Transport Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713766937

Automobile classification for choice and demand modelling

David A. Hensher[a]; Nariida C. Smith[a]
[a] School of Economic and Financial Studies, Macquarie University, North Ryde, NSW 2113, Australia

To cite this Article Hensher, David A. and Smith, Nariida C.(1984) 'Automobile classification for choice and demand modelling', Transport Reviews, 4: 3, 245 — 271
To link to this Article: DOI: 10.1080/01441648408716563
URL: http://dx.doi.org/10.1080/01441648408716563

TRANSPORT REVIEWS, 1984, VOL. 4, NO. 3, 245–271

Automobile classification for choice and demand modelling

By DAVID A. HENSHER and NARIIDA C. SMITH
School of Economic and Financial Studies, Macquarie University,
North Ryde, NSW 2113, Australia

ABSTRACT

Classification of automobiles into homogeneous groups is often required for modelling the choice between and demand for automobile types as well as assessing the energy consumption implications of changes over time in the composition of the vehicle stock. Automobiles can be grouped on technological criteria (supply side) and/or behavioural criteria (demand side), the selection of an approach being dependent on the purpose in hand. We propose and outline a method to classify vehicles in a systematic manner. Using a new data set of 827 representative vehicles and over 200 attributes, automobiles are grouped using component score and cluster analyses. Especial consideration is given to the appropriate index of similarity between pairs of attributes where the measurement scales are different. The results provide two sets of groupings of automobiles, one suitable for choice modelling and the other for monitoring trends in energy consumption.

§1. INTRODUCTION

Forecasts of automobile demand on a make and model basis have always been of interest to vehicle manufacturers and only in recent times have other organizations shown a serious concern for the structure of the vehicle stock. Although the fleet mix has always influenced the amount of revenue raised by government in the form of registration fees and petrol taxes, it appears that forecasts of revenue receipts based on a knowledge of total vehicle stock was deemed an adequate procedure. The relatively accessible models and forecasts of total automobile demand, and average vehicle usage figures collected for a wide range of uses, made this practice appealing.

When the need for more detailed knowledge of the composition of the vehicle fleet occurred, motivated primarily by energy considerations, resort was made to a simplified classification of vehicles based on vehicle mass in Australia by the Industries Assistance Commission (Filmer and Mannion 1980) and interior volume in the U.S.A. by the Environmental Protection Agency (Dubin *et al.* 1979). Readily available data on *new* vehicles, collected monthly by private organizations for use predominantly by motor manufacturers interested in trends in their market shares (overall and by model) enabled governments to identify the demand for new vehicles of various 'size' classes (Filmer and Mannion 1980, Hensher 1982, Dubin *et al.* 1979). Crude estimates of average annual kilometres travelled by vehicles in each class could be used, in conjunction with predicted vehicle class stock to identify the government revenue levels. Additional assumptions on average fuel efficiency of each vehicle class enables estimates of energy consumption changes over time as the market mix changed.

This analysis however emphasized the changing composition of the *new* vehicle stock. Although this is important as an indicator of the direction of energy consumption levels, it fails to take into account the manner in which vehicles are used as they age and the impact that the use of the total vehicle fleet has on energy consumption and government revenues. Furthermore, the classification of vehicles on an *ad hoc* (even if intuitively plausible) single criterion may not be the best way to group vehicles. Clearly, there is no one classification rule, it will depend on the use to which it is put.

In this paper we argue that, even if energy consumption and revenue raising are central policy concerns, that a broader set of criteria related to the influences on the dimensions of auto demand (number and types of vehicles, use of vehicles) should be used to classify vehicles†. Only then can we obtain improved explanations of automobile ownership and usage and hence better predictions of levels of energy consumption and government revenue. A fundamental condition in any grouping method is that the groups should be as homogeneous as possible with respect to those attributes that might create dependencies between marginal rates of substitution for vehicles within a group and attributes outside of the group.

We should not 'force' energy consumption-related attributes (e.g. fuel efficiency) to be dominant grouping variables if homogeneity is better satisfied on other criteria. Once grouped on systematic criteria, however, it is a straightforward exercise to identify the energy consumption levels of the vehicles in each group. We would, given knowledge of numbers of vehicles registered of each vehicle model, fuel efficiency of each model, and average kilometres travelled of each model obtain group energy consumption.

The main aim of this paper is the development of a procedure to systematically group vehicles, and the application of the approach to the full range of sedans, coupes, station wagons, hatchbacks and sports automobiles available in Australia during the period 1965 to 1982‡. Although the empirical evidence is specific to the Australian context, the approach developed to group vehicles has transnational relevance§. The concepts used herein are drawn from economics and psychology, providing an example of the way in which disciplinary tools can be combined to produce effective transdisciplinary methods. We also identify groupings where fuel efficiency is the only criterion, to be used in a descriptive monitoring of trends in energy consumption.

The paper is organized as follows. In the next section we discuss some of the theoretical issues relevant in determining the selection of attributes for grouping vehicles. Then we discuss suitable correlational indices of similarity for grouping

† Implicit in the discussion on grouping is the presumption that grouping is a desirable and/or necessary requirement. The number of unique vehicles and the high level of similarity of many of them makes maintenance of unique vehicles unwieldy in modelling *if* all are to be included as alternatives. Another approach however is to treat them uniquely but randomly select a subset. The alternative approaches are evaluated in Hensher (1983).

‡ Data on other vehicles (utilities, panel vans, light commerical, other trucks and pre 1964 cars and station wagons) was also collected; however it is not planned to use it in this grouping context. Such vehicles types will each be treated as predefined groups; because the household sector (the emphasis of the main study) has insufficient number of such vehicles.

§ Detailed background information is available from the authors as Working Paper No. 4, Dimensions of Automobile Demand Project. It includes the listing of specific vehicles in each cluster, the nature of distances between cluster centres, group means and standard deviations.

attributes. This is followed by the specification of measures of similarity when attribute pairs have differing scales of measurement. The data is then discussed, with particular attention given to the selection of attributes which enter into the procedure to group vehicles (i.e. definition of supply) as distinct from those attributes which have a *direct* influence on the individual's utility (i.e. definition of demand). The empirical framework is then presented, together with the results.

Comparison is made between the resulting groupings when the correct indices of similarity are used for each attribute pair in contrast to when a standard index is used; and a comparison of groupings based on varying subsets of the attribute set. We then identify, using the preferred vehicle groupings, the major attributes influencing the groupings and the relationship between the groups and other vehicle attributes not included in group definition. The manner in which the output can be used in choice modelling is outlined, followed by a conclusion.

§ 2. Some theoretical considerations

One of the most difficult topics in economics is the definition of a commodity, both theoretically and empirically. It is not an issue for academic abstraction alone; it is at the centre of all applied economic studies, yet frequently passed over as if any selected specifications of a consumption or production measure of output are indisputable. Classic examples of ambiguous measures include 'the vehicle', 'vehicle kilometres', 'tonne kilometres', 'tonnes carried' and 'annual income'. In the current context, the use of the 'automobile' as the unit of analysis is ambiguous. Such a construct is not homogeneous†; different types of automobiles substitute imperfectly for one another. However, each vehicle is not sufficiently distinct that it should be defined as a unique good. One procedure is to define automobile types (or classes) which are distinct goods, which involves aggregating the dependent variable in a choice/demand modelling framework.

Another procedure is to recognize explicitly the heterogeneity of the commodity, accounting for differences in quality by suitable specification of the independent variable set. This latter approach creates difficulties at the vehicle level in the modelling of a functional relationship between the quantity of the commodity and the variables influencing possession because the dependent variable is a constant. We would have to specify the demand model as an 'hedonic' cost or utility function (similar to the specifications in production duality theory) in which the dependent variable is the annual cost of possessing and using a vehicle. The explanatory variables include a measure of 'physical' output (such as annual vehicle kilometres travelled) and characteristics of that output (such as the quality of service provided by the travel in a particular vehicle—its fuel efficiency, interior dimensions, handling etc.). The 'effective' output (a construct used to account for heterogeneity in output definition) is a function of physical output and the characteristics of that output. Suitable flexible functional forms (e.g. the translog reciprocal indirect utility function) can be used to estimate the function. At the household level, however, summing the number of vehicles in the household as the dependent variable introduces the need for some procedure to aggregate the output levels and output characteristics, another contentious area in allowance for heterogeneity.

† It may be argued that homogeneity was appropriate up to the early 60s when quality differences between makes and models were relatively minor (with the few exceptions) for the great majority of registered automobiles. This is no longer the situation.

Overriding all these important considerations, however, is the need to apply the demand models to predict automobile possession and usage. The hedonic approach emphasizes the costs to the household of vehicle possession and usage, allowing for heterogeneity of vehicle output. As such we are not aggregating the commodities from the perspective of demand theory. Hence automobiles in any given group need not necessarily be homogeneous with respect to their attributes; it simply enables comparisons between dissimilar vehicles based on the amount of quality embodied in them. This approach is furthermore complicated by (*a*) the need to allow for varying utility functions and income of each household (in production duality theory it is reasonable to assume that all firms operate under the same technology) and (*b*) in contrast to production estimation where the output of each firm is directly observed, in demand estimation we often observe only aggregate *traded* quantities. Observed prices are assumed to be equilibrium prices, which raises the problems of identification and simultaneous equation bias.

The approach adopted here is to group the vehicles into classes, with each class a distinct good; then to use the groupings to either (*a*) model the household's choice between vehicle classes and choice within the vehicle class containing the chosen vehicle(s) or (*b*) to model automobile type choice as a single decision but use the grouping as a way of accounting for the heterogeneity in taste variation in the sample population (where the unobserved attributes of choice are assumed to be strongly correlated with the grouping criteria). We will seek groupings which result from an attribute set that is independent of the point in time, enabling the significance of each attribute to remain unchanged in subsequent applications of assigning new vehicles to a class. The alternative approach of including attributes that vary in level over time for the same vehicle (e.g. price) means that the groupings will need to be redefined at least annually. Both approaches have strengths and weaknesses. The former (preferred) approach as applied below is based on the premise that individuals' tastes for certain vehicle attributes should be expressed in the definition of the set of variables that enter directly into the individual's utility function associated with the 'consumption' (possession and use) of a vehicle; and that grouping of vehicles should be based on the objective characteristics of the vehicles which do not enter directly into the individual's utility function. This enables us to group the vehicles and then explain the choice of vehicle class in terms of attributes not used to define the group. The inclusion of the same or a subset of attributes in the classification of vehicles and the choice between classes will result in confoundment (the simultaneous equation bias dilemma). We acknowledge that many attributes are defined objectively (e.g. interior head and leg room, rear and front) and are not available in the dimensions relevant to the utility function (e.g. interior roominess) and so we have to decide whether they enter the classification analysis or enter as explanatory variables in the choice model (in the latter instance as proxies for unobserved attributes). Here is the element of personal judgement.

