RESIDENTIAL LOCATION AND THE
JOURNEY TO WORK

An Empirical Analysis

By Patrick S. McCarthy*

There has been a growing trend over the past few years to explore, at the
microeconomic level, the interrelationship between residential location and
transport decisions. This is exemplified in the studies of Senior and Wilson [31],
Beesley and Dalvi [4], Senior [30], Apps [2], Echenique, Feo, Herrera, and
Riquezes [10], and Vickerman [36]. It is in sharp contrast to previous urban
economic analyses, typified by Wingo [39], Alonso [1], and Muth [24], which are
primarily concerned with locational decisions and pay little attention to individual
decisions on transport.

The purpose of the present study is to formulate and test a behavioural model of
transport modal choice. However, unlike similar analyses, including those by
Warner [37], Quarmby [28], Lave [17], DeDonnea [8], and McGillivray [19], the
present model incorporates residential location decision making and specifically
examines its impact upon modal choice for the work trip.

THEORETICAL CONSIDERATIONS
Consider a standard residential land use model in which it is assumed that two
modes of travel—automobile and public transit—are available and equally ac-
cessible for the journey to work.1 An individual’s utility is assumed to be a function
of an \((I \times 1)\) vector of housing attributes \(Q\), a \((J \times 1)\) vector of transport attributes
\(T\),2 and a composite good, \(x\). It is further assumed that the rent per unit of
housing, \(R\), is a function of the vector of housing attributes \(Q\) and distance, \(k\), from
the CBD, and that in the transport cost function, \(C\), the vector of transport at-
tributes \(T\) is an explicit argument. Given these assumptions, an individual’s
Lagrangian function can be expressed as

\[
L = U[H(Q), T, x] + \lambda [Y - R(Q, k)H(Q) - C(T, k, Y) - px]
\]

1The theoretical model presented here is similar to the urban economic model developed by Muth
[24].
2The \(T_j\)'s (\(j = 1, \ldots , J\)) include those attributes on which an individual compares competing
modes.
where $H$ is the consumption of housing activity, $Y$ is total household income, and $p$ is the price of the composite good.

The standard implications for residential choice behaviour can be obtained from the model. They may be summarised as follows:

1. An individual locates at a distance from the CBD such that a small change in $k$ results in marginal housing expenditures which are equal but opposite in sign to marginal transport costs; his equilibrium location is assumed to be stable.

2. Any increase in an individual’s consumption of housing induces him to locate further from the CBD, i.e. $(\partial k / \partial Q) > 0$ if $(\partial H / \partial Q) > 0$.

3. The impact upon equilibrium distance arising from an increase in household income is ambiguous, because it affects both the consumption of housing (which induces an outward move) and the marginal value of time (which induces an inward move), i.e. $(\partial k / \partial Y) \approx 0$.

ESTIMATION PROCEDURE

This section explicitly analyses an individual’s modal choice problem in a binary modal setting, and derives an appropriate statistical estimating model.

Population Choice Behaviour$^1$

In the model posited, an individual’s utility function is expressed by

$$ U = U[H(Q), T, x] .$$  \hspace{1cm} (2)

Although they are not generally included, socioeconomic and unobserved characteristics affect an individual’s utility and, when explicitly incorporated into his utility function, enable equation (2) to be rewritten as

$$ U = U[H(Q), T, x, s, \varepsilon]$$ \hspace{1cm} (3)

where $s$ is a vector of observed socioeconomic characteristics and $\varepsilon$ is a vector of unobserved residential and mode-of-travel attributes as well as of unobserved socioeconomic characteristics. For each individual $i$ ($i = 1, \ldots, N$) in the population, the unobserved vector $\varepsilon_i$ will vary and will induce a variation in observed demands which will be influenced by the structure of tastes. When aggregated over the population, variations in $\varepsilon$ lead to systematic variations in aggregate demand. When the consumption good or activity being studied is discrete, as in the case of the mode of travel, systematic variations in aggregate demand occur at the extensive margin.$^4$ That is, individuals are changing from the consumption of one good to that of another. In the discrete commodity case, therefore, systematic variations in aggregate demand cannot be attributed to error

$^1$The task of generalising individual to population choice behaviour is undertaken in the analyses of Domenich and McFadden [9] and McFadden [18]. This section summarises and applies their research to the model described in the previous section.

$^4$In conventional demand theory analysis, the consumption good is divisible. Systematic variations in aggregate demand, therefore, occur at the intensive margin. That is, individuals are buying more or less of the good.
RESIDENTIAL LOCATION AND THE JOURNEY TO WORK

Patrick S. McCarthy

in measurement, but must be the result of varying tastes in the population. It follows that individual differences in tastes need to be incorporated into the model.

