

# Improvement of Decennial Census Small-Area Employment Data: New Methods To Allocate Ungeocodable Workers

---

Edward Limoges, *Southeast Michigan Council of Governments*

The Southeast Michigan Council of Governments (SEMCOG), like any other agency involved in general and transportation planning, makes extensive, continuous use of decennial census data products. Standard products meet most of SEMCOG's data needs. When they do not, special tabulations have been purchased from the Census Bureau. In connection with SEMCOG's latest forecast, special tabulations of 1980 and 1990 census data were obtained, each including a tabulation of housing units and households by traffic analysis zone (TAZ) of residence and a tabulation of workers by TAZ of work. The worker data included a cross-tabulation by industrial class and occupational class. In response to puzzling industrial class mixes at the TAZ level found earlier in the Census Transportation Planning Package (CTPP), the special tabulations separated geocoded workers from allocated workers. (The Census Bureau uses worker allocation to assign tract and block numbers to workers whose workplace addresses cannot be geocoded to tract and block.) Subsequent mapping of geocoded workers by TAZ evidenced high quality, but allocated workers showed questionable distributions. SEMCOG decided to develop its own reallocation method that accepted Census Bureau county assignment, geocoded or allocated, and place geocoding. The method was applied to workers needing allocation, using as a base the workers geocoded to tract and block by detailed industrial class and occupational class. A comparison of TAZ employment by detailed industrial class demonstrates that workers reallocated by SEMCOG replicate the Census Bureau geocoded pattern much better than do workers allocated by the Census Bureau.

A new, comprehensive allocation method is proposed for the 2000 census. The new method would combine features of the current Census Bureau allocation and SEMCOG reallocation methods, and add the demarcation of commuting origin and destination areas.

The overall purpose of this paper is to discuss worker allocation, the method used by the Census Bureau to assign tract and block numbers to workers whose workplace addresses cannot be geocoded to tract and block, and to present the reallocation method developed by the Southeast Michigan Council of Governments (SEMCOG) to deal with locational inaccuracies found in the allocated census numbers.

## DECENNIAL CENSUS PRODUCTS AND SEMCOG'S DATA NEEDS

SEMCOG, like any other agency involved in general and transportation planning, makes extensive, continuous use of decennial census products. These include printed reports, summary tape files (STFs), other computer files, maps, and various other products. Decennial census data are indispensable. Without such data, many planning applications that are now commonplace would become impossible.

There is no other data source with the combination of data dimensions provided by the decennial census: a

wide variety of data items about population and housing, spatially organized in a system of geographic areas that range from the nation as a whole down to individual city blocks, occurring as a data time series represented by the name "decennial."

Standard decennial census products meet most of SEMCOG's needs for demographic data, and meet them very well. However, in preparing data inputs for a recently completed Regional Development Forecast (RDF), two areas were encountered in which the standard products were inadequate: data availability and data quality.

Regarding availability, certain data combinations were needed that had not been produced by small-area geography corresponding to traffic analysis zones (TAZs). For example, such data were needed as input to SEMCOG's demographic forecasting process in order to group the households in each TAZ into income quartiles, each quartile subdivided into portions with children and portions without children. In addition, for the input required by SEMCOG's method of assigning land use classes to decennial census small-area employment data, cross-tabulations of worker data and industrial class by occupational class and by TAZ of work were needed.

In the past, SEMCOG had overcome such problems of data availability by using one of the cost-reimbursable "special services" provided by the Census Bureau, that is, a special tabulation of data collected in the decennial census. Because of the interest in data from both the 1980 and 1990 censuses, SEMCOG needed two tabulations for each year, one of housing unit and household characteristics by place of residence, the other of worker characteristics by place of work.

Questions of data quality related to place-of-work data in the Census Transportation Planning Package (CTPP). SEMCOG's TAZs had been demarcated wherever possible to isolate individual major facilities and establishments, including large shopping centers, hospitals, colleges and universities, major government buildings, corporation headquarters, and large industrial plants. It was anticipated that most, if not all, of the CTPP's place-of-work employment counts for such TAZs would be either of a single industrial class (for example, transportation, communications, and utilities in the TAZ that contained only the telephone company's headquarters office building) or of a characteristic combination of related industrial classes (e.g., a shopping center TAZ that would have mostly retail trade employment, but also small numbers of workers in such industrial classes as services, and finance, insurance, and real estate).

However, many such TAZs were found in which, although the expected industrial class made up the majority of workers, there was a substantial minority of workers in a variety of other industrial classes. These unexpected workers were a puzzle that could not be

solved using only the CTPP data. In addition, it could not be determined to what extent the industrial class mix of workers given in the CTPP represented the actual situation or reflected geocoding errors or shortcomings in the allocation procedure for nongeocodable workers.

Through discussions with the Census Bureau staff, it was agreed that a special tabulation of workers' characteristics by place of work could be structured in a manner that would separate geocoded workers from allocated workers. Ultimately, agreement was reached that the Census Bureau would produce special tabulations of 1980 and 1990 census data, one by place of work and one by place of residence.

## REALLOCATION: METHOD AND RESULTS

### Census Worker Special Tabulation

SEMCOG's 1990 census worker special tabulation is like the CTPP in that it

- Was derived from the same "parent file" as the CTPP—the Sample Edited Detail File,
- Includes the key employment-related variables of the CTPP,
- Includes household characteristics, and
- Includes TAZ of work.

However, the special tabulation is different in that it

- Is more detailed (e.g., 75 industrial classes and 40 occupational classes compared with 18 and 14, respectively, in the CTPP), and
- Includes flags indicating whether or not the workers were geocoded to tract as well as to place.

### Necessity of Allocation

In 1990 the Census Bureau geocoded workers to tract and block provided that the workplace addresses reported were sufficiently complete and that either the Topologically Integrated Geographic Encoding and Referencing (TIGER) File could identify the address or the Workplace File included the specific worksite. In 1980 the procedure for the Urban Transportation Planning Package (UTPP) was similar, although some of the details were different. However, some tract and block geocoding errors still occur. Such errors in the UTPP and CTPP are difficult to recognize and correct because geocoded workers and allocated workers are mixed.

Because the special tabulation included a flag that indicated whether or not the workers had been geocoded

to tract (in most cases such workers had also been geocoded to block), SEMCOG could not produce computer-generated maps of tract-geocoded employment by any of the 75 industrial classes of the special tabulation. Upon examination, the overall quality of tract and block geocoding appeared to be quite good. There were relatively few recognizable major errors, and these were corrected, resulting in an improved base for subsequent reallocation.

In the Census Bureau procedure for creating small-area employment data, workers who could not be geocoded to tract and block were instead assigned to tract and block by an allocation procedure (1, 2). In 1990 a considerable percentage of workers needed to be allocated. In the four most urban counties of Southeast Michigan, an average of over 30 percent of all workers needed to be allocated to tract and block. In Detroit City, nearly two-thirds of the tracts had more than 40 percent of their workers allocated by the Census Bureau. Some sort of allocation procedure is absolutely necessary so that workers with ungeocodable addresses can be assigned to an appropriate small area. However, the Census Bureau allocation procedure was believed to se-

riously reduce the overall accuracy of small-area place-of-work employment data. Therefore, it was decided to develop a reallocation procedure. That procedure accepted the Census Bureau's state and county assignments, as well as their geocoding to place, and then used the special tabulation's detailed breakdowns of industrial class and occupational class to match workers needing to be reallocated to TAZ with workers who had been geocoded to tract and block and thereby to TAZ by the Census Bureau.

### SEMCOG Reallocation Method

The reallocation method used by SEMCOG is illustrated in Figure 1. The procedure begins by grouping the workers to be reallocated by their workplace destination area, which is the county to which the Census Bureau assigned the worker (through either geocoding or allocation), subdivided into places for those workers who were so geocoded by the Census Bureau.

