

# Evaluation of small area population estimates produced by Housing Unit, Ratio-correlation, and Component Method II compared to 2000 Census counts

Nazrul Hoque

Department of Demography and Institute for Demographic and Socioeconomic Research, University of Texas at San Antonio  
Nazrul.hoque@utsa.edu

## Abstract

*This paper evaluates small area population estimates produced by the Housing Unit Method, Ratio-correlation Method, Component Method II, and the average of all three methods compared to the 2000 Census counts for the 254 counties and 1,279 places in Texas. This evaluation of three estimation methods shows the expected patterns of error by population size and population change. Of the three methods tested, no single method produced more accurate estimates than the average of two or three methods. The assessment of the accuracy of the place-level estimates show substantially higher levels of errors than those found for counties.*

**Keywords:** *small area population estimates, housing unit method, ratio-correlation method, component method.*

## Résumé

*Cet article évalue les estimations de populations vivant dans de petites régions produites par unité d'habitation, rapport de corrélation et méthode des composantes II ainsi que la moyenne des trois méthodes comparativement aux recensements de 2000 pour les 254 comtés et les 1 279 endroits au Texas. Cette évaluation des trois méthodes d'estimation démontre le modèle prévu d'erreurs par taille et changement de population. Des trois méthodes ayant fait l'objet de l'essai, aucune n'a produit des estimations plus précises que la moyenne de deux ou trois méthodes. L'évaluation de l'exactitude des estimations de lieu indique des taux d'erreurs significativement plus élevés que celles des comtés.*

**Mots-clés:** *estimations de populations vivant dans de petites régions, unité d'habitation, rapport de corrélation, méthode des composantes.*

## **Introduction**

The Population Census is one of the most important sources of demographic data. The aim of the census is establish the size, composition and distribution of the population through an accurate count of the number of people and households with their characteristics (Rowland 2003). However, in the U.S. the censuses are conducted once every ten years. In non-census years, population estimates provide demographic data regarding the size, distribution, and composition of the population by place of residence. Population estimates for the state, counties, and places are essential for planning different types of services, such as health care, schools, highways, water, and sewer. Planning for health services require accurate information on the number of persons by age (for services targeting children or elderly), sex, marital status, and place of residence. Population estimates provide a basis for allocating resources between areas in relation to population size in non-census years (Shryock and Siegal 1980; Bryan 2004). The federal government uses the Census Bureau's national and subnational population estimates for program evaluation, needs assessment and distribution of many billions of dollars to the states. The American Community Survey (ACS) and Current Population Survey (CPS) also use the Census Bureau's population estimates to control their surveys. Some state governments use State Data Center (SDC) population estimates to administer the state revenue sharing program. For example, the States of Florida, Arizona, Texas, and California allocate billions of dollars to county and local governments on the basis of their population estimates (Smith and Cody 2004; Murdock and Hoque 1995). Population estimates are also necessary to provide denominators to compute many types of rates and ratios, such as birth rates, death rates, labor force participation rates, school enrollment rates, dependency ratios, and sex ratios in non-census years. Population estimates play an important role in market analysis, public facility and environmental planning, and form a major basis for determining the present and future markets for a variety of goods, services, and other aspects of private-sector planning and marketing efforts (Murdock and Ellis 1991). These current figures are often critical elements in the analyses leading to decisions of whether or not to build a new school, fire station, library, hospital, a shopping mall, or highway (Siegel 2002). Thus, population estimates make an important contribution to the activities of governments, organizations, and businesses in non-census years.

Census data are normally used for reapportionment purposes. However, for some cases when census data were obsolete, estimated population figures were used for reapportionment. For example, in 1988 intercensal population estimates were used for redistricting Palm Beach County, Florida and in 1991 Los Angeles County, California, used population estimates for redistricting purposes (for detailed discussion, see Serow et al. 1997 and Bolton 1997).

Several methods have been developed to estimate population. However, population estimates are difficult to complete with accuracy for small areas because small areas can grow or decline rapidly, or may even undergo substantial changes in age, sex, and race/ethnicity, and other demographic characteristics, such as migration. All these factors increase the difficulty for making accurate estimates. As a result, it is essential that any ongoing program of population es-

timation periodically evaluate the results of past estimation against actual census counts of the population (Murdock and Ellis 1991). Only by assessing the accuracy of past efforts it is possible to know the nature of errors made and then take necessary steps to improve future estimates. This paper presents the results of the evaluation of population estimates produced by the Housing Unit Method (HUM), Ratio-correlation Method (RCM), and Component Method II (CMII) for 2000 against the 2000 Census counts for 254 counties and HUM and CMII for 1,279 places in Texas. The primary aim of this paper is evaluate the population estimates produced by the above three methods and Texas was selected for several reasons. First, Texas is one of the States that produce annual population estimates for 254 counties and more than 1,200 places on a regular basis. Second, the size of counties/places varies from less than 100 to more than 3.5 million people. Third, although Texas is one of the fastest growing States in the United States, population growth has not been distributed evenly throughout Texas. Some parts of the State have grown rapidly, some have grown slowly and others have declined. Migration also plays an important part in Texas population, more than 50 per cent of the 1990–2000 population change in Texas is due to net migration. All of the above criteria make the Texas population challenging to calculate and an excellent case study to evaluate estimates produced by different methods when compared with 2000 census counts.

In the following sections I will provide a brief description of each these three methods and then present the evaluation procedure and results of the analysis.