Varying the classes over time will make it difficult to compare the output over time. Inclusion in class definition of factors directly influencing utility will prevent an assessment of the relative influence of such attributes across the sample of households (such as between life cycle stage); it would certainly exaggerate the significance of any attribute included in both the class definition and the utility function. Our proposed classification approach may be viewed as technological as distinct from behavioural—the vehicles exist, we are not designing new products which maximize individual utility (as in Hensher 1981). Hence the interpretation of the groupings in terms of their use in choice/demand modelling is that of: given technologically similar vehicles in each class, what behavioural factors influence an individual's choice between vehicle classes and choice within vehicle class? It is likely that calculations of actual (as distinct from perceived) levels of energy consumption and revenue will be easier to identify using technologically-determined vehicle classes than 'behaviourally-determined' vehicle classes. There is a need for further research on this topic; our position represents only one possible approach.

Adopting the preferred approach raises the question of the appropriate set of attributes to use in vehicle classifications, and which should be used in explaining household choice. If we adopt Ohta and Griliches' (1975) distinction between physical and performance attributes, we would include the latter only in the utility function. They show however that physical attributes are highly correlated with performance attributes. Typical examples of physical attributes are vehicle mass, length, height, body type, number of cylinders; examples of performance attributes are acceleration, fuel economy, frequency of repair records. Data sets currently available often have an additional influence on the classification of attributes, and our study is no exception. Although we have one of the most comprehensive vehicle attribute files, the majority of the attributes are physical in the Ohta-Griliches' sense. The physical definition should not be equated with engineering specifications. Many of these physical attributes could be interpreted as proxy-performance attributes (e.g. headroom rear, legroom front) in the absence of strictly performance attributes. Since this is where we have to impose our own prejudices, the logical strategy is to conduct sensitivity analysis to assess whether such exclusions from the physical set affects the final vehicle classification. The full set of attributes and the proposed classification is discussed in a later section on data.

§3. Correlational indices of similarity

The majority of applied studies in numerical taxonomy use a correlation coefficient as a general index class for a measure of similarity between attributes or individuals. No explicit restrictions are made, and commonly the Pearson product moment correlation coefficient is used or in situations where the data is predominantly ordinal, Kendall's general correlation coefficient (Kendall 1948). Within the family of correlation indices, to which we restrict our consideration of a measure of similarity, it can be shown (e.g. Vegelius 1973) that the use of a single correlation index where the data is of more than one (appropriate) scale type can often result in indices with invalid magnitude and signs, producing perverse groupings in component/cluster analysis.

Many correlation indices even if applied correctly to the data do not satisfy the necessary conditions for it to be a measure of similarity and hence a valid metric for input into a component analysis and cluster analysis. The recognition and accommodation of this position has not been acknowledged in either the economic or transport literature. Since it can have serious implications for vehicle grouping, we must ensure that the correlation matrix used in clustering the vehicles is derived using the correct formulae. Standard statistical packages such as SPSS, BMDP and SAS use the simple Pearson product moment correlation (PPMC) formula in component and cluster analysis when the raw data is used as input. If all attributes are ratio or interval scaled the PPMC is appropriate, but not if attributes are nominal, ordinal and dichotomous.

The first requirement of a presentation of the correct correlation formulae for mixed scale types is a definition of the scale types. The lowest level of measurement is the *nominal* scale; attributes are categorized and numbered. Since the numbering is arbitrary we have to be careful to ensure that the correlation index is invariant under permissible transformations (one to one) of the variables compared. A special case of a nominal variable is the dichotomous variable. The more common presence of dichotomous variables compared to nominal variables has resulted in an extensive literature on possible correlation indices for two-level nominal variables. Some of the more popular formulae for dichotomised variables (e.g. the Phi coefficient) are not applicable to general nominal variables and not invariant when the categories are exchanged, the sign changing.

The ordinal scale categorizes the attributes and orders the categories. The attribute categories have a particular relation to each other, with a strictly increasing permissible transformation. The intervals between categories are unknown. The interval scale has all the characteristics of the ordinal scale and the intervals between any two numbers on the scale are known, but it has no true zero point as its origin. A correlation index for an interval scaled attribute should be invariant under the permissible transformation $X \rightarrow a*X+b$, $a>0$. The highest scale level is the ratio scale, which has interval scale properties plus true zero as the natural origin. Its correlation index should be invariant under the permissible transformation $X \rightarrow a*X$, $a>0$.

Examples of nominal attributes are vehicle type (where 1 = sedan, 2 = hatchback car, 3 = station wagon, 4 = sports car, 5 = coupe), driving wheels (0 = rear wheel drive, 1 = front wheel drive, 2 = four wheel drive), and engine cooling type (1 = water, 0 = air). An ordinal attribute is type-of-vehicle-finish available (1 = standard only, 2 = standard and metallic, 3 = metallic only). An example of an interval scaled attribute is warranty kilometres (1 = 10 000, 2 = 20 000, 3 = 50 000);

Table 1. Appropriate correlation formulae in definition of a correlation matrix for component and cluster analysis (R, ratio; I, interval; O, ordinal; D, dichotomous; N, nominal).

Scale pair	Formulae (X, Y) = two attributes) $1,\ldots,N$ observations, $1,\ldots,m$ levels	Name
R, R R, I	$r_p(X,Y)=\dfrac{\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{\left[\sum_{i=1}^{N}(X_i-\bar{X})^2\right]^{1/2}\left[\sum_{j=1}^{N}(Y_j-\bar{Y}^2\right]^{1/2}}$	Pearson product moment coefficient [P]
D,D	$r_G(X,Y)=\dfrac{A-B-C+D}{A+B+C+D}$ where A = sum of positive agreeing responses ($X=+$, $Y=+$) D = sum of negative agreeing responses ($X=-$, $Y=-$) B = sum of non-agreeing responses ($X=-$, $Y=+$) C = sum of non-agreeing responses ($X=+$, $Y=-$) when the dichotomous variable (0, 1) is converted to (-1, $+1$)	*G* index [G]

Table 1 *continued*

Scale pair	Formulae (*X*, *Y*) = two attributes) 1, …, *N* observations, 1, …, *m* levels	Name
N, N N, D	$r_J(X, Y)=\dfrac{\sum_{i=1}^{N}\sum_{s=1}^{N} x_i^{(s)}y_i^{(s)}}{\left[\sum_{i=1}^{N}\sum_{s=1}^{N} x_i^{(s)2}\right]^{1/2}\left[\sum_{i=1}^{N}\sum_{s=1}^{N} y_i^{(s)2}\right]^{1/2}}$ where $x_i^{(s)}=\begin{cases} dx-1 \text{ if } X_i=X_s \\ -1 \text{ otherwise}\end{cases}$ and d_x = number of categories for X	J index [J]
O, O	$r_{SR}(X, Y)=\dfrac{\sum_{i=1}^{m}(X_i-\bar{X})^2+\sum_{i=1}^{m}(Y_i-\bar{Y})^2-\sum_{i=1}^{m}d_i^2}{2\left[\sum_{i=1}^{m}(X_i-\bar{X})^2\right]^{1/2}\left[\sum_{i=1}^{m}(Y_i-\bar{Y})^2\right]^{1/2}}$ where $d_i^2=\sum_{i=1}^{m}(X_i-\bar{X})^2-2\sum_{i=1}^{m}(X_i-\bar{X})(Y_i-\bar{Y})+\sum_{i=1}^{m}(Y_i-\bar{Y})^2$	Spearman Rank correlation [SR]
D, R	$r_{PB}(X^D, Y^R)=\dfrac{\mu_1-\mu_0}{\sigma_Y}[P(1-p)]^{1/2}$ σ_Y is standard deviation of the ratio variable Y, μ_1 and μ_0 are the means of the values of the ratio variable, corresponding to the dichotomous X variables values 1 and 0. Note: requirement 3 is not satisfied (Vegelius 1973); however, since PB is a special case of P, it may under certain empirical conditions satisfy the requirements i.e. produce results identical to P. In our data this is the case.	Point Biserial correlation [PB]
N, I	$r_{CP}(X^I, Y^N)=\dfrac{d\overbrace{\sum_{i=1}^{N}\sum_{j=1}^{N}}^{Y_i=Y_j}(X_i-\bar{X}(X_j-\bar{X})}{\left[N^2+d(d-2)\sum n_r^2\right]^{1/2}\sum_{i=1}^{N}(X_i-\bar{X})^2}$ where n_r is the number of individuals with $Y=r$; d is the number of categories of the nominal attribute	CP-coefficient
I, I (P can also be used)	$r_H(X, Y)=\dfrac{\sum_{i=1}^{N}X_iY_i}{\left[\sum_{i=1}^{N}X_i^2\right]^{1/2}\left[\sum_{i=1}^{N}Y_i^2\right]^{1/2}}$	H-INDEX [H]

and of ratio-scaled attributes, engine capacity (in cubic centimetres) and luggage capacity (in cubic metres).

When attributes in a correlation analysis are on different scales, we need a family of coefficients, one for each scale *pair*, invariant over the permissible transformations of the attributes, which jointly satisfy the requirements of a scalar product between normalized vectors in a Euclidean space. Correlation measures satisfying this coodition are called E-correlation coefficients. This requirement is necessary when the correlation matrix is to be used as the data input matrix in component and cluster analysis which uses lengths of vectors and angles between them to identify the similarity of attributes and individuals (vehicles). Normalization (i.e. giving attribute vectors the length 1) is required when attributes are measured in incommensurable units. It also has the advantage that fitting a linear transformation of a scale does not affect the value of the coefficients.

Vegelius and his colleagues (1973, 1979, 1982) have evaluated a large number of correlation formulae and developed new formulae for the full range of scale pairs, identifying which formulae satisfy the requirements for their use in factor/component and cluster analysis as meaningful measures of association. The conditions which must be fulfilled for every vector pair (X, Y) to have a relationship r and to be an E-coefficient are (Janson and Vegelius 1979, 1982):

(1) its absolute value $(|r(X, \ldots, Y)|) \leqslant 1$,
(2) it is symmetric, $r(X, Y) = r(Y, X)$,
(3) each attribute correlates perfectly with itself, $r(X, X) = 1$,
(4) if an attribute correlates perfectly with both of two other attributes, those two latter attributes must correlate perfectly. If $r(X, Y) = 1$, and $r(X, Z) = 1$, then $r(Y, Z) = 1$,
(5) its correlation matrix is positive semi-definite,
(6) if two attributes have a correlation of unity, then the correlation between each attribute and a third should be equal. $r(X, Y) = 1 \rightarrow r(X, Z) = r(Y, Z)$,
(7) if the correlation between two attributes is minus 1, then the correlation between one of the attributes and a third should be minus the correlation of the other attribute and the third. $r(u, v) = -1 \rightarrow r(v, w) = -r(u, w)$.