To incorporate variations in tastes into the model, the population must consist of individuals with common socioeconomic characteristics and with the same available alternatives, so that the ε vector is random. The consequence of this is that an individual’s utility function becomes stochastic. An individual, chosen at random from the population, will select mode A in his journey to work if

\[ U[Q_A, T_A, x_A, s, \varepsilon] > U[Q_B, T_B, x_B, s, \varepsilon] \]

\[ = > U_A > U_B. \]

Since an individual’s utility is now stochastic, however, this will occur with some probability which is given by: \(^6\)

\[ P_A = Pr(U_A > U_B). \]  

**Derivation of the probit model**

Equation (5) gives the probability that a randomly selected individual will choose mode A in his journey to work. Specific implications cannot be derived from the analysis, however, until the form of the utility function is specified. To this end, let an individual’s utility function be given by:

\[ U(\cdot) = V(\cdot) + \eta(\cdot) \]

where \( V \) is the deterministic part of an individual’s utility and reflects representative tastes in the population; \( \eta \), on the other hand, is the stochastic part of an individual’s utility and reflects variations of taste in the population. The probability of taking mode A in the journey to work can now be written as:

\[ P_A = Pr(V_A + \eta_A > V_B + \eta_B) \]

\[ = Pr(\eta_A - \eta_B < V_A - V_B) \]

\[ = G(V_A - V_B) \]  

where \( G \) is the cumulative distribution function of the random variable \( (\eta_B - \eta_A) \). Therefore, the choice probabilities will depend upon specification of the cumulative distribution function \( G \) and the nonstochastic component \( V \).

Let the nonstochastic component of utility be specified as

\[ V(\cdot) = v_1(\cdot)\beta_1 + \cdots + v_r(\cdot)\beta_r \]

where the \( v_m \)'s \( (m = 1, \ldots, r) \) are empirical functions which depend upon no unknown parameters. The \( \beta_m \)'s \( (m = 1, \ldots, r) \) are unknown parameters. In vector notation, equation 8 can be written as

\[ V = \nu^T \beta \]

and the probability of taking mode A in the journey to work is

---

\(^6\) Systematic variations could also be caused by errors in optimisation. This, however, would bring into question the value of the behavioural model, in which a large part of the variation in observed demands is due to ad hoc error specification.

\(^6\) It is assumed that the utility associated with one mode does not equal the utility associated with the alternative mode, so that the problem of ties is obviated. Also, to simplify the presentation, \( U[H(Q), T, x, s, \varepsilon] \) will henceforth be written as \( U(\cdot) \).
If $\eta_a$ and $\eta_a$ are normally distributed, it can be shown that $G$ is the cumulative normal distribution function, and the binary probit model obtains. In testing the model, the results of which are presented below, probit analysis is the statistical estimating procedure.

**DATA**

The nature of the model developed requires that the data fulfill two fundamental criteria: it must be obtained from an area in which more than one mode is available for the journey to work, and it must be constrained to include individuals who have made a simultaneous residential-modal decision. Observations, therefore, must be obtained from individuals who have recently moved.

Since this data is not currently available, it was necessary to write a survey and administer it to newly moved residents. In all, 2,078 randomly selected individuals residing in 36 San Francisco Bay Area communities were sampled and asked a variety of questions pertaining to their residential, work trip, and socioeconomic characteristics. Of the 2,078, there were 118 usable observations, which constituted a usable response rate of 5.68 per cent.

In the set of usable observations, various pieces of information were unknown to some respondents. Before the observation could be used in the analysis, it was therefore necessary to estimate this information. The missing data and the manner in which it was estimated are given below.

a. **Income.** In the survey, an individual was asked to specify into which category family income fell when he first moved to his present residence. Four individuals specified family income as "over $55,000". To obtain an estimate of income for these individuals, income was assumed to follow a Pareto distribution whose parameter was estimated to be 4.5. This implied as estimate of family income of $70,713.50.

b. **Age of home.** For those individuals who did not indicate the age of their home, an estimate was obtained from the census data. From his address, an individual's census tract could be determined from which the average age of home could be calculated.

c. **Distance to public services.** Individuals were asked to estimate the distance

---

1See Domencich and McFadden [9], pp. 53-56.

2The San Francisco Bay Area was chosen for the survey because it offers its residents feasible alternatives to the automobile for the journey to work.

