two modes. According to the behavioral model, the individual will have a utility function of $u = U^k(x, s^k)$, which summarizes the desirability of a mode with attributes x , and will choose the alternative that gives the higher utility; i.e., mode a will be chosen if $U^k(x^{ak}, s^k) > U^k(x^{bk}, s^k)$. Not all attributes of alternatives and socioeconomic characteristics determining tastes can be measured. Consequently, the utility function of an individual drawn randomly from the population can be thought of as containing a random component reflecting his or her unmeasured idiosyncrasies in tastes. We therefore write the utility function in the form

$$U^k(x, s) = V(x, s) + \epsilon^k(x, s) \quad (1)$$

where $V(x, s)$ is common to all members of the population and can be interpreted as the representative utility, and $\epsilon^k(x, s)$ is the random component. The condition for mode a to be chosen by individual k can now be written

$$\epsilon^k(x^{bk}, s^k) - \epsilon^k(x^{ak}, s^k) < V(x^{ak}, s^k) - V(x^{bk}, s^k) \quad (2)$$

The unmeasured term on the left side of Eq. 2 has a statistical distribution in the population and a cumulative distribution function G . Consider an individual drawn randomly from the subpopulation who has socioeconomic characteristics s^k and who faces alternative trip attribute vectors x^{ak} and x^{bk} . The probability that this individual will choose mode a is given by

$$\begin{aligned} P &= \text{Prob}[V(x^{ak}, s^k) + \epsilon^k(x^{ak}, s^k) > V(x^{bk}, s^k) + \epsilon^k(x^{bk}, s^k)] \\ &= G[V(x^{ak}, s^k) - V(x^{bk}, s^k)] \end{aligned} \quad (3)$$

The methodology of disaggregate behavioral demand analysis is to specify parametric functional forms for the representative utility function V and the distribution function G . [In general, the parameters of the distribution function G will be functions of the measured trip attributes and socioeconomic characteristics of the subpopulation. We suppress these arguments and later assume G to be independent of these variables. This is a strong restriction which is inconsistent with some descriptions of the structure of taste variation in the population. For example, the disaggregated behavioral model of Quandt (8) is incompatible with this restriction. A more complete discussion of this specification is given in Domencich and McFadden (4).] Observed choices from a sample are treated as drawings from binomial distributions, and the probabilities are given in Eq. 3. Statistical methods such as the maximum likelihood procedure are used to calibrate the unknown parameters. Following this procedure, we first specify that the representative utility function be linear in unknown parameters, and the form is

$$V(x, s) = \sum_{\ell=1}^L \beta_{\ell} Z^{\ell}(x, s) \quad (4)$$

where the $Z^{\ell}(x, s)$ are numerical functions. [Numerical functions Z^{ℓ} can be simple or complex functions of the trip attributes and socioeconomic characteristics. For example, Z^{ℓ} may be a trip attribute such as trip cost or on-vehicle time or a transformation such as the logarithm or square of one of these variables. It may be a dummy variable that is one for mode a and zero for mode b, which corresponds to a pure mode preference effect. It may involve interactions of trip attributes and socioeconomic variables such as trip cost divided by trip time, trip cost divided by wage, or trip walking time times an index of physical health. It may involve interactions of trip attributes or socioeconomic variables with mode dummy variables, for example, a variable that is income for mode a and zero for mode b or on-vehicle time for mode b and zero for mode a. Because the utility comparison involves only differences of the Z^{ℓ} variables, choice will be influenced only by factors that vary between modes. This means a pure socioeconomic characteristic such as income, unless interacted with a pure mode preference dummy, will not influence tastes and should not be included as a Z^{ℓ} variable. However, it should be noted that an arbitrary smooth utility function $V(x, s)$ can be approximated as closely as desired by the form in Eq. 3. An important class of models will be those in which no pure mode effects appear either singly or in interactions with other variables. These generic models can be used to forecast new mode demand.]

Defining $z_1^k = Z^{\ell}(x^{ak}, s^k) - Z^{\ell}(x^{bk}, s^k)$ and $z^k = z_1^k, \dots, z_L^k$ and letting β denote the column vector of unknown parameters, we rewrite Eq. 3 as

$$P = G(\beta'z^k) \quad (5)$$

Equation 5, derived from the behavioral model, has a simple conventional interpretation. Vector z^k measures the differential attributes of the two modes, which are weighted to account for the effect of measured socioeconomic differences. For example, z_1^k might be the cost of mode a less the cost of mode b, and z_2^k the on-vehicle time of mode a less the on-vehicle time of mode b, which is multiplied by the individual's after-tax wage. The coefficient vector β weights the components of z^k into a single measure $\beta'z^k$ of the differential impedance of the two modes, and G represents the response curve that gives the proportion of individuals choosing mode a at each level of relative impedance. To simplify further analysis, we will assume that the distribution function G is standard normal. This condition allows us to obtain simple closed formulas but is not critical for our general conclusions. If Φ denotes the cumulative standard normal distribution, Eq. 5 has a final form

$$P = \Phi(\beta' z^k) \quad (6)$$

that determines the probability of the choice of mode a by an individual drawn randomly from the subpopulation of individuals facing a vector z^k of differential attributes of the two modes. This form in statistics and transportation demand analysis is known as the probit model.

