Analysis of the Influence of Certain Personal and Distance-Based Factors on the Tolerance of Circuity in Ridesharing

JON D. FRICKER and Md. GOLAM HABIB

ABSTRACT

An attempt is made to develop a regression model that will explain which factors influence the circuity associated with an individual's ride to work in a carpool or vanpool. The factors are of two types: those related to the trip and those that characterize the traveler. Although the resulting model was not extremely strong, certain of its elements indicate promise for variations and extensions of the work presented here. The factors included in the final model are reasonable and logically consistent. Trip-related factors, such as number of pickups, degree of common destination, and dropoff ride distance, emerge as more important explanatory variables than the personal variables, although job type also plays a useful predictive role. Application of this sort of model to ridesharing matching operations holds promise as a screening procedure for proposed matches.

Data collected for an earlier paper (see paper by Fricker in this Record) provided a rare opportunity to examine certain relationships that may exist between the distance traveled by carpoolers in excess of their drive-alone distance and (a) the personal makeup of the individuals involved and (b) the characteristics of the components of the shared-ride trip. The relationships between circuity and the personal and trip-related factors are investigated using a regression model. The idea is to construct a regression model

$$Y = b_o + b_1X_1 + \dots + b_mX_m \quad (1)$$

that makes possible the conversion of known or easily estimated X_j-values into a reasonable estimate of maximum tolerated circuity (Y). These X_j-values could be based on a proposed carpool matching. The Y-value that results from the use of Expression 1, when the b_j-values have been established, could serve as a measure of the probability of success for the proposed matching. In this paper are described the factors that were available in the data set, the relationships hypothesized, the analysis performed, and the prospects for possible application by ridesharing matching agencies.

CANDIDATE VARIABLES FOR REGRESSION MODEL

In this section, variables that describe the characteristics of individuals and their shared-ride work trips with respect to circuity in ridesharing are defined. Eighteen variables have been introduced for the analysis (Table 1). The variables can be classified as two basic types: trip-related or spatial variables and person-related or behavioral variables. The spatial types are those that mainly describe various components of the journey and their related measures--usually distances or counts. The variables "home-end walking distance" and "number of pickups" (defined later) fall into this group. The behavioral variables describe personal and qualitative aspects associated with each individual in the ridesharing context. "Sex," "sex mix," and "job type" are examples of behavioral variables in the analysis.

The contribution of each variable to the model requires careful interpretation. For example, consider the variable "pickup ride distance" in the circuity model. Spatially, as this distance goes up, the circuity is likely to go up because of the increased detour. But, from a behavioral point of view, this might not be true. Although a person may "suffer" more in the pickup distance, he may be compensated for it by some other components of the journey, which result in lower overall circuity. If this "compensation" were not possible, the overall circuity might be intolerable enough to cause the individual to

TABLE 1 Independent Variables for Regression Model

Symbol	Description
X_1	Drive-alone distance, d_o (mi)
X_2	Ridesharing distance, d_r (mi)
X_3	No. of pickups
X_4	No. of rendezvous points
X_5	No. of dropoffs
X_6	Carpool size, n
X_7	No. of sequences
X_8	Sex of carpoolers
X_9, X_{10}	Sex mix in carpool
X_{11}, X_{12}	Degree of common destination
$X_{13}, \dots, X_{18}$	Job types
X_{19}, X_{20}	Job mix in carpool
X_{21}	Home-end walking distance (mi)
X_{22}	Pickup ride distance (mi)
X_{23}	Line-haul distance (mi)
X_{24}	Dropoff ride distance (mi)
X_{25}	Work-end walking distance (mi)
X_{26}	Circuity ratio, CR or Y_1
X_{27}	Circuity distance, CD or Y_2 (mi)

School of Civil Engineering, Purdue University, West Lafayette, Ind. 47907.

forgo membership in the carpool. The individual's circuity threshold having been exceeded, this person would, of course, not be included in the ridesharing data set. This and other aspects of the search for an individual's maximum tolerable circuity are shown in Figure 1.

Figure 1 shows how an individual's circuity factor (CF) value changes from day to day as the driving responsibility rotates or the number of riders varies. If interest centers on the "circuity factor threshold" for an individual, the best available approximation is the maximum CF value experienced by that individual. For this reason, the Y- and X_i-values employed in the model correspond to the "max CF" day for each of the 206 individuals in the survey data. Because some individuals have two distinct days with the same maximum CF value, both days are included in the analysis, which brings the number of maximum CF observations to 215.

DEFINITIONS AND HYPOTHESES

The candidate variables and their contribution to the model must be viewed in a comprehensive way. The following subsections contain a description of each variable proposed for the current analysis, including a definition and hypothesized "circuity relationship" for each.

Y_1 or X_{26}, Circuity Ratio

The circuity ratio (CR) is the ratio of ridesharing distance (d_r) to drive-alone distance (d_o). This is the dependent variable Y_1 in the first CF model reflected in positive and negative coefficients.