Next, to identify all tract-geocoded workers in the destination area who match the worker being reallo-

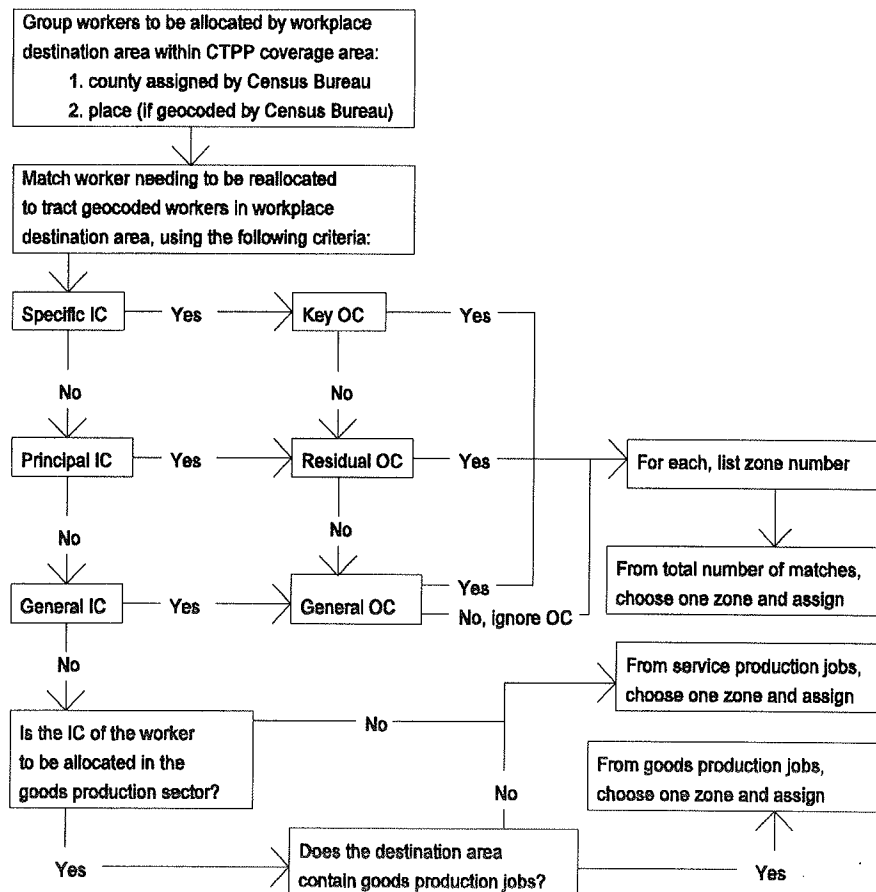


FIGURE 1 SEMCOG reallocation method.

cated, a decision tree consisting of a series of questions about industrial class followed by a series of questions about occupational class is run for each worker. These questions are based upon a four-tier system of industrial classes and a three-tier system of occupational classes, both systems going from less general to more general.

The system of industrial classes (ICs) consists of the following:

1. **Specific IC:** These are the 75 classes used in the 1980 and 1990 special tabulations obtained from the Census Bureau. The Specific ICs were designed to recognize ICs that made up at least 1 percent of Southeast Michigan's employment plus some smaller industries having distinctive land use characteristics.
2. **Principal IC:** These are the 18 ICs used in the 1990 CTPP. Specific ICs combine to form Principal ICs.
3. **General IC:** These are the 11 ICs used in Census Bureau printed reports to summarize employment by industry. The principal ICs combine to form General ICs.
4. Finally, if no match can be made by General IC, one is made using a distinction between goods-producing and services-producing classes.

The system of occupational classes (OCs) consists of the following:

1. **Key OC:** These are Specific OCs that form at least 10 percent of the region's employment within a Specific IC. (The 1990 special tabulation included 40 Specific OCs, whereas the 1980 tabulation had 39 because of the absence of military occupations.) It turns out that in 1990 every Specific IC has at least one Key OC and that the range extends up to six. Key OCs are used to identify those Specific OCs that occur relatively frequently in a Specific IC, and therefore can be expected in many zones, where they will provide potential matches. Conversely, OCs that are uncommon to a Specific IC will tend to occur in relatively few zones. If the reallocation procedure did not use Key OCs, workers of a given Specific IC and a given uncommon Specific OC needing to be allocated would be matched only to those few tract-geocoded workers.
2. **Residual OC:** To prevent such a biased allocation, Specific OCs that do not qualify as Key OCs within a given Specific IC are grouped into Residual OCs. A Residual OC consists of the Specific OCs of the special tabulations that occur within a given OC of the CTPP (that is, the OC equivalent of a Principal IC) less any relevant Key OCs of the given Specific IC.
3. **General OC:** These are the six OCs used in Census Bureau printed reports to summarize employment by occupation. The Key OCs plus the Residual OCs form General OCs.

The procedure begins with Specific IC and if necessary moves down the IC sequence until matches are found. It then moves over to the appropriate box in the OC sequence and down until a match on OC has been made. If no match can be made even on General OC, OC is ignored and all the IC-matching geocoded workers are used. Finally, a random process is used to choose a single matching tract-geocoded worker. That worker's zone of work is assigned to the worker needing reallocation, completing the process for that worker. The process is repeated for each worker needing reallocation.

## Reallocation Results

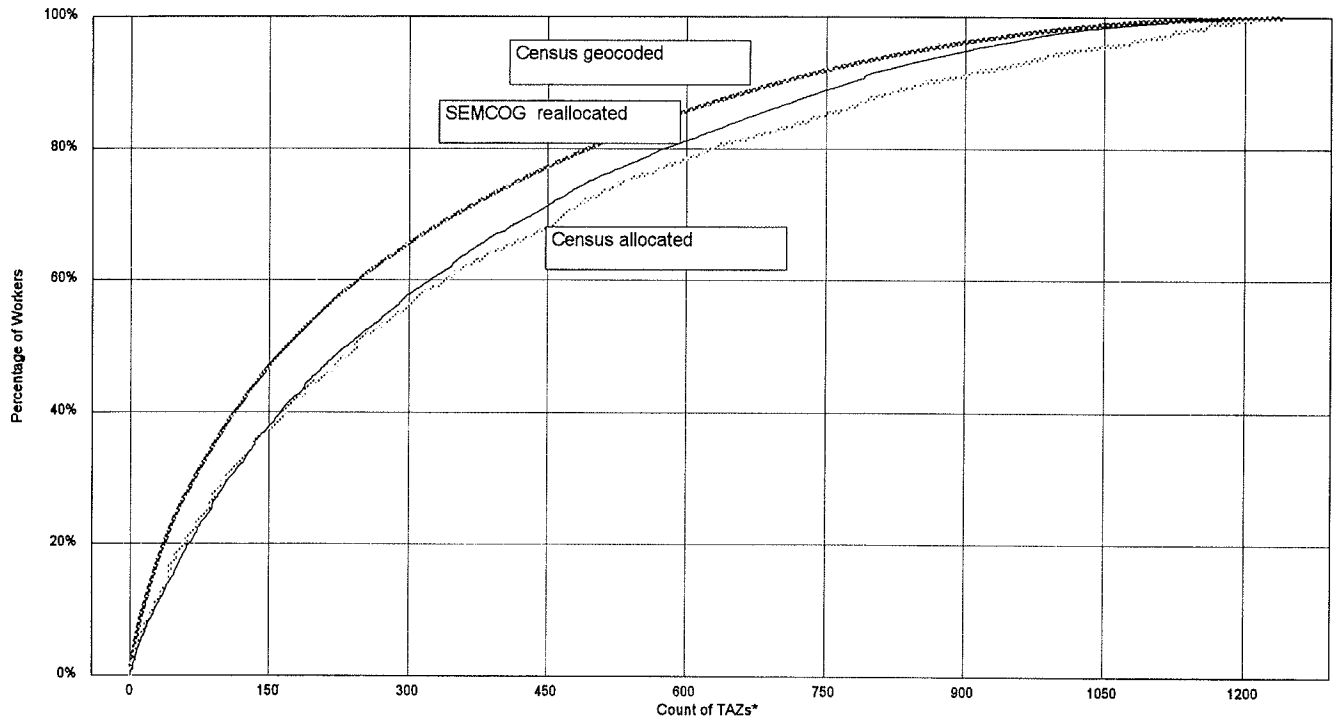
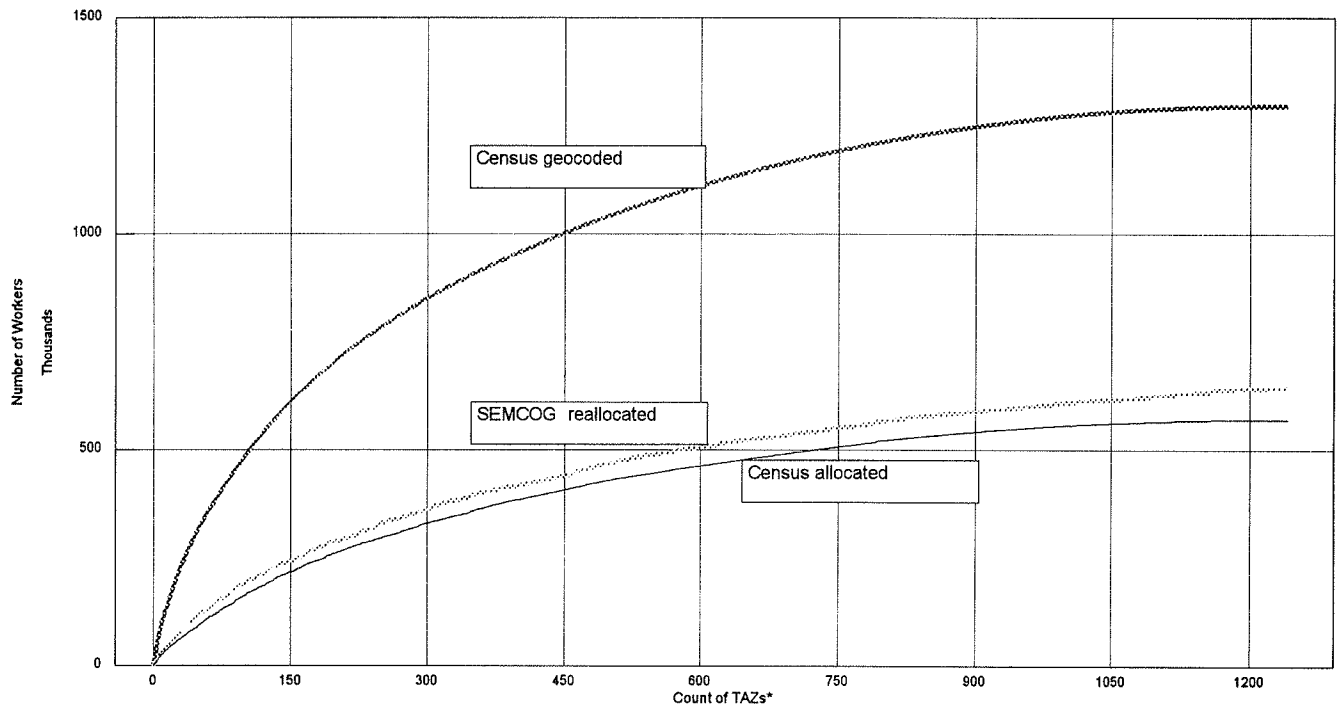
The results of applying the SEMCOG reallocation method to those workers who could not be geocoded to tract are illustrated in Figures 2 through 9. Figure 2 includes all workers, and the others present data on a selection of the specific ICs of the special tabulation. Each is divided into two graphs, the upper numerical graph giving a count of workers and the lower percentage graph giving a percentage distribution, both along the Y-axis.