AGGREGATE MODAL SPLITS FROM DISAGGREGATED MODELS

Suppose that the disaggregated behavioral model in Eq. 6 is a correct specification of the distribution of individual modal choices in the population and that the parameter vector β has been calibrated accurately in a statistical study. We now wish to determine the aggregate modal split between zones i and j . The formula for this aggregate is straightforward. Recall that $N_{i,j}$ is the set of individuals k traveling from i to j and let $n_{i,j}$ be the number of individuals in this set. Because $P_k = \Phi(\beta' z^k)$ is our best prediction that an individual with a measured choice environment described by z^k will choose mode a , the best prediction for the aggregate of individuals making the trip is

$$P_{i,j} = \frac{1}{n_{i,j}} \sum_{k \in N_{i,j}} \Phi(\beta' z^k) \quad (7)$$

Equation 7 is just the expectation of the response probability of the empirical distribution of the vector of the independent variables z . When $n_{i,j}$ is large, this formula is closely approximated by the expectation of the response probability of the underlying distribution of the independent variables. We assume that the distribution of z for individuals in $N_{i,j}$ is normal; the mean is $\bar{z}^{i,j}$ and the covariance matrix is $A^{i,j}$. (This is a plausible large sample approximation even when the individual observations are clearly nonnormal, e.g., discrete. It should be noted that some of the conclusions, e.g., the result concerning the consistency of disaggregated models calibrated from zonal averages, depend critically on the symmetry of this distribution and would change substantially if the distribution were skewed.) Then, the differential impedance $y = \beta' z$ is distributed normally with mean $\beta' \bar{z}^{i,j}$ and variance $\sigma_{i,j}^2 = \beta' A^{i,j} \beta$. Therefore, the expectation of the response probability is

$$P_{i,j} = \int_{-\infty}^{+\infty} \Phi(y) \frac{1}{\sigma_{i,j}} \phi\left(\frac{y - \beta' \bar{z}^{i,j}}{\sigma_{i,j}}\right) dy \quad (8)$$

Equation 8 can be simplified further by using the convolution properties of normal distributions. In general, if W_1 and W_2 are independent normal random variables with mean μ_1 and variance σ_1^2 for $i = 1, 2$, one has $W_1 - W_2$ distributed normally with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$, and this implies

$$\begin{aligned} \text{Prob}[W_1 - W_2 \leq w] &= \Phi \frac{w - \mu_1 + \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \\ &= \int_{-\infty}^{+\infty} \Phi\left(\frac{w + y - \mu_1}{\sigma_1}\right) \frac{1}{\sigma_2} \phi\left(\frac{y - \mu_2}{\sigma_2}\right) dy \end{aligned} \quad (9)$$

Setting $w = \mu_1 = 0$, $\sigma_1 = 1$, $\sigma_2 = \sigma_{1j}$, and $\mu_2 = \beta' \bar{z}^{1j}$, Eq. 9 implies a final form for the aggregate modal split between zones i and j :

$$P_{1j} = \Phi \left(\frac{\beta' \bar{z}^{1j}}{\sqrt{1 + \sigma_{1j}^2}} \right) \quad (10)$$

Equation 10 can also be obtained directly by returning to the condition given in Eq. 2 for an individual to choose mode a . Given that $V(x^{ak}, s^k) - V(x^{bk}, s^k) = \beta' z^k$, this condition becomes

$$\epsilon^k(x^{ak}, s^k) - \epsilon^k(x^{bk}, s^k) - \beta' z^k < 0 \quad (11)$$

The first two terms in Eq. 11 taken together are assumed to have a standard normal distribution, and $\beta' z^k$ is assumed to be normal with mean $\beta' \bar{z}^{1j}$ and variance σ_k^2 . Therefore, the left side of Eq. 11 is normal with mean $-\beta' \bar{z}^{1j}$ and variance $1 + \sigma_{1j}^2$. This implies that the probability of the event in Eq. 11 is given by Eq. 10.

By comparing Eqs. 10 and 6 for the individual response probability, one sees that the components of differential impedance have the same relative weights, given by the component of β , but that the effect of the mean differential impedance for the zones is attenuated by the factor $\sqrt{1 + \sigma_{1j}^2}$, which reflects the degree of heterogeneity of the variables facing the individuals in the zone. If the zones are homogeneous, so that each individual has the same socioeconomic characteristics and faces the same trip attributes yielding $\sigma_{1j}^2 = \beta' A^{1j} \beta = 0$, then the aggregate and disaggregate models coincide. This conclusion provides a condition under which the disaggregated model can be calibrated directly from interzonal data.