### Housing Unit Method

The Housing Unit Method (HUM) is regarded as one of the most reliable methods for making population estimates for small areas and is one of the easiest to apply. The Census Bureau uses the HUM to create population estimates for places and some state agencies, including ones in Florida and Texas, use it for population estimations. The logic of the HUM is that everyone lives in some type of household (Smith and Cody 2004). The HUM produces population estimates by taking into account the number of occupied households times the average number of persons per household. In terms of an equation it can be expressed as

$$P_t = (\text{OHU}_t \times \text{PPH}_t) + \text{GQ}_t$$

where  $P_t$  = total population at time of estimate;  $\text{OHU}_t$  = occupied housing units on the estimate date;  $\text{PPH}_t$  = household size or population per household on the estimates date; and  $\text{GQ}_t$  = the group quarters population at the time of estimate.

Each of the components of the HUM can be estimated using a variety of data sources, such as building permits and demolition data, or utility data based on active residential electric utility meters (for a detailed discussion, see Smith and Lewis 1980). The form of HUM used for this evaluation is:

$$P_t = (\text{OHU}_t + \text{BP}_t - \text{DU}_t - \text{VU}_t) \times \text{PPH}_t + \text{GQ}_t$$

where  $P_t$  = total population at time of estimate;  $\text{OHU}_t$  = occupied housing units counted in the most recent census (by type, e.g., single family, multifamily, mobile

home);  $BP_t$  = building permits issued by type between the most recent census and time of estimate (adjusted for time lag);  $DU_t$  = units reported demolished by type between the most recent census and time of estimate;  $PPH_t$  = household size or population per household by type at the time of estimates or most recent census;  $VU_t$  = Vacant Units by type between the most recent census and the time of estimate; and  $GQ_t$  = group quarters population at time of estimates.

Building permit data can be obtained from the U.S. Department of Commerce, which collects the data directly from counties and cities throughout the United States. The Texas State Data Center also collects residential building permit data, as well as data on vacancy rates and mobile homes, from counties and places in Texas. The Texas State Data Center collects additional data on vacancy rates and mobile homes, since the U.S. Department of Commerce does not collect these data.

There are specific problems associated with using Building Permit data for the HUM. Some counties and places neither issue building permits nor provide data to the U.S. Department of Commerce or Texas State Data Center (in 2005, 13 counties did not provide building permit data to the U.S. Department of Commerce and the Texas State Data Center). The U.S. Department of Commerce no longer collects data on demolition permits. Finally, most of the counties and places do not provide data on vacancy rates to the Texas State Data Center.

### Ratio-correlation Method

The Ratio-correlation Method (RCM) is a multiple regression-based technique which compares change in subarea to change occurring in a parent or super area. Such estimates are developed using the following multiple regression equation:

$$Y = B_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$$

where  $Y$  = the dependent variable to be estimated (e.g., population);  $B_0$  = the intercept to be estimated;  $\beta_j$  = the coefficient to be estimated;  $X_j$  = independent variables, such as births, deaths, voter registration, etc.; and  $e$  = error term.

The dependent and independent variables are expressed in the form of a ratio. For example, to obtain the estimate of population for a county in 2001, where the state population is known, the following equation could be applied:

$$\left( \frac{\text{County Pop., 2001/State Pop., 2001}}{\text{County Pop., 2000/State Pop., 2000}} \right) = \beta_0 + \beta_1 \left( \frac{\text{County Births, 2001/State Births, 2001}}{\text{County Births, 2000/State Births, 2000}} \right) + \beta_2 \left( \frac{\text{County Deaths, 2001/State Deaths, 2001}}{\text{County Deaths, 2000/State Deaths, 2000}} \right) + \beta_3 \left( \frac{\text{County School Enrollment, 2001/State School Enrollment, 2001}}{\text{County School Enrollment, 2000/State School Enrollment, 2000}} \right) + \dots$$

In the equation above, all of the indicator values are known except county population. In order to obtain the intercept and coefficients to use in solving the equation, estimates of the values must be obtained. This is done by solving the equation for past periods for which all the values are known. For example, the coefficients obtained by solving the equation for the past periods (e.g., 1990–2000)

can be used in the above formula for a 2001 estimate. Thus, 1990–2000 intercept and coefficients can be obtained by solving the equation for 1990–2000 period such as:

$$\left( \frac{\text{County Pop., 2000/State Pop., 2000}}{\text{County Pop., 1990/State Pop., 1990}} \right) = \beta_0 + \beta_1 \left( \frac{\text{County Births, 2000/State Births, 2000}}{\text{County Births, 1990/State Births, 1990}} \right) + \beta_2 \left( \frac{\text{County Deaths, 2000/State Deaths, 2000}}{\text{County Deaths, 1990/State Deaths, 1990}} \right) + \beta_3 \left( \frac{\text{County School Enrollment, 2000/State School Enrollment, 2000}}{\text{County School Enrollment, 1990/State School Enrollment, 1990}} \right) + \dots$$

The independent variables used in the RCM for 2000 population estimates for counties are births, deaths, school enrollment, voter registration, and vehicle registration.

## Component Method II

The Component Method II (CMII) depends on the use of three characteristics of population that directly determine population change: births, deaths and net migration. Thus, for any period, the population can be determined using the following equation

$$P_t = P_o + B - D + NM$$

where  $P_t$ =population for the estimate period;  $P_o$ =population at the base period;  $B$ =births between  $P_t$  and  $P_o$ ;  $D$ =deaths between  $P_t$  and  $P_o$ ;  $NM$ =net migration between  $P_t$  and  $P_o$ .