The X-axis gives a count of TAZs arranged in an order that corresponds to the occurrence of Census Bureau tract-geocoded workers in the specified IC. The leftmost TAZ has the greatest number of workers in that IC; the other TAZs are arranged in descending order.

Each graph shows three curves representing Census Bureau geocoded, Census Bureau allocated, and SEMCOG reallocated workers. The curves are cumulative, each corresponding to the distribution (either numerical or percentage) of workers of its type of workplace coding procedure. (It is important to keep in mind that the Census Bureau allocated and the SEMCOG reallocated curves both follow the X-axis TAZ sequencing established by the Census Bureau geocoded workers: descending order, highest TAZ first).

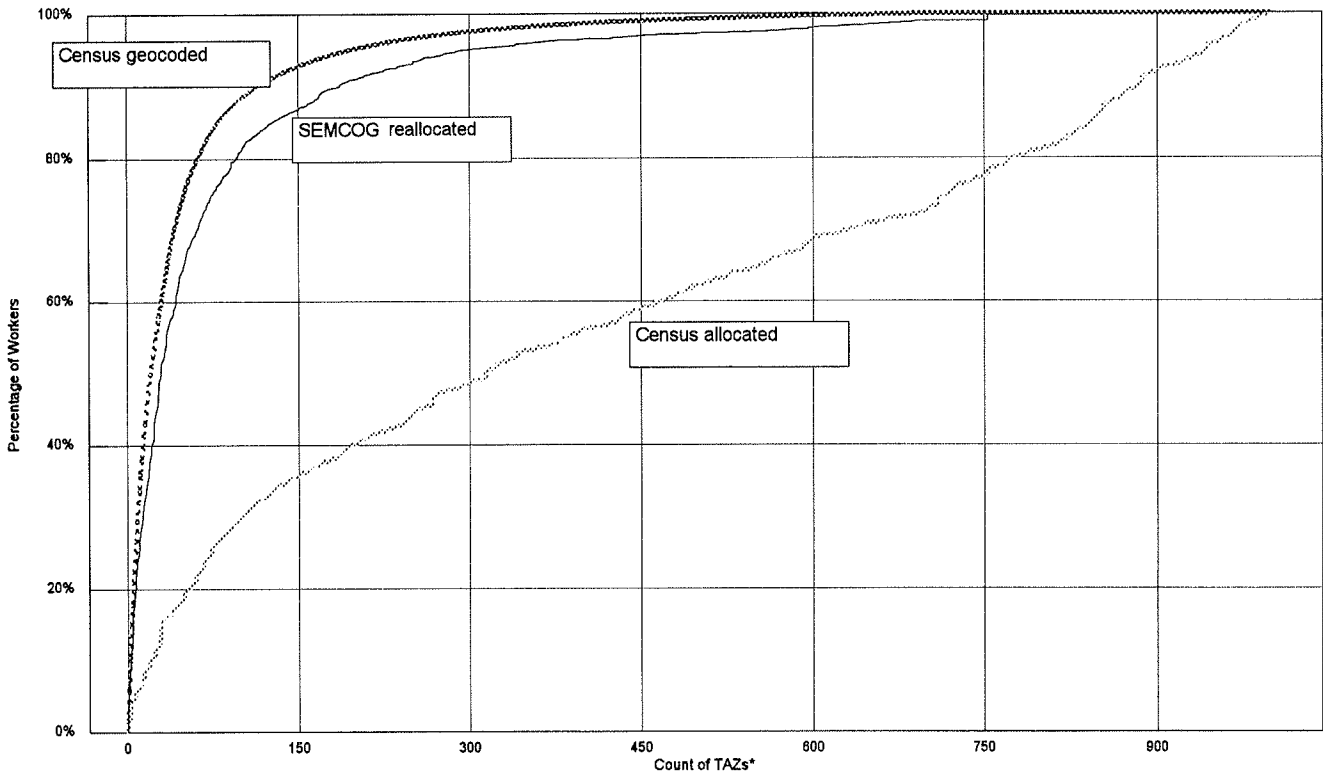
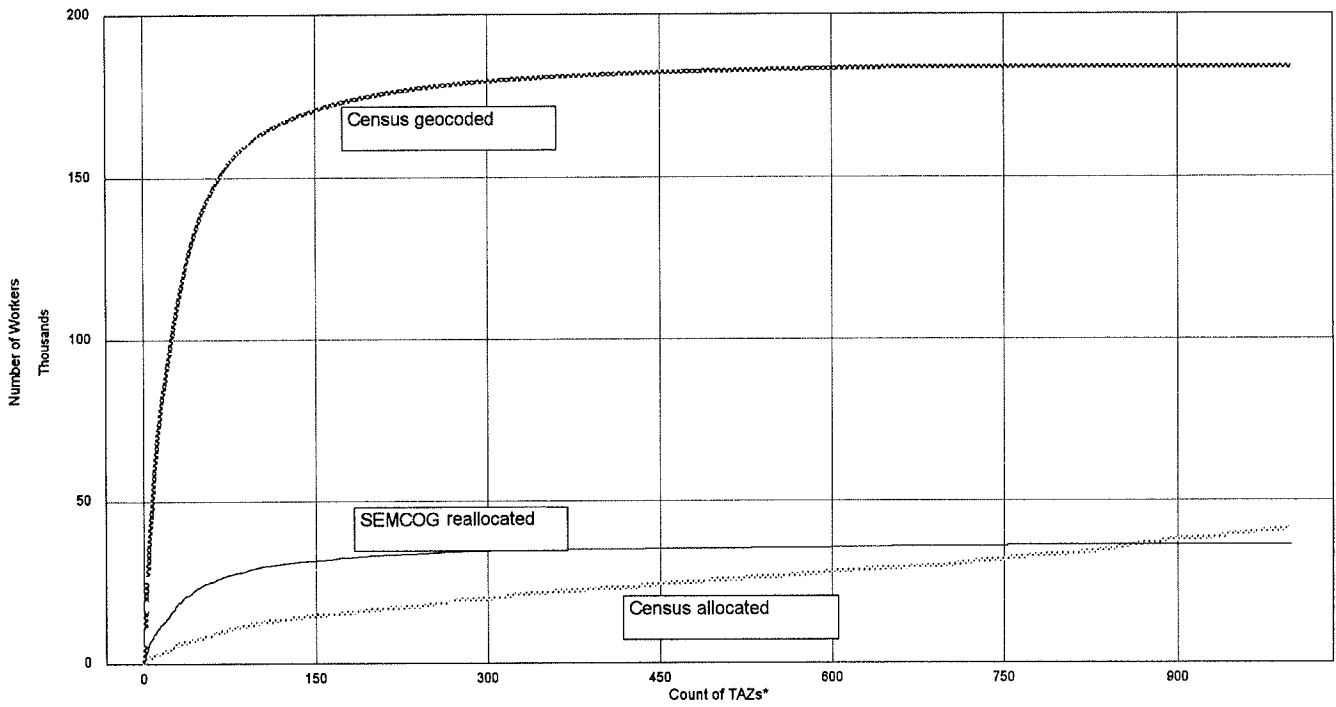
The curves demonstrate four key aspects of the specified IC:

1. **Number of workers:** Using the top graph, sum the Census Bureau geocoded plus the Census Bureau allocated. (Note that the SEMCOG-reallocated total tends to be somewhat lower than the Census Bureau allocated because of a difference in the two allocation procedures.)
  2. **Degree of concentration:** Using the lower graph, note that the steeper the curve, the higher the portion of workers located in a relatively few TAZs. The air transportation IC (Figure 4, *bottom*) shows an extremely high degree of concentration.
  3. **Geocoding rate:** Using the top graph, approximate the number of Census Bureau geocoded workers as a
- (text continues on page 45)*



