RURAL/URBAN RESIDENCE LOCATION CHOICE

Tubagus Feridhanusetyawan and Maureen Kilkenny

Working Paper 96-WP 157
March 1996

Center for Agricultural and Rural Development
Iowa State University
578 Heady Hall
Ames, IA 50011

Tubagus Feridhanusetyawan is a research associate and Maureen Kilkenny is an adjunct assistant professor, Department of Economics, Iowa State University.

This research is supported by a grant from the USDA National Competitive Research Initiative Grants Program, 1994-1996.

The authors thank Mary Henry and the other participants at the 1995 North American Regional Science Association International meetings, November 1995, in Cincinnati, Ohio.

For more information about the content of this paper, please contact (e-mail) tubagus@iastate.edu or kilkenny@iastate.edu.

RURAL/URBAN RESIDENCE LOCATION CHOICE

Motivation

People concerned about *rural development* need to know which effect is stronger: do "jobs follow people" or do "people follow jobs?" (Muth, Carlino and Mills; Knapp and Graves). This dichotomy has frequently been the basis of the analysis of *metro*politan population and employment growth. And, the evidence supports each hypothesis under alternative circumstances. Employment is always important. But since urban locations all offer higher probabilities of employment than rural locations, rural development economists really need to know what else determines the choice of a rural residence.

The rural economic development programs of "smokestack chasing," federal funds acquisition, and export promotion (Isserman 1994) all operate on the assumption that people-follow-jobs. On the other hand, strategies based on attracting entrepreneurs, tourists, or retirees, rely on the jobs-follow-people assumption. Different strategies are appropriate for different regions, depending on the characteristics of the desired population and the characteristics of the region. In this paper we report how rural versus urban residential locations are chosen as determined by personal and local characteristics. In doing so we shed some light on two questions: (1) What types of people choose to live in rural areas as opposed to urban ones; and (2) What types of local characteristics increase the probability that people choose to reside in a rural area rather than an urban one?

While there are a number of studies of intermetropolitan migration and intrametropolitan location choice, there are only a few that consider rural locations at all (Beladi and Ingene; Clark and Hunter; Plane). Clark and Hunter estimate population-weighted net migration as functions of labor market characteristics, amenities, location indicators, and fiscal policies using a sample of both metro and nonmetro counties. The estimated coefficient on the dummy variable for proximity to the central city indicates only that rural locations are not preferred by any age group. Plane describes how the changing age distribution affects the regional population distribution. As the baby-boomers retire, we should expect a reallocation of retiree population as well as working-age population, due to service-sector interdependencies.

We attempt to identify the things local governments can do to attract the economically active population to reside in nonmetropolitan areas. We find that there are some public expenditures, such as on infrastructure and highways, which successfully attract population to rural areas. Other types of government spending work in the opposite direction. And the net locational effects of government spending by program type depend on adjacency or nonadjacency and local population densities. Not all strategies can be used equally well by all nonmetro communities.

We use the same basic approach as those who have studied intrametropolitan migration using multinomial logit statistical analyses (Fox, Herzog, and Schlottman; McFadden 1978; Greene). This is a discrete choice problem: people choose to live in one type of area as opposed to another for some reason. What are the most likely reasons? We assume that this choice is a constrained optimum: people may choose their residence, but they take into account their own abilities and opportunities to make a living, and there are often family considerations, other personal reasons, and local economic conditions determining why their potential residential locations are limited. Furthermore, we are interested in focusing on rural as distinct from urban locations, so we adopt a typology to group all possible locations into a few alternative categories.

We provide a theoretical overview of residential location choice that accommodates most hypotheses and provides a model that can be estimated. Then we review the multinomial logit econometric modeling technique, discuss data problems and normalization issues, and explain our approach to typology. We show the results of the statistical analyses and interpret them.

Theoretical Framework: Residential Location Choice

The literature provides a wide range of relevant hypotheses about residence location choice. In the extreme, these may be summarized into five models: residence location depends on (1) workplace location, (2) local amenities or "quality of life," (3) life-cycle and other personal characteristics, (4) return to human capital accumulation, and, (5) real costs of living. Combinations of these models determine most household location decisions, as is argued in the theoretical papers (e.g., Turnbull 1992) and demonstrated in the empirical ones (e.g., DeSalvo 1985).

If the "workplace" model is correct, residence locations are chosen to maximize expected household earnings net of commute costs. Expected wage levels and employment probability in the chosen location relative to other locations, and distance to work, should be significant determinants of the choice. This model has been widely used to explain residential location choice in monocentric urban areas and to explain migration between metropolitan areas (Lerman; DeSalvo; Greenwood and Hunt; Nakosteen and Zimmer). In the "human capital" model, the optimal location choice maximizes the expected return on the individual's investment in human capital and job search (Schwartz; Rogerson). If the "quality of life" model is correct, relative wages lose explanatory power since individuals accept lower incomes in order to obtain higher amenities (Dickie and Gerking; Knapp and Graves; Roback). If the "personal characteristics" model is correct, demographic information such as stage in the life-cycle, marital status, number of school-age children, map significantly to chosen places (Clark and Hunter; Heckman,; Plane). If the "cost of living" model is correct, travel time and housing costs are the most significant determinants (Turnbull, Glascock, and Sirmans).

In general, people choose a residence location (r) among all potential locations (r 0 ***R***, ***R*** = {1, ... ,R}). Assume that households choose their residence location to maximize their expected utility over market goods and services and leisure, conditioned upon local amenities; given the income they earn and the expenses they must incur (including travel time) while living at the chosen place, and working and shopping in possibly different places:

$$U(\mathbf{C},l;S^r) + \lambda(\mathbf{E}[\, w_x^r \,(H\text{-}l\text{-}t^w\text{-}\mathbf{TC})] - \mathbf{PC}) \qquad (1)$$

where:

U(·) represents utility from the package of goods, leisure and amenities at r,

C = (c_1,c_2, ..., c_z) is a vector of market goods and services including housing, distinguished by location

l denotes leisure time

S^r is a vector of residence area characteristics measured in units of increasing preference.

λ is the Lagrange multiplier (the value of relaxing the full income constraint)

E[·] is the expected income that can be earned in the residence location. This depends on the unemployment rate in the location of employment (u^r) and job market characteristics around residential location r (**J**^{r})

w**x** denotes returns to labor in the market for the household's skills, determined by the household's characteristics (**X**)

H ! l is total time, net of:

t^w, travel time to work

l, leisure time

TC (a vector product) is travel time to obtain consumer goods in the various market locations from residence location

P = (p_1,p_2, ..., p_z) prices of goods and services distinguished by place of purchases.

The optimal choices of housing and goods and leisure (C*,l*) are functions of the exogenous variables: local amenities (**S**^{r}), distances to work and shop, unemployment rates and job market characteristics (u^r, $\mathbf{J}^r$), prices at various locations (including rents), ($\mathbf{P}$), and the household's characteristics, $\mathbf{X}$. The value function (indirect utility function), is thus a function of these variables:

$$V = U(C^*,l^*) = v(\mathbf{X},\mathbf{S}^r,u^r,\mathbf{J}^r,\mathbf{P},t^w,\mathbf{T}). \tag{2}$$

The observed location is the one that provides the largest V, chosen with some error. This is the basis of the random utility model we use for this analysis.

We estimate a model that determines the probability that location r is chosen as a function of characteristics of the chooser ($\mathbf{X}$) and the characteristics of the choice $\mathbf{L} = (\mathbf{S},u,\mathbf{J},\mathbf{P})$. We do not treat travel distances explicitly (because the data, as we show, do not allow it) but implicitly by using a location typology based on proximities to metropolitan areas.

The Econometric Model

While theoretical models of residential location decisions have been developed extensively, econometric studies are fewer for several reasons. First, micro data on persons and/or households, including personal attributes, residential locations, work locations, and/or commute times, are not publicly available. In particular, the exact residential locations of individuals are suppressed to maintain confidentiality. This makes it prohibitively difficult to associate local, personal, and workplace characteristics with an explicit consideration of travel time. Finally, the multinomial discrete choice model that is appropriate for the problem requires nonlinear optimization methods that, until recently, were cumbersome.