A population estimate is developed with CMII by updating the base population as enumerated in the most recent census, by adding to it the natural increase (births minus deaths) that occurred between the census and the estimate date, and estimating the amount of net migration in the area (Raymondo 1992). CMII typically takes direct account of natural increase through actual data on births and deaths while using symptomatic data for assessing net migration. CMII assumes that the rate of migration for school-age population can be used to assess the migration rate for the population 64 years of age and younger and that Medicare data can be used to estimate the migration rate for the population 65 years of age and older. There is some variation in terms of school-age population. Some prefer to use the elementary school enrollment for grades 1 to 8, while others prefer to use the elementary school enrollment for grades 2 to 8. For this exercise I use the elementary school enrollment data from grades 1 to 8 (both public and private schools).

CMII provides reliable results at the county level assuming that actual birth and death data are available at the county level. However, it is difficult to obtain births and death data at the place level (particularly small places). It is also difficult to collect private school enrollment data from private schools. Another concern is the implicit assumption that migration patterns of the elementary school-age population may be generalized for the population age 64 and under.

## **Factors limiting small area population estimates**

As mentioned in the previous section, different methods use different types of data to estimate population. The accuracy of the estimates depends on the accuracy or the quality of the data. If the input data is accurate, all of the methods are supposed to produce accurate estimates. However, it is hard to get reliable data particularly for the small areas, either for the small counties or small places.

Population estimates tend to be more accurate for large areas such as, an entire country or state rather than for subareas within the country or state because data from larger areas are generally more reliable compared with smaller areas. Some large-sized counties and places may as well be considered as large areas, such as counties with population of a million or more. In terms of population size, there are 22 counties and 758 places with population of less than 26,500 that may be considered as small areas, while there are 54 counties and 48 places with population of 50,000 or more that can be considered as large areas. As mentioned before, one of the simplest and accurate methods of population estimates is the HUM. With HUM, households can be estimated using building permit data. However, there are some problems collecting accurate data for small areas with HUM because some of the small counties and places do not issue building permits and are unable to provide any data on building permits. For example, there are more than 20 small counties in Texas, and more than half of the 1,279 places do not provide any data on building permits. For small areas, it also difficult to get any accurate data on demolition because county offices do not issue demolition permits as well.

With RCM it is better to include as many independent variables as possible. However, due to lack and quality of data, I include only five variables. I wanted to use driver's license data but I was not able to get those data for Texas. Same goes for CMII. I did not have private school enrollment data for small counties. However, in terms of availability of data, county level is a little better than place level. For example, I was able to get data for the RCM for county level estimates but did get similar data for the place level estimates. Therefore, I could not evaluate RCM at the place level.

## **Population change in counties and places in Texas in the 1990s**

Texas population has undergone dramatic change in the last 30 years. After rapid population growth during the 1970s and early 1980s, the rate of population growth fell to its lowest level in the years during the mid-1980s before beginning to show patterns of renewed growth during the late 1980s and the 1990s. Such a dramatic pattern of change in population makes it difficult to do accurate estimates of population. The population of Texas increased from 16,986,510 in 1990 to 20,851,820 in 2000. This increase of 3,865,310 persons or 22.8 per cent was the largest of any decade in state's history and was the second largest numerical increase of any state in the nation. Only California, which increased its population by 4,111,627, showed a larger numerical increase. However, the growth was not the same everywhere in the Texas. During the 1990s, 68 (26.8 per cent) of the state's 254 counties and 334 (26.1 per cent) of the 1,279 places lost population. During 1980s, 98 (38.6 per cent) of the 254 counties and 557 (46.1 per cent) of the 1,208 places lost population.

Overall, during the 1990s the State of Texas and its component areas showed not only extensive population growth but also dramatic changes in rates of growth from 3.5 per cent per year in the early 1980s to 0.5 per cent per year during the later part of the decade. The patterns of population change in many counties and places also changed from growth to decline or from decline to growth during the 1990s. There were 16 counties that changed direction from growth to decline and 46 from decline to growth. These changes make accurate estimates of the population for the state, counties and places within Texas very challenging.

## Methods for evaluation

Given the patterns and principles noted above, several widely used procedures were selected to evaluate the population estimates for Texas (Murdock and Ellis 1991; Siegel 2002). These methods generally rely on comparisons of values of error measures for the estimates, or projections being evaluated relative to expected patterns and relative to those for estimates or projections from other sources. The estimates were evaluated relative to the expected patterns of increased rates of error with decreased population size and increased rates of error with increased rates of population change. They were also evaluated relative to their tendency to underestimate or overestimate the population of different types of areas. Comparison of estimates to those from other sources assists in identifying which factors may be impacting the accuracy of estimates, because the assumptions can be compared to those used by other sources. Such a comparison often helps to determine which of the assumptions are increasing or decreasing the accuracy of the estimates.

Several error measures are used to assess the accuracy of estimates. The error of an estimate or projection is determined by subtracting the estimated or projected population value for an area from the census count (for purposes of this report, the 2000 census count) and dividing the difference by the census count. This proportion is then multiplied by 100 to produce a per cent rate of error.

Three error measures are used for this evaluation. The formulas for these measures are shown in Figure 1. They include the Mean Algebraic Per cent Error (MALPE), the Mean Absolute Per cent Error (MAPE), and the Mean Per cent Absolute Difference (MPAD). This later measure is also referred to as the weighted mean absolute per cent error.

The MALPE is simply the arithmetic average of the per cent errors for each area (county, place, etc.). This value is useful, but because positive and negative values cancel out one another in computation, it may provide somewhat misleading estimates of error. For example, if the population of one area was to be underestimated by 50 per cent and the population of the second area was to be overestimated by 50 per cent, the MALPE would be 0.0 per cent, suggesting that the estimates were perfect when in fact the two component estimates were quite inaccurate.

