Socioeconomic Forecasting Model for the Tri-County Regional Planning Commission

Susan C. Hendricks, KJS Associates, Inc.; Patrick J. Costinett, Parsons Brinckerhoff, Inc.; Maren L. Outwater, Cambridge Systematics; Paul T. Hamilton, Tri-County Regional Planning Commission; and Paul Hershkowitz, Michigan Department of Transportation

Abstract

The relationship between land use and travel demand forecasting is very strong; at the same time, the methods to represent these relationships in travel demand forecasting models has been cumbersome, complex and difficult to calibrate. The socioeconomic forecasting model was developed for the Tri-County Regional Planning Commission to support the development of the Regional 2015 Transportation Plan. This model created a process to represent the land use and transportation relationship using analytical allocation procedures and incorporating feedback from local jurisdictions. The socioeconomic modeling approach involved developing forecasts at three levels of geographic detail (by county, by Minor Civil Division and by Traffic Analysis Zone) and an interactive projection, review and refinement process.

The process to develop a socioeconomic forecasting model involved the following components: (1) develop base year socioeconomic data, (2) estimate MCD-level forecasts, (3) disaggregate forecasts from MCDs to TAZs, and (4) estimate household size and vehicles per household. The base year population and household data was developed primarily from Census data and the employment data was developed from the Regional Economic Model, Inc. (REMI) and Michigan Employment Securities Commission (MESC) sources. The MCD-level forecasting model relied on trend analysis using historical relationships of population and households, combined with data from local jurisdictions on post-1990 development. The household models (for household size and vehicles per household) were developed using the Census Transportation Planning Package (CTPP) and the Public Use Microdata Sample (PUMS). The household size model estimated household size from population by age group and the vehicles per household model estimated vehicle per household from household income.

The allocation for these socioeconomic forecasts from MCDs to TAZs is perhaps the most unique aspect of the forecasting model. The allocation model is based on local input where this was available, and a combination for accessibility and potential development where local input was not available. Accessibility was calculated as a function of existing population and employment and travel time along the transportation system. Potential development was calculated from the amount of develop able land by zoning classification. The combination for accessibility and potential development was defined as an allocation factor that could be applied to each TAZ. Allocated land uses were carried forward to the next period and incorporated into forecast allocation factors.

The process to develop a socioeconomic forecasting model resulted in several lessons learned about what worked and what didn’t work. The allocation of socioeconomic forecasts from Regional controls to traffic analysis zones can incorporate both local knowledge, accessibility and develop able land. Ideally, local knowledge would account for most of the near-term forecasts and the analytical procedures would be used for longer-term forecasts. The reliability for the base year data is of paramount importance to the reliability of the forecasts and should be accorded adequate resources to improve the process. Specifically, the employment data was troublesome and may be improved by conducting an employment survey. Finally, the process could be improved by developing and maintaining a GIS monitoring program for land use.

Socioeconomic data is a critical input to transportation planning and travel demand forecasting. Accurate estimates of existing population, incomes, employment and other socioeconomic characteristics are necessary for meaningful calibration of a travel demand forecasting model. Technically sound projections of these same data are essential inputs to applications of the travel models to assess future transportation needs and deficiencies. The Lansing Area Travel Demand Model Calibration project, developed for the Tri-County Regional Planning Commission (TCRPC) in Lansing, Michigan, addressed the need for good socioeconomic data. This project developed procedures to forecast the small area distribution of economic and demographic variables required for the TCRPC travel demand model to support Long Range Transportation Plan development.

To balance land use and transportation needs, there is emphasis on managing demand and improving efficiency rather than increasing system supply; on promoting land use patterns which are more conducive to public transportation, and on encouraging more travel by non-motorized modes. The work described here incorporates the interrelation of land use and transportation system characteristics in a simple yet effective way that avoids the problems of more complex land use allocation models.

Alternative Approaches

Many approaches are used to forecast land use and socio-economic variables, and most methods use a "top down" process. Control totals and other exogenous inputs are established at an aggregate level (region, state, city, etc.), and the land use or socioeconomic model is used to allocate activities among smaller areas. Hence, land use forecasting models are often also referred to as activity allocation models.