3The overall response rate was 28.32 per cent. However, there were several reasons for excluding an observation from the usable sample. These included length of residence, change of jobs or no job at all, and that car-public transport was not the modal pair specified. Also, since the response rate was so low, a procedure described in Scott [29] was used to determine whether the usable response was representative. The data was found to be representative on all criteria but one, work trip modal choice, in which it was found that non-automobile users were over-represented in the sample.

4See J. S. Cramer [7], pp. 51-58.

5The 1970 Census of Population and Housing provided census tract information on the year when a structure was built.
RESIDENTIAL LOCATION AND THE JOURNEY TO WORK

from their home to the nearest elementary, junior, or high school, the nearest police station, the nearest fire station, the nearest library, and the nearest park. Several people left all or part of these blank. An estimate was obtained by locating the residences and various public service facilities on a map and calculating the distance between the nearest structures and an individual's home. 12

ESTIMATION RESULTS

The model developed will be estimated for the car-public transport modal pair where public transport includes bus, BART, and the Southern Pacific commuter train. 13 The dependent variable in the probit equations is the probability, \( P_A \), that an individual takes an automobile on his journey to work. The explanatory variables used in estimating the model are given below.

Transport variables

Relative travel cost.

In the survey, individuals were asked to estimate, for both their primary and alternative modes of travel, weekly out of pocket costs. From their estimates a relative cost variable is formulated as follows:

\[
COSRAT = \frac{\text{travel cost on public transit}}{\text{travel cost on automobile}}
\]

Since the dependent variable is the probability of taking an automobile on the journey to work, COSRAT is expected to have a positive probit coefficient.

Relative travel time.

In the survey, individuals were asked to estimate, for both their primary and alternative work trip modes, portal to portal travel time on a typical morning. From these responses the following relative travel time variable is defined:

\[
TIMRAT = \frac{\text{work trip travel time on public transit}}{\text{work trip travel time on automobile}}
\]

It is expected that the probit coefficient of TIMRAT will be positive.

Residential variables

Residential location in this study is assumed to affect mode of travel indirectly, since a change in location, \( ceteris paribus \), will alter costs both on the mode taken and on its alternative in the journey to work. In other words, although a change in some \( Q_i (i = 1, \ldots, I) \) will not directly affect relative automobile costs, it will have some impact upon equilibrium distance from the CBD, thereby altering costs on

12 Maps from the Automobile Club of California were used for this.

13 Six individuals specified BART as their current or alternative mode of travel when in fact BART was not yet operative in their locale. It is assumed, therefore, that these individuals moved into their present homes with full knowledge that BART would be operating in the area and with the anticipation of either taking BART or having it as an alternative.
the available modes. The change in relative automobile costs will, in turn, affect an individual's modal choice behaviour. Equation (11) summarises this process.

$$\frac{\partial P_A}{\partial Q_i} = \frac{\partial P_A}{\partial C^r} \frac{\partial C^r}{\partial k} \frac{\partial k}{\partial Q_i} \quad i = 1, \ldots, I$$  \hspace{1cm} (11)

where $C^r$ is relative automobile costs. By assumption, the first partial derivative on the right-hand side of equation 11 is negative, and the third partial derivative is determinate once it is known whether an increase in $Q_i$ ($i = 1, \ldots, I$) increases or decreases the consumption of housing. Furthermore, the partial derivative on the left-hand side of equation 11 is simply the coefficient of the $i$th housing attribute in the estimating equation. Given this information, the sign of $(\partial C^r / \partial k)$ can be determined.

In this section, the impact, a priori, of $Q_i$ ($i = 1, \ldots, I$) upon equilibrium distance from the CBD is given. This will be combined, in a later section, with the calculated sign of $(\partial P_A / \partial Q_i)$ ($i = 1, \ldots, I$) to induce a sign on $(\partial C^r / \partial k)$.

**Number of rooms (RMNUM)**

In the survey an individual was asked the number of rooms, exclusive of bathrooms, in his current place of residence. An increase in this variable is expected to increase the consumption of housing, which induces a move outward from the CBD.

**Age of home (HSAGE)**

From survey information obtained, it is possible to determine the age of a respondent's home at the time when he moved into it. A priori, an increase in house age decreases an individual's consumption of housing, which induces a move toward the CBD.

**Density (DENSE)**

An individual was asked to note the types of housing available in his present neighbourhood, and from the response a dummy variable, DENSE, is defined. Let

$$\text{DENSE} = \begin{cases} 1 & \text{if an individual indicated that his neighbourhood consisted primarily of apartments;} \\ 0 & \text{otherwise.} \end{cases}$$

Since an increase in density is assumed to decrease the consumption of housing, it induces an individual to decrease his equilibrium distance from the CBD.