The MAPE is the mean of the absolute values of the errors, that is, ignoring the sign of the value. Given that the magnitude, rather than direction of the error is usually the major concern, the MAPE provides a more useful overall estimate of total error and is the most widely used measure of error in evaluations of popula-

$$(1) \text{ Mean Algebraic Percent Error (MALPE)} = \frac{\sum_{i=1}^n \left( \frac{\text{Estimated Value} - \text{Census Count}}{\text{Census Count}} \right)}{n} \times 100$$

$$(2) \text{ Mean Absolute Percent Error (MAPE)} = \frac{\sum_{i=1}^n \left| \frac{\text{Estimated Value} - \text{Census Count}}{\text{Census Count}} \right|}{n} \times 100$$

$$(3) \text{ Mean Percent Absolute Difference (MPAD)} = \frac{\sum_{i=1}^n |\text{Estimated Value} - \text{Census Count}|}{\sum_{i=1}^n \text{Census Count}} \times 100$$

Where: n = number of areas (counties, places)

Figure 1. Error measures used to evaluate the population estimates produced by Housing Unit, Ratio correlation and Component Method II.

tion estimates and projections. Both MALPE and MAPE, however, share a common weakness, in that errors for all places contribute equally to the overall error rate computed. Suppose the estimate for an area with 1 million people fell within two per cent of the actual count, and the estimate for an area of 100 people fell within 18 per cent. The MAPE for the two areas would be 10 per cent (2 plus 18 divided by 2), although the estimate for the area with most of the population was quite good. The problem is that neither MALPE nor MAPE take the size of the areas in the computation into account.

The MPAD or weighted mean absolute per cent takes the size of areas into account by weighting the values of areas proportionally to their size (population size of the area as a proportion of the sum of the populations of all the areas of interest). The MPAD is thus also widely used in evaluations of estimates.

The values of these three error measures are presented for each type of area (i.e., counties and/or places), for the areas grouped by population size in 2000 and by rate of population change from 1990 to 2000. Data are also shown for the number of overestimated and underestimated areas to indicate the extent to which the estimates tend to be biased either upward or downward. The number of areas estimated within certain ranges of error is also provided to indicate how many areas are estimated within specified levels of error. Finally, the errors in the estimates are compared to those from other sources.

### Evaluation procedures for county and place estimates for 2000

First, I evaluate the population estimates produced by each of these three methods. Second, I evaluate the estimates produced by the average of these methods both at the county and place-level. In this section, the results of the evaluation of these population estimates are presented for the counties and then for the places by each individual method and for an average of combined methods.

## Results of the evaluation of county-level estimates

Table 1 presents the three error measures for different estimation methods for counties in Texas. Error measures for each method are further detailed by providing population size and total for all counties. The results for the Housing Unit Method (HUM), on Panel I show an overall MALPE is 3.40, the MAPE is 8.94, and the MPAD is 5.01 for HUM (Table 1, Panel I). The MALPE, MAPE and MPAD for counties with population less than 1,000 are 20.17, 20.17 and 23.12, respectively. In general the highest MALPE, MAPE, and MPAD are for counties with population less than 5,000. The lowest MALPE, MAPE, and MPAD are for counties with population of 100,000 or more and are -2.12, 4.30, and 4.42, respectively.

For the Ratio-correlation Method (RCM), Panel II shows the overall MALPE is -4.13, and MAPE is 5.74, and the MPAD is 3.28. Again, the highest MALPE, MAPE and MPAD are for counties with small size of population. The lowest MALPE, MAPE, and MPAD are for counties with population of 100,000 or more and are -2.33, 2.80, and 2.91, respectively.

For the Component Method II (CMII), Panel III shows the overall MALPE is 1.84 per cent, a MAPE is 8.04 per cent, and the MPAD is 4.48. The highest MALPE, MAPE, and MPAD are for counties with population less than 2,500. The MALPE are negative for counties with population of 25,000 or more. The lowest MALPE is for counties with population of 10,000 or more and less than 25,000 while the MAPE and MPAD is lowest for the counties with a population 100,000 or more.

Averaging the CMII, RCM, and HUM produced an overall MALPE of -1.07, a MAPE of 5.18 per cent, and a MPAD of 3.73 per cent (Table 1, Panel IV). The MALPE, MAPE, and MPAD for counties with population of 1,000 or more and less than 2,500 are 3.98, 9.63, and 9.76, respectively. The MALPE, MAPE, and MPAD for counties with population of 2,500 or more and less than 5,000 are 0.33, 7.88, and 7.95, respectively. The MALPE, MAPE, and MPAD for counties with population of 5,000 or more and less than 10,000 are -0.11, 5.09, and 5.02, respectively. The MALPE, MAPE, and MPAD for counties with population of 10,000 or more and less than 25,000 are -1.62, 4.85, and 4.62, respectively. The MALPE, MAPE, and MPAD for counties with population of 25,000 or more and less than 50,000 are -3.06, 4.07, and 4.06, respectively. The MALPE, MAPE, and MPAD for counties with population of 100,000 or more are -2.05, 2.96, and 3.55, respectively.

In general, the data in Table 1 suggest the expected patterns, with error measures being larger for counties with smaller size of population and smaller for counties with larger size of populations. The data in Table 1 also suggest that using the average of the three methods is superior to the use of any single method of estimation.

Table 2 presents the same error measures by the rate of population change from 1990 to 2000. Panel I presents error measures for HUM, Panel II presents error measures for the RCM, Panel III presents error measures for CMII, and Panel IV presents error measures for the average of all three methods. As can be seen from Table 2, the MALPE was largest for counties with declining population for HUM and CMII. The MALPE was 27.06 for HUM, 11.66 for CMII, and -1.46 for RCM for counties with declining population of 10.0 per cent or more.