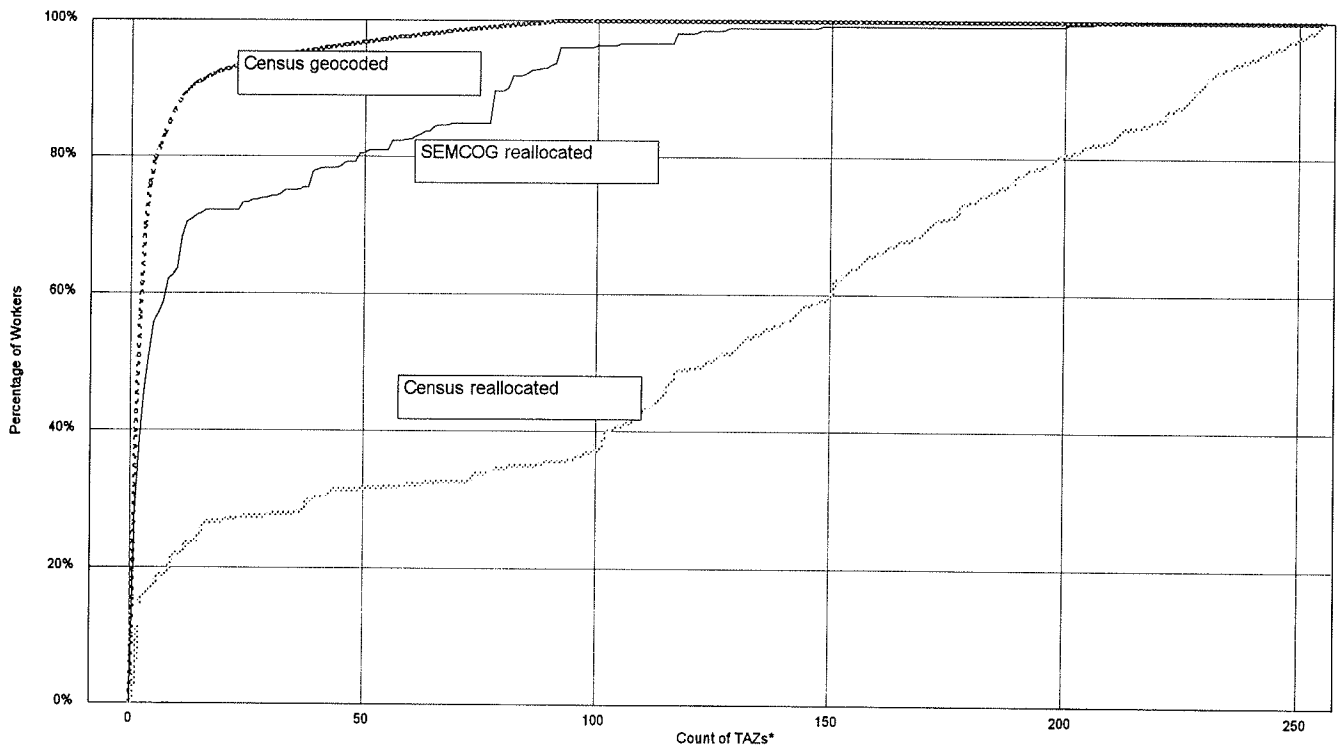
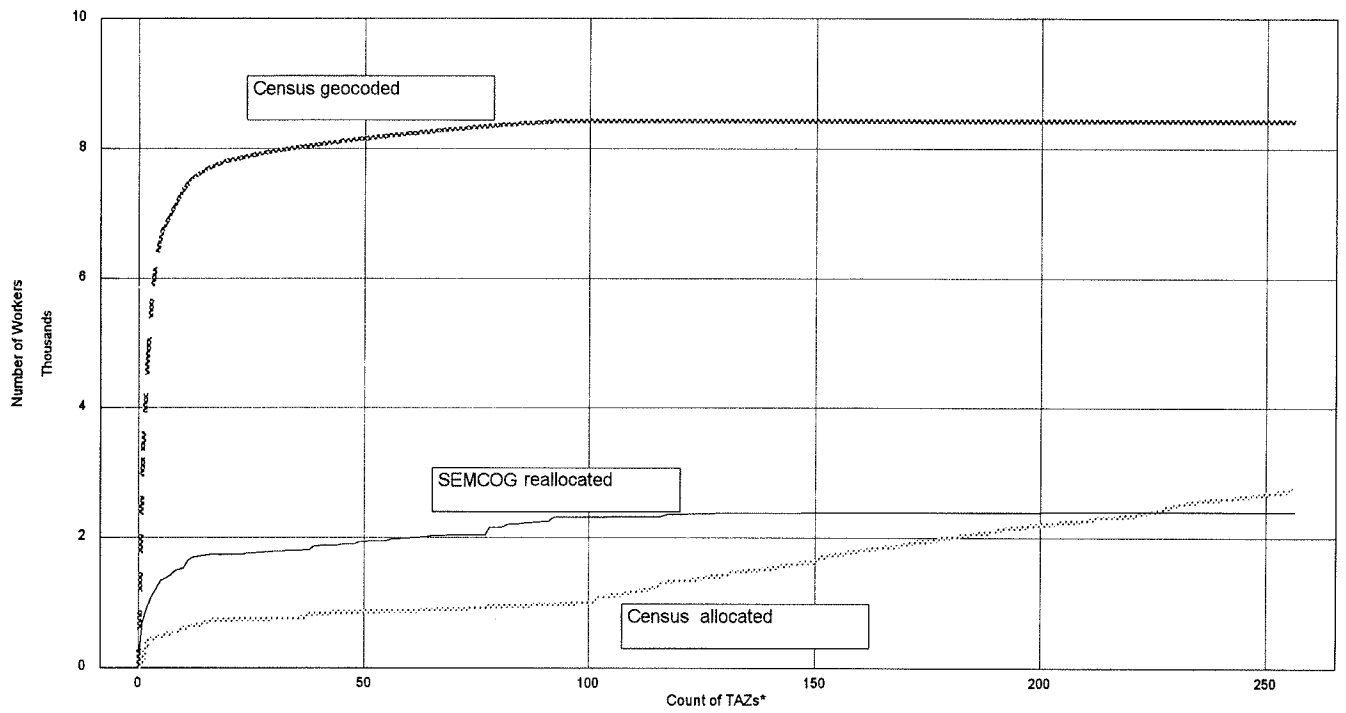
\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

FIGURE 2 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—all industrial classes.