Developing the data and estimating the diverse choice model, however, is worth the effort. The multinomial discrete choice approach provides more than descriptive analyses or discriminant analyses. With the former, we could simply describe the characteristics of people living in rural or urban locations. With the latter, we could estimate a discriminant function to indicate which factors were significant in classifying people into groups such as those who choose rural residences and those who don't. A multinomial discrete choice model estimates not only the significant factors classifying people into groups, but it also estimates the probability of being in a group, as well as the contribution of each explanatory factor to that probability. Furthermore, we can compute the rate at which increases in the explanatory variables affect the probability of choice (the "marginal effects"), which summarize the effects of personal and local characteristics on the probability of a particular residential location being chosen. (Maddala 1983; Greene 1990).

To do this, we estimate the following model. At any point in time, an individual faces R choices of r places to live that can be classified into J categories . The indirect utility of the i^{th} individual living in the r^{th} place of residence corresponding to j^{th} (V_{irj}) is represented by the random indirect utility model:

$$V_{irj} = \boldsymbol{\beta}'\mathbf{Z}_{irj} + \varepsilon_{irj}; \;\; i = 1,2, ..., N; r = 1,2, ..., R; \;\; j =1,..., J; \tag{3}$$

where

$\mathbf{Z}_{irj}$: individual and local attributes,
$\boldsymbol{\beta}$: vector of parameters, and
ε_{irj} : random disturbance.

The alternative locations are aggregated into j = 1, ..., J categories distinguished using rural/urban continuum codes (Butler, 1990). This categorization reduces the number of alternatives from 3,141 (the number of counties in the United States in this studies) to at most four.

The i^{th} individual chooses to live in place of type (j) that has local characteristics (r) if the expected V_{irj} is the highest. Thus, (j) is chosen to:

$$\text{Max } E[V_{ir1}, V_{ir2}, ..., V_{irJ}] \tag{4}$$

The probability that place type j is chosen depends on the probability that it provides the highest indirect utility:

$$\text{Prob } [V_{irj} > V_{irk}] \text{ œk...j} \tag{5}$$

Let Y_i denote a random variable indicating the choice made by individual i. In other words, Y_i is the outcome of the optimization problem, Equation (2). According to McFadden (1973), if the ε_{irj}'s are identically distributed Weibull among J alternatives, the probability that the choice j is made is:

$$\text{Prob } [Y_i = j] = e^{\boldsymbol{\beta}'\mathbf{Z}} / \sum_{j=1}^{J} e^{\boldsymbol{\beta}'\mathbf{Z}}, \tag{6}$$

where, as discussed above, $\mathbf{Z}_{irj}$ is a vector of attributes specific to the choosers and to the location chosen. In effect, the attributes of the chosen residential location reflect the preferences of the chooser. Thus, we express $Z_{irj} = [\mathbf{X}_i, \mathbf{L}_{irj}]$, where $\mathbf{X}_i$ is the vector of the characteristics of individual i, which remain the same regardless of their choice of residence location[1]. $\mathbf{L}_{irj}$ is the vector of attributes of the chosen location of type j, with local characteristics that matter in the individual's preferences. Because every individual does not face the same local characteristics r among J alternatives, we do *not* estimate a multinomial *conditional* logit, which relies on assumptions that observed local attributes are invariant to the chooser.

The estimated equations provide the set of probabilities for J location choices. To avoid indeterminacy, β_1 is normalized to zero. This normalization, however, renders the estimated β_j parameters un-interpretable. The interpretations are based on the computed "marginal effects" (**δj**) relative to sample averages, as explained by Greene (1991). The marginal effects in the model are partial derivatives of the probability with respect to the arguments:

$$
\begin{aligned}
\delta_j &\equiv \partial \mathrm{Prob}[Y=j] / \partial Z; \quad j = 1,2, ..., J \\
&\equiv \partial Pj / \partial Z \\
&= P_j \left[\beta j - \sum_{J=1}^{J} P_r \beta_r\right] \qquad (7)
\end{aligned}
$$

While the parameter vector β_1 is normalized to zero, the vector of marginal effects δj is constrained to sum to zero. This normalization means that δ_1 is calculated residually as the difference between the sum over $\delta_{j \ldots 1}$ and zero. Thus, the δ_j are interpretable as the *net push/pull effects* of an increase in the value of determinants **Z** on the decision to live in area j. With J choices of places to live, there are J marginal effects for every element of **Z** in the model. For example, while a good school system may in general increase the probability that a place will be chosen as a residence, it may be a relatively stronger determinant or more important to those who choose to live in a suburban area than for those who have

[1]To follow the standard notation in Multinomial Logit Model, Z_{irj} can be written as Z_{i*j}, where i* = ir. In this case, subscript i* represents personal characteristics of i^{th} individual and local characteristics of a chosen location r that matter to i^{th} individual's preference.

chosen to live in a metro area. In this case the δ_j for j = suburb on L_{irj} for the chosen place characteristic *education* would be positive while the one for metro would be negative.

As in most economic analyses, we are concerned mainly with the level of significance of a variable and the sign of the derivative of the dependent variable with respect to the driving one. The level of significance is estimated by the t-values for each βj, and these show which explanatory variables are important in the choice model. The derivative is given by the estimated marginal effects. And it is relevant to note that the signs of the estimated β_j and δ_j are not necessarily the same (except in the case where J = 2, or, binomial logit). The sign on the estimated δ_j's indicate the direction of the marginal effects of a change in explanatory variables on the probability of the place being chosen. These may be positive or negative, since they represent the net relative push/pull effects and must sum to zero.

The estimation procedure is standard maximum likelihood using LIMDEP version 6.0 (Greene). The likelihood function for the multinomial logit model is:

$$\text{Ln } L = \sum_{i=1}^{N} \sum_{J=1}^{J} d_{ij} \ln P_{ij} \tag{8}$$

where:

d_{ij} = 1 if Y_i = j, and 0 otherwise;

P_{ij} = Prob[Y = j].

Regional Typology, Data, and Normalization

We base the empirical work on individuals as the units of observation. We used a cross-section of data on individuals from the *Panel Study of Income Dynamics* (PSID) for 1987. Until 1988, the PSID data reported the place of residence of every individual in the sample up to a county level. The sampled population is male heads of households, aged 16 to 65, who lived in the United States in 1987. Individuals who were in school or lived in an institution such as armed forces, prison, and health care facilities were excluded from the sample. The final sample consists of 4,495 individuals.

The residential choice model assumes that each individual chooses among R distinct places or types of residence locations. Rather than consider as many alternative locations as there were individuals in the sample, we classified each county of residence into one of a few categories. Since there is more diversity among counties typically classified as "rural" than there is between "rural" and other types of counties, we experimented with different groupings.

The USDA's *Beale Code* is the primary county classification system we used.[2] It distinguishes ten types of counties by proximity to metropolitan areas and size (population density). It distinguishes four types of "rural" counties: adjacent and nonadjacent, with central places of populations of 20,000 or less, and without any central places. Based loosely on the Beale code, we investigated the characteristics of three different typologies. In all cases, we lump all counties containing metropolitan central places into a single "metro" category. The typologies we use differ in the distinctions among nonmetro counties.

The simplest is a metropolitan/nonmetropolitan dichotomy. The second distinguishes adjacent from nonadjacent nonmetro counties, for a three-way typology, also used by the Rural Policy Research Institute (RUPRI). The most complex distinguishes three types of nonmetro counties according to the proportion of the total county population that is "rural." "Rural" residents are those who live in communities with populations less than under 2,500 or, on farms. When the percentage of population that rural is used to distinguish counties, we obtain a progression from the more populous nonmetro counties where rural residences are in the minority, to counties where the rural residences are all there is. Table 1 presents the three typologies we used for this analysis.

The county data are drawn from the CD-ROM *Counties USA*-1994 (U.S. Department of Commerce, 1995). This data set contains information on all counties in the United States from the 1980s to the early 1990s. We merged the data on counties with the data on personal characteristics using the county where the individual lived in 1987. We also added some data on climatic characteristics at the state level, from the *Statistical Abstract of the U.S.A* (U.S. Department of Commerce 1989).

The local/county characteristics to be included in the residential choice model consist of amenities, government policies, employment, and local climatic conditions. Climate variables are included to

[2]The rural-urban continuum code was named after Calvin Beale, a USDA demographer who developed the classification system. An outline of the Beale Code used to aggregate our PSID Data is given in Appendix A of this paper. See Butler (1990) for a more detailed description of the original method.

represent amenities over which local authorities have no control, but which have already been shown to have significant explanatory power in residential location choice. The government policy variables are mostly expenditures, although taxes are also considered. Total and sectoral government expenditures, direct federal expenditures, and revenues transferred from state governments are also included in the model (see Table 2).

