Open Access

An Evaluation of Small Area Population Estimates Produced by Component Method II, Ratio-correlation and Housing Unit Methods for 1990

Md. Nazrul Hoque*

Department of Demography and Organization Studies and Institute for Demographic and Socioeconomic Research, University of Texas at San Antonio, 501 West Durango, San Antonio, Texas 78207-4415, USA

Abstract: Estimated population is one of the most widely used products of demographic analyses. Population estimates are difficult to complete with accuracy for small areas because small areas can grow or decline rapidly, can change directions from growth to decline or from decline to growth, or undergo substantial changes in age, sex, race/ethnicity and other demographic characteristics. As a result, it is essential any ongoing program of population estimation periodically evaluate the results of past estimates against actual census counts for the target population. Only by assessing the accuracy of past efforts, is it possible to know the nature of errors made and to take steps to improve future estimates. In this paper I present the results of the evaluation of the 1990 population estimates produced by Component Method II, Ratio-correlation, and Housing Unit Method compared to the 1990 Census counts for 254 counties and 1,210 places in Texas.

Three error measures are used to assess the accuracy of population estimates of Texas for 1990. They are the Mean Algebraic Percent Error (MALPE), the Mean Absolute Percent Error (MAPE), and the Mean Percent Absolute Difference (MPAD). The evaluation of population estimates presented here suggests that the estimates are generally adequate and show levels of error that, when compared to the 1990 Census counts, are within generally accepted ranges. They also show the expected patterns by population size and population change. Of the several methods tested, no single one 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: Population Estimates, Estimation Error, Evaluation, Component Method II, Ratio-correlation Method, Housing Unit Method.

I. INTRODUCTION

Population estimates and projections are among the most widely used products of demographic analyses [1-5]. 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 requires accurate information on the number of persons by age (for services targeting children or elderly), sex, marital status, distribution in different areas, and place of residence. Population estimates provide a basis for allocation of resources between areas in relation to population size. The federal government uses the Census Bureau's national and subnational population estimates in calculating the distribution of many billions of dollars in the form of block grants each year to states and jurisdictions within them. Some state governments use State Data Center (SDC) population estimates to administer the state revenue sharing program. For example, the State of Florida uses it's population estimates to distribute more than $1.5 billion each year to local governments [6]. 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 [4]. They 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 [7]. Thus, population estimates make an important contribution to the activities of governments, organizations, and businesses in non-census years. Intercensal estimates provide data for years between two existing censuses, such as 1980 and 1990, while postcensal estimates provide data for years since the last census [8].

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 [9] and [10]).

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. All these factors increase the difficulty for making accurate

*Address correspondence to this author at the Department of Demography and Organization Studies and Institute for Demographic and Socioeconomic Research, University of Texas at San Antonio, 501 West Durango, San Antonio, Texas 78207-4415, USA; Tel: 210-458-6533; Fax: 210-458-6541; E-mail: Nazrul.hoque@utsa.edu

estimates. As a result, it is essential that any ongoing program of population estimation periodically evaluate the results of past estimation against actual census counts of the population [3-5]. Only by assessing the accuracy of past efforts is it possible to know the nature of errors made and to take steps to improve future estimates.

II. BACKGROUND

Texas is one of the few states that have produced population estimates for counties and places since the mid-1980s. The Texas Population Estimates and Projections Program's population estimates for counties were calculated using an average of Component Method II and Ratio-correlation Method and for places using the Component Method II only. For this paper, population estimates for counties were calculated using the Component Method II, the Ratio-correlation Method, and the Housing Unit Method separately as well as estimates calculated using an average of the three methods. Population estimates for places were calculated using the Component Method II and the Housing Unit Method separately as well as an average of two methods and were evaluated against the actual 2000 census counts. Due to space limitation, estimated population produced by different methods for counties and places are not given here but are available from the author upon request. This paper will also recognize the basic principles of population estimates which are discussed following the section on specific methods. In addition, it is important to understand the changes in Texas population during the 1980s' since the 1990 estimates, I am evaluating were based on the 1980 Census Count.