The range of approaches can be broadly summarized into three categories:

- Models based on formal location theory,
- Analytical allocation procedures following no formal theory, and
- Judgmental or consensus-type procedures.

All three commonly used procedures can produce reasonable results. Location theory models explicitly include transport system costs as a fundamental factor in location choice, while other procedures consider transport factors in a more generalized or subjective manner. The current focus on transportation-system-sensitive land use forecasting, heavily influenced by ISTEA and CAAA, emphasizes location-based models, but this approach is considerably more complex than the other two and still requires many simplifying assumptions. Typically, such models focus on the potential effects of severe congestion on activity distribution but this is important only in the largest of urban areas. They do not capture important factors such as quality of schools, crime rates, and life-styles. The advent of two-worker households has greatly complicated the location decision process making it even more difficult to model. Further, formal models often are applied at an intermediate geographic level larger than traffic analysis zones (TAZs), requiring other disaggregation procedures to produce the final TAZ-level results.

There are many analytical allocation procedures for disaggregating land use activity measures or socioeconomic variables from one geographic level (e.g., minor civil division or MCD) to a more detailed level (e.g., TAZs). These procedures may employ a variety of factors, singly or in combination, such as: existing levels of activity, historical growth rates, knowledge of proposed developments, available developable land, zoning, and proximity to existing development. Often, sophisticated location theory models are used to produce results for subareas of a region and further disaggregation to TAZs is based on factors such as those above.

Largely intuitive or judgmental procedures are also popular. The most common of these is referred to as the Delphi method, which has seen widespread application. In this approach, a panel of relevant experts is used to allocate regional totals to smaller subareas. The process often involves analysis of future land use scenarios and related activity allocations as well as the disaggregation of base year data to TAZs. This procedure can be relatively informal or highly structured with specific forms, procedural steps, scoring schemes, and levels of interaction and reconsideration. Factors described above may also be used in the Delphi approach.

The Tri-County Approach

The relatively low pressures for development and low levels of congestion in the Tri-County region indicated that a sophisticated location theory approach was not appropriate for the Tri-County transportation project. The selected approach is, in effect, a combination of the analytical and judgmental procedures. It provides an automated process incorporating input from local jurisdictions and feedback from TCRPC staff and committees.

The TCRPC region expects only modest growth in population and employment over the next 20 years. The predicted 1990-2020 growth indicates less than a ten percent increase in population for the region and less than twenty percent increase in employment. These modest growth levels mean that current land use patterns will dominate the forecasts of travel demand. Thus, estimates of base year socioeconomic variables are relatively more importance compared to the forecast change in socioeconomic characteristics. Thus, the major focus of the effort was on preparing accurate 1990 base year data at the TAZ level. Base year estimates are critical to the calibration of the travel demand model and as a foundation for socioeconomic forecasts at TAZ level.

Base Year Socioeconomic Data

The base year socioeconomic data variables were selected based on the input requirements for the travel demand model. They were: total population, total households, average household size, average vehicles available per household, and retail and non-retail employment.

TCRPC’s travel demand forecasting model includes a process to cross-classify households by household size and number of vehicles available. This socioeconomic model was therefore developed to estimate relationships for estimating households and vehicle ownership, since the available county and MCD-level forecasts do not provide these data. Table 1 summarizes the data sources used in preparing the socioeconomic estimates presented herein. These included the 1990 Census of Population, the Census Transportation Planning Package (CTPP), the Michigan Employment Securities Commission (MESC) and the Public Use Microdata Sample (PUMS).

Estimates of 1990 population, households, mean household income, and number of vehicles were developed from the 1990 Census of Population and from the 1990 Census Transportation Planning Package (CTPP) using an aggregation of block level census data to Tri-County TAZs. Mean household income and vehicles available from CTPP data at CTPP TAZ level which in most cases correspond with the Tri-County TAZs. The reliability of the Census data and its compatibility with the Tri-County TAZ system produces highly reliable results.

