# SIMULATION OF CONTINUOUS MEASUREMENT FOR SMALL AREA ESTIMATION

Gregg Diffendal, Lynn Weidman
Gregg Diffendal, Bureau of the Census, 2A-SHEP, Washington, DC 20233

The Census Bureau is considering the replacement of the decennial census long form for the collection of the sample data with a continuous survey. The Continuous Measurement Survey (CMS) is designed to provide yearly estimates for large areas and multi-year averages for small areas. This will provide timely estimates for large areas that are currently produced only every ten years from the decennial long form. The tract and block group estimates (the smallest area that estimates will be produced) would accumulate sample over five years (three years at the start up) to provide rolling estimates. Currently very little, if any, data are available at these small areas except every ten years from the decennial census.

Considerable concern has been raised by the data-user community on the interpretation of multi-year estimates for small areas. In this paper, small areas are census tracts which consist of about 1000 housing units and block groups which consist of about 300 housing units. These are the smallest areas for which the Census Bureau provides small area sample data, usually only available from the decennial census long form. To assist data users to better understand the data that would be produced from the CMS, we have undertaken a simulation study of multi-year averages for census tracts. This paper reports on our current simulation work and the resulting estimates that the CMS would produced. Our hope is that the data users will be able to better understand the CMS and the resulting data products so that they can use the information that will be available from the CMS to fill their data needs.

## BASIC POPULATION FOR SIMULATION

The site chosen to do the simulation is the city of Oakland, Ca. This is one of the sites for the 1995 test census. The selection of the particular area was not necessarily important except to provide the basis for the simulations. We look at a variety of scenarios by changing the population and some characteristics over time. The simulations will allow us to measure how well the multi-year tract level estimates are tracking the actual population over time.

The simulations started with the census data from 1990 for Oakland. The 1990 short form was used to reflect the population. The long form data was added for those who had completed a long form in the census. A probabilistic imputation was used to complete the long form variables on the short form questionnaires for the other housing units. For the short form data households, we randomly select (conditional on some short form variables) from the observed long form data distribution. For example, the employment categories were imputed conditional on the age of the person on the short form household. To keep the number of variables manageable, only household income (in 25 categories), education (10 categories) and employment (4 categories) were used. The short form variables used were tenure, age, race, hispanic origin and information on the mail response. The employment categories were imputed conditional on the age of the person on the short form household. Income was imputed conditional on tenure and housing value (owners) or monthly rent (renters). Education was imputed conditional on age, tenure, race and hispanic origin.

The Oakland site consists of 105 census tracts, with 154737 housing units and 364371 persons. Group quarter persons were not included in this work.

## SIMULATION OF HOUSEHOLD AND HOUSING UNIT CHANGES OVER TIME

As a baseline for aging the population, we examined the change in households and the housing unit population from other surveys. The Annual Housing Survey indicated a 20% change in households over two years. We used a 9% change in occupied housing units per year. For the Oakland simulation, an occupied unit remained the same with probability 0.91. The occupied housing unit became a vacant housing unit with probability of .039 and was deleted with probability .051. A vacant housing unit became an occupied housing unit with probability 0.50. The number of persons and their characteristics for this converted housing unit was imputed from the last processed occupied housing unit. A vacant housing unit remained vacant with probability 0.40 and was deleted with probability 0.10. A new occupied unit was created with probability of .049. The previous occupied housing unit was used to impute the number of persons and their characteristics for this new housing unit. A new vacant unit was created with probability 0.0054. These new units were added to reflect new construction and business to residential conversions. The probabilities for new units were selected so that essentially no net growth would be achieved. Table 1 provides a summary of the city of Oakland from the 1990 census and from the simulations through 1995.

The deleted units were not carried on the file. In practice, when an address is sampled, some will be determined to be not a housing unit. Non-housing units in the sample frame were not accounted for in the simulations. Mostly, this effects our estimates of the total number of housing units for an area and whether to use this estimate as a housing unit control in the weighting.

## ALTERNATE SIMULATIONS FOR POPULATION AND EMPLOYMENT

To provide some alternative simulations, we examined changing the population and some characteristics over time. As a start we choose about one-fourth of the census tracts to have a sizable positive population growth over the years 1992 to 1995. For occupied housing units, the probability of remaining occupied is 0.95 (.91 under no growth). The probability of becoming vacant is 0.04 (0.039 under no growth) and the probability of being deleted is 0.01 (0.051 under no growth). For vacant housing units, the probability of becoming occupied is 0.58 (0.50 under no growth), the probability of remaining vacant is 0.40 (the same under no growth) and probability of being deleted is 0.02 (0.10 under no growth). New occupied housing units were created with probability 0.10 (0.049 under no growth) and new vacants were created with probability 0.0054 (the same under no growth). To increase employment, all new occupied housing units or converted from vacant housing units had all unemployed persons in these housing units changed to be employed. The not in labor force persons remained not in the labor force. The other three-fourth of the tracts changed with probabilities set under the no growth model and no change in the employment status.