Correlation formulae that fail in at least one of these requirements are unsuitable. In table 1 we summarize the formulae that satisfy the seven requirements and are valid for the given scale pair.

Proofs of these formulae satisfying the E-coefficient requirements are given in Vegelius (1973) and Janson and Vegelius (1982). The formulae are all programmed and applied in the development of a correlation matrix for all physical and performance attributes (Vegelius *et al.* 1983). The full matrix is also developed using the Pearson product moment formula, so that a comparison can be undertaken. In the remaining sections we describe the data base more fully, set out the procedure used to develop the correlation matrix and apply it to group the vehicles.

§4. The data base

During 1982 a data file of 210 variables relevant to automobile specification was developed. The producers used in compiling the data, obtained from disparate sources are summarized in table 2. 827 representative vehicles in the car and station wagon categories were selected covering the manufacture vintage period 1964 to the

Table 2. Dimensions of automobile choice and demand data acquisition and preparation—automobile attribute file.

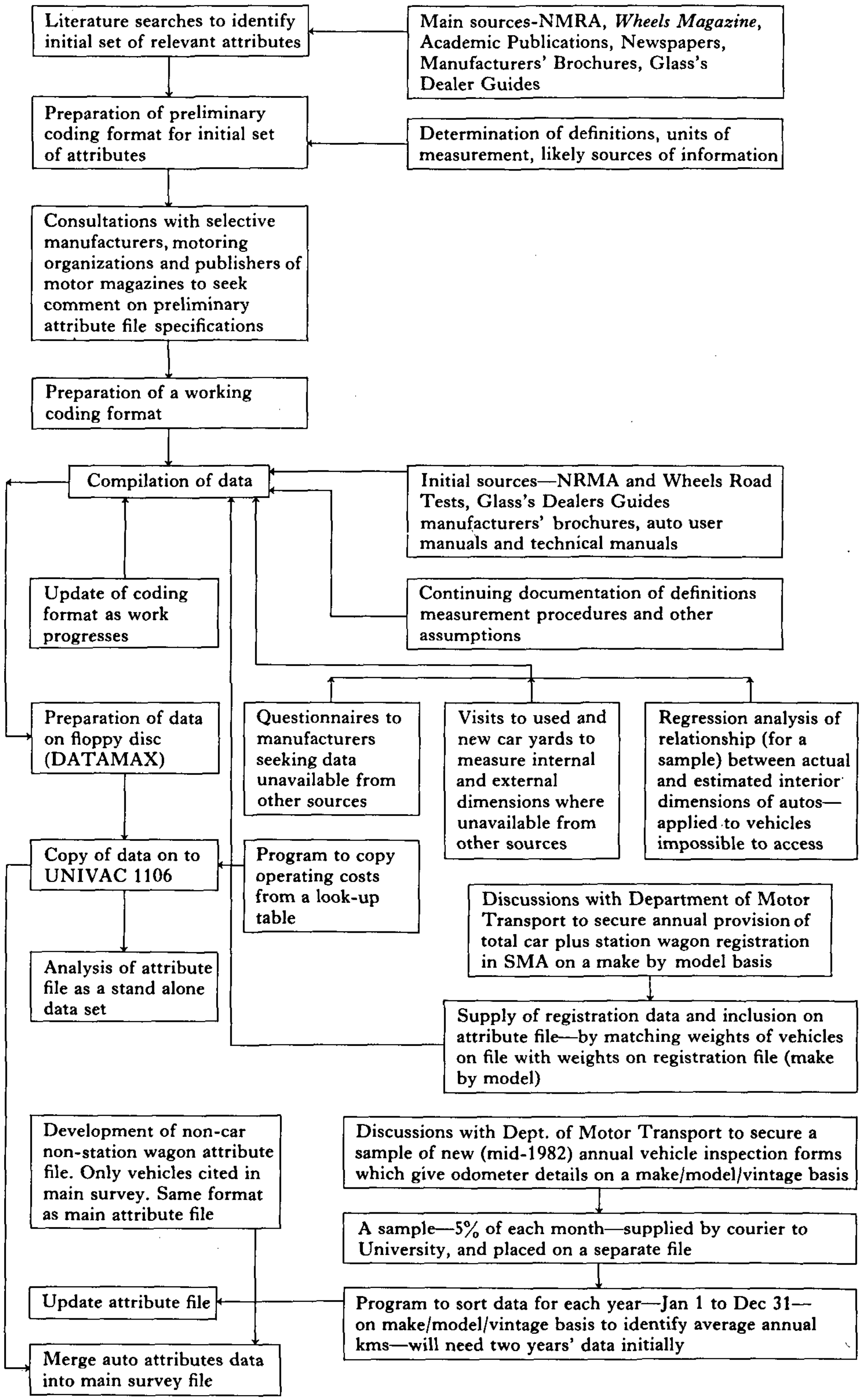


Table 3. Vehicle attributes evaluated in the search for vehicle classes (N, nominal; D, dichotomous; R, ratio).

Physical attributes [SET H]

Vehicle type (VEHTYPE) [N]
Number of cylinders (NCYLIND) [R]
Transmission type (TRANSMS) [D]
Number of gears (NGEARS) [R]
Engine location (ENGLOCN) [D]
Drive (DRIVE) [N]
Carburation type (CARBUR) [D]
Compression ratio (CMPRESS) [R]
Engine cooling type (ENGCOOL) [D]
Fuel tank capacity (NFLTANK) [R]
Reserve fuel tank capacity (RFLTANK) [R]
Wheelbase (WHLBASE) [R]
Vehicle length (VLENGTH) [R]
Vehicle width (VWDTH) [R]
Vehicle height (VHEIGHT) [R]
Gear ratios (1, 2, ..., 5) (GRRAT1, ..., GRRAT5) [R]
Final drive ratio (FINRAT) [R]
Front brake type (BRAKESF) [D]
Rear brake type (BRAKESR) [D]
Power brakes (PWBRAKE) [N]
Power steering (PWSTEER) [N]
Number of syncromesh gears (NSYNCRO) [R]
Anti-sway bar (SWAYBAR) [D]
Tyre type (TYRTYPE) [N]
Tyre diameter (TYRDIA) [R]
Rear suspension type (SUSPENR) [N]
Warning systems (oil pressure, door open, volts, choke, tachometer, water temperature, passenger side rear mirror, hazard warning lights) (WNO, WND, WNV, WNC, WNT, WNW, WNP, WHN) [D]
Vehicle mass (VMASS) [R]
Front seating type (STTYPF) [D]
Adjustability of height of driver's seat (STADJF) [D]
Number of doors (NDOORS) [R]
Front door aperture height (FDOORHT) [R]
Front door aperture width (FDOORWD) [R]
Back door aperture width (RDOORWD) [R]
Windscreen type (WNDTYPE) [D]
Number of vents (VNTUM) [R]
Window operation (WWOPER) [D]
Electronic ignition (ELECIGN) [D]
Front suspension type (SUSPENF) [N]
Number of headlights (NUMHL) [R]
Headlight type (HLTYPE) [D]
Rollbar/cage (ROLLBAR) [D]
Hardtop version available (HTOPOPT) [D]
Steering wheel turns lock to lock (SWTURNS) [R]
Theft deterrents (steering lock, locking fuel cap, central locking) (THEFPRS, THEFPRF, THEFPRC) [D]
Diameter of front brake drum/disc (BRDIAF) [R]
Diameter of rear brake drum/disc (BRDIAR) [R]

Performance attributes

Luggage capacity (with all seats in place) (LUGGCAP) [R]
Front headroom (HEADRMF) [R]
Rear headroom (HEADRMR) [R]
Front maximum legroom (LEGRMF) [R]
Rear minimum kneeroom (LEGRMR) [R]
Total seating capacity (SEATCAP) [R]
Front seating capacity (STCAPF) [R]
Front shoulder room (SHLRMF) [R]
Rear shoulder room (SHLRMR) [R]
Ventilation type (VNTYPE) [D]
Factory air conditioning (AIRCOND) [N]
Sound systems (SOUNDSY) [N]
Acceleration through the gears, 0–100 kmh (ACC0100) [R]
Fuel consumption on tour (FUELCN1) [R]
Fuel consumption in town (FUELCN2) [R]
Local *vs.* foreign production (LOCFOR) [N]
Prime country of manufacture (MNFORIG) [N]
Safety features (heated rear window, safety steering column, childproof locks) (SAFFEAW, SAFFEAS, SAFFEAC) [D]
Boot depth (BOOTDEP) [R]
Turning circle (TURNCIR) [R]
Engine capacity (ENCAP) [R]
Maximum net torque (MAXTORQ) [R]
Engine speed at maximum net torque (REVTORQ) [R]
Maximum net power (MAXPOWER) [R]
Engine speed at maximum net power (REVPOWR) [R]
Seating cover (SEATCOV) [N]

present. The car category includes sedans (saloons), coupes, sports and hatchbacks†. Each representative vehicle has assigned to it a set of represented vehicles. Within make and model categories, representation is defined on the basis of chassis type (car, station wagon), transmission type (manual, automatic), number of cylinders, engine type (conventional, rotary, petrol, diesel, turbo) and engine capacity (the smallest and largest for each available combination on the other four criteria, given the difference is greater than or equal to 100 cc). For popular models, an intermediate capacity is also selected and where the difference is in excess of 500 cc. The data base is updated annually. The vehicles on the file at December 1982 are used in the development of vehicle types; and subsequent inclusions will be either added as a representative vehicle or assigned to one of the existing representative vehicles. Full details of the data base are given in Hensher and Miller (1983).

A subset of the attributes are relevant in the current context. They are listed in table 3. The proposed classification into physical and performance (including proxy-performance) attributes is given, together with the measurement scale of each attribute. The classification is in part based on judgement, as discussed earlier in the paper.