Table 1: Socioeconomic data resources

Data source	Variables provided	Geographic level
Census of Population	Population By Age & Sex Group Quarters Population Households Housing Units by Type HH & Per Capita Income	STF1A: by block, STF3A: by block group
CTPP - Area of Workplace	Workers By Occupation, Industry, Class	655-TAZ System (preliminary version of current 704-TAZ System)
MESC	Employment by individual employer, including Address, Number of Employees, SIC Code	465-TAZ System, 704-TAZ System
PUMS	Population & Households cross-tabulated by one or more variables such as: Household Size, Household Income, #Workers, #Vehicles	Groups of jurisdictions with combined population of 100,000 or more persons
Individual Major/Special Employers	Employment by Type for General Motors Corporation, Michigan State University, local and state government, schools	704-TAZ system

Consolidated employment estimates were developed from MESC and CTPP estimates and incorporated additional refinements from local review. CTPP data, which is inclusive of all employees was used as control totals for most TAZs, while MESC is more accurate on a local level and was used wherever data is available.

Historically, MESC data has been the primary source of employment data by TAZ. Although it is a valuable source, the current files have significant problems with accurate address matching especially where the MESC report is filed at a centralized accounting location rather than actual work place. Also, MESC files only include “covered” employees which excludes most government workers and self-employed persons. It is, however, an important secondary source of employment estimates and the only source available for data on employment by employer (by SIC). A consolidated employment estimate was produced by using the MESC data as the basis for geographic distribution where CTPP was weak, but normalizing to CTPP values at the smallest geographic level for which they were available.

Socioeconomic Forecasting Process

Variety data resources were used in the development of TAZ-level socioeconomic projections:

- 1990 TAZ-level model input data
- 1990 Census data (PUMS and CTPP)
- County level REMI forecasts produced by the University of Michigan
- Population forecasts at MCD-level developed by TCRPC staff and historical data
- Historic employment data from MESC at MCD level for 1965, 1974, 1982, 1990
- The TCRPC inventory of prime industrial sites
- Area of developable land by TAZ based on physical and environmental constraints

The CTPP and PUMS data were used for the household classification models and related forecasting relationships. Data on household income and vehicles from the CTPP was used to model vehicles per household from forecasts of household income produced by the REMI model.

PUMS data provides the most detailed Census information on population, household, and labor force characteristics. It is a sample of the actual Census “long form” responses except residence and workplace locations are coded only to areas of 100,000 or more persons. These areas are termed PUMAs for Public Use Microdata Areas; the Tri-County Region contains three PUMAs. Thus, PUMS provides a valuable source for detailed cross-classifications such as population by age versus household size. In order to provide TAZ-level forecasts of household size tied to the REMI forecasts of population by age distributions, household cross classification relationships were developed from PUMS data to estimate 1990 household size by TAZ.

The University of Michigan produced a set of long-range forecasts of employment, income, and population for all eighty-three Michigan counties using the Regional Economic Model, Inc. (REMI) forecasting models. Population forecasts included a breakdown by age group and gender. Employment forecasts were provided by fourteen industrial divisions. These forecasts, generated for the 1995-2020 period by 5-year increment, are used as county-level control totals. The remaining sources are used to develop forecast data at the TAZ level.

The lack of reliable, consistent historical estimates of socioeconomic variables at TAZ level led to the adoption of a two-stage forecasting process. The first stage in the process is to forecast MCD-level population and employment based on a combination of historical MCD-level estimates and county-level population and employment forecasts. The second stage is to allocate the MCD-level forecasts to TAZ level. MCD-level population forecasts developed by TCRPC staff and MCD-level employment projections based on historical trends from MESC were used as controls for the TAZ-level allocations. This approach made the best use of available data and also provided forecasts that are more easily reviewed by local jurisdictions.

In general, historical trends in building permits and other socioeconomic characteristics provide a basis for developing relative growth rates throughout the region. Existing development patterns, plus recent trends in growth, can be among the strongest indicators of future growth patterns especially where growth rates are modest. While some developments are not well reflected by any of these factors, such developments will be hard to predict by any means unless they are already in the development pipeline. The socioeconomic estimates developed for 1990 provide a strong foundation for socioeconomic forecasts since overall levels of change are indeed modest.