Equation 10 provides a method of forecasting interzonal modal splits from a knowledge of the zonal average \bar{z}^{1j} of the variables entering differential impedance and of the intrazonal covariances A^{1j} of these variables. It is important to note that one does not require a sample of individuals going from i to j or observations on individual data points, although an alternative approach to computing the aggregate flows is to use Eq. 7 directly for a random sample from the population. The zonal means \bar{z}^{1j} are often available from transportation surveys; the covariances A^{1j} are usually not reported but could be constructed from the underlying data. The effect of a policy change can be forecast from Eq. 10, provided the effect of the policy on \bar{z}^{1j} and A^{1j} can be determined. The most straightforward case is a policy change that has a homogeneous impact on the zone, as for example a \$0.05 increase in the basic transit fare. This changes the cost component of \bar{z}^{1j} and leaves A^{1j} unchanged. A more complex example would be an increase in parking charges given that zones are heterogeneous with respect to the availability of free parking to different individuals. This increase would change the corresponding component of \bar{z}^{1j} but would also spread the distribution of parking charges more widely, which would increase the variance of this component in A^{1j} and possibly also change the covariance of this component and other variables. Specifying the precise effect of a policy change on A^{1j} may be challenging; various approximations may become necessary, including the extreme approximation implicit in conventional aggregate models that A^{1j} is always zero and is therefore unchanged by policy. The error introduced by this last approximation may be substantial.

An empirical example shows the order of magnitude of the factor $\sqrt{1 + \sigma_{1j}^2}$ in Eq. 10. We consider an automobile-bus modal-split model calibrated on 160 individual workers in the San Francisco Bay area (7, model 1). This model (7) has one pure socioeconomic variable (income), one pure transportation variable (cost difference), and four mixed variables defined by the after-tax wage times the time difference for on-vehicle, walk, initial wait, and transfer wait times. Table 1 gives means and standard errors of the

Table 1. Overall means and standard errors for population sample.

Variable				Normalized Standard Error of Intrazonal Variances	
Number	Description	Mean	Standard Error	Probit Coefficient	
1	Family income, ≤ \$10,000/year	8,670	2,238.3	0.0000391	1.46
2	Out-of-pocket cost per round trip ^a , cents	0.88	102.5	-0.00551	6.98
3	On-vehicle time ^a , one way × net wage, \$/hour	-51.7	65.0	-0.00514	1.40
4	Walk time ^a × net wage, \$/hour	-40.2	64.6	-0.000055	5.39
5	Initial wait time ^a × net wage, \$/hour	-37.4	33.2	-0.0103	1.39
6	Transfer wait time ^a × net wage, \$/hour	-28.3	42.1	-0.0114	3.32
7	Automobile dummy	1	—	0.0898	—

^aFor automobile and bus trips.

Table 2. Total correlation matrix.

Variable	Variable					
	1	2	3	4	5	6
1	1	0.224	-0.307	-0.372	-0.303	-0.27
2	0.224	1	0.195	-0.294	0.206	0.146
3	-0.307	0.195	1	0.0913	0.518	0.597
4	-0.372	-0.294	0.0913	1	0.0746	0.0597
5	-0.303	0.206	0.518	0.0746	1	0.642
6	-0.27	0.146	0.597	0.0597	0.642	1

Table 3. Average intrazonal covariance matrix as proportion of total covariance matrix for population sample.

Variable	Variable					
	1	2	3	4	5	6
1	0.91	-0.332	2.25	0.324	1.19	0.464
2	-0.332	0.136	0.156	-0.173	-0.14	-0.223
3	2.25	0.156	0.652	1.07	0.568	0.191
4	0.324	-0.173	1.07	0.168	0.88	0.756
5	1.19	-0.14	0.568	0.88	0.45	0.4
6	0.464	-0.223	0.191	0.756	0.4	0.26

variables and their probit coefficients β . Table 2 gives the overall correlation matrix of these variables. Using the MTC 440 zone network for the San Francisco Bay area, we grouped the sample into zones. Assuming a common intrazonal covariance matrix $A = A_{1j}$, we obtained the estimate of A (expressed as a proportion of the overall covariance matrix), which is given in Table 3. We note that 91 percent of income variation is intrazonal. The percentage is only 13.6 for cost difference variation but is as high as 65 for one of the variables for mixed wage times cost. For this A matrix, $\sigma_{1j}^2 = \beta' A \beta = 0.485$ and the factor $\sqrt{1 + \sigma_{1j}^2} = 1.22$. This implies that the elasticity of any independent variable of aggregate demand for the first mode between zones i and j will be 82 percent of the average of the corresponding individual demand elasticities for trips from i to j . [This is a conservative estimate of the bias in typical modal-split models because (a) the calibration on which our calculation is based underweights transit walk time because of a failure to distinguish walk access and automobile access transit trips and (b) a variety of additional automobile availability and socioeconomic variables are excluded from the model. In each of these cases, the contribution of intrazonal variance can be expected to be a high proportion of the total.] We conclude that direct application to aggregate modal splits of elasticities calculated from disaggregate models will tend to overestimate the magnitude of demand response, and

elasticities calculated from aggregate data will underestimate, on the average, individual demand elasticities.