§ 5. The empirical approach

The empirical procedure used to develop vehicle types is summarized in fig. 1. The correlation matrix is used as input into a principal component analysis. Since the interest is in grouping the vehicles, component analysis is the correct procedure. Component analysis is a data transformation technique‡ which takes the matrix of observations on a set of attributes and replaces the latter by a new set of variables—usually a smaller number—which are orthogonal. The principal components are thus no more than an exact mathematical transformation of the raw attributes. The scores are obtained by combining the raw attributes with weights proportional to their component loadings.

$$\text{component score} = \sum_{a=1}^{A} [(b_{ca}/\lambda_c)X_a]$$

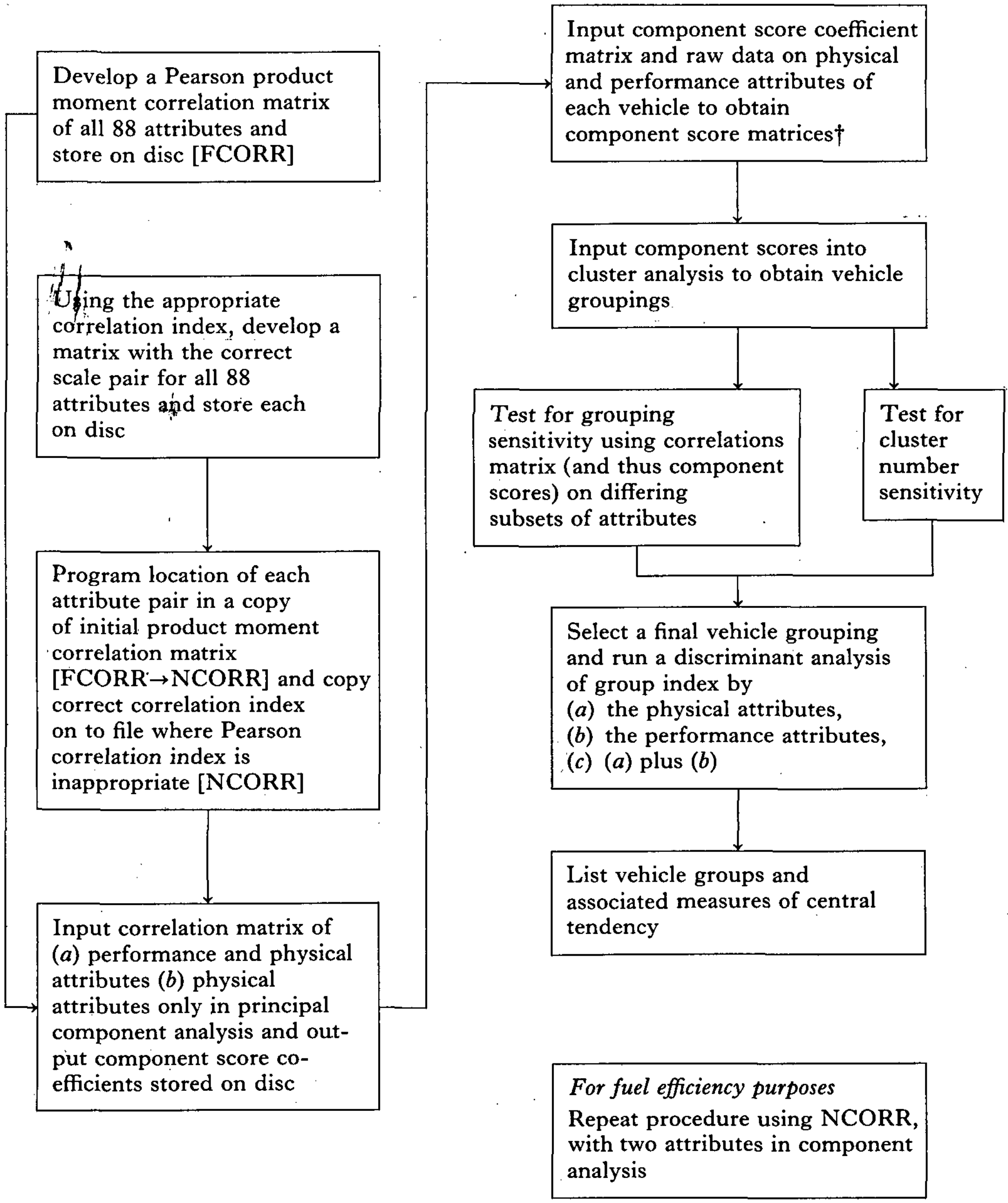
where b_{ca} is the component loading for the ath attribute on the cth component, and λ_c is the associated eigenvalue. The division by λ_c assures the resulting index has a variance equal to 1·0.

Since the scores for the observations on the components have a mean of zero and a standard deviation of 1·0, Pythagorean distances can be computed immediately in the cluster analysis. Another advantage of principal components analysis is that it removes redundancies in the original data matrix, replacing two variables which interact by a single variable, eliminating undue weighting if the grouping were to use both variables.

† The Dubin *et al.* (1979) study, the only known systematic study of vehicle grouping, limited its analysis to vehicles of the same vintage, 1975, and a very limited set of attributes. They used physical performance and price variables, which we argued is an invalid basis for classifications to be used subsequently in choice/demand modelling.

‡ Common factor analysis by contrast is not expressible by the combination of the observed attributes. Unique factors can be extracted which do not contribute to the covariation between attributes. Thus a certain mathematical function of the observed attributes, obtained by principal components analysis is not obtained by common factor analysis.

Fig. 1



Procedure to obtain vehicle groupings.

† When a correlation matrix is the data input in a component analysis, component scores cannot be automatically produced. While standard packages can produce component scores as output if the raw data is used as input, this only applies to situations where all attributes are at least interval scaled. Since NCORR requires mixed scales, the procedure outlined above is the only approach.

The use of component analysis rather than factor analysis has the added advantage of accommodating variables at any level of measurement, provided a correlation measure is used which satisfies the conditions pertaining to the family of E-coefficients. Factor analysis requires that the variables be measured at least at the interval level. Kim and Mueller (1978) outline some of the reasons for this restriction. For example, in general the greater the number of categories (for nominal and ordinal variables) the smaller the degree of distortion. Dichotomization of continuous variables attenuates the correlations, which is also further affected by the cutting points. However, although the grouping (i.e. allocating of levels of a variable) attenuates correlation in general, it does not affect the structure of clustering in the data because factor analysis is dependent on the *relative* magnitude of correlations. Thus, if the objective is the search for clustering patterns of attributes, factor analysis may be justified using measurement scales below the interval metric.

§ 6. Analysis

The cluster analysis uses an unweighted non-hierarchical technique to establish a fixed number of homogeneous groups of vehicles. The vehicles are initially defined as a single group and split into two groups in a stepwise manner until the predetermined number of clusters is obtained. When the number of clusters is reached, vehicles are iteratively reallocated into the group whose centre is the closest. The euclidean distance is used to measure the distance between each vehicle and the centre of each cluster (the latter calculated as the mean of vehicles in the cluster); the input data are the component scores associated with each vehicle. The clustering that has the minimum within-cluster sum of squares is finally chosen.

On *a priori* grounds we have argued above for a vehicle grouping based on the mixed-scale correlation matrix (henceforth called *N*CORR) and the physical attributes (henceforth called SNCOMPH). However, before committing the grouping to that obtained using SNCOMPH (regardless of the final number of clusters), cluster analyses was performed on component score matrices obtained from the all-Pearson and mixed-scale correlation matrices for both the physical attributes and the combined physical plus performance attributes. The results for 25 clusters are given in table 4, using all 60 physical attributes and all 28 performance attributes.

The data in table 4 refers to the number of vehicles with a common grouping for each of the other three cluster analyses. For example, of the eleven vehicles in cluster one of SNCOMPH, seven and four are grouped together where Pearson—physical and mixed-scale—all are used. This does not tell us what other vehicles may have been grouped with these 11 vehicles. The 7 and 4 vehicles in cluster 1 of SNCOMPH, associated with SPCOMPH happen to be the only vehicles in clusters 7 and 21. However the 8, 8, 1 (cluster 6 of SNCOMPH) are in clusters 2, 15, 19 which contain 67, 44 and 27 vehicles respectively. The sensitivity of grouping is clear; the results are with few exceptions highly variable. A closer study of the specific vehicles (available from the authors on request) shows that subsets of vehicles from each cluster do stay together. To bring the four approaches together would require a considerably larger number of clusters, well beyond a useful set (defeating the purpose of grouping). The result is important, demonstrating the invalidity of the groupings based on the all-Pearson correlation matrix—the use of the 'correct' correlation index does make a difference to the results. The selection of physical-only

Table 4. Grouping stability when the correlation matrix and the attribute set vary.

Cluster	No. in SNCOMPH	All Pearson, physical [SPCOMPH]	Mixed-scale physical and performance [SNCOMPA]	All Pearson, physical and performance [SPCOMPA]
1	11	7, 4	7, 4	11*
2	15	2, 2, 10*, 1	3, 10*, 1, 1	8, 2, 3, 2
3	38	21*, 3, 10, 3, 1	14, 3, 7, 4, 3, 3, 1, 3	9, 2, 11, 4, 1, 1, 6, 3, 1
4	9	9*	7, 2	7, 2
5	27	15, 3, 5, 1, 2, 1	5, 9, 3, 5, 4, 1	2, 17*, 3, 1, 1, 3
6	17	8, 8, 1	1, 1, 9*, 3, 1, 1, 1	1, 1, 1, 9*, 4, 1
7	12	4, 5*, 1, 2	2, 4, 2, 1, 2, 1	5*, 3, 2, 1, 1
8	49	4, 5*, 2, 1, 1	32, 9, 6, 2,	12, 9, 15, 11, 1, 1
9	60	22, 9, 2, 4, 7, 4, 12	18, 17, 7, 3, 5, 5, 2, 2, 1	33*, 10, 6, 5, 2, 1, 3
10	12	11*, 1	11*, 1	11*, 1
11	11	4, 2, 1, 4	11*	11*
12	29	5, 16*, 2, 2, 1, 1, 1, 1	8, 1, 4, 5, 6, 1, 1, 1, 1, 1	7, 1, 4, 5, 4, 2, 3, 1, 1, 1
13	40	19, 6, 2, 6, 5, 1, 1	22*, 1, 3, 2, 3, 4, 1, 2, 1, 1	14, 4, 4, 3, 3, 3, 3, 2, 1, 1, 1
14	46	1, 15, 25*, 1, 1, 3	1, 8, 4, 5, 8, 9, 1, 5, 4, 1	2, 12, 2, 15, 3, 6, 2, 3, 1
15	74	29, 34, 4, 4, 1, 1, 1	20, 38*, 9, 4, 1, 2	29, 12, 2, 2, 10, 1, 15, 2, 1
16	43	9, 28, 1, 2, 2	33*, 1, 1, 6, 2	13, 18, 2, 1, 7, 1, 1
17	50	6, 23, 13, 2, 3, 3	9, 17, 3, 9, 4, 2, 2, 1, 1, 1, 1	16, 7, 9, 5, 2, 4, 1, 5, 1
18	21	10, 9, 1, 1	12*, 2, 1, 3, 2, 1	11, 3, 5, 2
19	33	17*, 5, 2, 1, 2, 3, 1, 1, 1	5, 4, 1, 9, 3, 3, 2, 1, 2, 2, 1	7, 1, 8, 3, 5, 4, 1, 1, 1
20	52	10, 1, 29*, 2, 4, 2, 3, 1	25, 9, 4, 9, 2, 1, 1, 1	21, 5, 6, 1, 4, 5, 6, 2, 1, 1
21	27	18, 2, 3, 4	9, 16, 1, 1	20*, 1, 5, 1
22	51	34*, 12, 4, 1	6, 26, 15, 1, 2, 1	20, 8, 8, 12, 2, 1
23	46	23, 21, 1, 1	17, 21, 4, 1, 2, 1	31*, 3, 2, 1, 5, 4
24	36	8, 15*, 11, 2	5, 2, 14, 7, 5, 1, 1, 1	3, 10, 14, 6, 3
25	18	18*	18*	18*

The comparison is relative to the mixed-scale physical attribute set [SNCOMPH].
*, Highest number clustered as per SNCOMPH.