Beyond existing development and current growth patterns, the next most important factor was the potential for new development. Two important indicators are the availability of developable vacant land and proximity, or accessibility, to existing or future activity centers. Usually, there is far more land available for new development than can actually be absorbed by the market within the forecast horizon. Knowing the amount of developable land in each TAZ provides at least a crude check on the reasonableness of growth allocations. For example, a population allocation to a TAZ that yields residential densities well above existing levels is clearly suspect.

Additional data from local jurisdictions was used to enhance the MCD-to-TAZ allocation stage of the forecasting process. This data consisted of:

- Development that has occurred since 1990, by TAZ

- Pipeline developments, by TAZ
- Local policies and programs related to development potential

Where available, input from local jurisdictions on post-1990 development provided a sound basis for TAZ-level allocation of the MCD-level projections of population and employment. With several notable exceptions, relatively little input was available beyond year 2000. However, the data from local jurisdictions yielded population and employment growth that were much higher than the independently-derived MCD-level projections. TCRPC staff reviewed and in several cases modified the MCD-level forecasts of population.

Overview

Socioeconomic forecasts are produced at three geographic levels: County, Minor Civil Division (MCD) and TAZ. This approach takes advantage of MCD forecasts already available, ensures greater statistical reliability, and provides forecasts that can be more readily evaluated by local jurisdictions. Figure 1 shows the major steps in the socioeconomic forecasting process. The boxes at the top indicate the primary inputs to the socioeconomic forecasting process.

The first major step indicated is to develop MCD level forecasts of population and employment by 5-year interval. These forecasts are based on trend relationships from historical data and are described below under the heading *MCD Forecasts*. The second step is to disaggregate MCD population and employment to TAZs by 5-year increments. This is the most complicated step in the process and is described below under the heading *MCD to TAZ Disaggregation*.

Step three is to develop county-level household size and income forecasts by 5-year interval to reflect changes in these characteristics over time. No local forecasts are available for these variables. Although the REMI forecasts do not include these variables, they do include data from which these variables can be derived. Relationships were developed from a combination of 1990 CTPP and PUMS data for this purpose. These relationships are represented in Figure 1 by the box labeled *Household Models* and are described below. The forecasts for these variables are independent of the MCD forecasts at this point in the process.

The fourth step indicated in Figure 1 is to forecast changes in household characteristics by 5-year interval. County-level trends in household size and income developed in step three are applied to corresponding TAZ-level variables to produce future estimates for each forecast year. The effect of income growth on vehicle ownership is also reflected in this step. A key assumption is that the household size, household income, and vehicles per household *averages* estimated for each TAZ in 1990 will not change in a *relative* sense. In other words, the value of each variable will change in response to county trends but the relationship of each TAZ to the county average and the pattern of variation across TAZs within each county will remain relatively constant.

There is no practical way to predict how household size, income and vehicles per household will change for each individual TAZ over time. But, there is now considerable variation in these variables within the region, and they have a significant influence on travel demand levels. In general, these patterns will change only slowly. For example, today's high income areas will tend to remain high income areas in the future. At the same time, it is important to reflect the aggregate change in these variables indicated by past trends and forecasts of the future. The approach retains observed variations by TAZ but increments these to reflect forecast changes at the county level. If