Y_2 or X_{27}, Circuity Distance, $d_r - d_o$

This is the dependent variable in the second CF model. It is the difference between ridesharing distance (d_r) and drive-alone distance (d_o).

X_1, Drive-Alone Distance, d_o

This is the distance covered by an individual over his usual route from home to work when driving alone.

Drive-alone distance can be thought of as a proxy for various trip attributes, such as travel costs, travel time, and comfort. Individuals with large d_o-values could be considered prime candidates for ridesharing--to share travel costs, prevent boredom, provide company in remote areas, and so forth.

For smaller d_o-values, the net amount of detour (circuity distance, CD) to be tolerated by an individual in ridesharing is likely to be smaller, but the corresponding ratio measure of circuity (circuity ratio, CR) could still be about average or even larger. For large d_o's, the total deviation acceptable would be larger but, because it is the ratio of two large numbers, the CR is likely to be relatively close to unity. Hence, a positive slope for CD and a negative slope for CR when plotted against d_o.

X_2, Ridesharing Distance, d_r

This is the distance traveled by an individual in the journey to work as a carpool member. The distance components of the trip are home-end walking (X_{21}), pickup riding (X_{22}), line haul (X_{23}), dropoff riding (X_{24}), and work-end walking (X_{25}).

The effect of this variable on CD and CR is quite similar to that of d_o. It also gives an idea about the extent of various possible deterrents to ridesharing, such as extra distance and extra time. The variable d_r and its components might provide some clues about individual ridesharing decisions by revealing any strong correlations they might have with CR or CD, or both. Care should be exercised in using d_r, however. If d_o is already in the model, the definitions $CD = d_r - d_o$ and $CR = d_r/d_o$ make inclusion of d_r redundant and invite collinearity.

As d_r increases, CD is likely to increase, but CR may remain fairly stable. Hence, the coefficient of d_r would be expected to be positive when CD is the dependent variable but be approximately zero for CR.

X_3, Number of Pickups

This is the number of stops to pick up other carpool members experienced by a given individual from the time the individual in question enters the vehicle until the last person is picked up. In a five-person carpool, if person A drives and picks up members B, C, D, and E individually in that order, the X_3 count for person C on that day would be 2. For A, it would be 4. After E is picked up, the line-haul phase (X_{23}) begins.

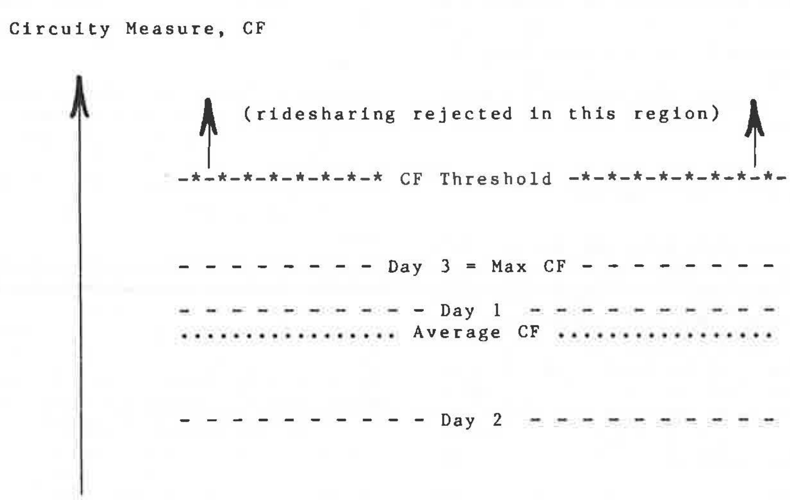


FIGURE 1 Seeking an individual's maximum tolerable circuity.

An increase in X_3 means reduced overall travel speed and may lead to unscheduled delays in excess of the standard boarding time or unnecessary extra distance if the person to be picked up is found to not be making the trip that day. Thus the variable X_3 is a measure of inconvenience (or ridesharing "cost") that is not reflected in distance values but may be a significant factor in the decision to rideshare. More stops mean more carpool members to be picked up and usually more detour. Hence, as the number of stops increases, both CR and CD are expected to increase, resulting in a positive coefficient for this variable in both the CD and CR regression equations.

X_4, Number of Rendezvous Points

Establishing a rendezvous point avoids the circuity associated with going door to door to pick up pool members. On the other hand, the typical rendezvous point involves the meeting of three or more individuals, each with his own potential for late arrival. It is expected that the use of rendezvous points leads to lower CR and CD values because purely distance values are involved. However, if rendezvous points are viewed quite negatively by a prospective carpooler, circuity elsewhere in the ridesharing route might have to be quite low to attract and retain that member. Thus only low-CF cases would appear in the data set when rendezvous points are involved. This notion of a CF threshold and its being influenced by rendezvous and other particular variables formed the basis for a subsequent analysis involving the individuals in the data set who use rendezvous points. (Note: a rendezvous point is not included in an individual's pickup count, X_3.)