We took the same set of tracts and allowed for a decrease in the population for these areas over the same time span. For occupied housing units, the probability of remaining occupied is 0.89. The probability of an occupied unit becoming vacant is 0.049 and becoming deleted is 0.061. For vacant units, the probability of becoming occupied is 0.30, the probability of remaining vacant is 0.50 and the probability of becoming deleted is 0.20. New occupied housing units were created with probability 0.02 and new vacant housing units were created with probability 0.01. For the declining population, all new housing units were converted to have only unemployed persons. The remaining three-fourths of the tracts used the same change in housing units and the population as simulated with no net growth.

Tables 1 and 2 contains a summary of these simulations for Oakland. Later we discuss the results at the tract level. For the positive growth simulation the results increase the population by 1 to 2 percent (2000 to 3000 persons) and the employment rate by about 2.0 to 3.0 per year. For the negative growth simulation, the population decreases by 1 to 2 percent per year and the employment rate decreasing by about 7.0 per year.

## SIMULATING THE CONTINUOUS MEASUREMENT SAMPLE

The CM sample will be a systematic sample from the Master Address File (MAF). The MAF will consist of the 1990 census address file with updates from the US Postal Services' Delivery Sequence File. Our current plans are to sample 15 percent of the housing units in most areas and 30 percent of the housing units in governmental units of fewer than 1000 housing units over a three year period from 1999 to 2001. The sample will be about 400,000 housing units per month. (This allows equivalent long form data products to be delivered in the same time frame as in 1990.) The sample will drop to about 250,000 housing units per month in 2002 and thereafter. Large area (population over 250,000) estimates will be available with an accumulation of a years worth of data. Small area estimates (block groups, census tracts, etc.) will average three years of data in the beginning. With the reduced sample size in 2002, the small area estimates will be a four year average in 2002 and five year average in 2003 and for then on. The samples for our simulations only examine the five year estimates for census tracts.

The data collection will be conducted in three phases. The first phase mails the CM form to every sample housing unit. The current mailing procedure uses a pre-notice letter, followed by a CM questionnaire and followed by a second CM questionnaire to all mail non-returns after about three weeks. Our current estimates are that about 60 percent of the occupied housing units will respond by mail. This corresponds to the observed rate in Oakland for the 1990 census.

For mail non-returns, about 40 percent are expected to respond by telephone. ( This involves telephone lookup of the sample cases.) The remaining sample cases (about 24 percent of the sample) will be sub-sampled at a one in three rate for a personal visit. The non-interview rate for the personal visit cases is estimated to be 40 percent. This corresponds to an estimated non-interview rate of 9.6 percent.

For our simulations of the CM sample and the resulting estimates, we used a 1 in 35 sample of the population per year. This is slightly less than the proposed sample plans, 14.3 versus 15.0 (versus 16.7 percent for the 1990 decennial census). The differences are small and help reflect some uncertainties in the estimates that are not accounted for in our simulations. ( For example the uncertainty from post-master returns which may be occupied housing units.) All housing units with 1990 census mail returns (short and long forms including Spanish forms) are assumed to return their CM forms by mail. The rates from a mailing test in 1993 of the 1990 census long forms using a similar mail scheme as is currently being proposed for CM obtained results similar to what is assumed here. (See Dillman, Clark and Treat 1994). Post-Master returns are assumed to identify all vacant units (no simulation of vacant housing unit information was done for this work). The post-master returns will be examined from the 1996 CM test.

Each mail non-return occupied housing units was randomly assigned to be a telephone responses with probability of 0.40. For the remaining cases, we take a systematic sample of one in three and randomly assign them as interviews with probability of 0.60. The corresponding yearly sample weights are 35 for mail returns, vacants and telephone cases, and 135 for all personal visit interviews. (The weights for all noninterviews from personal visits are spread over all personal visit interviews.) For the 5 year estimates the weights are divided by five. No housing unit or population controls were used. (It is expected that some form of controls would be used in the CMS, mostly at larger levels like the county. Possibly some form of housing unit controls could be used at the tract level.)

The multi-year estimates at the tract level are the sum of the observations over the five years. The employment is the weighted sum of the number of employed persons divided by the weighted sum of all employed and unemployed persons. The multi-year estimates should provide comparable estimates to the decennial census if the population is stable over time. The decennial census derives a reduction in variance due to a higher sample size and being able to use the census 100 percent data (such as age-race-sex, population and housing totals) as controls in estimation. However if the population and/or their characteristics are changing over time then the CMS should be able to track some of these changes with some lag due to the use of multiple years of data for estimates at the tract level. For the decennial census, it is only able to produce estimates once every ten years and would not be able to measure changes except every ten years.