III. ESTIMATING METHODS

In this paper I evaluate the 1990 population estimates produced by Component Method II, Ratio-correlation Method and Housing Unit Methods to the 1990 Census counts for counties and places in Texas. First I would like to give a brief description of each of these three methods then I will talk about the evaluation procedure and results of the analysis.

III.A. Component Method II

Component Method II 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 Component Method II 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 [11]. Component Method II typically takes direct account of natural increase through actual data on births and deaths while using symptomatic data for assessing net migration. Component Method II assumes that the rate of migration of 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. The Texas State Data Center's Population Estimates Program uses the elementary school enrollment for grades 1 to 8 (both public and private schools).

Component Method II 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 birth and death data at the place level (particularly small places). It is also difficult to collect private school enrollment for grades 1 to 8. 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.

III.B. Ratio-correlation Method

The Ratio-correlation Method is a multiple regression-based technique which compares change in one areal unit to change occurring in a parent area. Such estimates are developed using the following multiple regression equation:

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots \beta_n X_n + e$$

where: Y = the dependent variable to be estimated (e.g., population)

β_o = the intercept to be estimated

β_i = the coefficient to be estimated

X_i = independent variables, such as births, deaths, voter registration, etc.

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 1991, where the state population is known, the following equation could be applied:

$$\left(\frac{\text{County Pop., 1991/State Pop., 1991}}{\text{County Pop., 1990/State Pop., 1990}}\right) = \beta_o + \beta_1\left(\frac{\text{County Births, 1991/State Births, 1991}}{\text{County Births, 1990/State Births, 1990}}\right) +$$

$$\beta_2\left(\frac{\text{County Deaths, 1991/State Deaths, 1991}}{\text{County Deaths, 1990/State Deaths, 1990}}\right) +$$

$$\beta_3\left(\frac{\text{County School Enrollment, 1991/State School Enrollment, 1991}}{\text{County School Enrollment, 1990/State School Enrollment, 1990}}\right) + \ldots.$$

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., 1980-1990) can be used in the above formula for a 1991 estimate. Thus, 1980-1990 intercept and coefficients can be obtained by solving the equation for 1980-1990 period such as:

$$\left(\frac{\text{County Pop., 1990/State Pop., 1990}}{\text{County Pop., 1980/State Pop., 1980}}\right) = \beta_o + \beta_1\left(\frac{\text{County Births, 1990/State Births, 1990}}{\text{County Births, 1980/State Births, 1980}}\right) +$$

$$\beta_2\left(\frac{\text{County Deaths, 1990/State Deaths, 1990}}{\text{County Deaths, 1980/State Deaths, 1980}}\right) +$$

$$\beta_3\left(\frac{\text{County School Enrollment, 1990/State School Enrollment, 1990}}{\text{County School Enrollment, 1980/State School Enrollment, 1980}}\right) + \ldots.$$

The dependent variables used in the Ratio-correlation Method for population estimates for counties in Texas are births, deaths, school enrollment, voter registration, and vehicle registration.

III.C. Housing Unit Method

The Housing Unit Method is regarded as one of the most reliable methods for making population estimates for small areas and is one of the easiest to apply [12]. The Census Bureau uses the Housing Unit Method for population estimates for places [13], and some state agencies including ones in Florida and Texas use it for population estimations. The logic of the Housing Unit Method is that every one lives in some type of household [6]. The Housing Unit Method produces population estimates by taking into account the number of occupied housing units times the average number of persons per household. In terms of an equation it can be expressed as:

$$P_t = (OHU_t \text{ x } PPH_t) + GQ_t$$

where: P_t = total population at time of estimate

OHU_t = occupied housing units on the estimate date

PPH_t = household size or population per household on the estimates date

GQ_t = the group quarters population at the time of estimate

Each of the components of the Housing Unit Method 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 [12]). The form of Housing Unit Method used in Texas' population estimate is:

$$P_t = (OHU_t + BP_t - DU_t - VU_t) \text{ x } PPH_t + GQ_t$$

where: P_t = total population at time of estimate

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 the time of estimate (adjusted for time lag)

DU_t = units reported demolished by type between the most recent census and the time of estimate

VU_t = Vacant units reported by type at the time of estimate

PPH_t = household size or population per household on the estimate date

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, from the counties and cities in Texas. In addition, the Texas State Data Center collects data on vacancy rates and mobile homes, since the U.S. Department of Commerce does not collects data for either.