To determine whether the assumption of a common intrazonal matrix $A = A_{i,j}$ was a good one, we computed the standard errors over the origin-destination pairs in our sample of the diagonal variance elements in the $A_{i,j}$. These standard errors divided by their means (i.e., the diagonal elements of A) are given in Table 1. We conclude that there is substantial variation in the $A_{i,j}$ over origin-destination pairs and that it would be a poor practice in applications to assume a common A matrix. Finally, we note that over half the contribution to the variance $\sigma_{i,j}^2 = \beta A' \beta$ in our example is contributed by off-diagonal (covariance) terms. Thus, it is not sufficient to look only at the variance elements of A . Because of the small data base, these conclusions are necessarily tentative and are intended only to suggest orders of magnitude.

With any sigmoid-shaped choice function, attempts to represent aggregate modal splits by the disaggregated function with homogeneous zone assumptions will similarly bias the results. The bias will be toward the extreme possible values of the modal split—too low for values of Eq. 6 below 0.5, too high above 0.5. Only in the special cases where zones are homogeneous or where interzonal splits are equal to 0.5 and their differential impedances are distributed symmetrically about their mean would there be no bias. The magnitude of the bias found in the example is sufficient to require that the zone variances be accounted for in the aggregation process.

CALIBRATION OF AGGREGATE MODELS

We next examine the effects of calibrating an aggregate model directly from interzonal flows when, in fact, individual responses conform to the disaggregate behavioral model of Eq. 6. The typical aggregate modal-split model assumes that the frequency of choice of mode a for trips between i and j is given by a function of the differential impedance of the two modes, and differential impedance is measured in turn by a weighted combination of the zonal average differential attributes of the two modes. Given that the response curve is normal, this model is

$$P_{i,j} = \Phi(\gamma' \bar{z}^{i,j}) \quad (12)$$

where γ is the vector of unknown parameters weighting the components of differential trip attributes into a measure of differential impedance. Although Eq. 12 is identical to Eq. 6 for individual response probability, it would be justified in traditional aggregate demand analysis on the basis of its success as an empirical law rather than on behavioral grounds. Calibration of the model of Eq. 12 from data on interzonal flows would typically be done by applying least squares for origin-destination pairs to the equation

$$\Phi^{-1}(\bar{P}_{i,j}) = \gamma' \bar{z}^{i,j} + \eta^{i,j} \quad (13)$$

where Φ^{-1} is the inverse cumulative standard normal distribution, $\eta^{i,j}$ is an unobserved error, and $\bar{P}_{i,j}$ are the observed frequencies of choice of mode a between zones. Except for possible adjustments introduced by use of weighted least squares, the resulting estimator is

$$\hat{\gamma} = \left[\sum_{i,j} \bar{z}^{i,j} \bar{z}^{i,j'} \right]^{-1} \left[\sum_{i,j} \bar{z}^{i,j} \Phi^{-1}(\bar{P}_{i,j}) \right] \quad (14)$$

Now, suppose that travel demand behavior is in fact determined by the disaggregated model in Eq. 6. Aside from sampling error, which is inversely proportional to the number of trips between zones, and which can be ignored when the total number of trips is large, the observed interzonal modal-split frequencies \bar{P}_{ij} from Eq. 10 will satisfy the following:

$$\bar{P}_{ij} = \Phi\left(\frac{\beta' \bar{Z}^{ij}}{\sqrt{1 + \sigma_{ij}^2}}\right) \quad (15)$$

or

$$\Phi^{-1}(\bar{P}_{ij}) = \frac{\beta' \bar{Z}^{ij}}{\sqrt{1 + \sigma_{ij}^2}} \quad (16)$$

If one substitutes Eq. 16 in Eq. 14, the value of $\hat{\gamma}$ converges, with a probability of one as sample size increases, to

$$\gamma = \left[\sum_{i,j} \bar{Z}^{ij} \bar{Z}^{ij'} \right]^{-1} \left[\sum_{i,j} \bar{Z}^{ij} \bar{Z}^{ij'} / \sqrt{1 + \sigma_{ij}^2} \right] \beta \quad (17)$$