Table 5. Final vehicle groupings—summary description of key differences—57 physical attributes (SNCOMPH1).
Note: [the group numbers have no particular significance].

GROUP 1 (27 vehicles)
Predominantly 4 cylinder front engined, power brakes, sedans, 89% manual, relatively high compression ratio (11·63) 33% air cooled engine, average external dimensions and weight (1191 kg), high front door (94·25 cm), high gear ratios for gears 2 and 3, and final drive ratio, above average ventilation, vintages 65–82. EUROPEAN MEDIUM (2 LITRE) MANUALS.

GROUP 2 (51 vehicles)
Predominantly 4 cylinder manual coupes, roadsters and sedans, 92% manual, over 5% are rear wheel drive, relatively small on external dimensions, 857 kg, low wide front doors (coupes), relatively high final drive ratio, all drum rear brakes, narrow front brake diameter (21·26 cm), vintages 66–81. SMALL (1·5 LITRE) MANUALS.

GROUP 3 (30 vehicles)
4 and 6 cylinder manuals, average external dimensions, 1206 kg, high first gear ratio (3·38), power brakes standard, wide tyres (12·33 inches), vintages 63–77. SMALL-MEDIUM (2·5 LITRE) MANUALS, MIXED ORIGIN.

GROUP 4 (60 vehicles)
Mix of 4 cylinder sedans and station wagons, predominantly automatic (98·3%), all carburation, relatively short (4·15 m), 975 kg, all 4 doors, narrow front door, average number of vents, low 1st and 3rd gear ratio, no 4th gear, vintages 65–81. SMALL-MEDIUM (1·5–2·0 LITRE) AUTOMATICS, MIXED ORIGIN.

GROUP 5 (27 vehicles)
Mix of automatic sedans and station wagons, 40% fuel injection, slightly above average external dimensions, 1279 kg, mainly 6 cylinders, 97% laminated windscreen (vintage), well ventilated (10·15 vents) all windows open, low gear ratios, mainly steel-belted radials, relatively wide tyres (12·41 inches), excellent theft-proof devices and warning systems, vintages 74–82. LUXURY (2–5 LITRE) AUTOMATICS.

GROUP 6 (empty)
Reallocation of 1 vehicle to Group 16.

GROUP 7 (24 vehicles)
Mix of coupes, sedans and station wagons, 4–6 cylinder, 79% manual, all carburation, 1323 kg, slightly above average external dimensions, relatively well ventilated (9 vents), all windows open, relatively high gear ratio (all 4), all have anti-sway bar, relatively narrow tyres (10·33 inches), high incidence of independent suspension, above average warning systems, wide rear brake diameter (27·3 cm), vintages 67–82. MEDIUM-LUXURY (2 LITRE).

GROUP 8 (48 vehicles)
Mix of sedans and station wagons, 6–8 cylinders, 50% manual, all carburation, 75 litre fuel tank, long, wide and heavy (1453 kg) vehicles, all windows open, high gear ratios, all have anti-sway bar, narrow tyres (10·25 inches), average instrumentation, high steering wheelturns (4·72) and brake diameter (rear and front), vintages 70–82. LARGE (HEAVY) (4·1 LITRE) DOMESTIC.

GROUP 9 (49 vehicles)
Mainly 4 cylinder manual sedans and station wagons, all carburation, 879 kg, relatively small external dimensions and front door dimensions, below average number of vents, high first and second gear ratios, low final drive ratio, narrow brake diameters, vintages 65–81. SMALL (1·5 LITRE) MANUALS.

GROUP 10 (67 vehicles)
Full range of vehicle types, predominantly 4 cylinder, 94% automatic, average external dimensions, 1014 kg, well ventilated (8·14 vents), relatively low gear ratios, good warning systems and theft-proof devices, vintages 71–82. SMALL-MEDIUM (2 LITRE) AUTOMATICS, MIXED ORIGIN.

Table 5 *continued*

GROUP 11 (49 vehicles)
Mix of sedans and station wagons, 27% manual, mainly 6 cylinder, all rear wheel drive and carburation, relatively long and wide, 1389 kg, high and narrow front door, 22% windscreen laminated (vintage), poor ventilation (4 vents), 49% disc brakes on front, power brakes and steering mainly non-standard, crossply tyres (vintage), poor on theftproof and warning devices, high lock to lock steering turns (4·33), vintages 65–76. LARGE (HEAVY) (3·5 LITRES) DOMESTIC.

GROUP 12 (38 vehicles)
Mainly automatic (97%) sedans, coupes and station wagons, 4–6 cylinders, all 3 gears, 87% rear wheel drive, 34% fuel injection, average external dimensions, 1341 kg, high and narrow front door, 84% laminated windscreen, well ventilated (10·2 vents), low gear ratios, power brakes and steering standard, good theft-proof and warning devices, vintages 64–82. MEDIUM (2–3 LITRE) QUALITY.

GROUP 13 (28 vehicles)
Mix of sedans, coupes and station wagons, 6–8 cylinders, 60% manual, 86% rear wheel drive, 92% carburation, 1451 kg above average external dimensions, large front doors, well ventilated (9·7 vents), high gear ratios, all disc brakes, 12·04 inch tyres, best suspension, excellent theft-proof and warning system devices, vintages 66–82. LUXURY (2·5–3·5 LITRE) EUROPEAN.

GROUP 14 (empty)
Reallocation of 3 vehicles to Group 16. LUXURY SPORTS EUROPEAN.

GROUP 15 (52 vehicles)
Mainly coupes, 4–6 cylinder, 88% manual, 7% have rear engine, 79% front wheel drive, low vehicles (1·309 m), 1120 kg, low and wide front doors, 79% have laminated windscreen, power brakes standard, wide tyres (12·11 inches), good theft-prevention devices and warning systems, vintages 68–82. COUPES/SPORTS MIXED ORIGIN.

GROUP 16 (8 vehicles)
Mix of sedans, coupes and station wagons, 4 cylinders and 38% manual, low compression ratio (7·86), 969 kg, average external dimensions, all 2 doors, high and wide front door, relatively poor ventilation (5·5 vents), all have anti-sway bar, narrow tyre diameter (10·63 inches), partial warning systems, limited steering wheel turns (2·7), vintage 69–77. FASTBACK/SPORTS EUROPEAN.

GROUP 17 (9 vehicles)
Mainly manual coupes, rear wheel drive, relatively short and narrow, 824 kg only, all 2 door, lowest front doors (88·1 cm), only 3 vents, high gear ratios (especially first gear), high final drive ratio, wide tyres (12·44 inches), all hydro-elastic suspension, average instrumentation, narrow brake diameters, vintages 65–81. COUPES/EUROPEAN/MANUAL.

GROUP 18 (45 vehicles)
Predominantly sedans, 6 cylinders, all 4 gear 4 door manual, average external dimensions, 1219 kg, power brakes standard, close to average on most attributes, vintages 64–81. MEDIUM-LARGE (2–4 LITRE) MANUAL.

GROUP 19 (41 vehicles)
Predominantly automatic (71%) sedans and station wagons, all rear wheel drive, large fuel tank (73 litres), large external dimensions, heaviest vehicles (1573 kg), low ratio in gears 3 and 4, low final drive ratio, power brakes standard, good theft-proof devices and warning systems, relatively wide brake diameters, only 3·05 steering wheelturns, vintages 71–81. VERY LARGE (HEAVY)(4–5 LITRES) SEMI-LUXURY/AUTOMATIC.

GROUP 20 (39 vehicles)
Mainly 6 cylinder automatics (72%) sedans, 1330 kg, average external dimensions, power brakes standard, close to average on most attributes, vintages 65–78. LARGE (3–4 LITRES), AUTOMATIC SEMI-LUXURY.

Table 5 *continued*

GROUP 21 (16 vehicles)
all automatic (3 gears), 4 cylinder, mainly sedans, 75% rear wheeldrive, 1087 kg, 50% laminated windscreen, slightly less than average number of vents, relatively low gear ratios, 56% of rear brakes are disc, power brakes standard, poor theft-proof devices, above average warning systems, vintages 65–79. SMALL-MEDIUM (1·5–2 LITRE).

GROUP 22 (16 vehicles)
All coupes, 6–8 cylinders, 43% manual, all rear wheel drive, large fuel tank (73 litres), relatively long and wide but low, 1394 kg, all 2 doors, high and wide front doors, poor number of vents (4·7), low final drive ratio, 68% of front brakes are disc, rear are 100% drum, high incidence of optional power brakes, generally poor theft-proof and warning system devices, high number of steering wheelturns (4·04), vintages 65–77. COUPE (3·5–5·0 LITRES), DOMESTIC, LARGE.

GROUP 23 (27 vehicles)
Mix of all vehicle types, 89% manual, mainly 4 cylinders, 65% rear wheel drive, 7% fuel injection, small fuel tank (40 litres), relatively short and narrow, 854 kg, relatively small front and rear doors, poorly ventilated, relatively high gear ratios and final drive ratio, power steering not available, tyre diameter of 12·15 inches, relatively poor instrumentation, vintages 65–78. SMALL (1–1·5 LITRE) MANUALS.

GROUP 24 (54 vehicles)
Predominantly manual (100%), 4 cylinder sedans and station wagons, 54% rear wheel drive, all carburation, 1029 kg average external dimensions, above average ventilation (vintage), good instrumentation and security devices, vintage 70–82. MEDIUM-MEDIUM (1·5–2·0 LITRE) POPULAR MANUAL.