**Tax rate (TAX)**

For each usable observation, the property tax rate applicable to an individual's residence when he first moved into his home was obtained.\textsuperscript{14} An increase in the property tax rate constitutes an increase in the cost of housing and, ceteris paribus, a decrease in its consumption, which induces a move toward the CBD.\textsuperscript{15}

\textsuperscript{14}The tax rate data was obtained from the County Tax Collector's Office in the counties of Alameda, Contra Costa, San Francisco, San Mateo, and Santa Clara.

\textsuperscript{15}Although many usable observations were from respondents who rented rather than purchased their homes, the impact of increased property tax rates will be the same. Since the owner of a rental unit will not want to bear any or all of the burden of an increase in property taxes, it is reasonable to assume that he will pass all or part of the increase on to his tenants in the form of higher rents.
RESIDENTIAL LOCATION AND THE JOURNEY TO WORK

Patrick S. McCarthy

Public services

As previously described, individuals were asked in the survey to estimate the distances from their homes to various public service facilities. The proximity of an individual's home to public service facilities can be viewed as affecting his consumption of housing either by varying the amount of the public service he may perceive to obtain or by increasing the perceived negative externalities in the form of increased volume of traffic congestion and nuisances. Although the expected impact of these distances upon equilibrium distance from the CBD is unknown, it is possible to obtain this information from the estimation results. Noting that the impact of $Q_p$ ($p$ signifies a public service variable) upon $P_A$ is given by

$$\frac{\partial P_A}{\partial Q_p} = \frac{\partial P_A}{\partial C'} \frac{\partial C'}{\partial k} \frac{\partial k}{\partial Q_p}$$

(12)

the sign of $(\partial k / \partial Q_p)$ will be determinate, given the sign of $(\partial C' / \partial k)$ implied from the estimation results for the non-public service residential variables. Knowledge of $(\partial k / \partial Q_p)$, furthermore, enables a conclusion to be drawn regarding the impact of $Q_p$ upon the consumption of housing. Since distance to a facility is an extremely rough index of the public service provided, it will be included in the estimating equations only if the fit warrants.

Socioeconomic variables

Family income (FAMINC)

Theoretically an increase in income affects both the consumption of housing and the marginal value of time, with the result that its overall impact upon equilibrium distance from the CBD is uncertain. The impact of income upon modal choice can be expressed by

$$\frac{\partial P_A}{\partial Y} = \frac{\partial P_A}{\partial C'} \frac{\partial C'}{\partial k} \frac{\partial k}{\partial Y}$$

(13)

Analogous to the above case, $(\partial P_A / \partial Y)$ is the coefficient of income in the estimating equation, and knowledge of its sign combined with the implied sign of $(\partial C'/\partial k)$ determines the sign of $(\partial k/\partial Y)$. This provides information regarding the relative strength of the impact of income upon the consumption of housing and upon the marginal value of time. For example, assume that the coefficients of $Q_i$ $(i = 1, \ldots, I)$ imply $(\partial C'/\partial k) > 0$. If, in addition, $(\partial P_A / \partial Y)$ is positive, then the conclusion drawn is that $(\partial k/\partial Y) < 0$. In other words, the impact of income upon the marginal value of time more than offsets its impact upon the consumption of housing.

Sex and age were tested but found to be insignificant in the present study.

Goodness of fit measures

Since probit analysis is employed to test the model, a multiple correlation coefficient is not available as a measure of goodness of fit. Such a measure can be ob-

---

16 Initially it was hoped to obtain better indices of public services provided in the various communities, but much of the data was not available and it was necessary, therefore, to utilise inferior indices in the estimating equations.
tained, however, from the likelihood ratio statistic. Let $L(\beta)$ be the log-likelihood function where $\beta$ is the parameter vector, and define $b$ to be the estimate of $\beta$ which maximises $L(\beta)$. Also, let $b'_0 = (b_0, 0, \ldots, 0)$ be the coefficient estimate which maximises the log-likelihood function under the assumption that the dependent variable is independent of the explanatory variables in the model. Then

$$-2[L(b_0) - L(b)]$$

has an appropriate chi-square distribution with $(k - 1)$ degrees of freedom.\textsuperscript{17} The larger this figure, the higher will be the confidence level with which the hypothesis of independence can be rejected.