From Eq. 17 we can draw inferences on the biases introduced in the estimation of γ by the presence of heterogeneity within the zones. First, we note that, in homogeneous zones where $\sigma_{ij}^2 = 0$, $\hat{\gamma}$ is a consistent estimator of β . When zones are heterogeneous, the effect of the term $\sqrt{1 + \sigma_{ij}^2}$ will generally be to bias the estimates $\hat{\gamma}$ downward in magnitude. If large magnitudes of a component of \bar{Z}^{ij} tend to be associated with large values of σ_{ij}^2 , the bias in the corresponding coefficient will tend to be larger than for a coefficient for which the component of \bar{Z}^{ij} and σ_{ij}^2 tends to be uncorrelated. Given that σ_{ij}^2 is a constant σ^2 for all zones, Eq. 17 reduces to a simple expression for the bias

$$\gamma = \frac{1}{\sqrt{1 + \sigma^2}} \beta \quad (18)$$

The magnitudes calculated previously suggest that $\hat{\gamma}$ will underestimate β in magnitude by 18 percent or more in typical aggregate models.

The seriousness of the bias introduced in the calibration of $\hat{\gamma}$ from interzonal flows depends on the extent to which these biases introduce errors in demand forecasts. Consider, first, the relative bias in various components of $\hat{\gamma}$ and take the example of transit access time versus cost. In a typical transportation study, some zones will be densely serviced by transit, and this will result in low average transit access time and a low intrazonal variance for this variable. Other zones will be sparsely serviced, and this will lead to a high average transit access time and a high variance. This pattern will substantially bias the coefficient of transit access time; however, transit cost will be relatively homogeneous within zones, and this will result in less bias for the coefficient of cost. Therefore, the aggregate model will undervalue transit access time relative to cost and will therefore underpredict the net increase in transit demand, which would occur because of a policy change increasing the density of transit routes as well as transit cost.

Aside from the distortions introduced by differential biases in parameter estimates, the forecasting success of the calibrated aggregate model will depend on whether the policy change under study affects the intrazonal covariance matrix for the independent variables. Given that $\sigma_{i,j}^2$ is constant for all zones and that a policy change affects $\bar{Z}^{i,j}$ but leaves $A^{i,j}$ unchanged, the aggregate model forecasts correctly. Consider, for example, a uniform increase in the base transit fare by comparing Eqs. 6 and 10, in which γ is set equal to the calibrated value from Eq. 18. The aggregate model forecasts will err for any policy change that affects $A^{i,j}$; this will be the case particularly for policies that have a heterogeneous impact within the zone. For example, a policy change reducing access times by increasing the number of transit stops will reduce the access-time component in $\bar{Z}^{i,j}$ and the corresponding variance in $A^{i,j}$. The true response given by Eq. 10 will differ from the response predicted by the calibrated aggregate model by a term reflecting the effect of this policy on $\sigma_{i,j}^2$. If the differential impedances are distributed so that the frequency of extreme response probabilities in favor of transit is lower than that in favor of the alternative mode, the effect of decreasing $\sigma_{i,j}^2$ will usually be to decrease transit patronage. As a result, the calibrated aggregate model will overpredict the rise in transit patronage.

The preceding paragraphs have pointed out that, by using the aggregate model of Eq. 12, direct calibration and forecasting can result in substantial prediction biases. This is true except when the zones are homogeneous or when policy change has a homogeneous effect on the zone, which leaves the covariance matrix of the independent variables unchanged. However, when consistent estimates of the intrazonal variances $\sigma_{i,j}$ are available, the estimator of Eq. 17 can be modified to

$$\hat{\gamma} = \left[\sum_{i,j} \bar{Z}^{i,j} \bar{Z}^{i,j'} \sqrt{1 + \sigma_{i,j}^2} \right]^{-1} \left[\sum_{i,j} \bar{Z}^{i,j} \Phi^{-1}(\bar{P}_{i,j}) \right] \quad (19)$$

Equation 19 gives a consistent estimate of the parameter vector β . Then, Eq. 10 can be used with the estimates $\hat{\gamma}$ and $\sigma_{i,j}^2$ to correctly forecast the effects of policy changes that are either homogeneous or heterogeneous in their impact on zones.