X_5, Number of Dropoffs

This is the number of stops made to drop off other members before an individual's own dropoff point is reached.

The aversion to dropoffs may not be as severe as is that to pickups. There is little chance of unforeseen delay and, in most cases, the walk distance from dropoff point to workplace is much less than in the drive-alone case. Although CR and CD could be expected to increase with X_5, the coefficients should have smaller values than X_3. The hypothesis is that this variable is not a major factor and will not appear in the final model.

X_6, Carpool Size, n

This is the number of individuals who participate as members of a carpool on a particular day. In the data set, X_6 ranges from 2 to 6, with the one exception being a 9-member vanpool. The value of X_6 may change for a given carpool from day to day if, on a regular basis, some members ride only on certain days. For a given day, however, each of the n riding members of a given carpool is assigned the value $X_6 = n$.

Carpool size has both positive and negative ramifications. Larger carpools can mean greater savings in travel cost, with more persons to share costs that rise relatively little for each new member. If driving responsibilities are rotated, the number of days off between driving days increases as X_6 increases. For most people, this is desirable.

On the negative side, more members usually mean greater values for pickup and dropoff counts, pickup and dropoff distances, and the increased potential for incompatible personalities. The need for rendezvous points also becomes more likely. Again, if there were no threshold on the tolerance of various ridesharing characteristics, it could simply be hypothesized that CD and CF both go up with carpool size. However, the negative aspects just mentioned act in opposition to cost savings and relief from driving to complicate the relationship between X_6 and circuity. Examining the negative influences (X_3, X_4, X_5, X_{22}, and X_{24}) and their relationship to circuity separately may clarify the picture as the model is built. Until then, the expectation is that circuity increases with carpool size (i.e., the coefficient of X_6 will be positive).

X_7, Number of Sequences

This variable recognizes that a carpool may take on different configurations from day to day. On a regular basis, certain members may not participate on particular days of the week. Even if all members ride daily, the responsibility for driving may rotate, which in turn usually affects the sequence of pickups and dropoffs, as well as distance components associated with them. Each time some aspect of the carpool changes, there is a new sequence.

In a carpool in which the driving duty is uniformly rotated, an increase in X_7 means that an individual has more days as a rider, which is usually seen as desirable. However, more drivers increase the probability of one or more being unreliable, unsafe, or otherwise negatively viewed. Furthermore, any significant circuity may be less tolerated as a rider than as a driver. It is the hypothesis that the "signals" in the data may be so mixed that no significant relationship between the dependent variable and X_7 will be detected.

X_8, Sex of Carpooler

This is the sex of an individual in a carpool, indicated in the data set by introducing the dummy variables 0 and 1 for male and female, respectively.

Because there is no preconceived notion about whether men or women are more tolerant of circuity, the hypothesis states that there is no difference. The coefficient of X_8 will be tested to see if it can indeed be considered 0.

X_9, X_{10}, Sex Mix in the Carpool

This is the combination of the members of a carpool on the basis of sex.

A carpool could be formed by all males, all females, or a mix of the sexes. Indicator variables, given in Table 2, were used here. It is reasonable to expect that carpools of one sex, if due to common employment types and interests, would be more likely to form and be sustained. Although possible collinearity with X_8 is recognized, it is hypothesized that greater circuities will be tolerated by riders in carpools of one sex.

TABLE 2 Sex Mix in Carpools

Type	X_9	X_{10}
All male	0	1
All female	1	0
Mixed	0	0

X_{11}, X_{12}, Degree of Common Destination

The data in Table 3 indicate how indicator values X_{11} and X_{12} represent the degree of common destination for members of a carpool. Two or more riders share a common destination if they have the same dropoff points and work in the same (or adjacent) buildings.

TABLE 3 Degree of Common Destination

Destination	X_{11}	X_{12}
All same	0	1
Some same	0	0
All different	1	0

A ridesharing arrangement is more convenient if the participants have the same work place. Also, friendship and shared experiences at the work place could make for a more compatible membership. If this compatibility translates into a greater tolerated circuity at the (home) pickup end, it is to be expected that an increase in circuity would accompany an increase in the degree of common destination.

X_{13} Through X_{18}, Job Type

Each individual was placed in one of six categories on the basis of survey information and listings in the Purdue staff directory: faculty (X_{13}), administrative (X_{14}), clerical (X_{15}), maintenance (X_{16}), student (X_{17}), and other (X_{18}). In the input file prepared for data analysis, the indicator variables X_{13} through X_{18} represent these six categories by 10000, 01000, 00100, 00010, 00001, and 00000, respectively.