An alternative goodness-of-fit measure is the success of a model in classifying observed behaviour.\textsuperscript{18} In the present case, let

\begin{align*}
p_{it} & \quad = \text{probability that individual } i (i = 1, \ldots, N) \text{ selects an automobile for his journey to work;} \\
p_{jt} & \quad = (1 - p_{it}) = \text{probability that individual } i (i = 1, \ldots, N) \text{ selects public transport for his journey to work;} \\
c_i & \quad = \text{cost of misclassification if an individual selects mode } i \text{ and mode } j \text{ is forecast } (i, j = 1, 2); \\
f_j & \quad = \text{binomially distributed random variable which equals 1 if individual } i (i = 1, \ldots, N) \text{ selects mode } j (j = 1, 2) \text{ and equals 0 otherwise;} \\
\delta_{ij} & \quad = \text{forecasting rule which equals 1 if mode } j (j = 1, 2) \text{ is forecast for individual } i (i = 1, \ldots, N) \text{ and 0 otherwise.}
\end{align*}

The total cost of misclassification is

$$C = \sum_{i=1}^{N} \sum_{j=1}^{2} p_{ij}c_j(1 - \delta_{ij}) \quad (14)$$

where $p_{ij}$ is the probability that individual $i (i = 1, \ldots, N)$ takes mode $j (j = 1, 2)$ on his journey to work. It can be shown that the expected cost of misclassification is minimised if that alternative is forecast which maximises $p_{ij}f_j$. If, for example, mode $j$ is forecast, then $\delta_{ij} = 1$ and $p_{ij}f_j \leq p_{ik}f_k \quad (k \neq j)$, which implies that $E(C)$ is minimised. Furthermore, if equal costs of misclassification are assumed, the forecasting rule classifies individual $i$ into that group which maximises $p_{ij} (j = 1, 2)$. This will be the procedure followed below, where $p_{ij}$ is obtained from the maximum likelihood estimates of the coefficient vector.

Once the forecasting rule is specified, it is possible to evaluate how well the model predicts observed behaviour.\textsuperscript{19} Let

\begin{align*}
N_j & \quad = \text{number of individuals who are known to belong to population } i \text{ but who were classified in population } j (i, j = 1, 2); \\
N & \quad = \text{total sample size.}
\end{align*}

Then the matrix below summarises the correct and incorrect classifications made by the model:

\begin{tabular} { | c | c | c | c | c | c |} \hline
\text{Forecasted} & \text{Correct} & \text{Incorrect} \\
\text{Population} & \text{in} & \text{in} & \text{in} & \text{in} & \text{in} \\
\text{Response} & \text{Response} & \text{Response} & \text{Response} & \text{Response} & \text{Response} \\
\text{1} & \text{1} & \text{1} & \text{1} & \text{1} & \text{1} \\
\text{2} & \text{2} & \text{2} & \text{2} & \text{2} & \text{2} \\
\text{3} & \text{3} & \text{3} & \text{3} & \text{3} & \text{3} \\
\text{4} & \text{4} & \text{4} & \text{4} & \text{4} & \text{4} \\
\hline
\end{tabular}

\textsuperscript{17}See Henri Theil [32], pp. 396-97.

\textsuperscript{18}See Domencich and McFadden [9], pp. 124-25.

\textsuperscript{19}Since the same data is being used to estimate the model as well as for classifying individuals, there is an upward bias in the predictive ability of the model. See S. James Press [26], pp. 382-83.
RESIDENTIAL LOCATION AND THE JOURNEY TO WORK

Patrick S. McCarthy

<table>
<thead>
<tr>
<th>Actual Choice</th>
<th>Predicted Choice</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto</td>
<td>$N_{11}$</td>
<td>$N_{11}$</td>
</tr>
<tr>
<td>P.T.</td>
<td>$N_{11}$</td>
<td>$N_{21}$</td>
</tr>
<tr>
<td>Overall %</td>
<td>$N_{11} + N_{21}$</td>
<td>$N$</td>
</tr>
<tr>
<td>correct</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although the ratio \( (N_{11} + N_{21})/N \) gives the correct proportion of individuals classified, it does not indicate how well the model performs. In developing a model to predict behaviour, the underlying assumption is that knowledge of the important explanatory variables will enable an investigator to classify correctly a higher proportion of individuals than if he relied on chance alone. In evaluating a model's performance, therefore, it is necessary to know the \textit{a priori} probability that an individual is classified into a particular group.