GROUP 25 (17 vehicles)
Mainly manual (100%), 4 cylinder sedan, 42% rear wheeldrive, small fuel tank (36·6 litres), relatively small external dimensions, 776 kg (lightest vehicles), high gear ratios and final drive ratio, mainly drum brakes, power steering not available, slight below average instrumentation, small brake diameters (front and rear), vintages 65–78. SMALL-SMALL (1 LITRE) MANUAL.

---

or physical-plus-performance attributes is based on the theoretical separation of supply and demand influences and judgement. The evidence that the clustering is dependent on the variables included supports the finding of Dubin *et al.* (1979).

The final vehicle groupings are summarized in table 5. A complete list of the representative vehicles in each cluster, their intra-cluster distances, distances between cluster centroids, and the group means and standard deviations for all 88 physical and performance variables are available on request. Twenty-eight component scores on the 827 vehicles were used to group the vehicles. Twenty-five clusters were selected from analyses using 15, 20 and 25 groups. Twenty-five clusters were defined as the upper limit on practical grounds; in particular the subsequent use of the groupings in discrete choice modelling of auto-type choice. Initially all 60 physical attributes were used to obtain component scores; however assessment of the groupings resulted in concern about the allocation of some vehicles. Discriminant analyses of the resultant group index on the 60 attributes and an assessment of the means and standard deviations of all attributes within each group highlighted three attributes as possible influences on intuitively implausible groupings. The attributes are the fifth gear ratio, the availability of a hardtop option and the capacity of a reserve fuel tank. Many vehicles do not have a fifth gear, hence the ratio is zero. The existence or otherwise of a fifth gear is picked up by the NGEARS attribute. Given the need to exercise judgement in the final selection of clusters, we prepared new clusters based on 57 attributes. The resulting 25 clusters are accepted as intuitively plausible. The statistical differences of the components is confirmed in table 6. Because the groups are obtained empirically, the F-ratio should only be used to describe differences between components rather to test significance. All 28 components (and thus all 57 physical attributes) have an influence on the final groupings. However the small number of vehicles in clusters 6 (a 1978 VW Golf Diesel) and 14 (2 1979 VW Passatt Diesels and a 1981 Gemini Diesel) were assigned to cluster 16 on the basis of the levels of the 57 attributes. The subsequent discriminant analysis confirmed the reallocation, producing a 100% correct classification of vehicles in cluster 16. The closest clusters are 8 and 19, 4 and 10. Eight and 19 are large (over 4 litre) domestic vehicles of the three main manufacturers (Ford, General Motors, Chrysler); 4 and 10 are small-medium (1·5–2 litre) automatics of a wide range of manufacturers. The maximum distance occurs between clusters 4 and 16; group 16 is predominantly European sports and fastback autos and group 4 contains small-medium sized sedans and station wagons. Group 16 has the least number of vehicles (8), group 10 has the largest number (67).

The groups in table 5 show some key descriptors to distinguish them. Transmission (manual, automatic), number of cylinders, weight, vehicle type, location of predominant manufacture (domestic, imported) and gear ratios are key descriptive dimensions. Such descriptors are useful but could be potentially misleading, since they are based on the group mean levels of attributes. Discriminant analyses were undertaken to identify the relative importance of each physical attribute in vehicle classification. In addition we included the performance attributes to identify any possible systematic relationship between the groups and the full set of available attributes. The results are given in table 7. All 60 physical attributes were accepted in the analysis on the F-to enter criterion. Surprisingly, number of

Table 6. Analysis of variance—cluster analysis for choice model vehicle groupings [GPINDEX1]. (Cluster number in small print.) Degrees of Freedom = 24 802.

	1	2	3	4	5	6
	13	14	15	16	17	18
Mean squares	25	26	27	28		
between	5·3474	4·1421	6·9088	4·9466	4·5238	3·3965
	10·7622	12·0727	6·6569	25·9450	2·3683	12·3797
	8·1745	10·2239	31·5348	22·9424		
within	0·1983	0·2345	0·2528	0·2498	0·2965	0·2632
	0·2953	0·2459	0·2756	0·2267	0·2785	0·2545
	0·3451	0·3432	0·3212	0·2538		
F-Ratio	26·963	17·664	27·335	19·083	15·258	12·902
	36·445	49·089	24·153	114·427	8·504	48·610
	23·688	29·787	98·165	90·393		
	7	8	9	10	11	12
Mean squares	19	20	21	22	23	24
between	3·9851	5·8342	8·4114	3·7697	8·2919	4·1426
	5·8273	5·6975	7·7755	9·4556	9·9909	7·5914
Within	0·2146	0·1928	0·2932	0·2031	0·1815	0·2713
	0·3156	0·3134	0·3518	0·1875	0·1628	0·3190
F-ratio	18·573	30·260	28·688	18·560	45·687	15·270
	18·467	18·178	22·100	50·435	61·366	23·798

cylinders (NCYLIND), vehicle mass (VMASS) and transmission (TRANSMS) were well down the list of attributes, although statistically significant discriminators (entering at steps 22, 31 and 40). Many attributes are highly correlated as expected, hence given the statistical significance of all attributes the position of an attribute thought on *a priori* grounds to be low (or high) is a reflection of the significant contribution that one of the highly correlated attributes has already made, leaving residual discrimination powers to the other attribute. 86·94% of the grouped vehicles are correctly classified on the 60 attributes when applying the canonical discriminant functions (and a 0·5 cut off), giving confidence in the use of the canonical discriminant functions to classify additional vehicles into the Final 23 clusters. Eighteen discriminant functions are required to represent the observed differences

Table 7. Discriminant analysis for choice-model vehicle groupings [GPINDEX1]. (*a*) Physical attributes (60), (*b*) Performance attributes (28), (*c*) Combined attributes (88).

Step entered		Wilks' Lambda	Step entered		Wilks' Lambda
(*a*)					
1	GRRAT4	0·201476	31	VMASS	0·000024
2	VWIDTH	0·067477	32	WNC	0·000022
3	ENGLOCN	0·027250	33	TYRDIA	0·000019
4	RDOORWD	0·012643	34	STADJF	0·000018
5	BRAKESR	0·007388	35	FDOORHT	0·000016
6	BRAKESF	0·004589	36	NGEARS	0·000015
7	THEFPRC	0·003082	37	TYRTYPE	0·000014
8	SWTURNS	0·002189	38	ELECIGN	0·000012
9	CMPRESS	0·001565	39	DRIVE	0·000012
10	THEFPRS	0·001137	40	TRANSMS	0·000011
11	CARBUR	0·000838	41	VNTNUM	0·000010
12	ENGCOOL	0·000630	42	WND	0·000009
13	BRDIAF	0·000484	43	GRRAT2	0·000008
14	GRRAT1	0·000379	44	WHO	0·000008
15	HLTYPE	0·000297	45	GRRAT3	0·000007
16	VHEIGHT	0·000230	46	FINRAT	0·000007
17	PWBRAKE	0·000188	47	ROLLBAR	0·000006
18	STTYPF	0·000154	48	WNV	0·000006
19	SUSPENF	0·000127	49	PWSTEER	0·000006
20	THEFPRF	0·000105	50	WNDTYPE	0·000005
21	WNT	0·000087	51	VLENGTH	0·00005
22	NCYLIND	0·000074	52	BRDIAR	0·000005
23	VEHTYPE	0·000064	53	NDOORS	0·000005
24	WWOPER	0·000055	54	WHLBASE	0·0000004
25	NUMHL	0·000048	55	WNP	0·000004
26	FDOORWD	0·000043	56	HTOPOPT	0·000004
27	SUSPENR	0·000038	57	NSYNCRO	0·000004
28	GRRAT5	0·000033	58	NFLTANK	0·000004
29	WNH	0·000030	59	RFLTANK	0·000004
30	WNW	0·000027	60	SWAYBAR	0·000003
(*b*)					
1	VNTTYPE	0·289	14	REVPOWR	0·009
2	ENCAP	0·093	15	SAFFEAS	0·009
3	MNFORIG	0·063	16	LUGGCAP	0·008
4	STCAPF	0·046	17	LOCFOR	0·007

Table 7 *continued*

Step entered		Wilks' Lambda	Step entered		Wilks' Lambda
5	SHLRMF	0·035	18	BOOTDEP	0·007
6	SAFFEAC	0·029	19	LEGRMF	0·006
7	MAXPOWR	0·024	20	MAXTORQ	0·006
8	SOUNDSY	0·020	21	LEGRMR	0·005
9	TURNCIR	0·017	22	REVTORQ	0·005
10	SEATCAP	0·015	23	AIRCOND	0·005
11	HEADRMF	0·013	24	SHLRMR	0·004
12	SAFFEAS	0·012	25	HEADRMR	0·004
13	ACCO100	0·010	26	SEATCOV	0·004
(*c*)					
1	GRRAT4	0·201476	43	LEGRMF	0·000003
2	VNTTYPE	0·058357	44	WND	0·000003
3	VWDITH	0·019521	45	WNO	0·000003
4	ENGLOCN	0·007895	46	GRRAT2	0·000003
5	RDOORWD	0·003638	47	MAXPOWR	0·000003
6	BRAKESR	0·002144	48	MNFORIG	0·000002
7	THEFPRC	0·001461	49	SAFFEAS	0·000002
8	ENCAP	0·001031	50	TRANSMS	0·000002
9	BRAKESF	0·000735	51	GRRAT3	0·000002
10	CMPRESS	0·000526	52	HEADRMR	0·000002
11	SWTURNS	0·000380	53	DRIVE	0·000002
12	CARBUR	0·000281	54	STTYPF	0·000002
13	ENGCOOL	0·000213	55	WNV	0·000001
14	HLTYPE	0·000168	56	NCYLIND	0·000001
15	VHEIGHT	0·000132	57	PWSTEER	0·000001
16	STCAPF	0·000105	58	VLENGTH	0·000001
17	THEFPRF	0·000085	59	SHLRMR	0·000001
18	GRRAT1	0·000068	60	FINRAT	0·000001
19	PWBRAKE	0·000056	61	LOCFOR	0·000001
20	WNT	0·000046	62	BRDIAR	0·000001
21	BRDIAF	0·000039	63	WNDTYPE	0·000001
22	THEFPRS	0·000033	64	WNP	0·000001
23	VEHTYPE	0·000029	65	NFLTANK	0·000001
24	WWOPER	0·000025	66	SHLRMF	0·000001
25	SUSPENF	0·000022	67	MAXTORQ	0·000001
26	SUSPENR	0·000019	68	LEGRMR	0·000001
27	NUMHL	0·000016	69	SAFFEAC	0·000001
28	GRRATS	0·000015	70	SOUNDSY	0·000001
29	FDOORWD	0·000013	71	HTOPOPT	0·000001
30	WNW	0·000012	72	BOOTDEP	0·000001
31	WNC	0·000010	73	LUGGCAP	0·000001
32	TYRDIA	0·000009	74	VNTNUM	0·000001
33	VMASS	0·000009	75	RFLTANK	0·000001
34	WNH	0·000008	76	NDOORS	0·000001
35	NGEARS	0·000007	77	SEATCOV	0·000000
36	STADJF	0·000006	79	WHLBASE	0·000000
37	TYRTYPE	0·000006	79	WHLBASE	0·000000
38	FDOORHT	0·000005	80	NSYNCRO	0·000000
39	HEADRMF	0·000005	81	ROLLBAR	0·000000
40	TURNCIR	0·000004	82	AIRCOND	0·000000
41	REVPOWR	0·000004	83	SWAYBAR	0·000000
42	ELECIGN	0·000004	84	REVTORQ	0·000000

between the groups. Additional dimensions would not add any significant information about the group differences. For the Canonical Discriminant Functions; as Wilks' Lambda increases toward its maximum value of 1·0, it is reporting progressively less discrimination. Lambda equal to 1 indicates that the group centroids are identical.