There are several difficulties in dealing with local characteristics (L_{irj}) in the multinomial logit model. First of all, the attributes of the place are *chosen* attributes, and are thus subject to nonrandom selection bias. Second, some local attributes are by definition directly related to the dependent variable since they classify areas as rural or urban. These problems suggest a potential for simultaneity, endogeneity, and misspecification if raw local characteristics of the actual observed choice are included in the multinomial logit model.

An alternative approach, multinomial *conditional* logit, would explicitly include local characteristics as specific to the locations by type. One problem with this specification is that there is significant heterogeneity in the local characteristics within types of locations. For a conditional logit model, place characteristics should not vary from observation to observation for places of the same type. We could use the averages of each characteristic for each type, but this would mask the potential role that the actual variation in characteristics within types plays.

Furthermore, while we are looking for the relative significance that each local characteristic plays in residential location choice, it is *not* the characteristic of the location, but the preference of the chooser for the characteristic, that we are trying to discern. By categorizing chosen residential locations into two to four types, we implicitly allow for unobservable fixed effects specific to the locations by type that distinguish choice behavior. This is another reason why we do not use a conditional logit approach.

A second problem is how to incorporate the range of alternative, not chosen, local characteristics into the model without posing an unlimited alternative set. The typical approach is to normalize the observed characteristics relative to averages. In this way, the chosen place's characteristics are measured in a way that shows how the place compares with other places that could have been chosen. The problem we saw with this approach is that the preponderance of observations are of metro residences. To compare local to the average of characteristics of residential choices would thus be a comparison of local to metro characteristics. With that normalization each location choice would effectively be modelled as the result of a simple comparison between the chosen location and a *metro* location. It would not reflect that the choice is made according to a comparison of the chosen characteristic level relative to levels in *other nonmetro* as well as metro locations.

The third problem is to control for the fact that some local characteristics are defining characteristics for areas by type. For example, given the higher population density of cities, the raw data will suggest that persons who choose metro residences also prefer densely populated locations.

Our strategy to deal with these three problems is this. We (i) measure characteristics in the direction of increasing utility, (ii) normalize local relative to maximum levels of each characteristic, and (iii) identify the extent to which the observed, normalized measure deviates from the expected, by type. These residuals are used as the explanatory location variables in the multinomial logit model we estimate.

Our normalization rule is suggested by the theoretical model developed in the second section of this paper. By proper choice of units, indirect utility is increasing in each attribute (Roback):

$$\partial V_{irj}/\partial L_{irj} \geq 0 \tag{9}$$

This expression makes explicit our fundamental null hypothesis that the local attribute may have not relevance in the choice at all. By choice of units, the alternative hypothesis is that the larger the preference for the characteristic, the higher the utility, the more likely the place is chosen.

The steps in our normalization procedure are detailed as follows. Local characteristics of each county are each expressed relative to the maximum value among all the counties in the United States. This converts the measure of each local characteristic to a number ranging from zero to one. Letting L_r^* denote the normalized local attribute:

$$L_r^* = L_r/\mathrm{Max}(L_r) \tag{10}$$

where

$r\ \varepsilon$ All 3,141 counties in the United States.

Then, we regress normalized local attributes on metro/nonmetro, rural/urban, and state specific characteristics to control for potential typology bias. The regression model for local characteristics is:

$$L_r^* = \alpha_1\ \mathrm{METRO}_r + \alpha_2\ \mathrm{RURAL}_r + \alpha_3\ \mathrm{PLAND}_s + \sum_{s=1}^{50} \delta_s\ \mathrm{STATE}_s + \varepsilon_r \tag{11}$$

where

METRO_r : dummy variable, equals to 1 if the county r is a metropolitan area and equal to 0 otherwise;

$RURAL_r$: percentage of rural population in county r;

$PLAND_r$: average price of land in state s in 1987;

$STATE_s$: state dummy variable for state s;

α, δ: parameters to be estimated; and

ε_r : random disturbance.

The objective of this step is to obtain the residuals, which are thus cleaned of any simultaneity bias due to ruralurban, metro/nonmetro, or other state-specific patterns. This predicted residual is the variable that actually enters the equation for residential choice.

Notice that this is the only role for land rents in our model. We maintain the hypothesis that nonmetro land rents reflect the opportunity cost of land in *extensive* uses rather than the capitalization of population-attracting amenities (Henderson 1982).

We use the predicted residual $\Delta L_r = L_r^* - PL_r^*$, where PL_r^* is the predicted relative local characteristic level estimated using Equation (11). By construction, the predicted residual is not correlated with the rural/urban and metro/nonmetro typology. We interpret this version of the local characteristic to represent the #unanticipated& part of the local characteristic. A positive (negative) ΔL_r means that the actual L_r^* is higher (lower) than the expected L_r^* *for that location type.*

Personal Characteristics

The probability of a place being chosen is modeled as depending on personal as well as local area characteristics (see Table 3). Among personal characteristics, life cycle and family characteristics usually have good explanatory power. For example, younger people are expected to prefer cities while older people might favor a rural environment. Human capital is also expected to be strongly related to the choice of location. Educated people are expected to choose residences in or near places with proportionately more jobs requiring higher education. In general, high-skill jobs are more prevalent in cities. Urban areas are also often preferred by educated people because there is greater employment variety.

People who are married probably display higher tendencies to live in nonmetro areas for several reasons. First, married people often have families, and the lower land rents in non-metro areas mean more affordable housing for their bigger families. Second, amenities such as lower crime rates and lower population densities describe preferred environments in which to raise children. Thus, the choice to live in non-metro areas is expected to be positively related to the individual characteristics of being married and having children.

Some people are apparently more geographically mobile than others. And mobile people tend to choose cities. There are several indicators of mobility. One is experience of living in more than one state prior to 1987, and the second is presence of a physical disability. A third is habit, instrumented by the father's education level. Since those with higher educations are more likely to choose cities, a father's educational attainment is expected to have a positive effect on the probability of chosing a metro residence as well.

Personal occupational characteristics are also expected to be significant in determining the chosen residence location. It will not surprise anyone to find that being a farmer increases the probability of choosing a rural residence. Being retired is also expected to increase the probability of choosing a nonmetro area, especially since amenities are unlikely to be capitalized into nonmetro rents (Knapp and Graves). On the other hand, being a professional, sales, or service worker is expected to have a positive effect on choosing a metro residence. Finally, we expect unemployed people to prefer metro residences due to the wider job opportunities near there.

Estimation and Results

The full residential location multinomial logit model we estimate explains the probability of the choice as dependent upon personal and personally preferred local characteristics. The complete model is:

$$(\beta'Z) = \beta_0 + \beta_1 AGE_i + \beta_2 EDU_i + \beta_3 MARRIED_i + \beta_4 NCHILD_i + \beta_5 STLIVED_i + \beta_6 DISABL_i + \beta_7 FEDU_i + \beta_8 PUNEMP_i + \beta_9 RETIREDi + \beta_{10} PROF_i + \beta_{11} SELFEMP_i + \beta_{12} FARMER_i + \beta_{13} SALES_i + \beta_{14} SERVICE_i + \beta_{15} DENSITY_{ir} + \beta_{16} CRIME_{ir} + \beta_{17} EMPGR_{ir} + \beta_{18} UNEMP_{ir} + \beta_{19} GEXP_{ir} + \beta_{20} GEDU_{ir} + \beta_{21} GHLTH_{ir} + \beta_{22} GHWAY_{ir} + \beta_{23} GFIRE_{ir} + \beta_{24} GWELF_{ir} + \beta_{25} FED_{ir} + \beta_{26} STATE_{ir} + \beta_{27} DEBT_{ir} + \beta_{28} TAX_{ir} + \beta_{29} PHYS_{ir} + \beta_{30} SUN_{ir} + \beta_{31} JAN_{ir} + \beta_{32} JULY_{ir} \quad (12)$$

We approached the estimation problem in two steps. First we investigated the explanatory power of personal characteristics. Then we estimated the fully-specified model. This allowed us to observe the additional explanatory power of local characteristics. We discuss the fairly descriptive information relating personal characteristics to residential location first.

The estimated equations representing residential choice decisions as a function of personal characteristics alone are presented in Table 4. By construction, the parameters in the equation for a metro choice are normalized to zero, so they do not appear in the table. In general, these are high quality parameter estimates. The null hypothesis that all parameters equal zero is rejected at 1 percent level. The chi-square statistics are well above 300.