Job type may serve as a proxy for the socioeconomic variables about which information was not otherwise available. Higher income individuals may be less tolerant of circuity and less likely to appear in the data set at all. Also, the work hours for maintenance personnel are not the same as those for faculty members and are not likely to be as flexible. The tolerance of extra time spent in travel and the ability or willingness to accommodate someone else's schedule may not be the same for students and administrators.

Incompatibility in schedules would be minimized if all pool members held similar jobs. This may be picked up by the "common destination" and "mix of job type" variables. If the more flexible jobs also tend to be higher paying, the income effect may be detected as a lower circuity accepted when $X_{13} = 1$ or $X_{14} = 1$, and coefficients are approximately 0 for the variables X_{15} through X_{18}.

X_{19}, X_{20}, Mix of Job Types in Carpool

As mentioned previously, the mix of job types in a carpool may indicate the degree of work schedule compatibility among its members. The holders of similar kinds of jobs might also have a greater degree of compatibility in life-style, interests, habits, and values that would overshadow moderate increases in circuity (Table 4).

TABLE 4 Job Mix Variable Values

Type	X_{19}	X_{20}
All same	0	1
Mixed	0	0
All different	1	0

X_{21}, Home-End Walking Distance

This is the distance an individual must walk to reach the pickup point in the carpool.

Quite often this value is 0 because doorstep pickups at one's home are considered to involve about the same (insignificant) distance as does the drive-alone case. However, if $X_{21} >> 0$, this adds to the negative aspects of ridesharing. As the walking distance goes up, both the tolerated circuity and the actual associated d_r's might be expected to decrease.

X_{22}, Pickup Ride Distance

This is the distance traveled by a pool member from his point of pickup until the last member has been picked up.

Pickup distance adds to travel time and cost and involves the stops and starts described under the "pickup count" variable. Because line-haul portions of the d_r and d_o routes (X_{23}) are usually quite similar, the pickup and dropoff phases of the ridesharing journey contribute the most to circuity. Thus a close, positively signed relationship between X_{22} and circuity can be expected.

X_{23}, Line-Haul Distance

This is the roadway distance traveled between the last pickup at the home end and the first dropoff at the work end in ridesharing.

This is usually the largest component of the shared work trip. If the pickup and dropoff phases of a carpool are quite localized, the in-vehicle drive-alone distance will be close to the value of the shared-ride line-haul distance. For this reason a high correlation between X_{23} and d_o can be expected. If X_{23} remains in the model and the dependent variable is circuity ratio (CR), a reasonable hypothesis is that the ratio does not change with line-haul distance: $b_{23} = 0$. If circuity distance (CD) is the dependent variable, a strong positive correlation is expected.

X_{24}, Dropoff Ride Distance

This is the distance from the end of the line-haul segment at the work end (the first dropoff point) to the point of one's own dropoff.

The variable can be interpreted in two opposite ways. It could be said that, as X_{24} increases, so does circuity. However, if X_{24} is large enough to make a significant difference in circuity, it may also be significant enough to "drive away" a carpool member. Thus any high X_{24} values in the survey might have to be compensated by low values of other distance components, leading to a negative or zero coefficient for X_{24}. On the other hand, a positive coefficient might be evidence of increased tolerance of circuity as the work trip nears its completion.

X_{25}, Work-End Walking Distance

This is the distance one must walk after being dropped off from the carpool to get to one's work place.

As with home-end walking access distance (X_{21}), a doorstep dropoff is represented by $X_{25} = 0$. This zero value happens frequently and is the main reason for the surprising number of instances in which d_r is less than d_o. A doorstep dropoff eliminates a walk from the parking location used on drive-alone days. If this walking distance is greater than the circuity involved in the other shared-ride distance components, the $d_r < d_o$ case results. Because X_{25} is typically zero or small relative to other components, $b_{25} \approx 0$ can be expected in the final model.

METHODOLOGY AND ANALYSIS OF RESULTS

Statistical Analysis

In the previous section, the variables that affect circuity in the ridesharing trip to work were defined. They are either trip related, socioeconomic factors or related to subjective measures like comfort, convenience, and reliability. Through regression analysis, the effect of these factors on circuity factors (CFs)--circuity ratio (CR), and circuity distance (CD)--may be determined. An attempt is made to determine the factor or set of factors that best explains traveler response to this negative aspect of ridesharing.

Although not obvious from the scatter plots of Y versus the individual Xs listed in Table 1, subtle relationships between some factors and CF can be detected. By using analysis of variance (ANOVA), the consequences of certain demographic variables can be identified. This is done later in this section. In attempting to develop a meaningful regression model, it should be possible to determine whether CR or CD is the more useful measure of detour in carpooling.

Modeling of Circuity Factors

Simple Linear Regression and Scattergrams

CR and CD are the two candidates for the dependent variable. First, simple linear regression models for CR and CD versus each of the independent variables listed in Table 1 were formulated. At the same time, the scatter plots of all of the independent variables were generated and studied to detect any apparent trends (including nonlinearity) among the variables. The results of these procedures, especially with respect to the relationships hypothesized in the previous section, are summarized in Table 5. No noticeable nonlinear trend in any of the independent variables with respect to the circuity factors was found.