**Table 1. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 2000 census counts and population estimates produced by Housing Unit, Ratio correlation, and Component Method II by population size.**

Population Size	Number of Counties	HUM Method			Ratio-correlation Method		
		MALPE	MAPE	MPAD	MALPE	MAPE	MPAD
Under 1,000	7	20.17	20.17	23.12	0.79	13.17	11.60
1,000- 2,499	15	14.40	14.94	12.96	-5.68	7.67	7.53
2,500- 4,999	29	14.36	16.55	16.78	6.04	7.70	7.56
5,000- 9,999	38	6.74	10.58	10.57	5.17	6.70	6.99
10,000-24,999	69	4.04	8.61	8.18	4.86	6.38	6.25
25,000-49,999	42	-3.84	5.78	5.85	3.14	4.00	3.87
50,000-99,999	20	-3.54	6.33	6.20	2.51	3.47	3.29
100,000+	34	-2.12	4.30	4.42	2.33	2.80	2.91
All Counties	254	3.40	8.94	5.01	4.13	5.74	3.28
					Average of HUM, RCM, and CMII		
Population Size	Number of Counties	CMII			Average of HUM, RCM, and CMII		
		MALPE	MAPE	MPAD	MALPE	MAPE	MPAD
Under 1,000	7	15.97	20.61	19.24	4.07	9.15	6.32
1,000- 2,499	15	19.27	22.94	25.34	3.98	9.63	9.76
2,500- 4,999	29	7.07	14.56	14.08	0.33	7.88	7.95
5,000- 9,999	38	1.91	6.95	6.80	-0.11	5.09	5.02
10,000-24,999	69	0.19	5.60	5.57	-1.62	4.85	4.62
25,000-49,999	42	-2.06	5.45	5.42	-3.06	4.07	4.06
50,000-99,999	20	-3.46	5.67	5.65	-2.76	3.95	3.97
100,000+	34	-2.03	4.07	4.07	-2.05	2.96	3.55
All Counties	254	1.84	8.04	4.48	-1.07	5.18	3.73

\* For counties for which housing data were not available to implement the Housing Unit Method, only the Component Method II estimate was employed.

The MALPE, MAPE, and MPAD were lowest for counties with growth rates of 10.0 to 20.0 per cent, the MALPE was  $-0.43$  for HUM,  $-2.87$  for RCM, and  $0.40$  for CMII. As can be seen from Panel IV, the overall MALPE for counties with declining population of 10.0 per cent or more is  $9.35$  and MAPE of  $10.29$  and MPAD of  $10.43$ . The MALPE, MAPE, and MPAD for counties with increasing population of 40.0 to 50.0 per cent are  $-6.11$ ,  $6.73$ , and  $6.56$ , respectively. Overall, Table 2 suggests that the error measures are higher for the fastest declining and fastest growing countries. Table 2 also suggests that the error measures are smaller for moderate growing counties.

The ranges of error for the estimates are presented in Table 3. The data in Table 3 also provide general support for the relative accuracy of the methods with almost 75 per cent of all county estimates calculated using the CMII, almost 84 per cent of all county estimates calculated using the RCM, and 65 per cent of all county estimates calculated using the HUM being estimated within 10 per cent of the actual 2000 Census counts. Using the average of three methods (HUM, RCM, and CMII), 87.40 per cent of all counties are being estimated within 10 per cent of actual counts, and only 5 of the 254 counties have a 20 per cent or more error from the actual 2000 Census counts.

The results in Table 4 show that the RCM tended to be biased downward with 74.80 per cent of the counties being underestimated and 25.20 per cent of the counties being overestimated. With CMII 51.18 per cent of the counties were overestimated and 48.82 per cent underestimated. HUM produced estimates that tended to overestimate the population with 58.09 per cent of the counties being overestimated and 41.91 per cent of the counties being underestimated. An average of the three methods produced estimates that tend to underestimate the population of the counties with 62.20 per cent for counties being underestimated and 37.40 per cent being overestimated.

## Results of the evaluation of place-level estimates

Table 5 presents error measures for place estimates for HUM, CMII, and the average of HUM and CMII. Values are provided by population size and for the total of all places. For the HUM the overall mean per cent error was 2.19, the mean per cent absolute error was 13.55, and the mean per cent absolute difference was 7.85 (Table 5, Panel I).

**Table 3: Range of percent error for differences between 2010 census counts and estimated population produced by the Ratio–correlation Method for 254 counties in Texas**

Range of Error	Frequency	Per cent
00.0–04.9	94	37.01
05.0–09.9	80	31.50
10.0–14.9	40	15.75
15.0–19.9	18	7.09
20.0–24.9	8	3.15
25.0–29.9	5	1.97
30.0+	9	3.54

**Table 2. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 2000 census counts and estimates produced by Housing Unit, Ratio correlation, and Component Method II by per cent population change.**

Population Size 2000	Number of Counties	HUM Method			Ratio-correlation Method		
		MALPE	MAPE	MPAD	MALPE	MAPE	MPAD
< - 10.00	22	27.06	27.06	26.62	-1.46	5.66	4.68
-10.0 - 00.00	46	13.38	13.38	12.28	-3.27	5.75	3.98
00.00 - 09.99	61	4.26	4.59	3.49	-4.49	5.83	3.73
10.00 - 19.99	54	-0.43	3.52	3.35	-2.87	4.03	1.74
20.00 - 29.99	34	-3.48	5.31	3.70	-7.62	8.55	3.89
30.00 - 39.99	15	-9.92	11.09	9.02	-6.08	6.57	3.53
40.00 - 49.99	10	-9.92	11.19	7.14	-5.11	5.29	6.21
50.00 +	12	-12.10	12.43	10.34	-2.89	4.43	2.75
All Counties	254	3.40	8.94	5.01	-4.13	5.74	3.28