We estimated the three models corresponding to the three versions of the rural typology explained in the previous section. Notice that although all three models generally provide correct predictions 75 percent of the time, Model 1, the binomial choice model, has a highly significant intercept (or fixed effect), while Model 2 doesn't, and the intercept is significant in Model 3 only for suburban nonmetro county residence locations.

Otherwise, the differences between the models are most apparent in the coefficients about individual characteristics related to labor-force participation. In particular, the adjacent/nonadjacent distinction (Model 2) improves the significance of MARRIED in explaining location choice. The finding is that being married is significantly negatively correlated with choosing a nonadjacent nonmetro location. This contrasts with the simple hypothesis that married couples with families choose non-metro residences to reduce housing costs since the adjacent areas chosen are typically higher rent areas. It is consistent, however, with a hypothesis that married couples prefer adjacent nonmetro locations since both may not be employed in the same area and one may need to commute to work. The point is that a range of employment opportunities are important for married households in their residential location choice. And, these should be within the commute range, not necessarily inside the county of residence.

Second, the estimates confirm our prior hypothesis that unemployed individuals are more likely to remain in adjacent nonmetro areas, since there's more opportunity to find work in the adjacent metro community. An unemployed person in nonadjacent, nonmetro area may not "hang out" there. Both of these aspects are significant when Models 1 or 2 are used, but they are not significant using the typologies in Model 3.

All three models provide significant evidence confirming our expectations that the higher the education of the head of household or of his/her father (EDU,FEDU), the less likely a nonmetro residence is chosen. And, no surprise, if someone is a farmer, then it is likely that the residence is nonmetro.

Model 3, which distinguishes nonmetro counties on the basis of population densities, is the appropriate frame for recognizing a significant correlation between being retired and the probability of choosing a nonmetro residence. Adjacency doesn't matter in this choice, evidenced by insignificance of this element in Models 1 and 2.

Table 5 presents the marginal effects of personal characteristics on the probability that a nonmetro location is a chosen residential location. In every significant case, the signs of the marginal effects are the same as on the estimated coefficients. The estimated marginal effects thus provide additional evidence supporting most of the hypotheses. For example, consider the effect of personal unemployment on the choice between adjacent and nonadjacent nonmetro residences. The beta is significant and positive for the nonmetro adjacent location. And the marginal effect is positive at .083, indicating that being unemployed significantly increases the probability of having an adjacent nonmetro residential location by 8.3 percent, while it decreases the probability of a nonadjacent residence by 0.2 percent (insignificant).

Since Model 3 performs better in showing the effect of AGE on the probability of choosing the nonmetro residence, we look more carefully at the marginal effects of AGE in Model 3. The net effects of AGE is a pull towards more urbanized nonmetro areas, and away from the very rural areas. The signs on the estimated marginal effects are the same as those on the betas. It is surprising that our findings are significant since the sample is from economically active individuals who are between 16 and 65 years of age. Having truncated the distribution of age in the sample suggests that the ability to estimate its effect on residential choice is reduced.

Local Characteristics

The fitted equations for the complete residential choice model, which include both personal and local characteristics, are presented in Table 6. The coefficients for metro group in each model are not presented because they are constrained to be zero. The fitted equations show good quality parameter estimates reflected by large t-values. The null hypothesis that all the coefficients are zero is rejected at the 1 percent level

By comparing the full model with the model fit to personal characteristics alone (Table 6 and Table 4), we find that personal characteristics alone provide correct predictions for 75 percent of the sample. Local characteristics alone provide correct predictions for 77 percent of the sample. The addition of local characteristics to the model did not alter the signs nor the significance of the coefficients on personal characteristics. In fact, the addition of local characteristics increased the significance of personal characteristics in residential location choice.

The coefficients on local characteristics are also strongly significant. This is important evidence that local characteristics do matter in residential location choices, not just a person's job.

The marginal effects of the explanatory variables on residential choice associated with the results in Table 6 are presented in Table 7. Notice that the signs of the marginal effects are consistent with those on the coefficients. We focus on significant ones.

Population density is a significant determinant of residential location. The higher the *unanticipated* population density is relative to the maximum population density across all sampled counties, the less likely a nonmetro location is chosen. We interpret this to suggest that those who choose to live in nonmetro areas actually do prefer low population densities. In contrast, the effect of crime rates on residential choice is not significant.

Labor market characteristics also play important roles in rural/urban residential choice. The coefficients on EMPGR and UNEMP are significantly different from zero at the 5 percent level. A higher than expected relative unemployment rate works against attracting people to a nonmetro residential location. High relative employment growth rates appear to pull most effectively only to metro areas. This may suggest that attempts to attract population by using economic development strategies to raise expectations of employment growth are less effective in nonmetro areas than metro ones. It may also suggest that in nonmetro areas an unexpectedly high relative rate of employment growth is observed precisely where people have not been choosing to live in the past, but may choose to live in the future. This hypothesis must be further investigated using data on migration with known origin and destination information about the sampled individuals.

We also find significant evidence that government expenditures at the local level play an important role in the probability that the place will be chosen as a residence. Consider total government expenditure (GEXP). It is interesting that the signs of the coefficients vary across our regional typologies. According to the binomial Model 1, higher-than-expected relative total government expenditure makes nonmetro areas more attractive places to live. The marginal effect is about 0.3 percent, and positive. This means government expenditures as attractive factors are worth more to nonmetro areas than to metro ones. According to Model 2, we find that the effects, however, do depend on proximity to metro areas. There is a net pull toward adjacent areas, away from nonadjacent, nonmetro areas. Finally, Model 3 shows that total government expenditures provide the strongest net pull effects for 50 to 75 percent rural, nonmetro areas, regardless of whether or not the area is adjacent.

In particular, government expenditure on education (GEDU) is significant in all models. The net pull, however, is toward metro areas. In other words, metro areas that display higher than predicted relative spending on education pull residents from other types of locations. Thus the marginal effects of education spending in *nonmetro* locations are consistently negative. This result echoes by Fox, Herzog, and Schlottman (1989), who found that higher government expenditures on education significantly reduced the probability of moving out of a metro location.

The effects of distance are highlighted by comparing effects across typologies. Consider the estimated effects of government expenditure on health and hospitals (GHLTH) on residential location. First notice that it is not significant when nonmetro areas are all lumped together, as in Model 1. This is explained by the highly significant, but offsetting, effects in adjacent relative to nonadjacent nonmetro areas. Higher-than-expected relative government per capita expenditures on health make the nonmetro/nonadjacent areas more attractive, while adjacent nonmetro areas are less attractive. An explanation is that since adjacency means metro health care services are more accessible, people who live in adjacent nonmetro areas care less about the local government expenditure on health. This interpretation is further supported by the estimated marginal effects in Model 3. The positive effect of government expenditure on health service is significant only for the very rural locations. The implication is that government expenditure on health service is important in making nonadjacent and very rural areas attractive places to live.

Likewise, government expenditure on highways provides a significant net pull to the most rural areas, away from the adjacent and less rural nonmetro areas, as well as away from metro areas. The marginal effect of this expenditure on the probability of living in rural areas (Model 3) is one of the largest and it is positive. Again, the implication is that highway expenditure is an important factor in making the nonadjacent and very rural locations attractive for residents.

Government expenditure on welfare programs is significantly negatively related to the attractiveness of a nonmetro residential location. The net-pull effect of higher-than-predicted relative welfare payments is toward the metro area and away from the nonmetro ones. All models concur. By the same token, direct federal government expenditures (FED), which are mostly social security transfer payments (to elderly) and farm program payments, are negatively related to the probability that people choose the location. The implications are that localities that are unusually attractive to welfare and social security payment recipients are not attractive to the economically active segment of the population. No surprise! And imagine what this means when spending on welfare programs comes under local control. Communities interested in attracting the economically active segments would not favor generous welfare spending.

An important and significant contributor to the probability that a nonmetro residential location is chosen is the level of general government revenues from the state (STATE). This characteristic is significant in all models and signs of all marginal effects are consistently positive. In many cases, the majority of state funds are used to improve infrastructure and other services. In other cases, it is for education, when local taxes are insufficient. Most of these types of state funds will be anticipated and thus are controlled for by our local characteristics normalization procedure. Thus, it is very interesting that our multinomial logit model estimates suggest that unanticipated, relatively high state-to-local revenues are positively correlated with residential location choice. Higher than expected state revenue transfers are particularly effective at making the most rural nonmetro areas attractive places to live.