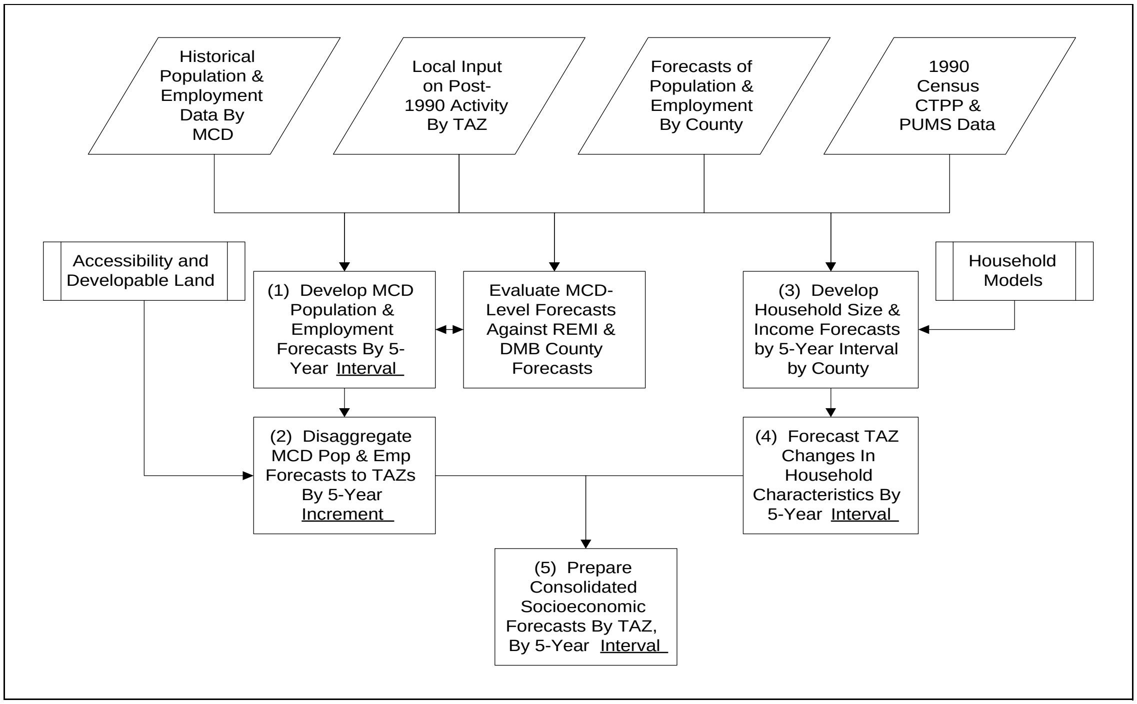

Figure 1: Overview of socioeconomic forecasting process

there is a reasonable basis to estimate changes from base year characteristics, as a result of major new development or redevelopment for example, then such changes could be incorporated in the last step of the process.

The final step is to consolidate socioeconomic forecasts by TAZ for 5-year intervals. The results from Steps (2) and (4) in Figure 1 are combined and any problems, exceptions, or inconsistencies reconciled. The number of households for each TAZ is estimated from the population and the household size forecast independently for each TAZ (Population divided by household size equals households).

The aggregation of TAZ-level households to county level will yield a different total than one produced by dividing county population by the average household size. This is due to the diversity of TAZ household size within each county and the uneven allocation of population among TAZs. For example, if population is allocated primarily to TAZs which have a higher than average household size, then summation of the TAZ household estimates will yield a lower number of households than would be estimated based on county values.

MCD Forecasts

MCD-level forecasts an important part of the overall process because they are a logical step between county-level forecasts and TAZ-level forecasts. The basic steps in the MCD-level forecasting process are:

- *Convert local input to population and employment equivalents*. Population estimates are based on local estimates of housing units multiplied by the TAZ household size. Employment estimates are based on average conversion factors between floor area or acres by type of development from the ITE Trip Generation Manual.

- *Develop trend relationships for MCD population and employment*. Nonlinear regression analysis was used to develop trend relationships using data for 1965, 1970, 1974, 1980, and 1990. The variations in employment data made it necessary to group MCDs into five groups with similar growth patterns.

- *Apply trend relationships to MCDs to obtain forecast population and employment by five-year interval*. Employment trends are expressed as the percent change from 1990 for each five-year interval for each of the five MCD groups. The same percent change was assumed for all MCDs in each group. These relationships derived from historical data for the five MCD groups were used as defaults. Local input were substituted for the default values where they were available.

- *Adjust MCD forecasts to county control totals as necessary*.