  

Population Size 2000	Number of Counties	CMII			Average of HUM, RCM, and CMII		
		MALPE	MAPE	MPAD	MALPE	MAPE	MPAD
< - 10.00	22	11.66	13.67	13.09	9.35	10.29	10.43
-10.0 - 00.00	46	10.77	12.45	7.13	2.66	5.17	4.28
00.00 - 09.99	61	-0.20	5.76	3.49	-1.82	3.76	2.53
10.00 - 19.99	54	0.40	6.36	4.09	-2.33	3.81	3.76
20.00 - 29.99	34	-3.53	6.18	3.65	-5.96	6.38	3.53
30.00 - 39.99	15	-6.24	7.98	6.73	-5.05	6.26	3.83
40.00 - 49.99	10	-7.38	8.31	8.72	-6.11	6.73	6.56
50.00 +	12	-0.59	5.08	3.97	-1.97	3.23	2.71
All Counties	254	1.84	8.04	4.48	-1.07	5.18	3.73

\* For counties for which housing data were not available to implement the Housing Unit Method only the Component Method II and Ratio-correlation estimate was employed.

**Table 4. Number and per cent of counties, Mean Algebraic Percent Error (MALPE) and Mean Percent Absolute Difference (MPAD) for counties with estimates above and below the 2000 census counts.**

	Housing Unit Method		Ratio-Component		Correlation Method II		Average of three methods	
	Below	Above	Below	Above	Below	Above	Below	Above
Number*	101	140	190	64	124	130	158	95
Per cent	41.91	58.09	74.80	25.20	48.82	51.18	62.20	37.40
MALPE	-6.61	10.63	-6.59	3.95	-6.35	9.65	-5.02	3.98
MPAD	5.10	4.73	3.20	0.97	4.61	3.99	5.49	2.52

\* One county's 2000 population was exactly estimated by the base estimates and so is not included in the comparisons shown here.

**Table 5. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 2000 census counts and estimates produced by Housing Unit and Component Methods II for Places, by population size.**

	Population size (2000)	Number of places	MALPE	MAPE	MPAD
Panel I: HUM					
	Under 1,000	406	7.61	18.57	15.34
	1,000–2,499	320	2.53	11.42	11.40
	2,500–4,999	186	-2.62	13.49	14.10
	5,000–9,999	132	-0.49	11.66	11.83
	10,000–24,999	104	-3.02	9.30	8.84
	25,000–49,999	50	-2.92	7.01	6.86
	50,000–99,999	24	-4.00	7.92	7.37
	100,000+	24	-2.47	5.40	6.72
	All Places	1,246	2.19	13.55	7.85
Panel II: CMII					
	Under 1,000	425	19.60	29.08	24.15
	1,000–2,499	324	6.81	17.71	17.32
	2,500–4,999	187	-0.21	18.13	18.62
	5,000–9,999	132	1.13	16.14	16.35
	10,000–24,999	104	-0.49	15.57	15.17
	25,000–49,999	50	-7.61	16.94	16.88
	50,000–99,999	24	-1.67	11.47	10.47
	100,000+	24	-0.73	6.67	7.09
	All Places	1,270	8.00	20.88	10.80
Panel III: Average of HUM and CMII					
	Under 1,000	432	10.86	21.01	17.51
	1,000–2,499	326	4.92	13.39	13.23
	2,500–4,999	187	-0.82	12.80	13.35
	5,000–9,999	132	2.47	11.93	12.00
	10,000–24,999	104	1.53	9.57	9.35
	25,000–49,999	50	2.00	6.80	6.51
	50,000–99,999	24	2.18	6.22	5.64
	100,000+	24	-1.86	4.82	4.46
	All Places	1,279	5.26	14.86	6.52

For the CMII, the overall mean algebraic per cent error was 8.00, the mean per cent absolute error was 20.88, and the mean per cent absolute difference was 10.80 (Table 5, Panel II). The average of HUM and CMII produced the MALPE of 5.26, MAPE of 14.86 and MPAD of 6.52.

The overall levels of error shown on this table are within the higher end of the expected 10 to 15 per cent error range, when estimates for a relatively large number of places with small population size are involved; however, they are higher than is desirable. For example, the MPAD for the average of the HUM and the CMII for places with population of under 10,000 varies from 12.0 to 17.51, while for places with population over 10,000, it varies from 4.46 to 9.35 (Table 5, Panel III).

Table 6 presents the same error measures by the rate of population change from 1990 to 2000. Differences in population growth rates had the same impact on errors for places as for counties (i.e., the fastest declining and growing places

**Table 6. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 2000 Census Counts and Estimates Produced by Housing Unit and Component Methods II for Places by Population Change, 1990–2000.**