The effects of total outstanding debt are significant only when adjacency or non-adjacency is the fixed effect (Model 2). We found that higher than predicted debt is positively related to the choice of adjacent nonmetro locations while it is significantly negatively related to the choice of nonadjacent ones. The fact is that both growing and declining localities may be in debt, with opposite implications about attractiveness as residences. Nonadjacent areas in debt are more likely the declining ones, and thus the negative marginal is a reasonable reflection of that unobserved fixed effect distinguishing adjacent from nonadjacent nonmetro areas. The lack of significance in Model 3 on DEBT further supports this interpretation.

Higher than predicted tax rates would hardly be considered attractive, and most of the estimated marginal effects reflect that. Nevertheless, the marginal effect of TAX on the probability that a mildly rural location is chosen is significantly positive. This means that unexpectedly higher relative tax rates provide a net pull to the least rural of nonmetro locations; in other words, to the suburbs. This important finding suggests that there are nonmetro residents willing to pay the taxes needed to support a high quality suburban lifestyle.

By the same logic, we interpret the significantly positive effect of statewide sunny days (SUN) as meaning that if someone can afford to be a resident of a nonadjacent, nonmetro area, sunny days matter. This finding is also likely associated with the population growth of the Sun Belt in the late 1980s. The rural areas of the southwest and the mountain states (Texas, Arizona, New Mexico, Colorado, Montana, and Wyoming) are also characterized by higher than predicted numbers of sunny days relative to the sunniest places. Furthermore, nonmetro residential locations appear significantly more desirable than metro ones is the weather is hot. Better to swelter in the countryside or by a suburban swimming pool than in a city.

Finally, we also have evidence of the "icefishing" effect. The estimated marginal effects of January temperatures are significantly negative. Thus, the relatively warmer January is, the less likely a nonmetro residential location is chosen. Note, however, that these weather effects enter the model without controlling for predictable January temperatures by type of location. Thus, in general, nonmetro residents appear to prefer to experience temperature extremes from a nonmetro residential base. This can mean that people who choose nonmetro residences really do prefer extensive outdoor activities like skiing, snowmobile riding, and icefishing that people just don't get to do frequently if they live in a metro area.

Summary and Conclusions

In order to identify which rural development strategy is more appropriate for a place, we have taken a rural/urban perspective on the second part of the old conundrum: 'do people follow jobs or do jobs follow people?' We have estimated a multinomial logit model explaining the probabilities that people choose a residential location according to their own and the location's characteristics.

This is the first empirical discrete choice study of rural/urban residential location choice of which we are aware. Empirical investigation is critical since, as Turnbull (1992) has shown, in the model of household behavior with the time allocation problem explicit and where both wages and rents may vary across space, there are few unambiguous theoretical implications relating residential location to income or to the distance from employment location.

Our discrete choice approach, pioneered by McFadden, is not the typical approach applied to issues of regional population growth. One precedent is the work by Fox, Herzog, and Schlottman (1989). They estimated the probability that a person has moved between county groups and show how personal and local characteristics explain whether the movement was into or out of a metropolitan area. Our findings concerning *nonmetro areas*, especially about the roles of fiscal variables, are consistent with theirs. Most other studies use least-squares models to explain changes in net in-migration or the change in population density. Our findings are also consistent with those interpreted as supporting the "jobs follow people" hypotheses. We find (as did Carlino and Mills) that amenities (such as roads) do matter.

We can predict with 77 percent accuracy the residential location choice based on local characteristics (amenities) alone. In particular, we find that the low population density of a nonmetro residential location contributes significantly and positively to the probability that a nonmetro residence location is chosen. Thus, low population density is an attractive amenity for nonmetro residents. (Carlino and Mills found that nonadjacent locations attracted employment while repelling population.) In contrast, crime rates have no significant explanatory power in our model of rural/urban residential location choice.

Very interesting is the *lack of evidence* that people follow general job growth to nonmetro areas. Our estimates show the higher the employment growth rate, the stronger is the net pull towards metro away from nonmetro residential locations. The coefficient is consistently significantly negative across all model specifications. We are not surprised to find major differences between what explains intermetropolitan versus urban/rural patterns. For example, Greenwood and Hunt found that for *metropolitan migration*, employment opportunities are far more important than location-specific amenities.

Furthermore, unexpectedly high unemployment rates don't reduce the probability that a nonmetro residence location is chosen. On the contrary, it is significantly and positively affecting the probability that a nonmetro residence is chosen.

We also find significant evidence that people are attracted to rural residences by other things governments can do. The higher government spending is in a nonmetro area, the higher the probability it is chosen. This is particularly significant for adjacent nonmetro areas, but it is insignificant for most rural areas. The government instrument that has significant net positive pull effects into nonadjacent, nonmetro locations is the level of state government transfer to the local government (typically related to roads, utilities, and schools). In general, state subsidies for local infrastructure are positively correlated with increased probability that the nonmetro location is chosen. Government spending on highways also contributes significantly and positively to the probability that the most rural areas, and nonadjacent areas in general, are chosen for residence.

Government expenditures on health care are insignificant in the choice of an adjacent nonmetro location, but they are significant and positive in determining the choice to live in nonadjacent or in the most rural locations. Also, the higher the local dependence on welfare and social security transfers, the lower the probability that the place is chosen as a residence. The implications are that localities that are unusually attractive to welfare and social security payment recipients are not attractive to the economically active segment of the population. (Get ready for the "race to the bottom" on welfare spending.)

Finally, it appears that relatively high taxes do not necessarily reduce the probability that a nonmetro residence is chosen. The empirical work provides evidence that suburban-type residential locations are attractive even with relatively higher taxes per person than expected among counties of the same type.

We can also predict with 75 percent accuracy the residential location of a person according to their personal characteristics alone. As expected: educated, mobile, professionals with fewer kids are more likely to reside in a metro area than a nonmetro one. If the person is a farmer, retired, has lots of kids, or is unemployed, they are more likely to have a nonmetro residence. If the male head-of-household is married, the more likely the household has a metro or at least an adjacent nonmetro residence. This suggests that opportunities for both spouses to work within a commute range is a significant constraint on where people choose to live. There are also significantly fewer self-employed people with adjacent nonmetro residences than with metro or nonadjacent, nonmetro ones.

In further research we would like to analyze the residential location choice by considering complete information about employment location and commuting choice. At present, these data are not available to the public except in the form that identifies locations as PUMA (Public Use Microsample Areas), which are aggregates of metro and nonmetro counties for confidentiality purposes. Less aggregated data would also be useful to focus more closely on the effects of local characteristics on residence location by particular types of persons, such as the "entrepreneurial" class. And we would like to do a better job of looking at retirees and the implications of family ties.

Furthermore, we are surprised that local characteristics at the county level of regional aggregation were significant at all, given the variations in fiscal patterns across communities within counties. We would like to estimate the model with local data at a much finer geographic level. We are also concerned about the possibility of Type I error given that the arguments in our estimated models are estimated as well.

Finally, some aspects of the broader hypothesis we are investigating—about the relative efficacy of smokestack chasing as opposed to investing in providing a better quality of life—should be investigated in an explicitly dynamic migration choice model.

Table 1. Rural/urban typologies and distribution in the sample

Classification	Frequency	(Percent)
Model 1: Metro/Non-Metro Dichotomy (J = 2)		
Define Y1 = 1 if a county is metro (Beale code = 1, 2, 3, 4)	3,395	(74.7)
= 2 otherwise.	1,136	(25.3)
Model 2: RUPRI classification (J = 3)		
Define Y2 = 1 if a county is metro (Beale code = 1, 2, 3, 4)	3,395	(74.7)
= 2 if a county is non-metro, adjacent (Beale Code = 5, 7, 9)	482	(10.7)
= 3 if a county is non-metro, non-adjacent (Beale Code = 6, 8, 10)	654	(14.5)
Model 3: Classification based on percentage of rural population (J = 4)		
Define Y4 = 1 if a county is metro (Beale code = 1, 2, 3, 4)	3,395	(74.7)
= 2 if a county is non-metro (Beale Code > 4) and rural population is less than 50 percent[a]	296	(6.6)
= 3 if a county is non-metro (Beale Code > 4) and rural population is equal or greater than 50 percent, but less than 75 percent.	510	(11.3)
= 4 if a county is non-metro (Beale Code > 4) and rural population is equal or greater than 75 percent.	330	(7.3)
Total	4,495	(100.0)

[a] The portion of the sample who live in non-metro areas with rural populations of less than 25 percent is very small (21, or 0.5 percent of total sample).