In practice, it may be feasible to obtain a consistent estimate of the intrazonal covariance matrix $A^{i,j}$ from external sources, but it may not be feasible to obtain the initial consistent estimators of β necessary to construct estimators of $\sigma_{i,j}^2 = \beta' A^{i,j} \beta$. An alternative is to consider the estimator in Eq. 19 as an implicit function of γ :

$$\left[\sum_{i,j} \bar{Z}^{i,j} \bar{Z}^{i,j'} / \sqrt{1 + \gamma' A^{i,j} \gamma} \right] \gamma = \sum_{i,j} \bar{Z}^{i,j} \Phi^{-1}(\bar{P}_{i,j}) \quad (20)$$

Solution of this system of equations by iterative methods provides a consistent estimator γ^* of β . Given that $A^{i,j}$ is uniform across zones, this equation has the solution

$$\gamma^* = \frac{M_{zz}^{-1} M_{zy}}{\sqrt{1 - M_{zy}' M_{zz}^{-1} A M_{zz}^{-1} M_{zy}}} \quad (21)$$

where

$$M_{zz} = \sum_{i,j} \bar{Z}^{i,j} \bar{Z}^{i,j'}$$

and

$$M_{zy} = \sum_{i,j} \bar{z}^{ij} \Phi^{-1}(\bar{P}_{ij})$$

The difference in the denominator of Eq. 21 is likely to result in rather unsatisfactory statistical properties of this estimator in small samples. In particular, this estimator may fail to exist. Equation 21 may provide a useful initial step for iterative solutions of Eq. 20 for unequal A^{ij} if we initially use some average A of the A^{ij} . We conclude that an efficient and practical procedure for demand forecasting may be (a) to obtain consistent estimators of the A^{ij} covariance matrices and the behavioral parameters β , (b) to use these estimates to obtain estimates of $\sigma_{ij}^2 = \beta' A^{ij} \beta$ and $\sqrt{1 + \sigma_{ij}^2}$, (c) to use interzonal flows in the adjusted aggregate model estimator in Eq. 19 to obtain a more precise estimator of β , and (d) to use these estimates in Eq. 10 to obtain consistent forecasts of the effect of a transportation policy.

CALIBRATION OF DISAGGREGATED MODELS WITH AGGREGATE DATA

Given data for a sample of individuals, which include accurate measurements of the attributes of the transportation alternatives faced by each person, the disaggregated model of Eq. 6 can be calibrated by straightforward application of a variety of statistical techniques. These techniques include maximum likelihood procedures and, under suitable circumstances, minimum chi-square procedures (3, 6). A second approach to calibration is to use the conclusions of the preceding discussion to estimate the behavioral model from zonal data when the intrazonal covariance matrix of the independent variables is known or can be estimated consistently from external sources.

In practice, a third approach is important. Data on choices, socioeconomic characteristics, and some attributes of alternative trips are collected for a sample of individuals, and the remaining attributes of trips are measured only by zonal averages. For example, individual data may be collected on income, age, and travel costs, and zonal averages obtained from transportation grids may be used for access and on-vehicle travel times. This introduces a measurement error in the independent variables, which may bias the estimates obtained by applying the statistical methods ordinarily used in calibrating disaggregated models. (The reason calibration of a mixed model with individual response frequencies and interzonal average explanatory variables gives consistent estimates and calibration of an aggregate model with zonal average response frequencies does not give such estimates is that an arithmetic average of inverse cumulative normal transformations of frequencies does not equal the inverse cumulative normal transformation of the arithmetic average of the frequencies. It is this nonlinearity of the response curve that makes it necessary to distinguish disaggregate and aggregate models.)

We will now determine the conditions under which biases will occur and derive correction formulas. The first useful conclusion is that replacing all independent variables by zonal averages yields consistent estimates of the parameters of the behavioral model when the usual statistical methods are applied. (This conclusion depends critically on the assumption that the intrazonal distributions of the independent variables are not skewed. If, to the contrary, these distributions are skewed, and modes lie between the zonal mean vectors and the total population mean vector, then the use of zonal averages introduces a regression-to-the-mean effect, which generally biases estimates downward in magnitude.) (Note that individual observations are used as the dependent variables and, thus, differ from the zonal average frequencies used in the previous section.) The loss of efficiency in estimation resulting from use of this procedure may be substantial because not all information is being used. However, in very large

samples, this loss may be offset by the saving in cost of providing accurate measures of the travel attributes for each individual. The second result provides a straightforward linear transformation of the estimates obtained from the usual statistical procedures, which makes the resulting estimators consistent. These correction formulas require external estimates of the intrazonal variance of the independent variables.

The behavioral model of Eq. 6 can be written as

$$\Phi^{-1}(P_k) = z^k \beta \quad (22)$$

Partition $z^k = (z_1^k, z_2^k)$, where z_1^k is the (possibly empty) subvector of components that can be measured for each individual and z_2^k is the subvector of components that will be approximated in the estimation process by the zonal averages. Then,

$$\Phi^{-1}(P_k) = z_1^k \beta_1 + z_2^k \beta_2 \quad (23)$$

where β is partitioned commensurately. It is convenient to cast the calibration process into a regression format by assuming there are a large number of individuals of each type k and that \bar{P}_k is the response frequency of this homogeneous group. Then,

$$y^k \equiv \Phi^{-1}(\bar{P}_k) = z_1^k \beta_1 + z_2^k \beta_2 + \xi^k \quad (24)$$

where ξ^k are error terms with a zero mean, which converge to zero with a probability of one as the numbers of members of the groups increase. Suppose now that z_2^k is replaced by the zonal average \bar{z}_2^j and apply ordinary least squares to the equation.

$$y^k = z_1^k \alpha_1 + \bar{z}_2^j \alpha_2 + \epsilon^k \quad (25)$$

This general estimation procedure is known as Berkson's method, which under a correct specification of the independent variables provides estimates of the parameters, which are consistent and equivalent in asymptotically large samples to the maximum likelihood estimator.