## SOME RESULTS

Table 2 shows the continuous measurement civilian employment estimate for Oakland for each year under the three simulations. Employment is the civilian employment, not including employment in the military. We can observe the variability in the continuous measurement estimates. Even at the city level, it is difficult to determine the trend for the positive or negative growth scenarios in any one year. As a comparison, the 1990 decennial census employment estimate for the city was 90.6 percent.

Table 3 contains a summary of tract level estimates under the three census scenarios and Table 4 contains the Decennial Census estimates and the continuous measurement multi-year estimates. The tract summaries present the quartiles to assess the simulations. The distribution for the years other than 1990 under the no growth model are essentially the same as the 1990 distribution. The distributions of the CM multi-year estimates are similar to those of the simulated census.

Table 5 provides a comparison of the census simulations and the CM estimates over time for the no growth scenario. The results for the multi-year averages for CM show some of the effects of sampling error on the estimates for these small areas. Under the positive or negative growth simulations (not shown), the CM estimates do not give any significant improvements that are observable in the quartile summary of the data. This is not surprising since the CM estimates are measuring the midpoint of the years being averaged, 1992 for the CM 90-94 estimates and 1993 for the CM 91-95 estimates. This is the start of the changes in the 25 tracts with change.

Looking only at the 25 tracts with change, the simulated 1990 census measured a mean tract employment rate of 88.4 percent. Under the positive growth scenario these same tracts had an employment rate of 92.6 percent in 1994 and a rate of 93.6 percent in 1995. The CM 90-94 mean estimate for these tracts is 90.9 percent and the CM 91-95 mean estimate is 92.3 percent. Clearly CM is picking up the change in these tracts (that is masked by sampling variability overall). For the negative growth scenario the mean tract employment rate is 73.6 percent in 1994 and 69.4 percent in 1995. The corresponding mean tract estimates from CM over the five years are 84.8 percent for 90-94 and 81.5 percent for 91-95. CM is capturing part of the decline for the small areas with some lag. Table 7 gives the error distribution for the 25 tracts that changed under positive growth and Table 8 gives the error distribution for the 25 tracts that changed under the negative growth scenario.

CONCLUSIONS

These simulations look at the multi-year estimates for census tracts. Continuous Measurement will be able to supply more up-to-date estimates for small areas for data users. These simulations begin to address the interpretation of these data and should assist data users to more fully interpret this data.

REFERENCES:

Dillman, Don A., Clark, Jon R. and Sinclair, Michael D. (1993), " The 1992 Simplified Questionnaire Test: Effects of Questionnaire Length, Respondent-Friendly Design, and Request for Social Security Numbers on Completion Rates." Proceedings of the Bureau of the Census 1992 Annual Research Conference, U.S. Department of Commerce, Washington DC, 8-17.

This paper reports the general results of research undertaken by Census Bureau Staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

Table 1. No Growth Simulation - Census

| Year | Vacant Housing Units | Occupied Housing Units | Number of Persons | Employment Rate(16+) |
|---|---|---|---|---|
| 1990 | 10216 | 144521 | 364371 | 90.4 * |
| 1991 | 10392 | 144350 | 364767 | 90.3 |
| 1992 | 10605 | 144073 | 364349 | 90.3 |
| 1993 | 10697 | 144318 | 365147 | 90.4 |
| 1994 | 10611 | 144022 | 365117 | 90.4 |
| 1995 | 10770 | 143890 | 364923 | 90.4 |

* The 1990 decennial census employment rate using the official weighting was 90.6%.

Summary of Alternate Simulations

| | Positive Growth | | | Negative Growth | | |
|---|---|---|---|---|---|---|
| Year | Pop. | Occ HU | Vac HU | Pop. | Occ HU | Vac HU |
| 1992 | 370610 | 146561 | 10618 | 360082 | 142229 | 11134 |
| 1993 | 376901 | 149079 | 10744 | 356543 | 140711 | 11518 |
| 1994 | 383856 | 151799 | 10979 | 353708 | 139510 | 11533 |
| 1995 | 391725 | 155114 | 11021 | 351044 | 138353 | 11314 |

Table 2. Employment for Oakland

| Year | No Growth | | Positive Growth | | Negative Growth | |
|---|---|---|---|---|---|---|
| | Census | CM Estimate | Census | CM Estimate | Census | CM Estimate |
| 1990 | 90.4 | 89.8 | same | same | same | same |
| 1991 | 90.3 | 90.9 | same | same | same | same |
| 1992 | 90.3 | 91.0 | 90.5 | 89.5 | 89.7 | 89.4 |
| 1993 | 90.4 | 90.9 | 90.8 | 91.0 | 89.0 | 88.5 |
| 1994 | 90.4 | 90.7 | 91.0 | 91.3 | 88.3 | 89.6 |
| 1995 | 90.4 | 91.2 | 91.2 | 93.1 | 87.8 | 86.9 |