Table 2. County local characteristics, descriptions, sample means, and maximum values

Variable	Description	Sample Mean	Max. Value
DENSITY	Population/square mile	219	58,405
CRIME	Serious crime known to police/ 100,000 population in 1987	2,826	19,259
UNEMP	Unemployment rate, 1986 (percent)	8.69	37.9
EMPGR	Employment growth, 1986-87 (percent)	3.08	191.42
GEXP	General government expenditure/capita ($)	1,400.52	15,267.00
GEDU	Government expenditure on education/capita ($)	687.66	6,719.28
GHLTH	Government expenditure on health and hospitals/capita ($)	124.39	2,528.23
GHWAY	Government expenditure on highways/capita ($)	122.12	3,534.93
GFIRE	Government expenditure on fire protection/capita ($)	19.85	239.29
GWELF	Government expenditure on welfare/capita ($)	37.00	876.30
FED	Direct federal government expenditure/capita (thousand $)	253.38	29,952.53
STATE	General government revenue from state government/capita, 1987 ($)	527.85	5,908.13
DEBT	Total debt outstanding; $/capital (1987)	1,381	207,906
TAX	Estimated tax rate, calculated as total revenue from tax per capita divided by per capita income, 1987 (percent)	5.36	31.75
PHYS	Active nonfederal physician/ 100,000 population, 1986	88.60	1,938
SUN	Number of sunny days/year, state average, 1987	253	328
JAN	January temperature, state 30-year average (degrees Fahrenheit)	33	73
JULY	July temperature, state 30-year average (degrees Fahrenheit)	77	92

Note: **(N = 3,141 counties)**

Table 3. Personal characteristics, descriptions, sample means, and standard deviations

Variable	Description	Mean	Std.
AGE	Age of individual in years	37.160	11.964
EDU	Education of individual in years	12.232	2.785
MARRIED	Dummy variable equals to 1 if individual is married, and 0 otherwise	0.836	0.370
NCHILD	Number of children in the family	1.088	1.216
STLIVED	Number of states individual has lived before	1.365	1.938
DISABL	Dummy variable equals to 1 if individual experience physical disability that limits his mobility, and 0 otherwise	0.036	0.186
FEDU	Father's education in years	9.103	3.580
PUNEMP	Personal unemployment, dummy variable equals to 1 if individual is unemployed during the survey and 0 otherwise	0.040	0.195
RETIRED	Dummy variable equals to 1 if individual is retired, and 0 otherwise	0.066	0.249
PROF	Dummy variable equals to 1 if individual is professionals, technical, or kindred workers, and 0 otherwise	0.286	0.452
SELFEMP	Dummy variable equals to 1 if individual is self-employed, and 0 otherwise	0.126	0.332
FARMER	Dummy variable equals to 1 if individual is a farmer or working in a farm, and 0 otherwise	0.027	0.163
SALES	Dummy variable equals to 1 if individual is a sales worker, and 0 otherwise	0.012	0.107
SERVICE	Dummy variable equals to 1 if individual is service workers, and 0 otherwise	0.068	0.252

Note: (N = 4,495 individuals)

Table 4. Multinomial logit estimates of residential choice model, personal characteristics

	Model 1	Model 2		Model 3		
	Non-metro	Non-metro Adj.	Non-metro Non-adj.	Non-metro Rural<50	Non-metro 50<Rural<75	Non-metro 75<Rural
	Y1=2	Y2=2	Y2=3	Y3=2	Y3=3	Y3=4
Constant	0.703 *	-0.050	0.038	-2.004 *	0.081	0.608
	(3.016)	(-0.164)	(0.129)	(-5.093)	(0.259)	(1.616)
AGE	0.002	-0.001	0.004	0.015 *	0.003	-0.014*
	(0.631)	(-0.154)	(1.068)	(3.023)	(0.606)	(-2.662)
EDU	-0.098 *	-0.127 *	-0.073 *	-0.050 *	-0.113 *	-0.123*
	(-6.859)	(-6.984)	(-4.034)	(-2.107)	(-6.076)	(-5.523)
MARRIED	-0.136	0.128	-0.350 *	-0.030	-0.222	-0.115
	(-1.381)	(0.926)	(-2.869)	(-0.183)	(-1.668)	(-0.680)
NCHILD	0.123 *	0.118 *	0.128 *	0.159 *	0.093	0.131*
	(3.639)	(2.669)	(2.946)	(3.008)	(1.958)	(2.350)
STLIVED	-0.049 *	-0.034	-0.064 *	-0.102 *	-0.036	-0.019
	(-2.506)	(-1.338)	(-2.497)	(-2.954)	(-1.344)	(-0.604)
DISABL	0.171	-0.288	0.490	0.203	-0.168	0.706*
	(0.762)	(-0.857)	(1.850)	(0.573)	(-0.504)	(2.002)
FEDU	-0.048 *	-0.023	-0.071 *	-0.015	-0.046 *	-0.090*
	(-4.322)	(-1.508)	(-5.006)	(-0.849)	(-2.954)	(-4.749)
PUNEMP	0.497 *	0.756 *	0.096	0.350	0.542	0.687
	(2.320)	(3.002)	(0.299)	(1.005)	(1.896)	(1.859)
RETIRED	0.088	0.093	0.083	-0.652 *	0.187	0.667*
	(0.559)	(0.447)	(0.406)	(-2.181)	(0.889)	(2.489)
PROF	-0.373 *	-0.282 *	-0.458 *	-0.210	-0.344 *	-0.732*
	(-3.619)	(-2.040)	(-3.376)	(-1.308)	(-2.358)	(-3.599)
SELFEMP	-0.072	-0.337 *	0.121	0.107	-0.222	-0.116
	(-0.686)	(-2.225)	(0.950)	(0.695)	(-1.462)	(-0.637)
FARM	1.712 *	1.173 *	1.951 *	1.956 *	1.463 *	1.711*
	(8.587)	(4.265)	(9.168)	(7.909)	(5.646)	(6.227)
SALES	-1.094 *	-2.089 *	-0.743	-1.971 *	-0.735	-0.965
	(-2.797)	(-2.188)	(-1.754)	(-2.061)	(-1.459)	(-1.331)
SERVICE	-0.243	-0.103	-0.394	-0.073	-0.191	-0.554
	(-1.527)	(-0.506)	(-1.809)	(-0.288)	(-0.876)	(-1.810)
Likelihood	-2,461	-3,261		-3,724		
Accuracy (%)	74.97	74.95		74.39		
Chi-Square	303.8 (14)	367.87 (28)		381.10 (42)		

Note: (t-values are in parentheses)

*****Denotes significance at the 5 percent level.

Table 5. Marginal effects of personal characteristics on the probability of Y = r

	Unit	Model 1	Model 2		Model 3		
		Non-metro	Non-metro		Non-metro		
			Adj.	Non-adj.	Rural<50	50<Rural<75	75<Rural
		Y1=2	Y2=2	Y2=3	Y3=2	Y3=3	Y3=4
AGE	year	0.000	-0.000	0.001	0.001	0.000	-0.001
EDU	year	-0.019	-0.013	-0.006	-0.002	-0.010	-0.007
MARRIED	(0, 1)	-0.026	0.021	-0.044	0.000	-0.022	-0.006
NCHILD	1, 2, ...	0.023	0.011	0.013	0.010	0.007	0.007
STLIVED	1, 2, ...	-0.009	-0.003	-0.007	-0.007	-0.003	-0.000
DISABL	(0, 1)	0.033	-0.041	0.064	0.012	-0.025	0.046
FEDU	year	-0.009	-0.001	-0.008	-0.000	-0.004	-0.005
PUNEMP	(0, 1)	0.095	0.083	-0.002	0.017	0.047	0.039
RETIRED	(0, 1)	0.017	0.009	0.008	-0.053	0.020	0.046
PROF	(0, 1)	-0.071	-0.023	-0.050	-0.008	-0.027	-0.044
SELFEMP	(0, 1)	-0.014	-0.040	0.021	0.011	-0.023	-0.006
FARM	(0, 1)	0.326	0.097	0.213	0.120	0.119	0.089
SALES	(0, 1)	-0.208	-0.221	-0.052	-0.132	-0.050	-0.046
SERVICE	(0, 1)	-0.046	-0.004	-0.046	-0.000	-0.014	-0.034

Note: Marginal effects relative to a metro location are not presented, but can be calculated residually given that $\Sigma_r \delta_r = 0$.