TABLE 5 Effect of Independent Variables on Circuity Factors (simple linear regression)

	b-coefficient				
	Sign				
Variable	Expected	Actual	Magnitude	R^2	Significance
For Circuity Ratio					
X_1	-	0	-0.002	.0151	.07
X_2	0	0	-0.0003	.0005	.75
X_3	+	+	0.034	.0758	.00
X_4	-	0	-0.013	.0014	.59
X_5	+/0	-	-0.020	.0198	.03
X_6	+	0	0.010	.0086	.17
X_7	0	0	0.001	.0002	.85
X_8	0	-	-0.050	.0205	.03
X_9	Large	0	-0.032	.0094	.15
X_{10}	Large	0	-0.088	.0003	.81
X_{11}	-	0	0.008	.00003	.93
X_{12}	Large	0	0.044	.0088	.17
X_{13}		0	0.012	.0003	.81
X_{14}		0	-0.060	.0101	.14
X_{15}		0	0.002	.00003	.93
X_{16}		0	0.021	.0017	.55
X_{17}		0	0.075	.0058	.26
X_{19}		-	-0.004	.0001	.89
X_{20}	Large	-	-0.019	.0032	.40
X_{21}	-	0	-0.224	.0062	.25
X_{22}	+	+	0.007	.0376	.00
X_{23}	0	0	-0.002	.0149	.07
X_{24}	+/-	+	0.023	.3160	.00
X_{25}	+	0	0.047	.0012	.61
X_{27}	-	+	0.048	.3903	.00
For Circuity Distance					
X_1	+	+	0.051	.1020	.00
X_2	+	+	0.068	.2037	.00
X_3	+	+	0.743	.2082	.00
X_4	-	+	0.676	.0208	.03
X_5	+/0	+	0.833	.2010	.00
X_6	+	+	0.318	.0544	.00
X_7	0	+	0.521	.1203	.00
X_8	0	0	-0.320	.0049	.30
X_9	Large	0	-0.290	.0046	.32
X_{10}	Large	0	-0.500	.0049	.30
X_{11}	-	0	0.274	.0042	.34
X_{12}	Large	-	-0.925	.0224	.03
X_{13}		0	-0.007	.0000	.99
X_{14}		-	-1.144	.0214	.03
X_{15}		0	-0.214	.0025	.46
X_{16}		0	0.135	.0004	.76
X_{17}		+	2.402	.0351	.00
X_{19}		0	-0.758	.0151	.07
X_{20}	Large	-	-0.978	.0517	.00
X_{21}	-	0	-4.102	.0122	.10
X_{22}	+	+	0.210	.2026	.00
X_{23}	+	+	0.060	.1073	.00
X_{24}	+/-	+	0.733	.1880	.00
X_{25}	+	0	-0.714	.0016	.55
X_{26}	-	+	8.137	.3903	.00

Multiple Linear Regression Analysis: Correlation Matrix

The presence of collinearity among the variables (a high correlation coefficient) was also a concern. For example, the variables "drive-alone distance" (X_1) and "ridesharing distance" (X_2) are quite highly correlated ($r = 0.989$), as might be expected. High correlation was also observed between "line-haul distance" (X_{23}) and "drive-alone distance" ($r = 0.953$) and "ridesharing distance" ($r = 0.947$), respectively. These variables, and other pairs with high r-values ($r > 0.60$ was used as the criterion), should not be placed in a regression model at the same time. A more detailed discussion of this topic is given elsewhere (1).

After repeated applications of the surviving variables, using stepwise multiple linear regression (2) to determine the impacts of adding and removing promising independent variables, certain models emerged as the most meaningful. They are given in Table 6. The resulting performance of the models ($R^2 = 0.219$ for CR and 0.467 for CD) is lower than is normally acceptable in regression analysis. However, it must be remembered that what is being examined is not the behavior of a physical specimen in a laboratory but rather the complex behavior of a group of individual travelers in a variety of trip-making environments. After the basic stepwise regression package (2) had been used to determine the impacts of adding and removing variables to identify the probable most explanatory independent variables, each of the variables listed in Table 6 was examined, before its inclusion in or exclusion from the model, for its effect on the overall model. The formula used to determine each variable's contribution to the F-value of the model was based on the

TABLE 6 Multiple Linear Regression Results After Stepwise Approach at 5 Percent Significance Level

Dependent Variable	Circuity Ratio, Y_1	Circuity Distance, Y_2
Variables not used	X_2, X_{23}, X_{27}	X_2, X_{23}, X_{26}
Variables in the model	X_1, X_3, X_8, X_{22}, X_{24}	X_3, X_{11}, X_{12}, X_{17}, X_{22}, X_{24}
Variables close to being accepted	X_{14}(0.07), X_{25}(0.08), X_{13}(0.08)	X_6(0.10), X_4(0.13), X_7(0.14)
R^2	0.219	0.467

partial f-test statistics (3) of the successive sets of variables:

$$\Delta F_{1,\nu} = \left[(R_p^2 - R_{p-1}^2)/1\right]/(1 - R_p^2)\nu \quad (2)$$

where

$\nu = n - p$,
n = data points, and
p = number of parameters in the model (4).