MCD-to-TAZ Disaggregation

The TAZ allocations are controlled to the MCD forecasts in a way that makes maximum use of the local input at TAZ level, and reflects the relative attractiveness and development capacity of competing TAZs in a rational way. The steps in the MCD-TAZ disaggregation process are:

- *Step 1: Develop accessibility factors*. Accessibility factors represent the tendency for new development to occur near existing activities and where the transportation system is (or will be) adequate. These factors are calculated as functions of existing population and employment, and measures of travel time. A zone-to-zone travel time matrix and the Gamma function parameters from the HBW trip distribution model were used to derive a friction factor matrix. The friction factors were multiplied by the employment of the destination zone and totaled by the origin zone to represent the origin zone's aggregate accessibility to employment. Likewise, the origin zone population * f-factor values were summed by the destination zone to represent the destination zone's aggregate accessibility to population. Finally, the resulting accessibility measures were divided by the maximum accessibility to yield accessibility factors that range from zero to one.

- *Step 2: Estimate potential development*. The maximum potential population and employment growth in each TAZ was calculated from available developable land by zoning category using average rates of development per acre. These rates, estimated from the ITE *Trip Generation Manual*, *5th Edition*, are 10.5 population/acre; 20 retail employees/acre and 37 non-retail employees/acre.

- *Step 3: Calculate allocation factors*. TAZ allocation factors are calculated by multiplying the accessibility factors and potential development factors to produce a measure of the probability that development will occur in a particular TAZ.

- *Step 4: Compare local input data to MCD forecasts*. The population and employment changes indicated by the local input do not necessarily agree with the incremental changes produced by the MCD forecasts, even though the MCD forecasts have been influenced by the local data as indicated earlier. A key feature of the MCD-to-TAZ disaggregation process is that it deals explicitly and logically with any differences between the two. If the MCD forecast shows a greater change than the local input for a given 5-year period, then the difference is allocated to

TAZs based on measures of development potential derived independently from the local input (Step 5a). If the MCD forecast changes is lower than local data, then the difference is carried over to the next 5-year period (Step 5b).

- *Step 5a: Use TAZ allocation factors to normalize MCD forecasts minus local input data*. The allocation factors are used to allocate any growth forecast by MCD that is not accounted for by local input.
- *Step 5b: Normalize local input data to MCD estimates for current forecast interval; carried over residual local input to subsequent period*. If the local input data for a 5-year interval exceeds the MCD forecasts for that 5-year forecast year, the local input data is still assumed to occur but may take longer than the 5-year interval and will be shifted to the next 5-year interval. In this manner, all local input data is incorporated into the forecasts by TAZ, but the forecast years may be extended if the local data exceeds the MCD forecasts.
- *Step 6: Reduce potential development for subsequent periods by the amount of growth allocated to the current period*. Allocation factors for subsequent years are adjusted to reflect areas where growth has been forecast to occur in earlier periods.
- *Step 7: Iterate for each 5-year forecast period*. These steps are repeated for each 5-year forecast interval until the disaggregation process has completed all forecast years.

Household Models

Forecasts of households, average household size, and average vehicles per household are not available even at county level. Therefore, basic models were developed estimate these variables.

The household models are used to forecast the *change* in corresponding characteristics by TAZ. The models are applied at the county level to estimate *synthetic* values for 1990 and for each forecast year; these values are then used to calculate a ratio or increment of change for each variable for each forecast year. These ratios are then applied to *actual* 1990 values for each TAZ to get the estimated TAZ values for each forecast year. This process retains the pattern of variation throughout the region but adjusts it to reflect trends indicated by county-level forecasts.

Average Household Size

Fortunately, household size is closely related to population by age and the REMI forecasts provide population by age group. A simple model was developed to relate household size to population by age group. As the distribution of population by age group shifts, corresponding changes are reflected in household size. The following steps describe the process:

- *Step 1*: PUMS data is used to cross classify household population by age group versus household size.
- *Step 2*: In order to calculate an average household size by age group, the household population is divided by the household size argument in the first column to get households. Total household population is divided by the total households in each age group to get the average household size.
- *Step 3*: The forecast population for each age group is then divided by the corresponding average household size to obtain estimated households.