We rewrite Eq. 25 in vector notation as

$$y = z_1 \alpha_1 + \bar{z}_2 \alpha_2 + \epsilon \quad (26)$$

where y and ϵ are column vectors with components y^k and ϵ^k respectively, and z_1 and \bar{z}_2 are matrices with rows z_1^k and \bar{z}_2^j respectively. The ordinary least squares estimates from Eq. 26 then satisfy

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} z_1' z_1 & z_1' \bar{z}_2 \\ \bar{z}_2' z_1 & \bar{z}_2' \bar{z}_2 \end{pmatrix}^{-1} \begin{pmatrix} z_1' y \\ \bar{z}_2' y \end{pmatrix} \quad (27)$$

But Eq. 24, written in the vector notation used above, is $y = z_1 \beta_1 + z_2 \beta_2 + \xi$, and this implies

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} z_1' z_1 & z_1' \bar{z}_2 \\ \bar{z}_2' z_1 & \bar{z}_2' \bar{z}_2 \end{pmatrix}^{-1} \begin{pmatrix} z_1' z_1 & z_1' z_2 \\ \bar{z}_2' z_1 & \bar{z}_2' z_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} z_1' z_1 & z_1' \bar{z}_2 \\ \bar{z}_2' z_1 & \bar{z}_2' \bar{z}_2 \end{pmatrix}^{-1} \begin{pmatrix} z_1' \xi \\ \bar{z}_2' \xi \end{pmatrix} \quad (28)$$

The last term in this expression converges to zero with a probability of one as the number of individuals of each type becomes large. Then $\hat{\alpha}_1$, $\hat{\alpha}_2$ converge (with a probability of one) in this limit to values α_1 , α_2 and satisfy

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} z_1' z_1 & z_1' \bar{z}_2 \\ \bar{z}_2' z_1 & \bar{z}_2' \bar{z}_2 \end{pmatrix}^{-1} \begin{pmatrix} z_1' z_1 & z_1' z_2 \\ \bar{z}_2' z_1 & \bar{z}_2' z_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad (29)$$

We will now consider further the structure of the matrices in this expression. Define E to be a square block-diagonal matrix with a block for each zone pair i, j of the form $(1/n_{i,j}) e^{i,j} e^{i,j}'$, where $e^{i,j}$ is a column vector of ones of length $n_{i,j}$. Then, E satisfies

$$\bar{z}_2 = E z_2 \quad (30)$$

and

$$E^2 = E$$

The average of the intrazonal covariance matrices $A^{i,j}$ introduced previously will be denoted by A and can be defined in terms of the matrix E :

$$\begin{aligned} nA &= \sum_{i,j} n_{i,j} A^{i,j} = \sum_{i,j} \sum_{k \in N_{i,j}} (z^k - \bar{z}^{i,j})(z^k - \bar{z}^{i,j})' \\ &= (z_1 - \bar{z}_1, z_2 - \bar{z}_2)' (z_1 - \bar{z}_1, z_2 - \bar{z}_2) \\ &= (z_1, z_2)' (I - E)(z_1, z_2) \end{aligned} \quad (31)$$

Partition

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad (32)$$

commensurately with (z_1, z_2) . Then, define M_{kr} from Eq. 31 as

$$\begin{aligned} nM_{11} &\equiv z_1' z_1 & nM_{12} &\equiv z_1' z_2 = z_1' z_2 - nA_{12} \\ nM_{21} &\equiv \bar{z}_2' z_1 = z_2' E z_1 = z_2' z_1 - nA_{21} \end{aligned} \quad (33)$$

$$nM_{22} \equiv \bar{z}'_2 \bar{z}_2 = \bar{z}_2 E z_2 = z'_2 z_2 - nA_{22} \quad (34)$$

Equation 29 can be written

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} z'_1 z_1 & z'_1 z_2 \\ \bar{z}'_2 z_1 & \bar{z}'_2 z_2 \end{pmatrix}^{-1} \begin{pmatrix} z'_1 z_1 & z'_1 \bar{z}_2 \\ \bar{z}'_2 z_1 & \bar{z}'_2 \bar{z}_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad (35)$$

$$= \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} - \begin{pmatrix} z'_1 z_1 & z'_1 z_2 \\ \bar{z}'_2 z_1 & \bar{z}'_2 z_2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & z'_1(z_2 - \bar{z}_2) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad (36)$$

$$= \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} - \frac{1}{n} \begin{pmatrix} M_{11} & M_{12} + A_{12} \\ M_{21} & M_{22} \end{pmatrix}^{-1} \begin{pmatrix} 0 & nA_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad (37)$$