Table 7 gives a summary of the regression models that were developed by this method. The residual plots of the models showed no noticeable indications of nonnormality.

Despite the low R^2 values, the models are still useful. The standard deviation of circuity ratio values ($\sigma = 0.154$) is surprisingly small (see paper by Fricker in this Record), given (a) the large variability in the components that make up d_r and d_o and (b) the low R^2 values in the regression models developed. A plausible explanation is that the degree of tolerance of circuity in ridesharers is quite consistent from person to person, but the way in which the distance components and nondistance factors contribute to this tolerance level varies widely among individuals. Furthermore, factors not available in this study's data set--travel budgets, stage in life cycle, relative modal attributes, perspectives of nonridesharers--all have an important influence on the decision to join a particular carpool. Nevertheless, the well-behaved nature of the CF values indicates a consistent level of tolerance of circuity among current ridesharing participants.

Between the two measures of CF, CR and CD, CD produces far better results as far as R^2 is concerned. So, CD appears to be a more desirable definition of circuity factor, if an explanatory model of maximum individual CF values is desired.

Independent Variables Accepted and Model Performance

The variables accepted in the CF models are shown schematically in Figure 2. The variables "number of pickups" (X_3), "degree of common destination" (X_{11}, X_{12}), "pickup ride distance" (X_{22}), and "dropoff ride distance" (X_{24}) are common to both equations. Other variables included are, in CR: "drive-alone distance" (X_1), "sex of carpooler" (X_8), "administrator job type" (X_{14}), and "work-end walking distance" (X_{25}), and in CD: "number of sequences" (X_7) and "student" as "job type" (X_{17}). Note that, among the variables, trip-related factors dominate the list of common variables as well as the complete set of variables in both models.

The coefficients of the variables (b) and the relationships among them (sign) in the models are summarized in Table 7. Also presented there are the beta (β) coefficients and t-statistics for the respective variables (i.e., for each b-value) in the models. In the case of the CD model, the coefficients are satisfactory, with the exception that the value of X_7 is not close to zero. But, from the β-values in Table 7, X_7 is the least important variable in the model.

According to the beta coefficients in Table 7, X_3 and X_{22} are among the most valuable of the factors common to both models. For CR, X_1 is the most important variable, whereas for CD the strongest variable is X_{24}. Analysis of the beta coefficients indicates that the trip-related variables are the most important variable types in determining the level of circuity that is tolerated.

The F-test for the regression relation indicates whether the variables in the model have any statistical relationship to the dependent variable. The hypotheses are

$$C_1 : \beta_1 = \beta_2 \ldots = \beta_{p-1} = 0$$

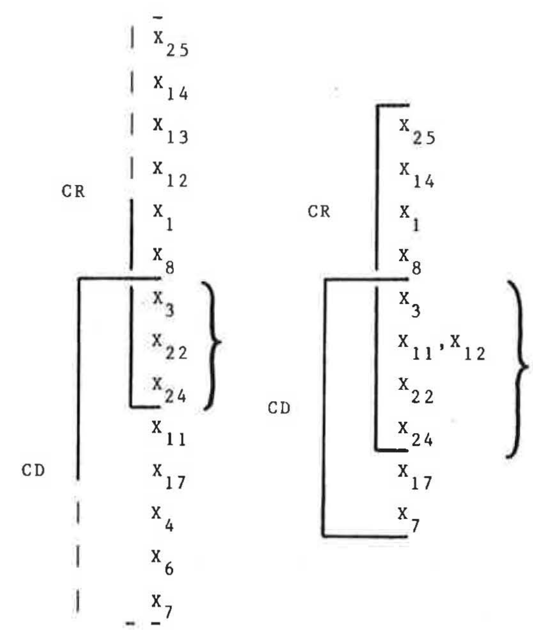
$$C_2 : \text{all } \beta_k \; (k = 1, \ldots, p-1) \neq 0$$

The test results show that the hypothesis that the relationships among the variables in the model exist (C_2) cannot be rejected at an α-level of as low as 0.1 percent for both the CR ($F^* = 8.65$) and