Table 3
Summary of Tract Level Employment Rates
Simulated Census Tract Statistics

| | | Positive Growth | | Negative Growth | |
|---|---|---|---|---|---|
| | 1990 | 1994 | 1995 | 1994 | 1995 |
| Maximum | 97.1 | 100.0 | 100.0 | 97.1 | 97.6 |
| 3rd Quartile | 92.7 | 93.5 | 93.8 | 92.2 | 92.1 |
| Median | 89.3 | 90.5 | 91.0 | 88.0 | 88.0 |
| 1st Quartile | 87.0 | 87.7 | 87.8 | 84.2 | 84.1 |
| Minimum | 66.7 | 60.9 | 60.1 | 55.0 | 57.9 |

Table 4
Decennial Census and Continuous Measurement Estimates

| | | CM No Growth | | CM Positive Growth | | CM Negative Growth | |
|---|---|---|---|---|---|---|---|
| | Decennial Cen90 | 90-94 | 91-95 | 90-94 | 91-95 | 90-94 | 91-95 |
| Maximum | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.1 | 99.2 |
| 3rd Quartile | 95.1 | 94.0 | 94.3 | 93.8 | 94.5 | 93.0 | 92.7 |
| Median | 89.4 | 90.8 | 90.1 | 90.6 | 91.4 | 89.2 | 89.2 |
| 1st Quartile | 85.0 | 86.8 | 87.5 | 87.1 | 87.7 | 86.2 | 85.3 |
| Minimum | 69.3 | 74.2 | 77.3 | 60.0 | 60.0 | 55.0 | 54.2 |

Table 5
Comparison of Simulated Census and Continuous Measurement
No Growth

| | Cen90-Cen94 | Cen90 - CM90-94 | Cen94 - CM90-94 | Cen90 - Cen95 | Cen90 - CM91-95 | Cen95 - CM91-95 |
|---|---|---|---|---|---|---|
| Maximum | 4.2 | 15.8 | 12.4 | 5.8 | 15.8 | 9.0 |
| 3rd Quartile | 0.3 | 1.7 | 1.7 | 0.4 | 1.7 | 1.4 |
| Median | -0.1 | -1.0 | -0.8 | -0.1 | -1.0 | -1.1 |
| 1st Quartile | -0.4 | -2.4 | -2.5 | -0.4 | -2.4 | -2.3 |
| Minimum | -3.4 | -11.9 | -16.1 | -3.0 | -11.9 | -20.9 |

Table 6
Comparisons of Mean Employment for 25 Tracts with Change

| | Mean Employment Rate | | | |
|---|---|---|---|---|
| | Cen 94 | Cen 95 | CM 90-94 | CM 91-95 |
| Positive | 92.6 | 93.6 | 90.9 | 92.3 |
| Negative | 73.6 | 69.4 | 84.8 | 81.5 |

The Census employment rate for these 25 tracts was 88.4 percent.

Table 7
Summary of Employment Rate For 25 Positive Growth Tracts

| | Cen 90 - Cen 94 | Cen 90 - Cen 95 | Cen 90 - CM90-94 | Cen 90 - CM91-95 | Cen 94 - CM90-94 | Cen 95 - CM91-95 |
|---|---|---|---|---|---|---|
| Maximum | -0.6 | -0.7 | 4.5 | 3.5 | 8.6 | 9.8 |
| 3rd Quartile | -3.3 | -4.2 | -0.7 | -1.8 | 4.2 | 2.3 |
| Median | -4.3 | -5.4 | -1.6 | -3.6 | 1.0 | 0.8 |
| 1st Quartile | -5.2 | -6.3 | -5.4 | -6.2 | 0.0 | -0.1 |
| Minimum | -6.6 | -8.1 | -12.3 | -11.7 | -5.7 | -3.8 |

Table 8
Summary of Employment Rate for 25 Negative Growth Tracts

| | Cen 90 - Cen 94 | Cen 90 - Cen 95 | Cen 90 - CM90-94 | Cen 90 - CM91-95 | Cen 94 - CM90-94 | Cen 95 - CM91-95 |
|---|---|---|---|---|---|---|
| Maximum | 38.8 | 35.9 | 14.2 | 17.1 | 1.9 | 0.9 |
| 3rd Quartile | 15.3 | 19.8 | 7.5 | 12.1 | -5.8 | -5.8 |
| Median | 13.9 | 8.9 | 4.6 | 8.3 | -10.9 | -10.6 |
| 1st Quartile | 12.8 | 6.6 | 0.1 | 5.0 | -15.5 | -17.9 |
| Minimum | 4.2 | 6.1 | -10.4 | -5.3 | -43.2 | -40.4 |