Discriminant analyses on the 28 performance attributes and the total attribute set (88) found that 49·09% of the grouped vehicles could be correctly classified using the performance attributes, and 90·45% using all attributes. The performance attributes in general add little to the overall classification power of the discriminant functions derived from the physical attributes. The interdependence between physical and performance attributes is not sufficiently great to view performance attributes as proxies for physical attributes.

Physical characteristics produce groupings that are unlikely to result if performance and cost were included. For example, in group 5 we have a mix of vehicles ranging from a 1978 Mercedes 450 SEL to a 1980 Toyota Corona. We can justify the heterogeneity on non-physical criteria in terms of the use of the clusters in choice modelling where *direct* influences on preferences should not be used in vehicle grouping. If our interest is on monitoring price changes over time then we would have developed clusters on price alone. Price is a difficult attribute to use, except where all vehicles have the same vintage.

§7. Use of groupings in choice modelling

An important reason for grouping automobiles is to provide an alternative approach to modelling household auto-type choice. The extant literature has used one of four data specifications:

(1) The household's chosen vehicles are identified, and a set of unranked alternatives randomly selected from the universal finite choice set [Manski and Sherman 1980].
(2) The household supplies the set of ranked or unranked alternatives in addition to the chosen [Hensher and Manefield 1981, Hensher and LePlastrier 1982].
(3) A fixed set of representative ranked alteratives are defined and the chosen auto is defined by its representative auto or itself [Lave and Train 1979, Berkowitz *et al.* 1983].
(4) A global choice set of vehicles is defined by vintage and class and a randomly selected set of unranked alternatives [Train and Lohrer 1983, Hensher 1983].

Specifications 2 and 3 are especially problematic. Approach (2) exogenizes a decision which is endogenous, namely the choice of choice sets. By asking a household to indicate the alternatives considered but rejected implies a prior decision whose outcome was the rejection of the majority of the vehicles in the universal finite choice set and the selection of a subset for further evaluation. Hensher and LePlastrier (1982) concluded that for variables thought *a priori* to have an important effect this strategy is likely to produce a set of influences on type choice that have an intuitively implausible sign and are statistically non-significant. For example, when the choice set is the reported choice set, price is likely to have a positive sign (and often non-significant); the vehicles are so similar in price that when it is significant it is likely to be accounting for a quality effect, with households tending to prefer the vehicle with a higher price (within the perceived set). The role of price has been predominant in screening out those vehicles not in the perceived set; thus we only have explained part of the type choice decision. Specification (3) defines a representative vehicle for each alternative in a choice set which is invariant across the sampled households. In addition to the aggregation bias (loss of information), variability in the vehicle attributes can only be achieved by interacting them with socio-economic variables. This interaction amounts to a model in which socio-economic (alternative-specific) variables are the determinants of type choice, scaled by the particular representative vehicles attributes.

Using specification (1), Manski and Sherman 1980, defined the vehicle universe as all makes and models of 1967–76 vintage domestic passenger autos, all makes of 1967–76 vintage light trucks and foreign passenger cars, and dummy pre-1967 passenger auto and pre-1976 light truck alternatives. By assuming that a household annually reviews its holdings, the full range of vehicle vintages can be assumed to be available. To keep the analysis manageable each household's alternatives were 25 randomly selected autos.

Approach (4) is a modification of Approach (1). Instead of defining a unique make-model vehicle as Manski and Sherman did, a vehicle vintage-class is defined (120 in the Train-Lohrer study), and each household is randomly assigned a subset of the vintage-class groups as alternating to the chosen vehicle-mix. Hensher (1983) derived his class groups using the approach outlined above, but in terms of weight and cylinders, and then cross-tabulated the result by vintage. Discriminant analysis also used to identify the discriminating ability of different vintage groupings. The main levels of each attribute were obtained for each vintage-class and a 'size' variable included in the choice modelling to account for within vintage-class heterogeneity. The size variable is defined in terms of the number of make-model vehicles in a vintage-class. The advantage of this approach over Approach (1) is that it simplies the forecasting task.

The grouping of vehicles can assist in the attainment of estimates of parameters that are less likely to be associated with violation of the independence from alternatives (IIA) property of the multinomial logit model (MNL). This needs explanation. The simplest model is a single level MNL model in which we randomly select a set of alternatives (plus the chosen vehicle) from all attribute files (the 827 cars and station wagons and vehicles referred to in footnote ‡ on p. 246). Use of the MNL model requires acceptance of the IIA property. This property is almost surely to be violated where we have differential correlation patterns between pairs of alternatives.

One mechanism for partially overcoming this problem is to structure the choice as a nested choice in which separate choices are made within subsets of vehicles (i.e. choice conditional on subset or branch in the nest) and then choice between subsets (marginal decision on subset selection). This approach can be estimated as a sequence of a conditional choice and a marginal choice and viewed as one decision if the expected maximum utility of each lower level of the tree are the only explanatory variables in the upper level or as two decisions if the upper level also includes additional variables. Consistent but not efficient parameter estimates are obtained because a sequential approach does not use full information; information about higher-level choices (i.e. between subsets) is not used in estimating the lower-level parameters, and since the estimates at each level depend on parameters estimated at lower levels, errors may accumulate up the tree.

If the nested model, sequentially or simultaneously estimated, is a means of minimizing violation of the IIA property, we have to decide on a definition of subsets of vehicles. This is where the clustering applies. By grouping vehicles on physical attributes we are assuming that the unobserved attributes which influence preferences (and relate to unexplained taste variation) are correlated strongly with the physical attributes. We know from the analysis above that the performance attributes used as observed attributes influencing preferences are not strongly correlated with the physical attributes. By grouping vehicles on physical attributes and running a full information maximum likelihood nested-logit model we are assigning the same index of dissimilarity to vehicles within a branch but allowing the index to vary between branches. Since a source of IIA violation is the presence of different degrees of dissimilarity between pairs of alternatives, with violation increasing, *ceteris paribus*, as the unobserved attributes increase, clustering on systematic criteria postulated to be related to the unobserved attributes of choice is a means of minimizing violation of IIA. The 23 type groups plus the truck file and pre-1964 vintage file can define 25 vehicle groups. The sequential nested logit model is

$$P_{t\in TG} = P_{t/g} \cdot P_g \quad \forall \quad t \in T,\ g \in G,\ t \neq t',\ g \neq g'$$

where

$P_{t|g}$ is the probability of selecting vehicle t given group (choice set) g

P_g is the probability of selecting group (choice set) g out of the set of groups (choice sets).

This approach does not necessarily imply a sequentially recursive *behavioural* process; it could simply be a means of minimizing violation of IIA. The final vehicle groupings are defined on 57 physical attributes. This set of criteria are consistent with increasing the similarity of vehicles within a branch and decreasing it between branches, but there is no presumption that the groupings reflect the determinants of consumer choice (or are necessarily correlated with all such influences). As mentioned earlier, the grouping criteria are supply-side (technology) based. Thus we would suggest that a sequential nested-choice model using our groupings is designed primarily to improve the parameter estimates. However, the parameter estimates are not efficient; the final step is to either (1) use the consistent estimates from the sequential model as starting values in a Full Information Maximum Likelihood (FIML) estimation which simultaneously (in a single equation) estimates the nested structure or (2) assume no knowledge of consistent estimates and use the FIML procedure to obtain the only, but efficient, parameter estimates.

§8. Conclusion

The grouping of vehicles on a systematic basis is designed to improve the explanatory and predictive power of vehicle type-choice models, a central module in the development of a set of models of automobile demand (which includes automobile number, vehicle usage and mode choice for the work trip). The groupings reported in this paper can be used in the search for suitable type-choice holdings and transactions models.

Furthermore, the approach can be used in other transportation contexts where an improved definition of grouped commodities is desired; examples are categories of freight consigned and public passenger services. This practice should assist in an understanding of the commonality of commodities/services and hence provide a richer basis for decision making. While the value of this approach is clear in forecasting demand, it can be used in exploratory policy analysis in the search for sufficiently differentiable options/strategies; when mapped into the population we have a basis for developing market-specific promotional campaigns.

Acknowledgment

This paper is Working Paper No. 4 one in a series prepared for a study on the Dimensions of Automobile Demand, supported financially by the Australian National Energy Research Development and Demonstration Program, the Australian Research Grants Scheme, and a Macquarie University Research Grant. The assistance of study members in data development has been substantial, especially that of Tony Miller, Woh Choo and Otto Hanzel. The support of the National Roads and Motorists Association and *Wheels Magazine* in giving access to their files on automobile tests is very much appreciated. We are indebted to Jan Vegelius of the University of Uppsala who has contributed many useful ideas on correlation indices of similarity, to George Cooney and Les Sullivan for advice on cluster analysis, and the Editor in Chief for suggestions on the final form of this paper.

Foreign summaries

Il est souvent nécessaire de classer les automobiles en groupes homogènes pour modéliser le choix et la demande des divers types disponibles ainsi que pour évaluer les incidences des modifications structurelles du parc sur la consommation d'énergie. Ces classifications peuvent se fonder sur des critères techniques (caractéristiques d'offre) ou de comportement (caractéristiques de demande); cela dépendra surtout du but poursuivi. Les auteurs proposent et décrivent une méthode de classification systématique des véhicules. Sur la base de données nouvelles correspondant à deux cents caractéristiques de 827 types représentatifs, les automobiles ont été groupées par les méthodes de composantes principales et de classification. Une attention particulière a été apportée à un index approprié de similarité entre paires d'attributs mesurés par des systèmes d'unités différents. On aboutit ainsi à deux systèmes de classification, l'un adapté à la modélisation des choix, l'autre au suivi de l'évolution de la consommation énergétique.