TABLE 7 Regression Results for the Two Models

Independent Variable	CR (Y_1) β	CR (Y_1) b	CR (Y_1) t	CD (Y_2) β	CD (Y_2) b	CD (Y_2) t
Work-end walking distance, X_{25}	0.134	0.185	2.13	–	–	–
Administrative personnel, X_{14}	-0.14	-0.083	-2.24	–	–	–
Sex of the carpooler, X_8	-0.17	-0.060	-2.78	–	–	–
Drive-alone distance, X_1	-0.38	-0.005	-4.83	–	–	–
No. of pickups, X_3	0.27	0.033	4.11	0.30	0.485	5.18
Degree of common destination, X_{11}	0.12	0.04	1.78	0.20	0.832	3.35
Degree of common destination, X_{12}	0.19	0.089	2.66	0.14	0.896	2.41
Pickup ride distance, X_{22}	0.33	0.012	4.10	0.30	0.140	5.34
Dropoff ride distance, X_{24}	0.17	0.022	2.61	0.33	0.560	6.25
Student personnel, X_{17}	–	–	–	0.16	1.995	3.06
No. of sequences, X_7	–	–	–	0.12	0.185	2.06
Intercepts, b_0	–	1.094	32.93	–	-0.926	-2.78
Coefficient of determination, R^2	–	0.275	–	–	0.478	–

Note: In the CR model, $F^* = 8.65$ and $F(.001, 9, 205) = 3.10$; in the CD model, $F^* = 27.05$ and $F(.001, 7, 207) = 3.47$.



(a) Models after Stepwise Regression:

——— Variables included in the model

- - - Variables close to being included in the model

} Variables common to both models

(b) Final Models.

FIGURE 2 Schematic representation of the two models (CR and CD).

the CD ($F^* = 27.05$) models. Hence, valid regression relationships exist (Table 7).

The t-statistics in Table 7 offer the opportunity to test the statistical relationships individually for each variable in the model. The test is whether there exists any relationship between the dependent variable and the variable in question. The hypotheses in this case are

$$C_1 : \beta_k = 0$$

$$C_2 : \beta_k \neq 0$$

For the CR model, the hypothesis that the statistical relationship does not exist (C_1) cannot be accepted for all the b's at an α-level of 2 percent ($t = 2.33$) except for X_{25}, X_{14}, and X_{11}. X_{25} and X_{14} are significant at the 5 percent α-level ($t = 1.96$) and X_{11} is significant at 10 percent. For the CD model, except for X_7, the null hypothesis of $\beta_k = 0$ can be rejected at the 2 percent significance level, and for X_7 rejection at 5 percent is justified. Therefore, for all of the variables as a whole, C_2 cannot be rejected for both the CR and CD models at α-levels of 0.10 and 0.05, respectively. The variables in the CD model, therefore, can be accepted more confidently in the context of the statistical relationships tested by the t-statistic than by examination of the model's beta coefficients. In either case, the variables for the model, as portrayed in Table 7 and Figure 2, represent a reasonable basis for interpreting those aspects of ridesharing behavior that concern trip-related and person-related factors.

Analysis of Variance (ANOVA)

Although no well-defined trend in CFs was observed in the scatter diagrams mentioned earlier, some trends in the mean of the maximum values of CF ("Mean Max CF") could be detected as certain independent variables took on different values. Figure 3

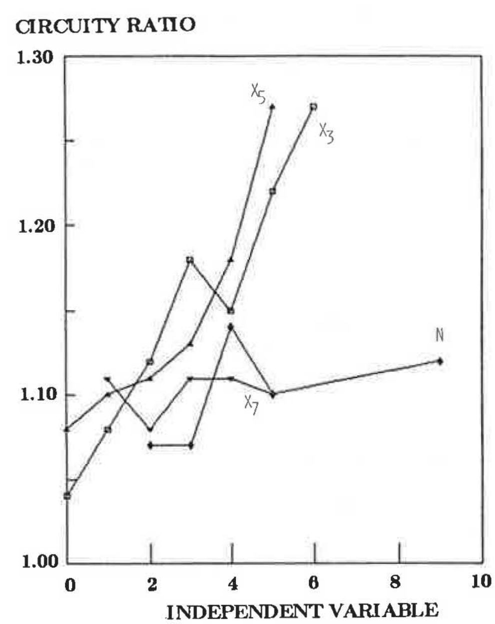


FIGURE 3 Mean of maximum circuity ratios.

TABLE 8 Effect of Job Type on the Mean of the Maximum Values of Circuity Ratio, Circuity Distance, Drive-Alone Distance, and Ridesharing Distance