There are some problems associated with using building permit data for the Housing Unit Method. 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.

Other potential sources of housing unit data are problematic as well. Since utility data or other residence information is not collected by any control source in Texas, these data would be expensive to acquire and verify.

IV. PRINCIPLES OF POPULATION ESTIMATION AND PROJECTION

The history of population estimates and projections suggests certain basic findings from past analyses that merit recognition in an evaluation of any estimation or projection. These basic findings or principles, as outlined by [4, 11, 14, 15, and 16] show that no single method has been found to consistently produce more accurate estimates and projections than any other method, and that population estimates and projections are generally more accurate:

1. For geographic areas with larger populations than for those with smaller populations. For example, population estimates tend to be more accurate for an entire country or state rather than for subareas within the country or state. Data from larger areas are generally reliable compare with smaller areas. Natural disasters in local areas may have a major impact on the popula-

tion in the immediate area but will have virtually no impact on the national population. For example, the city of Galveston in Texas lost most of its population due to Hurricane Ike but the State of Texas was not impacted because the dispersed populations were absorbed by neighboring cities in Texas. Likewise, the city of New Orleans lost most of its population immediately following Hurricane Katrina and the State of Louisiana was impacted to some extent but the national (U.S.) population was not influenced. All of the dispersed populations were absorbed by the neighboring states. Similarly, a military base closing would have almost no impact on the population at a state or national level but it would definitely reduce the population of a small county or place significantly. Likewise, a newly established industrial base may increase the population significantly for a small county or place but not for the population at state or national level.

2. For total populations rather than for population subgroups because estimates of such characteristics involve additional assumptions that may prove to be in error. Population estimates that bear a greater level of detail such as age, sex, race and ethnicity thus produce greater levels of error.
3. For short rather than long periods of time past the reference date for the base data used in the estimates (i.e. last census.) Population estimates that bear a longer duration between the base population and the estimation date, may result in a greater degree of error. For example, estimates for 1991 may be more accurate than estimates for 1998 or 1999, based on 1990 Census population.
4. For areas that show consistency in the direction of population change during the estimation period compared to the period from which the base data are derived. If the direction change varies from growth to decline or from decline to growth then accurate estimates of the population are difficult.
5. For areas that experience slow rather than rapid change. A period of dramatic population change (growth or decline) makes population estimates more difficult.
6. If completed with data that directly determine population change (such as data on births, deaths, and migration) rather than with data that employ indirect or symptomatic indicators of population change (such as voter registration, vehicle registration, and school enrollment).
7. No single method of population estimation will always be the best choice. The average of methods may be employed as a basis for improving the accuracy of population estimates. Using multiple techniques will provide a means of checking the validity of the estimates since similar results obtained from a variety of different methods tend to suggest the overall accuracy of the result.

In sum, an estimate is likely to be most accurate if it is based on direct birth, death, and migration data and is for the total population of a large area showing slow change that is in the same direction and form as the change in the recent past. In general, the greater an area's departure from these conditions, the greater the error in estimates.

V. POPULATION CHANGE IN TEXAS IN THE 1980s

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 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 trends make it difficult to accurately estimate the population. The population of Texas increased from 14,229,191 in 1980 to 16,986,510 in 1990. This was an increase of 2,757,319 persons or 19.4 percent. Only California and Florida had larger numerical increases than Texas, 6,092,119 and 3,191,602, respectively. However, the growth was not the same everywhere in the Texas. During the 1970s, 44 (17.3 percent) of the state's 254 counties and only 20 percent of the 1,279 places lost population. During 1980s, 98 (38.6 percent) of the 254 counties and 557 (46.1 percent) of the 1,208 places lost population.

Overall, during the 1980s 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 percent per year in the early 1980s to 0.5 percent per year during the later part of the decade. The patterns of population change in many counties and places changed from growth to decline or from decline to growth during the 1980s. These changes make accurate estimates of the population for the state and for counties and places within Texas very challenging.