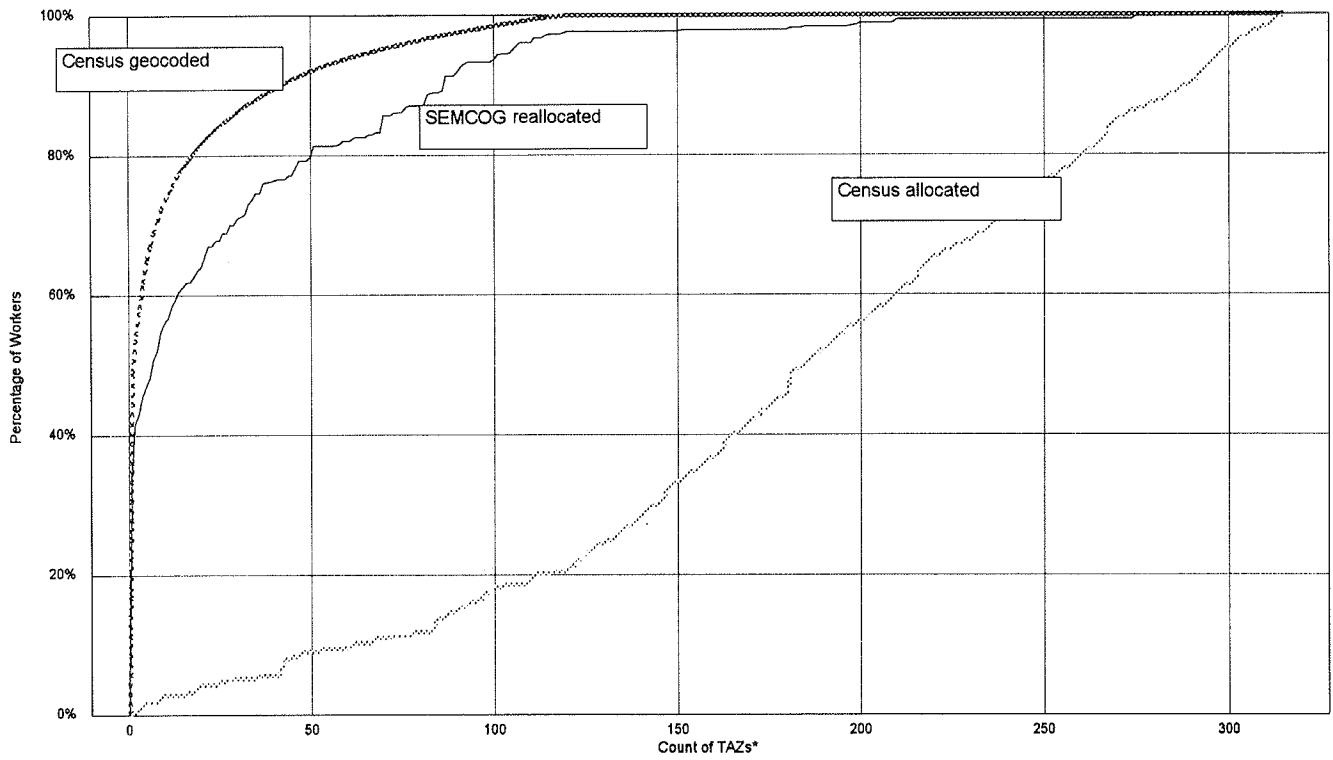
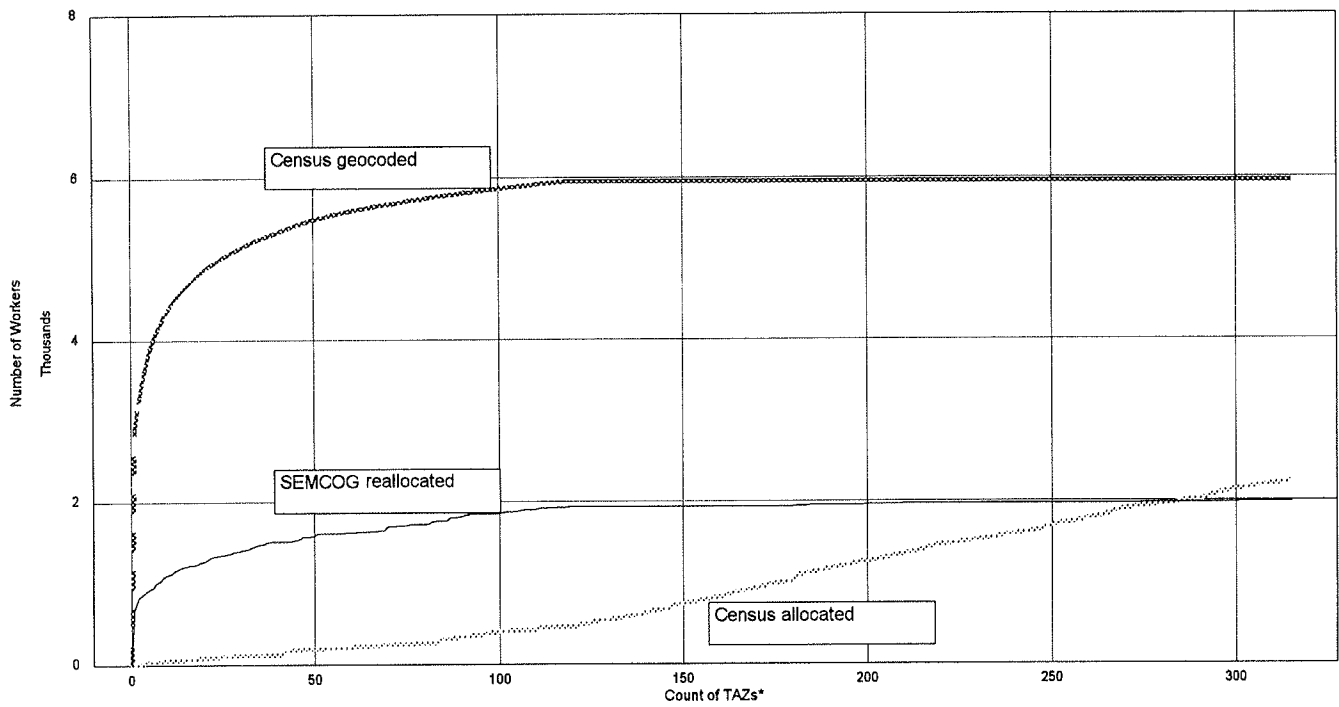
\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

FIGURE 3 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—manufacturing: transportation equipment.



\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

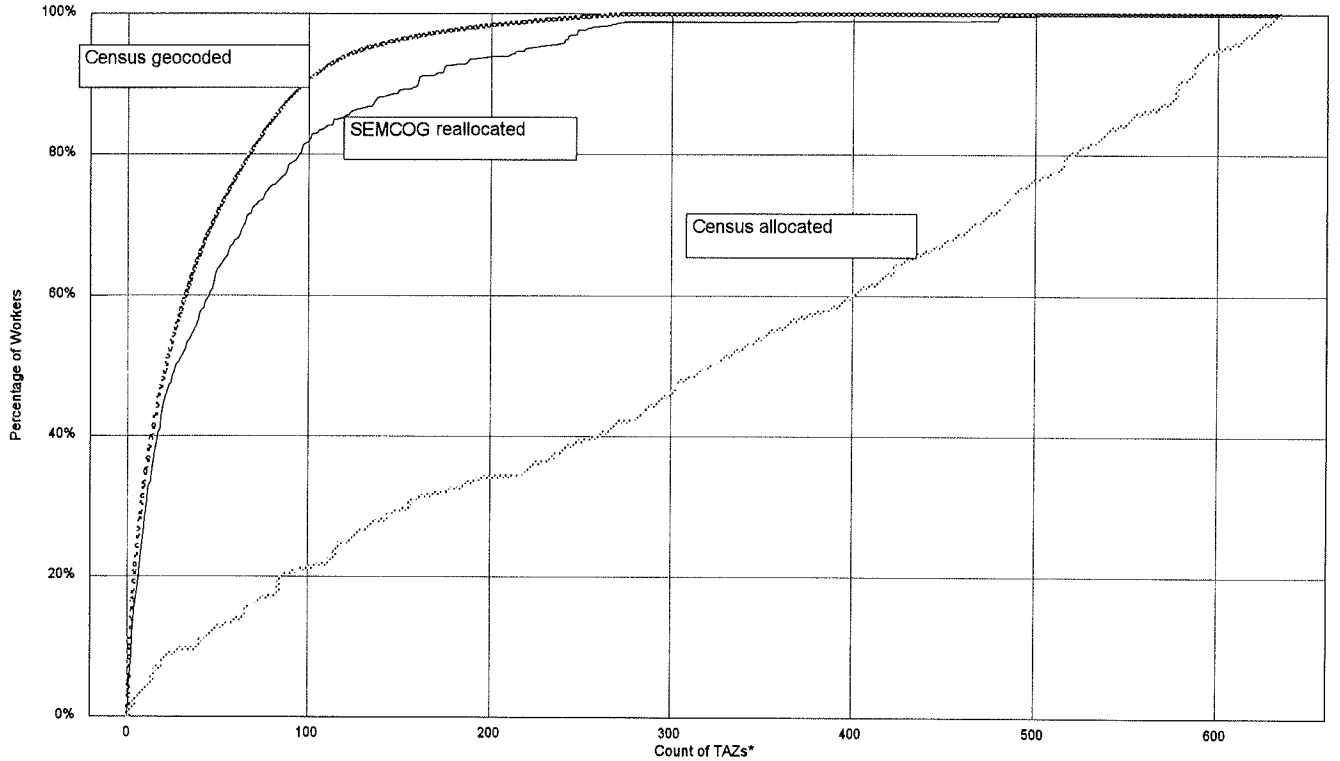
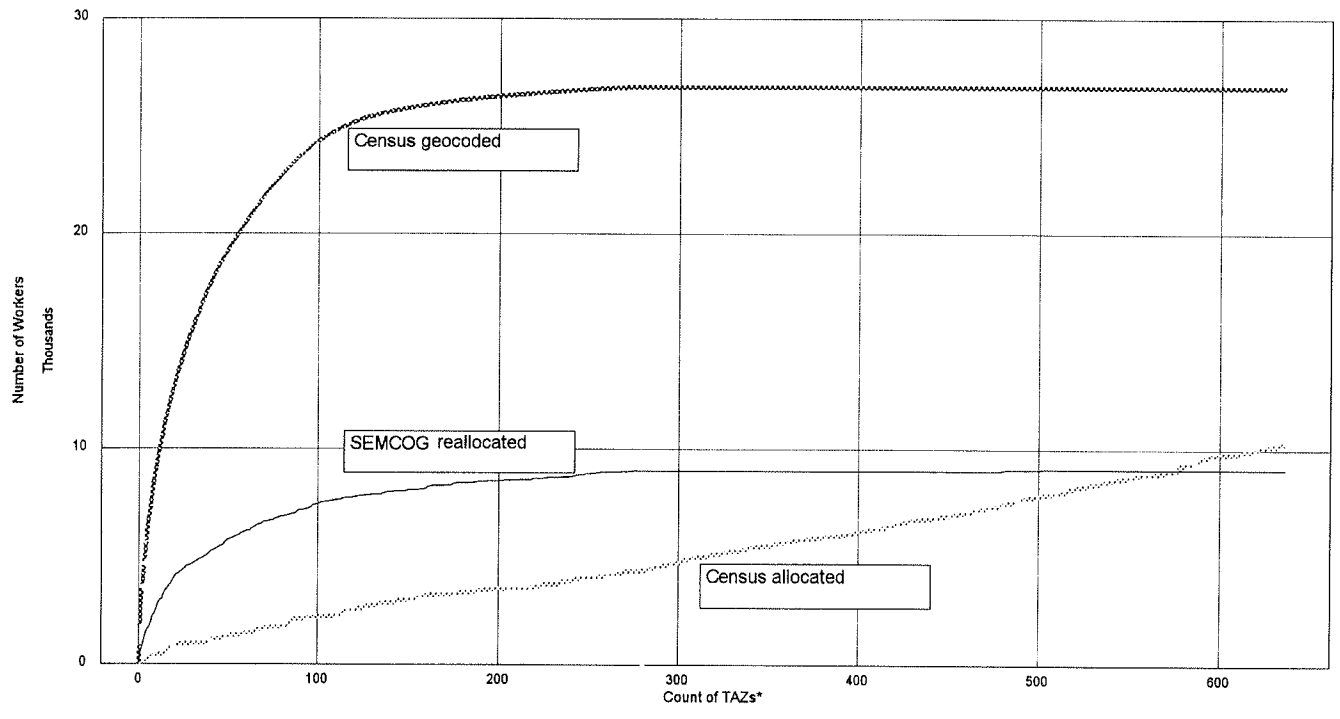
FIGURE 4 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—transportation: air.