		CR		CD		d_o		d_r	
Job Type	Sample Size	μ	σ	μ	σ	μ	σ	μ	σ
Faculty	10	1.11	0.10	1.73	1.43	29.77	21.25	31.50	21.60
Administrative	17	1.04	0.06	0.69	1.28	13.33	14.25	16.02	14.29
Clerical	124	1.10	0.19	1.65	2.06	19.29	11.15	20.94	12.09
Maintenance	24	1.12	0.15	1.86	1.88	18.54	6.33	20.40	6.40
Student	6	1.17	0.15	4.07	3.39	25.26	19.34	29.33	20.74
Other	34	1.09	0.10	2.10	2.43	23.15	17.53	25.25	18.33

contains several examples, each of which is associated in some way with carpool size. "Mean Max CR" is seen to grow as the number of dropoffs increases, but the relationship between each individual's Max CR and his X_5 value was not so well defined in the regression analysis: X_5 was excluded from the model. The trace for "number of pickups" straddles the "dropoffs" line in Figure 3 and has a dip at $X_3 = 4$. However, the relationship of X_3 with CR on an individual basis is among the strongest in the model. X_7, "number of sequences," exhibits almost no trend at all in Figure 3, as might be expected. The plot for "carpool size" has a peak at $X_6 = 4$, with lower values on either side of this point. This might be taken as an indication of optimal carpool size, but there is no pattern that can be seen in the scatter plot to support this view.

Perhaps more interesting at this stage of the analysis are the factors that are not easily quantified. Table 8 gives Mean Max CR values by job type. Although their small sample sizes make drawing conclusions risky, several job types attract attention. Although the faculty members have the largest mean d_o- and d_r-values in the table, their Mean Max CR and Mean Max CD values are not among the largest. Instead, the students are seen to tolerate the largest Mean Max CF values. This is not at all surprising. Meanwhile, administrative personnel tend to live closer and have the lowest Mean Max CF values. The other job types are not statistically distinguishable from one another.

In general, CD has demonstrated a greater ability than CR to establish statistically justified relationships with independent variables. While Mean Max CF was not sensitive to any of the qualitative variables, more detailed analysis of Mean Max CD revealed that carpools with "some common" job type mix have significantly longer mean maximum circuity distances than the mix types "common" and "different," which are statistically indistinguishable. The factor "sex" is insignificant for both circuity factors at an α-level of 0.05, as is the variable "sex mix."

CONCLUDING REMARKS

Summary

A regression analysis of the factors that may affect circuity in ridesharing was attempted for three reasons:

1. The existence of a data set of reasonable size and good accuracy built from the perspective of the individual ridesharing participant (see paper by Fricker in this Record);
2. The suprisingly small variance in the circuity values for individuals included in an earlier study; and
3. The potential for development of a screening model whereby potential carpool matches could be evaluated for likelihood of success or markets could be targeted for more intense promotion of ridesharing.

The results were mixed. The R^2-values for the regression models developed were lower than for "textbook cases," but they were not unreasonable for a first attempt at modeling a complex human decision-making process. The regression models that were developed were quite stable (Figure 2), with variables the behavior of which was compatible with logical expectations (Table 5).

Extensions

Several obvious alternatives and extensions to the work described in this paper are possible to better develop the relationship between easily measurable independent variables and dependent variables that represent the potential for a carpool's success. Some of these are:

1. More data. The data set used was developed with distance-based measures in mind. Some nondistance variables were easily generated, but acquiring others that would enhance model development would require a new survey. Also, it must be acknowledged that, although these data are representative in some ways of nationwide values (see paper by Fricker in this Record), data from other locations should be assembled.
2. Separate models for trip-related and person-related variables. Some independent variables may be positively correlated with Y out of the necessity of geography and geometry, and others deal more directly (but vaguely) with tolerances of circuity. The steps taken in the present statistical analysis could be expected to sort out these variable types, but a more explicit implementation of this philosophy may prove useful.
3. Instead of individuals' Max CF values, use their weighted average CFs or all of their daily CFs to develop a regression model. Although Max CF values were used here to seek a CF threshold (Figure 1), using a more central or exhaustive set of data might yield a model with greater explanatory power.
4. Focus on carpools using rendezvous points. Look at Figure 1 from a new perspective. Days 1, 2, and 3 differ only in who drives and, therefore, in the dropoff phase. Each day the same access routes and meeting point are used. This must be because a doorstep pickup arrangement for this group of individuals is intolerable to them. That is, any doorstep configurations would have CFs above the CF threshold level in Figure 1. If the best possible doorstep route or routes could be constructed, an upper bound on the elusive circuity threshold would be established. The Max CF day's level used in this paper can serve as the lower bound.

ACKNOWLEDGMENTS

The work described in this paper was carried out under an UMTA grant. The authors also thank Virgil Anderson for his advice on the statistical analysis. The authors, of course, are responsible for the way the analysis was actually carried out.

REFERENCES

1. J.D. Fricker and Md. Golam Habib. Circuity Factors in Ridesharing: The Individual's Travel Decision. Final Report. UMTA, U.S. Department of Transportation, June 1986.
2. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent. SPSS: Statistical Package for the Social Sciences. McGraw-Hill Book Company, New York, 1975.
3. V.L. Anderson and R.A. McLean. Design of Experiments: A Realistic Approach. Marcel Dekker, Inc., New York, 1974.
4. J. Neter, W. Wasserman, and M.H. Kutner. Applied Linear Statistical Models, 2nd ed. Richard D. Irvin, Inc., Homewood, Ill., 1985.