Table 6. Multinomial logit estimates of residential choice model, personal and local characteristics

	Model 1	Model 2		Model 3		
	Non-metro	Non-metro		Non-metro		
	Y1=2	Adj. Y2=2	Non-adj. Y2=3	Rural<50 Y3=2	50<Rural<75 Y3=3	75<Rural Y3=4
Personal Characteristics						
Constant	-6.283 *	-2.096	-13.466 *	-16.769 *	-5.465 *	-6.556 *
	(-6.93)	(-1.79)	(-11.45)	(-11.50)	(-4.11)	(-4.57)
AGE	0.002	-0.000	0.006	0.017 *	0.004	-0.012 *
	(0.63)	(-0.09)	(1.24)	(2.99)	(0.72)	(-2.02)
EDU	-0.091 *	-0.110 *	-0.076 *	-0.051	-0.110 *	-0.103 *
	(-5.16)	(-5.05)	(-3.56)	(-1.80)	(-4.91)	(-4.07)
MARRIED	-0.397 *	-0.094	-0.620 *	-0.338	-0.476 *	-0.348
	(-3.25)	(-0.59)	(-4.18)	(-1.75)	(-3.04)	(-1.83)
NCHILD	0.094 *	0.075	0.114 *	0.101	0.066	0.110
	(2.28)	(1.45)	(2.22)	(1.60)	(1.18)	(1.73)
STLIVED	-0.028	-0.003	-0.054	-0.092 *	-0.007	0.002
	(-1.21)	(-0.10)	(-1.82)	(-2.46)	(-0.23)	(0.06)
DISABL	0.675 *	0.307	0.851 *	0.703	0.288	1.071 *
	(2.44)	(0.79)	(2.62)	(1.67)	(0.73)	(2.72)
FEDU	-0.040 *	-0.019	-0.054 *	-0.008	-0.039 *	-0.079 *
	(-2.90)	(-1.10)	(-3.26)	(-0.37)	(-2.16)	(-3.72)
PUNEMP	0.573 *	1.044 *	-0.073	0.571	0.684 *	0.701
	(2.19)	(3.45)	(-0.20)	(1.42)	(2.05)	(1.69)
RETIRED	0.232	0.355	0.134	-0.523	0.379	0.698 *
	(1.23)	(1.49)	(0.55)	(-1.56)	(1.55)	(2.29)
PROF	-0.373 *	-0.304	-0.417 *	-0.282	-0.451 *	-0.663 *
	(-3.04)	(-1.92)	(-2.66)	(-1.54)	(-2.68)	(-2.93)
SELFEMP	0.099	-0.092	0.234	0.235	0.068	-0.128
	(0.78)	(-0.53)	(1.56)	(1.30)	(0.40)	(-0.61)
FARM	1.632 *	1.181 *	1.718 *	1.844 *	1.743 *	1.375 *
	(6.80)	(3.76)	(6.73)	(6.13)	(5.82)	(4.18)
SALES	-1.398 *	-2.344 *	-1.054 *	-2.456 *	-0.875	-1.360
	(-3.18)	(-2.40)	(-2.18)	(-2.49)	(-1.63)	(-1.78)
SERVICE	-0.085	0.041	-0.185	0.055	-0.017	-0.385
	(-0.46)	(0.18)	(-0.76)	(0.19)	(-0.07)	(-1.19)
Local Characteristics						
DENSITY	-0.175 *	-0.154 *	-0.197	-0.262 *	-0.223 *	-0.105 *
	(-8.15)	(-5.79)	(-5.91)	(-4.31)	(-6.85)	(-3.76)
CRIME	0.001	-0.012	0.016	0.022 *	-0.004	0.009
	(0.10)	(-1.38)	(1.92)	(2.04)	(-0.43)	(0.81)

Table 6. (Continued)

	Model 1	Model 2		Model 3		
	Non-metro	Non-metro		Non-metro		
		Adj.	Non-adj.	Rural<50	50<Rural<75	75<Rural
	Y1=2	Y2=2	Y2=3	Y3=2	Y3=3	Y3=4
EMPGR	-0.131 *	-0.152 *	-0.116 *	-0.113 *	-0.192 *	-0.094 *
	(-8.34)	(-7.61)	(-6.18)	(-4.95)	(-8.91)	(-4.07)
UNEMP	-0.107 *	-0.165 *	-0.071 *	-0.142 *	-0.087 *	-0.086 *
	(-13.33)	(-14.43)	(-7.69)	(-10.96)	(-7.92)	(-7.14)
GEXP	0.048	0.259 *	-0.167	-0.328 *	0.131	0.017
	(0.96)	(3.58)	(-1.94)	(-3.03)	(1.75)	(0.18)
GEDU	-0.251 *	-0.327 *	-0.195 *	-0.089	-0.261 *	-0.226 *
	(-6.37)	(-6.38)	(-3.55)	(-1.20)	(-4.58)	(-3.51)
GHLTH	0.011	-0.062 *	0.062 *	0.033	-0.013	0.052 *
	(1.03)	(-3.90)	(4.04)	(1.67)	(-0.80)	(2.94)
GHWAY	-0.045	-0.256 *	0.037	-0.179 *	-0.326 *	0.220 *
	(-1.33)	(-4.77)	(0.85)	(-2.67)	(-5.46)	(4.73)
GFIRE	0.029 *	0.009	0.037 *	0.012	0.047 *	-0.003
	(5.91)	(1.23)	(6.82)	(0.92)	(7.20)	(-0.27)
GWELF	-0.094 *	-0.061 *	-0.122 *	-0.043 *	-0.077 *	-0.108 *
	(-7.34)	(-3.72)	(-6.95)	(-1.98)	(-3.72)	(-4.85)
FED	-0.295 *	-0.288 *	-0.284 *	-0.470 *	-0.260 *	-0.218 *
	(-15.03)	(-10.38)	(-11.45)	(-9.82)	(-8.89)	(-8.52)
STATE	0.242 *	0.289 *	0.256 *	0.183 *	0.124 *	0.256 *
	(9.28)	(8.27)	(7.72)	(2.94)	(2.52)	(6.69)
DEBT	-0.013	0.158 *	-0.476 *	-0.091	0.067	0.039
	(-0.24)	(2.45)	(-3.79)	(-0.67)	(0.98)	(0.54)
TAX	-0.614 *	-1.009 *	-0.044	0.500 *	-0.719 *	-0.914 *
	(-4.04)	(-4.79)	(-0.23)	(1.98)	(-3.29)	(-4.01)
PHYS	-0.036 *	-0.067 *	-0.011	-0.057 *	-0.007	-0.070 *
	(-3.44)	(-4.39)	(-0.92)	(-3.20)	(-0.58)	(-3.61)
SUN	0.448	-4.804 *	4.907 *	6.352 *	-7.732 *	-0.174
	(0.65)	(-4.90)	(5.64)	(5.74)	(-6.42)	(-0.16)
JAN	-2.020 *	-1.174 *	-2.702 *	-4.058 *	-2.504 *	-0.705
	(-5.56)	(-2.30)	(-6.11)	(-7.36)	(-4.26)	(-1.20)
JULY	9.265 *	7.308 *	13.238 *	14.063 *	15.095 *	8.928 *
	(9.36)	(5.17)	(10.93)	(9.08)	(7.54)	(5.70)
Likelihood	-1,773	-2,405			-2,784	
Accuracy (%)	78.73	76.88			76.48	
Chi-Square	1,679.3 (32)	2,078.9 (64)			2,259.7 (96)	

Note: (t-values are in parentheses)
*Denotes significance at the 5 percent level.