\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

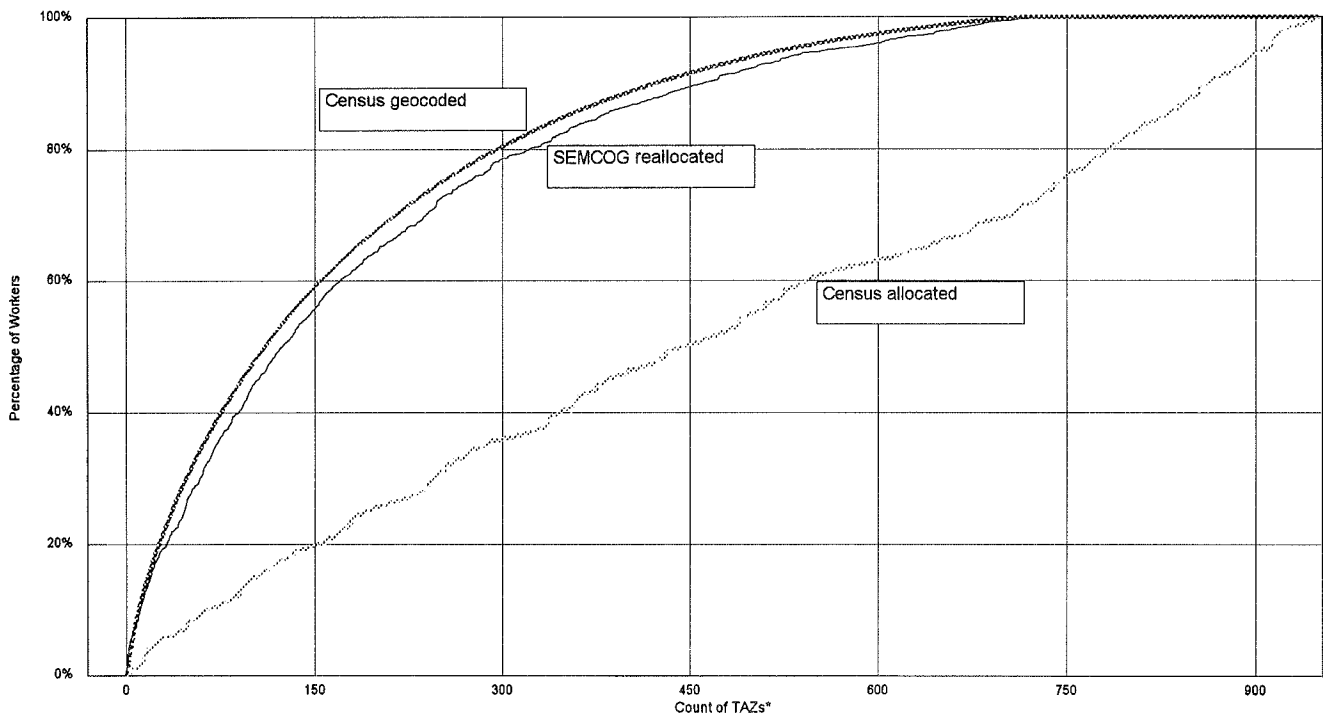
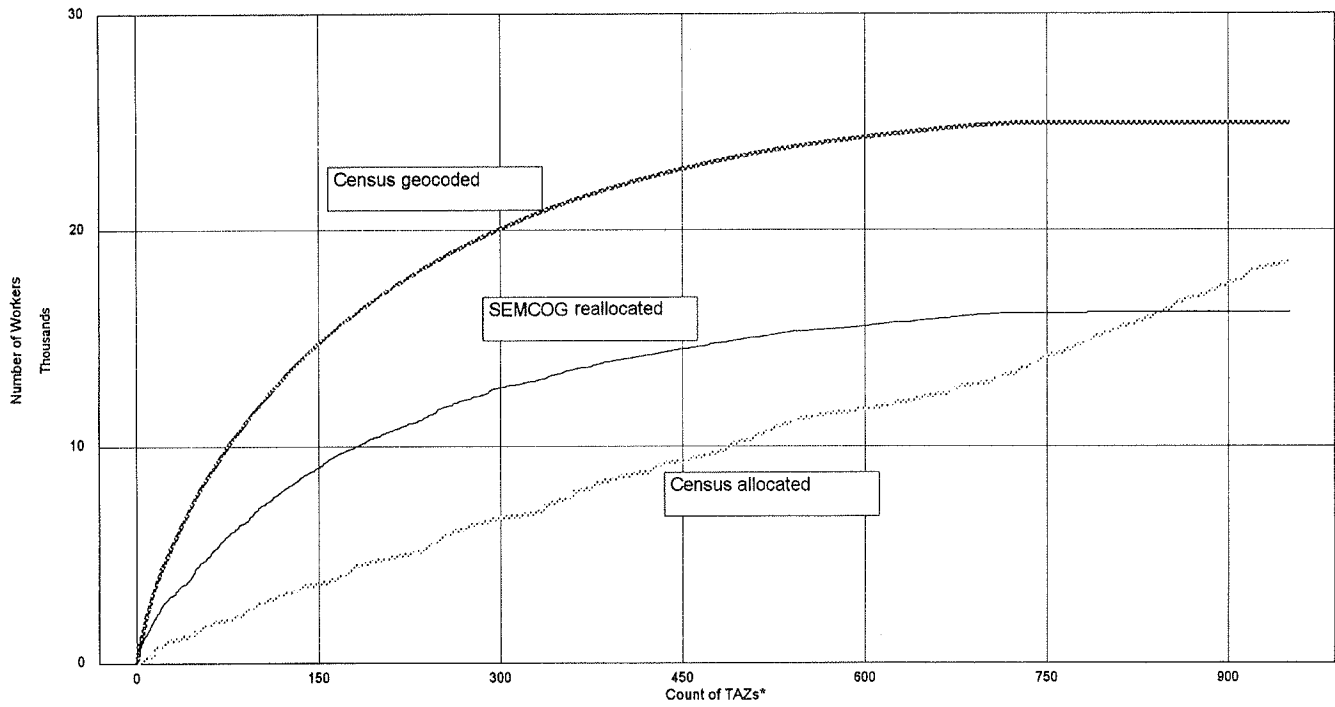
FIGURE 5 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—utilities: electric.