	Per cent population change (1990–2000)	Number of counties	MALPE	MAPE	MPAD
Panel I: HUM					
	<-10.00	109	40.10	40.10	31.87
	-10.00–00.00	214	12.99	13.00	8.43
	00.00–09.99	284	3.97	5.98	4.66
	10.00–19.99	221	-1.42	5.45	6.10
	20.00–29.99	126	-3.33	7.36	6.35
	30.00–30.99	83	-5.97	12.21	7.56
	40.00–49.99	53	-12.19	12.96	11.94
	50.00+	156	-23.50	26.89	17.04
	All Places	1,246	2.19	13.55	7.85
Panel II: CMII					
	<-10.00	114	49.98	51.07	28.33
	-10.00–00.00	218	15.13	18.09	11.52
	00.00–09.99	290	9.30	15.79	7.83
	10.00–19.99	225	7.95	16.45	10.66
	20.00–29.99	130	6.92	15.33	9.35
	30.00–30.99	84	-1.19	14.12	10.72
	40.00–49.99	53	-6.14	16.26	4.93
	50.00+	156	-24.36	28.42	22.03
	All Places	1,270	8.00	20.88	10.80
Panel III: Average of HUM and CMII					
	<-10.00	114	37.91	38.09	24.47
	-10.00–00.00	218	11.68	13.31	7.65
	00.00–09.99	290	6.66	10.17	5.84
	10.00–19.99	227	3.77	9.68	5.43
	20.00–29.99	133	3.74	11.31	4.73
	30.00–30.99	83	-3.09	11.14	8.85
	40.00–49.99	53	-3.59	9.94	6.15
	50.00+	161	-18.48	22.79	11.89
	All Places	1,279	5.26	14.86	6.52

**Table 7. Range of per cent error for differences between 2000 census counts and estimates produced by Housing Unit and Component Methods II, and average of Housing Unit and Component Method II for places.**

Range of Error	Housing Unit Method		Component Method II		Average of two methods	
	Number	Per cent	Number	Per cent	Number	Per cent
00.0–04.9	400	32.10	241	8.98	373	29.16
05.0–09.9	314	25.20	231	18.19	286	22.36
10.0–14.9	176	14.13	196	15.43	188	14.70
15.0–19.9	98	7.87	142	11.18	136	10.63
20.0–24.9	80	6.42	97	7.64	92	7.19
25.0–29.9	41	3.29	84	6.61	55	4.30
30.0+	137	11.00	279	21.97	149	11.65

\* One county's 2000 population was exactly estimated by the base estimates and so is not included in the comparisons shown here.

have higher error rates than slowest declining or growing places), but the patterns were more clearly visible for the places than the counties. As can be seen from Table 6, there was a tendency to overestimate the fastest declining places and underestimate the fastest growing places. The mean per cent error was 40.10 for HUM and 49.98 for CMII for counties with a declining population of 10.0 per cent or more. The mean per cent absolute difference was 31.87 for HUM and 28.33 for CMII for counties with declining population of 10.0 per cent or more. The MALPE, MAPE, and MPAD were higher for the fastest growing counties as well. This was evident from both HUM and CMII. Again, the average of HUM and CMII produced better results than each of the individual methods (Panel III).

Table 7 shows results of the evaluation of estimates in terms of range of errors. Overall (average of the HUM and the CMII) 51.52 per cent of the places (compared to 76 per cent of the counties) were estimated within 10 per cent, 25.33 per cent within 10 to 20 per cent, and another 23.14 per cent more than 20 per cent of 2000 population count. For the HUM 57.30 per cent of the places were estimated within 10.00 per cent while for CMII only 27.17 per cent of the places were estimated within the 10.00 per cent of the actual census count.

The results in Table 8 point to a tendency for the population of places to be underestimated when averaging the HUM and CMII. This underestimation may be due to the undercount in the 1990 census, resulting in a base population that was lower than it should have been. Murdock and Hoque (1995) evaluated the im-

**Table 8. Number and per cent of places, Mean Algebraic Percent Error (MALPE) and Mean Percent Absolute Difference (MPAD) for places with estimates above and below the 2000 census counts.**

	Housing Unit Method		Component Method II		Average of two methods	
	Below	Above	Below	Above	Below	Above
Number*	546	698	515	754	794	482
Per cent	43.89	56.11	40.58	59.42	62.08	37.69
MALPE	-12.95	14.05	-15.89	24.32	16.21	-12.74
MPAD	8.41	6.70	9.70	12.01	7.26	6.64

\* 2000 population was exactly estimated by the base estimates and so is not included in the comparisons shown here.

fact of undercount on the accuracy of small-area population estimates and found significant differences in using adjusted and non-adjusted population for 1990 (for detailed discussion see Murdock and Hoque 1995). Another possible explanation is that there might have been an over estimation of population in small places in 2000. To my knowledge, there is no study that has looked into the adjustment of the 2000 census.

The data in Tables 6 through 8 also suggest that in nearly all cases, the average of the HUM and CMII improves the accuracy of place level estimates. The results of the HUM averaged with CMII appear to lead to a reduction in the error of estimate for place level estimates.

## **Discussion**

Demographers use different methods to produce population estimates and often there is discussion regarding one method producing better estimates than others. The primary objective of this research was to evaluate the population estimates produced by Housing Unit, Ratio-correlation and Component Method II for 254 counties and almost 1,300 places in Texas compared with census counts for 2000 to evaluate the accuracy of these methods for small areas.

At the county-level the evaluation included all three methods; however, place-level estimates were only evaluated for HUM and CMII. At the place-level, I could not evaluate ratio-correlation due to the lack of available data. The RCM uses birth, death, school enrollment, voter registration and vehicle registration data and none of these data are available at the place-level. Although some of the counties are small in terms of population size, I was able to acquire most the data for RCM for county-level estimates. For CMII, birth and death data were not available for almost 40 per cent of the places. In that case, I estimated and distributed the birth and death data from the remainder of the county. At the county level, the RCM provided better estimates than HUM and CMII, both by the size of the population and also by the population change between 1990 and 2000. The MALPE, MAPE and MPAD are much higher for small counties for HUM and CMII. The average of the three methods was consistently better at the county levels.