If a researcher's objective is to classify correctly members of both populations, the appropriate criterion is the probability of an individual's being classified correctly.\(^{20}\) Let

\[
C = \text{event that an individual is correctly classified;} \\
\hat{p} = \text{observed proportion of individuals taking an automobile on the journey to work;} \\
(1 - \hat{p}) = \text{observed proportion of individuals taking public transit on the journey to work;} \\
\alpha = \text{proportion of individuals classified by the model as automobile takers;} \\
(1 - \alpha) = \text{proportion of individuals classified by the model as public transport riders.}
\]

The probability of an individual's being classified correctly can be expressed as:

\[
P(C) = P(C \mid \text{classed as auto taker}) P(\text{classed as auto taker}) + P(C \mid \text{classed as P.T. taker}) P(\text{classed as P.T. taker})
\]

which implies that

\[
C_{\text{pro}} = \hat{p} \alpha + (1 - \hat{p}) (1 - \alpha)
\]

where \( C_{\text{pro}} \) is the proportionate chance criterion. \( C_{\text{pro}} \) will be reported in the estimation results presented below.

\textbf{Empirical results}

The model was estimated for the total set of usable observations, 118 in all, and yielded the best results when the respondent's sex and age were not included and, of the public service indices, only distance to the nearest park was included in the set of explanatory variables. Table 1 presents these results.\(^{21}\) Both transport variables,

\(^{20}\)See Donald G. Morrison [23], pp. 156-63.

\(^{21}\)In the estimation results reported, the appropriate data set was examined for the presence of multicollinearity. In all cases, there was little collinearity among the explanatory variables.
TABLE 1
Effect of Residential Location upon Modal Choice for the Entire Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>M.L.E.</th>
<th>t-statistics†</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTN</td>
<td>-2.4405</td>
<td>-1.70</td>
</tr>
<tr>
<td>FAMINC</td>
<td>0.000023</td>
<td>1.74</td>
</tr>
<tr>
<td>DENSE</td>
<td>0.8405</td>
<td>1.86*</td>
</tr>
<tr>
<td>HSAGE</td>
<td>-0.0169</td>
<td>-2.03*</td>
</tr>
<tr>
<td>RMNUM</td>
<td>-0.0923</td>
<td>-0.99</td>
</tr>
<tr>
<td>DISPRK</td>
<td>-0.0200</td>
<td>-1.55</td>
</tr>
<tr>
<td>TAX</td>
<td>0.0856</td>
<td>0.95</td>
</tr>
<tr>
<td>TIMRAT</td>
<td>0.8844</td>
<td>5.45***</td>
</tr>
<tr>
<td>COSRAT</td>
<td>0.8070</td>
<td>3.04***</td>
</tr>
<tr>
<td>$\chi^2(8) = 33.06^{***}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***significant at the 0.01 level
**significant at the 0.02 level
*significant at the 0.05 level
†On the basis of asymptotic theory, the t-statistics reported are asymptotic. For details, see Theil [32], pp. 584-96.

TIMRAT and COSRAT, are significant at the 0.01 level and have the signs postulated by the model. Of the two significant residential variables, the density index carries a positive sign but the house age variable does not. Given the assumed impacts of density and house age upon equilibrium distance from the CBD, the positive sign of DENSE implies that $(\partial C^*/\partial k) > 0$, whereas the negative sign on HSAGE implies that $(\partial C^*/\partial k) < 0$. Two possibilities could account for this result:
1. $(\partial C^*/\partial k)$ is positive and an increase in house age is perceived as increasing the consumption of housing;
2. $(\partial C^*/\partial k)$ is negative and an increase in density is perceived as increasing the consumption of housing.

Of the two possibilities, the first would appear to be the more likely, since it is not unreasonable that individuals may to some extent perceive a positive relation between age of home and quality of construction. Notwithstanding this, the results are inconclusive and the sign of $(\partial C^*/\partial k)$ cannot be determined without further analysis.

As an indicator of the model's performance, the chi-square statistic is significant at the 0.01 level and denotes a good fit. The model does not fare so well, however, when it is used to predict observed behaviour. It is seen in Table 2 that, although the model correctly classifies 71.67 per cent of the observations (and, by comparison with the proportionate chance criterion, this constitutes an improvement), its predictive ability is limited.

Although these results support the hypothesis that location has a significant impact upon mode of travel in the journey to work, the model performed rather
### Table 2

**Predictive Success Table for the Entire Sample**

<table>
<thead>
<tr>
<th></th>
<th>Actual Choice</th>
<th>Predicted Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Auto</td>
<td>P.T.</td>
</tr>
<tr>
<td>Auto</td>
<td>79</td>
<td>8</td>
</tr>
<tr>
<td>P.T.</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>Overall % correct</td>
<td>71.67</td>
<td></td>
</tr>
</tbody>
</table>

\[C_{pro} = 65.27\]

poorly. A possible explanation is that residents in the East and West Bays do not constitute, in terms of their travel choices, a homogeneous population. Bus and BART are the public transport modes available on the East Bay, whereas on the West Bay the public transit modes include bus and the Southern Pacific commuter train. Since the same public transport modes are not available to residents of both bays, it seems appropriate to estimate the model for populations with similar modal choices. Accordingly, the data is disaggregated and the model is separately estimated for residents of the East and West Bays.