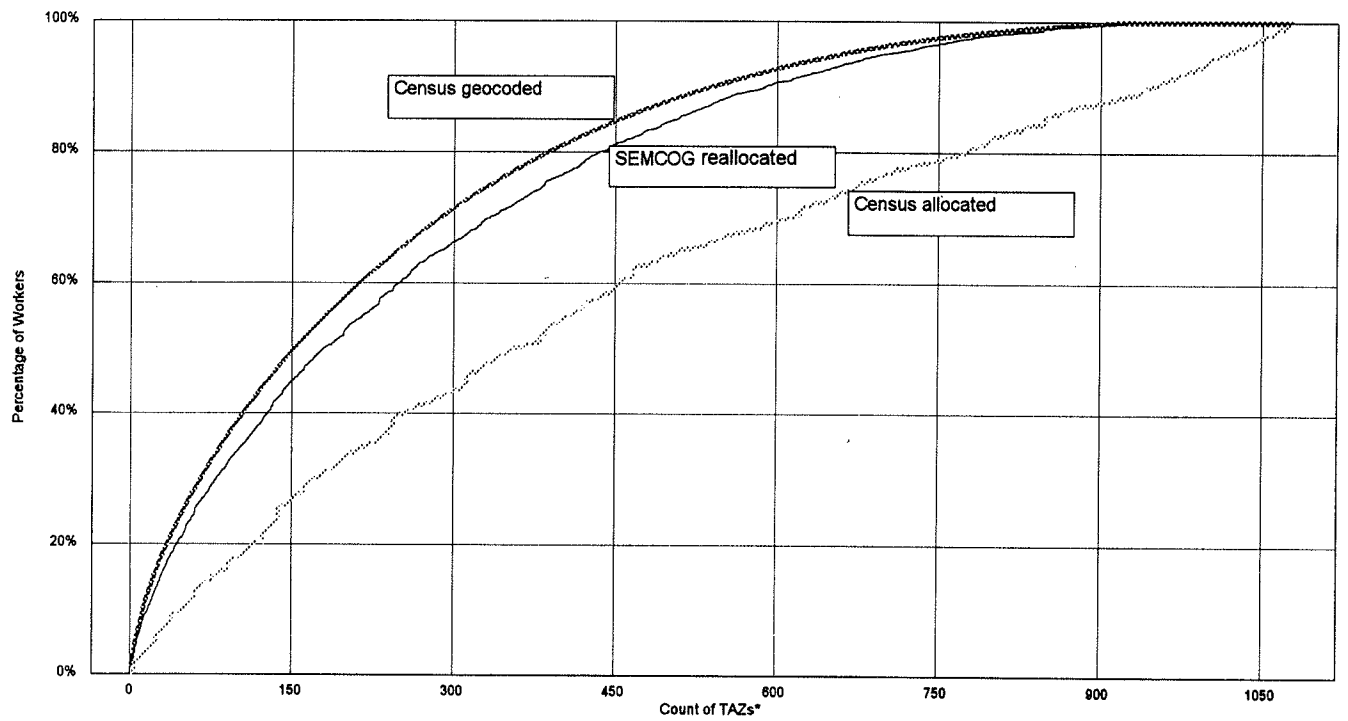
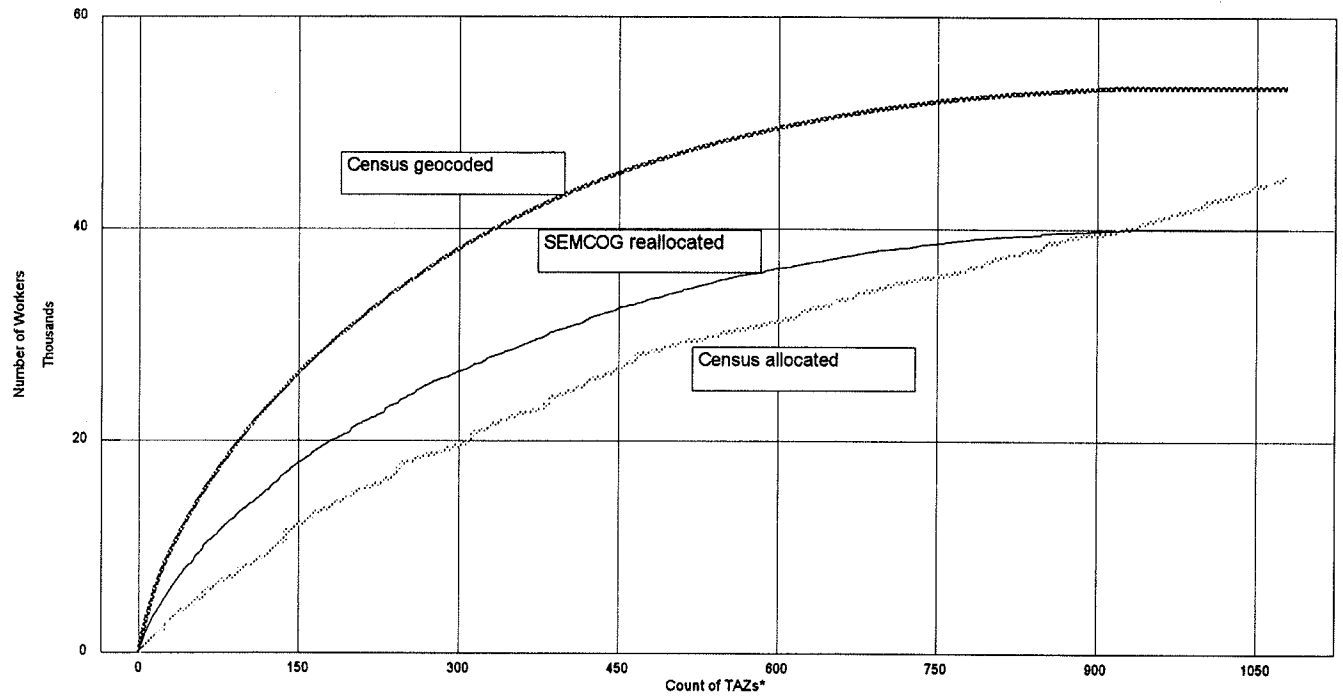
\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

FIGURE 6 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—retail trade: general merchandise stores.



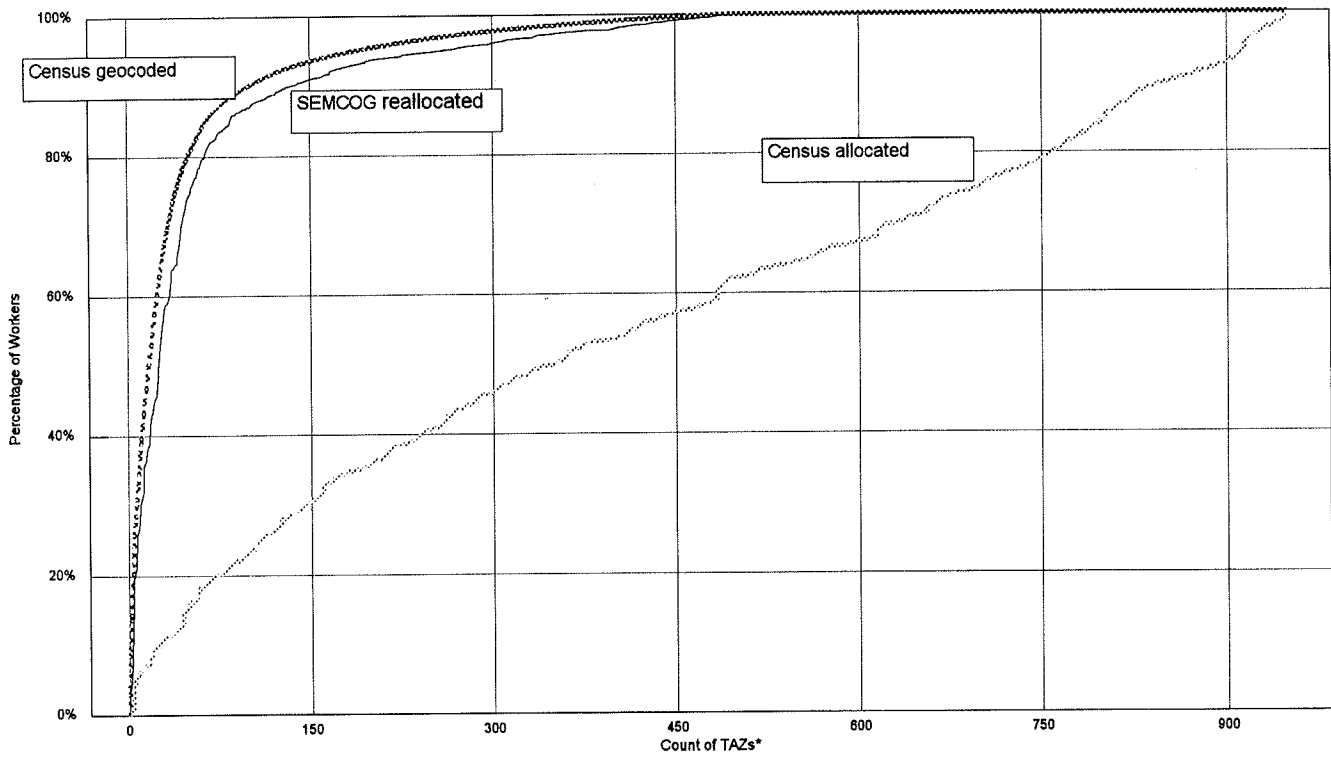
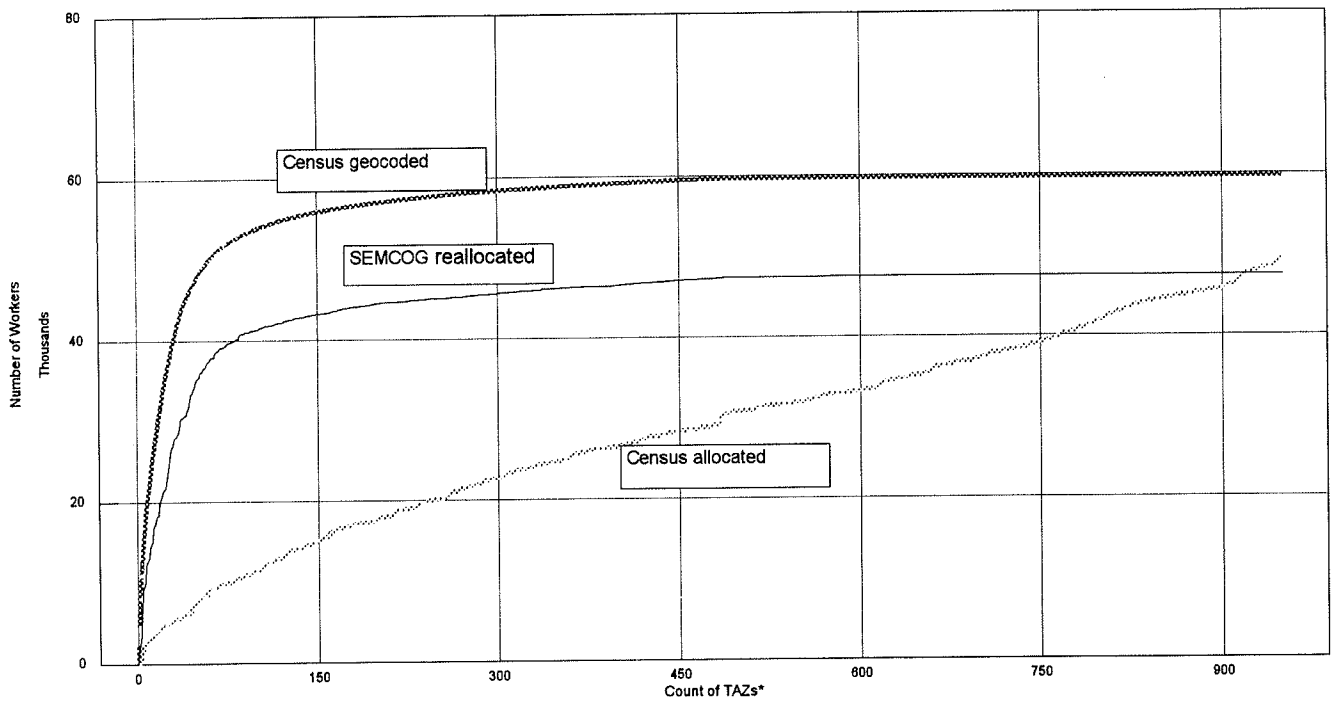
\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