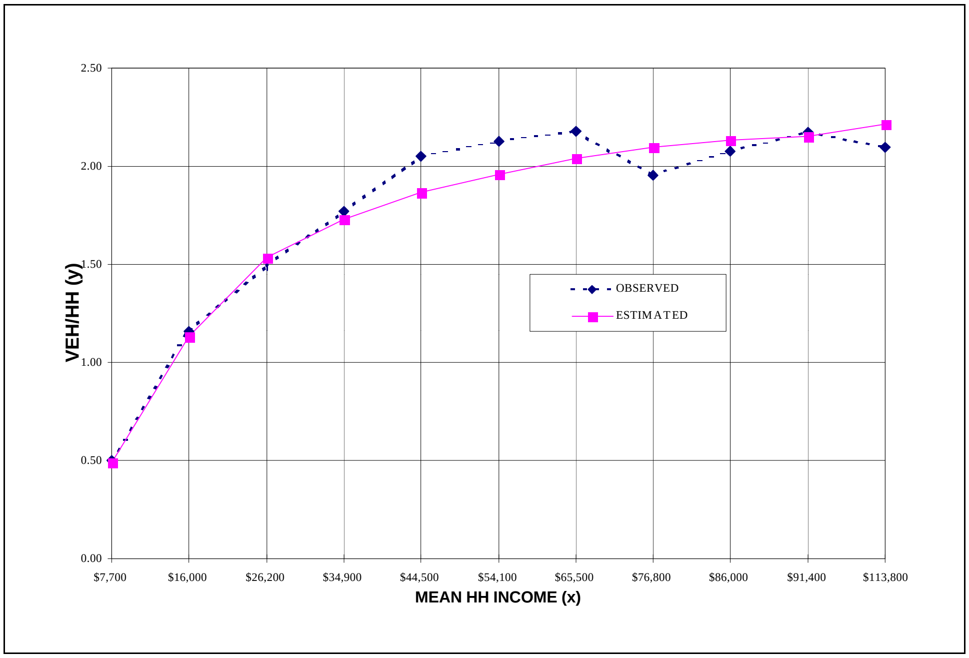

Figure 2: Vehicles per household as a function of household income

- *Step 4*: The households estimated in this way for each forecast year, divided by the corresponding estimate for 1990 yields a *ratio of change* for each county.
- *Step 5*: This *ratio of change* is applied to the actual 1990 average household size for each TAZ to yield the forecast household size.

Average Vehicles per Household

The forecasts of vehicles per household are derived from a relationship of vehicles available as a function of projected household income. The REMI forecasts include income but not vehicles; however, the two are very closely related as indicated in Figure 2 which shows the relationship based on 1990 data. However, that it is not a straight-line relationship. As average household income increases, the average vehicles per household also rises, but at a declining rate. The decline is related to the saturation level of vehicle ownership, about one vehicle per legal driver.

Forecasts of household income are derived from a combination of base year household income, forecasts of personal income from REMI, and forecasts of household size discussed above. Estimates of 1990 household income at both TAZ and county level are based on statewide CTPP data. Forecasts of household income at county level are based on the following relationship:

$$HHINC_y = (PER\_INC_y/PER\_INC90) * (HHSIZ_y/HHSIZ90) * HHINC90$$

where:

HHINC = Average household income for forecast year y / 1990
PER_INC = REMI average personal income for forecast year y / 1990
HHSIZ = Average household size for forecast year y / 1990

Ongoing Model Improvements

TCRPC and KJS are currently working to improve the reliability of the socioeconomic forecasts described in this paper. The current focus is on improving the base year estimates. Data on existing development is critical to the accuracy of the socioeconomic forecasts; and data problems (primarily in base year employment estimates) are the greatest source of inaccuracies by TAZ. The efforts underway to improve the base year data include:

- Improving the completeness and accuracy of the MESC data at the employer level, and
- Correction of geocoding errors affecting the TAZ assignment of specific employers.

The MESC data were compared with data from Dun and Bradstreet, an independent source of disaggregate employment estimates. Use of *Digital Yellow Pages* file helped to determine the existence of a business. Inconsistencies between these data sources are currently being reviewed to assess ways that missing and/or inaccurate employment information can be corrected.

TCRPC staff is also working on a comparison of TIGER and Caliper address files to improve the accuracy of the street base used for address matching purposes. The improved geocoding and address matching process will help ensure that the employers in the MESC data are allocated to the correct TAZs.