When the model was estimated for East Bay residents, who constituted a sample size of 61, the best results were obtained by omitting the respondent's age and sex and including distances to the nearest fire station and park. Table 3 presents the results, which provide strong evidence that residential location is a significant determinant of modal choice for the work trip. The relative travel time and relative travel cost variables are significant and carry the hypothesised signs. Each of the residential variables, DENSE, RMNUM, and TAX, is significant and, given the expected impact of each upon equilibrium distance, it is implied that \(\frac{\partial C'}{\partial k} > 0\). The positive sign on the income variable, assuming \(\frac{\partial C'}{\partial k}\) is positive, implies that, for residents on the East Bay, an increase in income has a greater impact upon the marginal value of time than upon the consumption of housing. Finally, given the tenuous interpretation of distance to a public facility as an index of an individual's perceived attainment of the public service and assuming that \(\frac{\partial C'}{\partial k}\) > 0, the results in Table 3 suggest that East Bay residents perceive proximity to a fire station as decreasing their housing consumption and proximity to a park as an increase in it.

Once again the chi-square statistic is significant at the 0.01 level, indicating a lack of independence between the dependent and independent variables. Table 4 summarises the predictive ability of the model, and the present model correctly classifies a much greater proportion of observed behaviour than the model estimated for the entire sample. This tends to substantiate the belief that, because

---

22 The cities of San Francisco and Daly City are exceptions, as BART runs through San Francisco and has its terminal station in the northern part of Daly City.
### Table 3
Effect of Residential Location upon Modal Choice for East Bay Residents

<table>
<thead>
<tr>
<th>Variable</th>
<th>M.L.E.</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTN</td>
<td>-8.5861</td>
<td>-2.78***</td>
</tr>
<tr>
<td>FAMINC</td>
<td>0.00014</td>
<td>2.99***</td>
</tr>
<tr>
<td>DENSE</td>
<td>1.5750</td>
<td>1.77*</td>
</tr>
<tr>
<td>HSAGE</td>
<td>0.0371</td>
<td>1.45</td>
</tr>
<tr>
<td>RMNUM</td>
<td>-0.6274</td>
<td>-2.60***</td>
</tr>
<tr>
<td>DISFIR</td>
<td>0.0982</td>
<td>2.41**</td>
</tr>
<tr>
<td>DISPRK</td>
<td>-0.0964</td>
<td>-2.52**</td>
</tr>
<tr>
<td>TAX</td>
<td>0.4800</td>
<td>2.45***</td>
</tr>
<tr>
<td>TIMRAT</td>
<td>1.0986</td>
<td>2.35**</td>
</tr>
<tr>
<td>COSRAT</td>
<td>1.1312</td>
<td>2.00*</td>
</tr>
</tbody>
</table>

$\chi^2(9) = 29.50$

***significant at 0.01 level  
**significant at 0.02 level  
*significant at 0.05 level.

### Table 4
Predictive Success Table for the East Bay

<table>
<thead>
<tr>
<th></th>
<th>Actual Choice</th>
<th>Predicted Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Auto</td>
<td>P.T.</td>
</tr>
<tr>
<td>Auto</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>P.T.</td>
<td>5</td>
<td>11</td>
</tr>
</tbody>
</table>

Overall % correct $= 88.52$

$C_{pro} = 65.63$

of the different types of public transport modes available, the East and West Bay populations are not homogeneous.

When the model was estimated for West Bay residents, who constituted a sample size of 57, the best results were obtained when a respondent's sex and age were excluded and distance to the nearest fire station was the only public service index included. It is seen from Table 5 that relative travel time and relative travel cost are again significant and carry the hypothesised signs. The opposite signs on DENSE and HSAGE may again imply opposite signs for $(\partial C''/\partial k)$, but, given the results stated, a plausible explantion is that $(\partial C''/\partial k) > 0$ and individuals on the West Bay
perceive an older home as increasing their consumption of housing. Finally, the negative sign on DISFIR, assuming \( \frac{\partial C^*}{\partial k} \) is positive, suggests that West Bay residents view proximity to a fire station as increasing their consumption of housing.