FIGURE 7 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—retail trade: food stores.



\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

FIGURE 8 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—retail trade: eating and drinking places.



\*Note: TAZs are sorted by number of Census geocoded workers, largest first.

FIGURE 9 Cumulative distribution of workers by workplace coding procedure by TAZ, Southeast Michigan, 1990—hospitals.

portion of total workers (geocoded plus allocated). The transportation equipment manufacturing IC (Figure 3, *top*) shows roughly 185,000 geocoded and 40,000 allocated workers, equal to a geocoding rate of roughly 82 percent. In contrast, the graph for retail trade: eating and drinking places (Figure 8, *top*) shows roughly 55,000 geocoded and 45,000 allocated, giving a geocoding rate of only 55 percent. If the allocation procedure does have the problems that it appears to have, it is obvious that the lower the IC geocoding rate, the greater will be the negative impact and the greater the percentage of workers who will be assigned to TAZs that are inappropriate for that industry. Further, it should be borne in mind that incorrect workplace assignments produce incorrect home-to-work commuting trips.

4. **Degree of replication:** This is the central concern of this paper, namely, how well do Census Bureau allocated workers replicate the small-area locational pattern of Census Bureau geocoded workers? Further, can there be a better way to allocate ungeocodable workers? To answer both these questions, examine the percentage graphs (Figures 2 through 9, *lower graphs*).

The graph of all ICs (Figure 2, *bottom*) shows that both the Census Bureau allocated and the SEMCOG reallocated curves track the Census Bureau geocoded curve fairly well. However, such is not the case with the selected Specific ICs; the Census Bureau allocated curve does not do a good job of tracking the Census Bureau geocoded curve. For example, the Census Bureau allocated curve of the electric utilities IC (Figure 5, *bottom*) has allocated less than 20 percent of its workers by the first 100 TAZs, whereas by that point the Census Bureau geocoded curve has accounted for 98 percent of its workers. That geocoded curve hits 100 percent at around the 130th TAZ, but at that point the allocated curve still has 80 percent of its jobs yet to be assigned to TAZs. In other words, that 80 percent will go into TAZs that have no geocoded workers for that industry, or into TAZs where, for the most part, they do not belong. The SEMCOG reallocated curve replicates the geocoded curve well, but not exactly. Some degree of discrepancy is to be expected. For one thing, if a worker in a given IC was geocoded to place but the place had no geocoded workers for that industry, the reallocation method accepts the Census Bureau's geocoding and then matches on Principal IC or, failing that, General IC, as illustrated in Figure 1.

A comparison of the Census Bureau geocoded and SEMCOG reallocated curves by Specific IC demonstrates that the reallocation method, even without using means of transportation and travel time, produces a better replication of the small-area pattern of geocoded workers than does the current Census Bureau allocation method. However, a full allocation method should include both means of transportation and travel time. Time demands

plus the complexity of the task kept SEMCOG staff from attempting to include them in their reallocation method. In the final section of this paper a proposal for a comprehensive allocation method is presented.

## RECOMMENDATION FOR CENSUS 2000

An improved allocation method for use in Census 2000 could combine features from both the current Census Bureau allocation method and the SEMCOG reallocation method, and in addition could incorporate several new features. Means of transportation and trip duration would be adopted from the Census Bureau method. However, there would be at least four separate means, namely, walk, bicycle, car/truck/van, and transit (with transit being split into separate types in metropolitan areas in which transit types other than bus are significant). Walk and bicycle are recognized as separate means because their reach is so restricted compared with car or transit. The current Census Bureau method uses only two means of transportation: transit and all other.

Both existing methods use ICs. However, the Census Bureau method uses up to 11 classes, whereas the SEMCOG method uses up to 75. In addition, the improved method would adopt OCs from the SEMCOG method. The Census Bureau method makes no use of OCs, whereas the SEMCOG method uses up to 40. The improved method would use detailed breakdowns for both industry and occupation, with the exact number of classes to be determined.

The new method would introduce the concepts of "commuting origin areas," representing areas of residence, and "commuting destination areas," representing areas of workplace. Each TAZ would have an origin area, consisting of the TAZ itself as a central zone plus a ring of immediately surrounding TAZs. Place-of-work data for tract-geocoded workers residing in each origin area would be used to demarcate a set of destination areas, which would then be assigned to the origin area's central TAZ. (By assigning the destination areas of the entire origin area to the central TAZ, the sample size of each central TAZ is increased.)

There would be a separate destination area for each travel time interval of the car/truck/van and the transit means of transportation. Each particular destination area would represent the TAZs that were the work ends of commuting trips made by workers residing in the origin area who were geocoded to tract and block of workplace and who made a trip of the specified length by the specified means. The allocation of workers whose workplaces had been geocoded to a place but not to a tract would be restricted to the portion of the commuting destination area that fell within that place.

Because walk and bicycle trips are relatively infrequent, their destination areas should not be demarcated

using data for tract-geocoded workers residing in the origin area. Instead, the destination area representing the first or shortest travel time interval would always include the entire origin area, plus any other TAZs that could be reached within the time interval. It would be assumed that there was one average travel speed for walk trips and another for bicycle trips, and straight-line centroid-to-centroid distances, central TAZ to each TAZ beyond the origin area, would be used. Subsequent travel time intervals for walk and bicycle trips would also use average travel times and zone-to-zone straight-line distances. In the improved allocation method, following the demarcation of an origin area and destination area for each central TAZ, workers needing to be allocated would be assigned to tract and block using the previously described procedure illustrated in Figure 1.

The proposed allocation method represents a systematic effort to make maximum use of all that is known about workers needing to be allocated so that they can be assigned to a tract and block that best correspond to their place of residence, their work-trip characteristics, and their specific industry and occupation. The nationwide TIGER File and the current capabilities of geographic information systems bring such an improved allocation system within reach.

#### ACKNOWLEDGMENT

The technical work described in this paper was financed in part through grants from the Federal Transit Administration and Federal Highway Administration, U.S.

Department of Transportation; through the Michigan Department of Transportation; and through local membership contributions. The specifications of the worker data file were developed through consultation with Census Bureau staff, who then created the special tabulation. Phillip Salopek of the Census Bureau and Philip Fulton of the Bureau of Transportation Statistics both provided valuable information and advice throughout this project. Processing and manipulation of the special tabulation data were performed by former SEMCOG staff member George Janes, planning analyst, and current SEMCOG staff member Ming H. Tsai, planning analyst. In addition, Tsai did additional analytical work, designed and produced the figures, and gave other important assistance in the production of the final paper.

Finally, special thanks are due the Southeast Michigan Council of Governments for generous willingness to provide financial and staff support for this project. Without SEMCOG's allocation, there would be no reallocation.

#### REFERENCES

1. Fulton, P.N. Allocating Incomplete Place-of-Work Responses in the 1980 Census Urban Transportation Planning Package. In *Transportation Research Record 981*, TRB, National Research Council, Washington, D.C., 1984, pp. 15-21.
2. *Census Transportation Planning Package, Part 7: Characteristics of Workers by Place of Work*. File documentation, compact disc. Bureau of the Census, U.S. Department of Commerce, Dec. 1995.