Define

$$C = M_{22} - M_{21} M_{11}^{-1} (M_{12} + A_{12}) \quad (38)$$

Then, writing out the expression for the inverse of a partitioned matrix, we obtain

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} - \begin{pmatrix} M_{11}^{-1} (M_{12} + A_{12}) C^{-1} M_{21} + I_{11} \\ -C^{-1} M_{21} \end{pmatrix} M_{11}^{-1} A_{12} \alpha_2 \quad (39)$$

The second term in Eq. 39 reflects the bias in the estimators of the disaggregated model introduced by using zonal average values for the variables z_2 . First, we note that, if the intrazonal correlations of the variables measured by zonal averages are all zero, i.e., $A_{12} = 0$, then the usual estimators are consistent, and no correction is necessary. This is true, in particular, if all variables are measured as zonal averages so that z_1 is empty. Second, we note that the magnitude of the bias is determined by the degree of intrazonal variability relative to total variability. If, for example, $A_{12} = \theta M_{12}$ for some θ with $0 \leq \theta < 1$, then

$$\text{bias} = -\theta \begin{pmatrix} (1 + \theta) M_{11}^{-1} M_{12} \tilde{C}^{-1} M_{21} + I_{11} \\ -\tilde{C}^{-1} M_{21} \end{pmatrix} M_{11}^{-1} M_{12} \alpha_2 \quad (40)$$

where $\tilde{C} = M_{22} - (1 + \theta) M_{21} M_{11}^{-1} M_{12}$. Third, Eq. 39 provides a correction for the usual disaggregate estimators, which makes them consistent for this problem. Application of this correction requires that consistent estimates of A_{12} be obtained from external sources. Estimates of A might be obtained from previous transportation surveys or from more limited data sets when it is possible to limit, a priori, the structure of A . An example of the latter construction would be an estimate of access-time covariance with income in each zone based on the geometry of the zone, the location of transit routes, and census block statistics on income.

SUMMARY

This paper has established links between aggregate travel demand models and disaggregate behavioral models. Equation 10 provides a formula for computing interzonal flows, given a calibrated disaggregate model, and zonal averages and intrazonal variances for the independent variables. Biases resulting from direct calibration of an aggregate model when individual behavior conforms to the disaggregate model are derived. The implications of these biases for forecasting are discussed. Equation 20 provides a method for obtaining consistent estimates of the parameters of the disaggregate model from interzonal flow data. Biases introduced in calibration of disaggregate models when some independent variables are approximated by zonal averages are also discussed. It is shown that the use of zonal averages for all independent variables results in consistent estimates of the parameters of the behavioral model. Equation 39 gives a formula for correcting the bias in the estimates of the behavioral parameters when a mix of individual and zonal average variables is used, and an estimate of the intrazonal covariance matrix can be obtained from external sources.

ACKNOWLEDGMENT

This research has been supported by a grant from the National Science Foundation. This study has not been carried out as part of the research program of the Metropolitan Transportation Commission or the U.S. Department of Transportation, and the views expressed herein do not necessarily reflect the opinions of these agencies.

REFERENCES

1. M. Ben-Akiva. Structure of Travel Demand Models. Transportation Systems Division, Department of Civil Engineering, Massachusetts Institute of Technology, 1972, unpublished.
2. D. Brand. The State of the Art of Travel Demand Forecasting: A Critical Review. Harvard Univ., 1972, unpublished.
3. D. Cox. Analysis of Binary Data. Methuen, 1970.
4. T. Domencich and D. McFadden. Urban Travel Demand—A Behavioral Analysis. North Holland, 1974.
5. T. Lisco. The Value of Commuters' Travel Time—A Study in Urban Transportation. PhD dissertation, Univ. of Chicago, 1967.
6. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers of Econometrics (P. Zarembka, ed.), Academic Press.
7. D. McFadden. The Measurement of Urban Travel Demand. *Journal of Public Economics*, 1974.
8. R. Quandt. *The Demand for Travel: Theory and Measurement*. D. C. Heath and Company, 1970.
9. P. Stopher and T. Lisco. Modelling Travel Demand: A Disaggregate Behavioral Approach—Issues and Applications. Proc., Transportation Research Forum, 1970, pp. 195-214.
10. F. Reid. Travel Time Data Base for the BART Impact Study. BART Impact Program, Metropolitan Transportation Commission, Working paper 2-1-74, 1974.