In terms of goodness of fit, this model performs well. As in the previous cases, the chi-square statistic is significant at the 0.01 level; this rejects the hypothesis of independence between the dependent and explanatory variables. Table 6 summarises the model's ability to predict observed behaviour and indicates that disaggregating the total set of observations and estimating the model separately for each bay improves the model's performance in significance and goodness of fit.

**Table 6**

**Predictive Success Table for the West Bay**

<table>
<thead>
<tr>
<th>Actual Choice</th>
<th>Predicted Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Auto</td>
</tr>
<tr>
<td>Auto</td>
<td>39</td>
</tr>
<tr>
<td>P.T.</td>
<td>5</td>
</tr>
<tr>
<td>Overall % correct</td>
<td>82.78</td>
</tr>
</tbody>
</table>

\[ C_{pre} = 62.87 \]
SUMMARY AND CONCLUSIONS

The analysis in this study constituted a test of the hypothesis that residential location is a significant determinant of modal choice for the work trip. Implications from a standard residential land use model, generalised to incorporate transport attributes, were exploited to obtain the impact, *a priori*, exerted by residential location upon an individual's choice of mode in the journey to work. The data needed for testing the model was obtained from a mail survey which was administered to 2,078 new residents in 36 San Francisco Bay Area communities. Probit analysis was employed to estimate the model. From the estimation results yielded, the conclusion can be drawn that residential location is a significant factor in an individual's modal choice.

The primary policy implication of the analysis pertains to modal choice forecasting. Since the underlying population is that of new residents, a large advantage of the present model is its ability to forecast the impact which changes in the spatial structure of a community will have upon modal split. If, for example, transport planners anticipated urban sprawl in the foreseeable future, various residential variables (such as average number of rooms and density) would change. The impact of these changes upon modal split could be estimated from the model. Alternatively, if government agencies decided to alter the provision of public services or change the property tax rate, the transit authorities could estimate the influence of these changes upon modal split and, consequently, implement appropriate policies to meet the varying demands.

Notwithstanding the results obtained, this analysis can be viewed as a pilot study which presents preliminary results and suggests areas for further study. Four areas of immediate concern are briefly discussed.

1. Although the total sample size numbered 2,078 individuals, the set of usable responses was distressingly low. As a result, the modal pair considered for the analysis was car-public transport, where public transport included BART, bus, and the Southern Pacific commuter train. The paucity of usable data points obviated estimating the model separately for each public transport mode; since the public transport modes are dissimilar in their attributes, it can be suspected that residential location will not affect every mode of travel in the same manner. What is needed, therefore, is a larger survey of new residents to obtain a sufficient number of observations for the modal pairs car-BART, car-bus, and car-Southern Pacific commuter train, so that the impact of residential location upon each can be estimated.

2. An additional extension of the analysis is to obtain from each individual the necessary information on travel time and travel cost for every mode available to him. This would enable a researcher to use a multinomial logit program which calculates the probability of selecting each mode available to an individual.2) Modelling the problem in this manner is more realistic, since an individual is not constrained to an "either-or" choice.

3. Considerable research has been devoted to estimating a complete system of demand equations, including Barten [3], Parks [25], Theil [33], and Kraft and

---

2) See McFadden in Zarembka [40], pp. 113-19.
RESIDENTIAL LOCATION AND THE JOURNEY TO WORK

Patrick S. McCarthy

Kraft [15]. An extension of the present study would be to develop and estimate a system of demand equations for housing and mode of travel. One obstacle to this approach, however, is the definition of housing. Since housing constitutes a bundle of services rather than a quantifiable commodity, a suitable proxy must be found for housing consumption.

4. The model presented in this analysis considers only the demand side of modal choice in the journey to work. A natural extension of the study is to incorporate supply characteristics, both for the modes available and for other variables utilised in the analysis.

APPENDIX

The definitions of the variable names referred to in Tables 1 through 6 are:

CONSTN — constant term;
FAMINC — family income at the time a respondent moved into his current residence;
DENSE — dummy variable which equals one if an individual indicated that his neighbourhood consisted primarily of apartments and equals zero otherwise;
HSAGE — age of a respondent's home at the time he moved into the residence;
RMNUM — number of rooms, exclusive of bathrooms, in a respondent's current home;
DISPRK — distance from a respondent's home to the nearest park;
DISFIR — distance from a respondent's home to the nearest fire station;
TAX — property tax rate corresponding to an individual's residence and obtained for the year in which he moved into the residence;
TIMRAT — ratio of public transit to automobile travel time in the journey to work on a typical morning;
COSRAT — ratio of public transit to automobile weekly costs in the journey to work trip.

REFERENCES


183
May 1977

JOURNAL OF TRANSPORT ECONOMICS AND POLICY


184