At the place-levels the HUM provided more accurate estimates for small areas compared with CMII. For areas with large population size, CMII proved to provide the most reasonable estimates. It should be noted that the accuracy of the estimates is not solely dependent on the method utilized but may also be due to the quality of data. As mentioned before, almost 40 per cent of the places do not have reported data for births and deaths. Similarly school enrollment data are not available for all areas.

## **Conclusion**

Accurate estimates are difficult for small areas and for areas showing inconsistency in the direction of change during the estimation period. Many places in Texas have experienced rapid population change making it challenging to do accurate estimates of population. For this analysis, population estimates for counties calculated using the HUM, the RCM, and the CMII separately as well as estimates

calculated using an average of the three methods were evaluated against the actual 2000 census counts. For the 1279 places in Texas, population estimates produced by HUM, CMII, and an average of Housing Unit and CMII were evaluated against the actual 2000 census counts.

Three error measures are used to assess the accuracy of population estimates produced by HUM, RCM and CMII compared to the 2000 census counts. They are the MALPE, the MAPE, and the MPAD. At the county level, CMII provided more accurate estimates than the HUM and RCM. The RCM proved to be more reliable for small areas and the areas with rapid population growth or decline. At the place-level, HUM estimates did better than those produced using CMII. However, as mentioned earlier, the HUM could not be used for all counties and places due to lack of data. Some smaller counties and places do not issue building permits nor provide data to the U.S. Department of Commerce or Texas State Data Center. Recent data on vacancy rates is similarly unavailable.

The evaluation of the population estimates produced by HUM, RCM and CMII, and the average of three methods presented here suggest that the average of two or three methods performed better than a single method. The estimates also show the expected patterns of error by population size and population change. That is, population estimates are more accurate for large counties and places than small counties and places. Of the several methods tested, no single method produced more accurate estimates than the averages of two or three methods. The assessment of the accuracy of the place-level estimates showed substantially higher levels of errors than the levels found for county-level estimates. This higher rate of error for places results due to the large number of places with small population size and the inconsistency in the direction of change during the estimation period. For future research, one way to improve the accuracy of place-level estimates is to add RCM and to collect symptomatic data for RCM. For county estimates, I would recommend adding the Administrative Record Method provided that input data are available to do so and compare the results with the census counts.

## Acknowledgements

The author would like to thank Beverly Pecotte, the editor and reviewers for their comments and suggestions. Special appreciation is due to Lisa Espinoza who corrected repeated drafts of this work.

## References

- Bolton, N. 1997. The use of intercensal population estimates and projections in the court-mandated reapportionment of Los Angeles County, in *Demographics: A Casebook for Business and Government*, edited by H.J. Kinter, T.W. Merrick, P.A. Morrison, and P.R. Voss. Rand Publication, pp. 55–69.
- Bryan, T. 2004. Population estimates, in *The Methods and Materials of Demography*, edited by J. Siegel and D.A. Swanson. 2nd edn. San Diego, CA: Elsevier Academic Press, pp. 523–60.
- Hoque, N. 2008. An evaluation of population estimates for counties and places in Texas for 2000, in *Applied Demography in the 21<sup>st</sup> Century*, edited by S.H. Murdock and D.A. Swanson. Dordrecht (Netherlands): Springer Science and Business Media B.V., pp. 125–48.

- Murdock, S.H. and D. Ellis. 1991. *Applied Demography: An Introduction to Basic Concepts, Methods, and Data*. Boulder, CO: Westview Press, Inc.
- Murdock, S.H. and M.N. Hoque. 1992. *Evaluation of Texas Population Estimates and Projections Program's Population Estimates and Projections for 1990*. Departmental Miscellaneous Report No. 92–1. College Station, TX: Department of Rural Sociology, Texas A&M University.
- . 1995. The effect of undercount on the accuracy of small-area population estimates: Implications for the use of administrative data for improving population enumeration. *Population Research and Policy Review* 14(2):251–71.
- Murdock, S.H., C. Kelley, J. Jordan, B. Pecotte, and A. Luedke . 2006. *Demographics: A Guide to Methods and Sources of Data for Demographic Analysis in the Media, Business, and Government*. Boulder, CO: Paradigm Publishers.
- Raymondo, J.C. 1992. *Population Estimation and Projection: Methods for Marketing, Demographic, and Planning Personnel*. New York: Quorum Books.
- Rowland, D.T. 2003. *Demographic Methods and Concepts*. New York: Oxford University Press.
- Serow, W., J.W. Terrie, B. Waller, and R.W. Wichmann. 1997. The use of intercensal population estimates in political redistricting, in *Demographics: A Casebook for Business and Government*, edited by H.J. Kinter, T.W. Merrick, P.A. Morrison, and P.R. Voss. Rand Publication, pp. 33–54.
- Shryock, H. and J. Siegel. 1980. *The Methods and Materials of Demography*. Washington, DC: Government Printing Office.
- Siegel, J.S. 2002. *Applied Demography: Applications to Business, Government, Law and Public Policy*. San Diego: Academic Press.
- Smith, S.K. and B. Lewis. 1980. Some new techniques for applying the Housing Unit Method of local population estimation. *Demography* 17:323–39.
- Smith, S.K. and S. Cody. 2004. An evaluation of population estimates in Florida: April 1, 2000S. *Population Research and Policy Review* 23:1–24.
- Smith, S.K. J. Tayman, and D. Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.
- U.S. Census Bureau. 2001. Methodology for Estimates of State and County Total Population. Available online at <http://www.census.gov/popest/topics/methodology/state.php>.