Table 7. Marginal effects on the probability of Y = r

	Unit	**Model 1**	**Model 2**		**Model 3**		
			Nonmetro		Nonmetro		
		Nonmetro	Adj.	Nonadj.	Rural<50	50<Rural<75	75<Rural
		Y1=2	Y2=2	Y2=3	Y3=2	Y3=3	Y3=4
Personal Characteristics							
AGE	year	0.000	-0.000	0.000	0.000	0.000	-0.000
EDU	year	-0.005	-0.002	-0.002	-0.000	-0.002	-0.002
MARRIED	(0, 1)	-0.022	-0.001	-0.013	-0.001	-0.008	-0.006
NCHILD	1, 2,..	0.005	0.001	0.002	0.000	0.001	0.002
STLIVED	1, 2,..	-0.002	-0.000	-0.001	-0.000	-0.000	0.000
DISABL	(0, 1)	0.037	0.005	0.018	0.002	0.005	0.019
FEDU	year	-0.002	-0.000	-0.001	-0.000	-0.001	-0.001
PUNEMP	(0, 1)	0.031	0.019	-0.002	0.001	0.012	0.012
RETIRED	(0, 1)	0.013	0.006	0.003	-0.001	0.006	0.012
PROF	(0, 1)	-0.021	-0.005	-0.009	-0.001	-0.008	-0.011
SELFEMP	(0, 1)	0.005	-0.002	0.005	0.001	0.001	-0.002
FARM	(0, 1)	0.090	0.021	0.037	0.004	0.030	0.023
SALES	(0, 1)	-0.077	-0.042	-0.022	-0.006	-0.015	-0.023
SERVICE	(0, 1)	-0.005	0.001	-0.004	0.000	-0.000	-0.007
Local Characteristics[a]							
DENSITY		-0.010	-0.003	-0.004	-0.001	-0.004	-0.002
CRIME		0.000	-0.000	0.000	0.000	-0.000	0.000
EMPGR		-0.007	-0.003	-0.002	-0.000	-0.003	-0.002
UNEMP		-0.006	-0.003	-0.001	-0.000	-0.001	-0.001
GEXP		0.003	0.005	-0.004	-0.001	0.002	0.000
GEDU		-0.014	-0.006	-0.004	-0.000	-0.004	-0.004
GHLTH		0.001	-0.001	0.001	0.000	-0.000	0.001
GHWAY		-0.002	-0.005	0.001	-0.000	-0.006	0.004
GFIRE		0.002	0.000	0.001	0.000	0.001	-0.000
GWELF		-0.005	-0.001	-0.003	-0.000	-0.001	-0.002
FED		-0.016	-0.005	-0.006	-0.001	-0.004	-0.004
STATE		0.013	0.005	0.005	0.000	0.002	0.004
DEBT		-0.001	0.003	-0.010	-0.000	0.001	0.001
TAX		-0.034	-0.018	-0.001	0.001	-0.012	-0.016
PHYS		-0.002	-0.001	-0.000	-0.000	-0.000	-0.001
SUN[b]		0.025	-0.090	0.108	0.015	-0.135	-0.001
JAN[b]		-0.111	-0.020	-0.058	-0.009	-0.043	-0.011
JULY[b]		0.510	0.128	0.284	0.032	0.260	0.150

[a]The units of local characteristics in the model are deviation from the predicted-normalized values, multiplied by 100 (see Equation 10 and 11)

[b]SUN, JAN and JULY are measured only as relative to their maximum values, but not the deviation from predicted values multiplied by 100 (see Equation 10).

APPENDIX A. BEALE RURAL-URBAN CONTINUUM CODES

Code	Description
1	Central counties including metropolitan areas of 1 million population or more
2	Fringe counties of metropolitan areas of 1 million population or more
3	Counties in metropolitan areas of 250 thousand to 1 million population
4	Counties with metropolitan areas with less than 250 thousand population
5	Counties with urban populations of 20,000 or more, adjacent to a metropolitan area
6	Counties with urban populations of 20,000 or more, not adjacent to a metropolitan area
7	Counties with urban populations of less than 20,000, adjacent to a metropolitan area
8	Counties with urban populations less than 20,000, not adjacent to a metropolitan area
9	Completely rural counties, adjacent to a metropolitan area
10	Completely rural counties which are not adjacent to a metropolitan area

Source: Institute for Social Research (1989).

REFERENCES

Beladi, Hamid, and Charles Ingene. 1994. "A General Equilibrium Analysis of Rural-Urban Migration under Uncertainty." *Journal of Regional Science* 34:91-103.

Butler, Margaret A. 1990. *Rural-Urban Continuum Codes for Metro and Nonmetro Counties*. Staff report No. 9028, Agricultural and Rural Economy Division, Economic Research Service, United States Department of Agriculture.

Carlino, Gerald and Ed Mills. 1987. "The Determinants of County Growth." *Journal of Regional Science* vol. 27 pp. 39-54.

Clark, David E. and William J. Hunter. 1992. "The Impact of Economic Opportunity, Amenities and Fiscal Factors on Age-specific Migration Rates." *Journal of Regional Science*, 32:349-65.

DeSalvo, Joseph. 1977. "Theory of Locally Employed Household." *Journal of Regional Science* 17(3):345-55.

____________. 1985. "A Model of Urban Household Behavior with Leisure Choice." *Journal of Regional Science* 25(2):159-74.

Dickie, Mark, and Shelby Gerking. 1987. "Interregional Wage Differentials: An Equilibrium Perspective." *Journal of Regional Science*, 27:571-585.

Fox, William, H. Herzog, and A. Schlottman. 1989. "Metropolitan Fiscal Structure and Migratio." *Journal of Regional Science* 29(4):523-36.

Greene, William H. 1990. *Econometric Analysis.* New York: Macmillan Publishing Company.

Greene, William H. 1991. *LIMDEP Reference Guide Version 6.0*. Bellport, New York:: Econometric Software, Inc..

Greenwood, Michael and Gary Hunt. 1989. "Jobs versus Amenities in the Analysis of Metropolitan Migration." *Journal of Urban Economics* 25:1-16.

Heckman, John. 1980. "Income, Labor Supply, and Urban Residence." *American Economic Review* 70(4):805-11.

Henderson, J. Vernon. 1982. "Evaluating Consumer Amenities and Interregional Welfare Differences." *Journal of Urban Economics* 11(3):32-59.

Institute for Social Research. 1989. *A Panel Study of Income Dynamics: Procedures and Tape Codes 1987 Interviewing Year*. Ann Arbor: University of Michigan.

Isserman, Andrew. 1994. "State Economic Development Policy and Practice in the United States: A Survey Article." *International Regional Science Review* 16(1&2):49-100.

Knapp, Thomas and Philip Graves. 1989. "On the Role of Amenities in Models of Migration and Regional Development." *Journal of Regional Science* 29:71-87.

Lerman, Steven. 1976. "Location, Housing, Auto-Ownership, and Mode to Work: A Joint Choice Model." *Transportation Research Record* 610:6-11.

Maddala, G. S. 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." Chapter 4 in P. Zarembka, ed. *Frontiers in Econometrics*, New York: Academic Press. pp. 105-42.

____________. 1978. "Modelling the Choice of Residential Location." In *Spatial Interraction Theory and Planning Models*, edited by A. Karlquist, L. Lundquist, F. Snickars, and J. Weibull, North Holland; pp. 75-96.

____________. 1984. "Econometric Analysis of Qualitative Response Models." In Z. Griliches and M. Intriligator, eds. *Handbook of Econometrics*, (2). Amsterdam: North Holland.

Muth, Richard F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago: University of Chicago Press.

____________. 1971. "Migration: Chicken or Egg?" *Southern Economic Journal* 37:295-306.

Nakosteen, Robert and Michael Zimmer. 1982. "The Effects on Earnings of Interregional and Interindustry Migration," *Journal of Regional Science* 22(3):325-41.

Plane, David. 1992. "Age-Composition Change and the Geographical Dynamics of Interregional Migration in the U.S." *Annals of the Association of American Geographers* 82(1):64-85.

Roback, Jennifer. 1982. "Wages, Rents and the Quality of Life." *Journal of Political Economy*. 90:1257-78.

Rogerson, Peter. 1982. "Spatial Models of Search." *Geographical Analysis* 14(3):217-28.

Schmidt, P. and R. Strauss. 1975. "The Prediction of Occupation Using Multiple Logit Models," *International Economic Review* 16:471-86.

Schwartz, Aba. 1976. "Migration, Age, and Education." *Journal of Political Economy* 84(4):701-19.

Turnbull, Geoffrey K. 1992. "Location, Housing, and Leisure Demand under Local Employment." *Land Economics* 68(1)62-71.

Turnbull, Geoffrey K., John Glascock, and C.F. Sirmans. 1991. "Uncertain Income and Housing Price and Location Choice." *Journal of Regional Science* 31(4):417-33.