Die Einteilung der Automobile in homogene Gruppen wird häufig zur Modellierung der Auswahl von Automobiltypen und der Nachfrage nach diesen ebenso benötigt wie zur Abschätzung des Energieverbrauchs unter Berücksichtigung der Veränderungen in der Zusammensetzung der Fahrzeugflotte im Laufe der Zeit.

Automobile können nach technischen Kriterien (Angebotsseite) und/oder nach Verhaltenskriterien (Nachfrageseite) eingeteilt werden; die Auswahl eines entsprechenden Ansatzes hängt von der jeweils aktuellen Aufgabe ab. Wir schlagen eine Methode vor und stellen sie dar, wie man Fahrzeuge systemanalytisch klassifiziert. Zur Verfügung stand eine Datenmenge von 827 repräsentativen Fahrzeugen und mehr als 200 Attributen, aufgrund derer die Automobile mit Hilfe der Hauptkomponenten- und Clusteranalyse gruppiert wurden.

Spezielles Augenmerk wird auf den angenäherten Ähnlichkeitsindex zwischen Attributepaaren, bei denen die Meßskalen unterschiedlich sind, geworfen. Als Ergebnis ergeben sich zwei Sätze gruppierter Automobile: einer eignet sich für Modellierung der Verkehrsmittelwahl, der andere, um Trends in der Entwicklung des Energieverbrauchs zu verdeutlichen.

A menudo se requiere clasificar a los automóviles en grupos homogéneos a fin de modelar la elección y demanda por distintos tipos de auto, así como para estimar las implicancias, en términos de consumo de energía, producto de cambios en el tiempo en la composición de la flota de vehículos.

Los autos se pueden agrupar de acuerdo a criterios tecnológicos (oferta) y/o criterios de comportamiento de los usuarios (demanda); la selección de uno u otro enfoque depende del problema a resolver. Se propone y reseña un método sistemático de clasificación de vehículos. Mediante métodos estadísticos sofisticados (cluster analysis y component score) se clasificó una nueva base de datos consistente en 827 vehículos representativos y más de 200 atributos. Se dió especial consideración a índices de similitud adecuados entre pares de atributos, cuando las escales de medición era diferentes.

Los resultados consistieron en dos conjuntos de agrupamientos de vehículos, uno adecuado para modelación de demanda y el otro para seguimiento de tendencias en cuanto a consumo de energía.

References

Berkowitz, M. K., Gallini, N., Miller, E. J., and Wolfe, R. A., 1983, *Modelling, Vehicle Type Choice, Mode Choice and Vehicle Kilometers Travelled for One Vehicle Households*, University of Toronto (mimeo).

Dubin, R., Greene, D. L., and Begovich, C., 1979, Multivariate Classification of automobiles by use of an automobile characteristics data base, *Transportation Research Record*, **726**, 29–37.

Filmer, R. J., and Mannion, B., 1980, An econometric model of passenger motor vehicle usage in Australia, *Australian Road Research Board Proceedings*, **10** (5), 154–164.

Hensher, D. A., 1981, Towards a design of consumer durables, *Journal of Economic Psychology*, **1** (2), 135–164.

Hensher, D. A., 1982, The automobile and the future: some issues, *Transport Policy and Decision Making*, **2** (2), 93–128.

Hensher, D. A., and LePlastrier, V., 1982, *A dynamic discrete-choice model of household automobile fleet size and composition* (mimeo).

Hensher, D. A., and Manefield, T., 1981, *A structured logit model of automobile acquisition and type choice*, Proceedings of 7th Australian Transport Research Forum, Hobart, March, 625–654.

Hensher, D. A., and Miller, A., 1983, *Vehicle Attribute Files, Documentation Report No. 1, Dimensions of Automobile Demand Project*, School of Economic and Financial Studies, Macquarie University, Sydney.

Hensher, D. A., 1983, *Automobile Type Choice: Alternative Specifications for Discrete-Choice Modelling*. Working Paper No. 7, Dimensions of Automobile Demand Project, School of Economic and Financial Studies, Macquarie University, Sydney.

Janson, S., and Vegelius, J., 1979, On the generalizations of the G INDEX and the PHI coefficient to nominal scales, *Multivariate Behavioural Research*, **14**, 255–269.

Janson, S., and Vegelius, J., 1982, Correlation coefficients for more than one scale type, *Multivariate Behavioral Research*, **17**, 271–284.

Kendall, M. G., 1948, *Rank Correlation Methods* (London: Charles Griffin).

Kim, J. O., and Mueller, C. W., 1978, *Factor Analysis: Statistical Methods and Practical Issues* (London: Sage Publication).

Lave, C. A., and Train, K., 1979, 'A behavioural disaggregate model of automobile-type choice', *Transportation Research*, **13** (A), 1–9.

Manski, C. F., and Sherman, L., 1980, An empirical analysis of household choice among motor vehicles, *Transportation Research*, **14A** (5), 349–366.

Miller, A., and Hensher, D. A., 1983, 'Allocation of model registrations to vehicle makes', *Documentation Report 6*, Dimensions of Automobile Demand Project, School of Economic and Financial Studies, Macquarie University, Sydney.

Ohta, M., and Griliches, Z., 1975, Automobile prices revisited: extensions of the hedonic hypothesis, *Household Production and Consumption*, edited by N. E. Terleckj (National Bureau of Economic Research, New York), pp. 375–397.

TRAIN, K., and LOHRER, M., 1983, *Vehicle ownership and usage: an integrated system of disaggregate demand models*, paper presented at Transportation Research Board Annual Meeting, Washington D.C., January.

VEGELIUS, J., 1973, Correlation coefficients as scalar products in euclidean spaces, *Report 145*, Department of Psychology, University of Uppsala, Sweden.

VEGELIUS, J., DAHLQUIST, B., and SMITH, N. C., 1983, TCORALL and TEFAM—Correlation measures with varying scales, *Documentation Report No. 2*, Dimensions of Automobile Demand Project, School of Economic and Financial Studies, Macquarie University, Sydney.

EDITORIAL SUGGESTIONS FOR FURTHER READING

DUBIN, R., GREENE, D. L., and BEGOVICH, C., 1979, Multivariate classification of automobiles by use of an automobile characteristics data base, *Transportation Research Record*, **726**, 29–37.

Interest in forecasting the fuel efficiency of the automobile population has led to the development of automobile market-shares demand models. The validity of these models depends on the automobile classification used, yet little rigorous attention has been given to the problem of classifying automobiles for demand analysis. All existing models use classifications that are heavily subjective and rely on only one or two vehicle characteristics for classification. A cluster analysis of 125 models of 1975 automobiles was conducted in order to aggregate the vehicles into homogeneous groups suitable for modelling the demand for automobiles by vehicle type. Eight variables extracted from an automobile characteristics data base developed by the U.S. Department of Transportation were employed: curb weight, wheelbase, engine displacement, roominess, passenger capacity, fuel economy, list price, and power-to-weight ratio. Several weighting schemes, two distance metrics, and hierarchical as well as nonhierarchical clustering techniques were used. The analysis strongly indicated that two- and six-group configuration were important. Within the six-group clustering, the three groups that had the highest average seat kilometres per litre and seats per initial cost comprised more than 80% of sales in 1975. A comparison of the cluster-analysis grouping with another classification used in a recent econometric automobile-demand model showed that the multivariate clustering did a consistently better job of accounting for the variability of vehicle characteristics.

(Authors)

LAVE, C. A., and TRAIN, K., 1979, A behavioural disaggregate model of automobile type choice, *Transportation Research*, **13A** (1), 1–9.

Previous models of auto-type choice have not been able to disentangle very much of the structure of the household's auto-choice decision: the models assumed that very few auto characteristics affect choice, and often these few parameters were estimated with low precision. Hence the models had only limited use in forecasting the effects of government policies to influence transportation energy consumption. The present paper introduces a multinomial logit model for the type of car that households will choose to buy. The model includes a large variety of auto characteristics as explanatory variables, as well as a large number of characteristics of the household and the driving environment. The model fits the data quite well, and all of the variables enter with the correct signs and plausible magnitudes.

(Authors)

MANSKI, C. F., and SHERMAN, L., 1980, An empirical analysis of household choice among motor vehicles, *Transportation Research*, **14A** (5), 349–66.

This paper presents an econometric model explaining the make, model and vintage composition of individual household motor vehicle holdings. We view the household as yearly evaluating its current holdings of vehicles and updating these as desired. The utility of any vehicle, or vehicle pair, to a household is assumed a function of vehicle seating capacity, luggage capacity, weight, acceleration time, noise level, scrappage probability, price, operating cost and of a search-transactions cost associated with entering the vehicle market. Household size, age, education, income, number of workers and residential location condition the utility function. A multinomial logit model probabilistically describes each household's choice among vehicle alternatives. The empirical analysis is based on a random sample of households drawn from a nationwide rotating consumer panel. Two models are estimated, one explaining vehicle choices in households holding a single vehicle, the other explaining the composition of holdings in two-vehicle households. Among the many empirical findings, one prominent result is that each household has an optimal vehicle seating capacity which varies directly with household size. We find that most aspects of vehicle performances have little effect on choices but, counter-intuitively, sluggish vehicles appear strongly prefered to quick ones. Vehicle price, operating costs and transactions costs are all important determinants of vehicle utility but the influence of price and operating cost varies considerably among socioeconomic and demographic groupings.

(Authors)

This was one paper in a special issue of *Transportation Research* on Automobile Choice and Its Energy Implications.

MOGRIDGE, M. J. H., 1983, *The Car Market:* A study of the statics and dynamics of supply-demand equilibrium (London: Pion).

The aim of this book is to examine the U.K. car market as an example of car markets in any country, and more generally of any economic market for durable goods, in order to show how market processes work. Specifically, the U.K. statistics on cars give details both of the number and of the prices of cars in total and actually traded in the market, and of the expenditure by households and companies on cars together with information about their ownership over a long period when the growth of car ownership fell from its intial high rates to its current low rates. In particular, the shocks to the car market caused by the rise in petrol prices in the 1970s can be traced in detail—how they caused falls in demand for new and large cars, extended the lives of old cars, but had little effect on the total number of cars. The static and dynamic equilibria can therefore be traced in detail, and the consequences for the kind of modelling required to reproduce this detail are set out, together with the author's conclusions for policy directions for national governments.

(Author)