VI. METHODS FOR EVALUATION

Given the patterns and principles noted above, several widely used procedures were selected to evaluate the population estimates for Texas [3,7]. These methods generally rely on comparisons of values of error measures for the estimates being evaluated relative to expected patterns and relative to those for estimates 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 is determined by subtracting the estimated population value for an area from the census count (for purposes of this paper, the 1990 census count) and dividing the difference by the census count. This proportion is then multiplied by 100 to produce a percentage rate of error.

Three error measures are commonly used in such assessments. The formulas for these measures are shown in

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

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

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

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

Fig. (1). Error Measures Used to Evaluate the Population Estimates Produced by Component Method II, Ratio-correlation and Housing Unit Methods.

Fig. (**1**). They include the Mean Algebraic Percent Error (MALPE), the Mean Absolute Percent Error (MAPE), and the Mean Percent Absolute Difference (MPAD). This later measure is also referred to as the weighted mean absolute percent error. The Mean Algebraic Percent Error (MALPE) is also known as Mean Percent Error (MPE).

The Mean Algebraic Percent Error (MALPE) is simply the arithmetic average of the percent 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 populations of one area were to be underestimated by 50 percent and the populations of the second area were to be overestimated by 50 percent, the MALPE would be 0.0 percent, suggesting that the estimates were perfect when in fact the two component estimates were quite inaccurate.

The Mean Absolute Percent Error (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 population 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 percent of the actual count, and the estimate for an area of 100 people fell within 18 percent. The MAPE for the two areas would be 10 percent (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 Mean Percent Absolute Difference (MPAD) or weighted mean absolute percent 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 population estimates and projections.

For this paper, 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 1990 and for rates of population change from 1980 to 1990. 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 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.

VII. RESULTS

1990 population estimates produced by Component Method II, Ratio-correlation Method, and Housing Unit Method are evaluated against the 1990 actual census count. The results of the evaluation of these population estimates are presented first for counties and then for places by each individual method and for an average of combined methods.

VII. A. The Results of the Evaluation of County-Level Estimates

Table **1** presents the three error measures for different estimation methods for counties in Texas. The results in this table show an overall Mean Algebraic Percent Error of -1.27 percent, a Mean Absolute Percent Error of 6.50 percent, and a Mean Percent Absolute Difference of 3.65 for Component Method II (Table **1**, Panel I). The MALPE, MAPE, and MPAD for counties with populations less than 1,000 are 2.98, 19.85, and 13.99, respectively. The highest MALPE, MAPE, and MPAD are for counties with population less than 2,500. There are 23 counties in Texas with population less than 2,500. The lowest MALPE is for counties with popu-

Table 1. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 1990 Census Counts and Estimated Population Produced by Component Method II, Ratio-correlation, and Housing Unit Methods for Counties in Texas

Method	Population Size 1990	Number of Counties	MALPE	MAPE	MPAD
Panel I: Component Method II					
	< 1,000	5	2.98	19.85	13.99
	1,000- 2,499	18	4.27	13.37	14.41
	2,500- 4,999	29	-6.81	8.54	8.16
	5,000- 9,999	43	-1.00	6.36	6.46
	10,000-24,999	70	-2.54	5.54	5.52
	25,000-49,999	38	0.35	4.82	4.64
	50,000-99,999	23	1.18	4.14	4.05
	100,000+	28	-1.28	4.43	3.18
	All Counties	254	-1.27	6.50	3.65
Panel II: Ratio-correlation Method					
	< 1,000	5	-1.81	17.74	10.72
	1,000- 2,499	18	-1.35	4.70	4.65
	2,500- 4,999	29	-4.59	7.28	6.95
	5,000- 9,999	43	-2.01	4.05	4.02
	10,000-24,999	70	-1.03	5.36	5.20
	25,000-49,999	38	-0.59	3.15	3.25
	50,000-99,999	23	1.47	3.51	3.40
	100,000+	28	-0.16	2.92	2.08
	All Counties	254	-1.25	4.79	2.50
Panel III: Housing Unit Method*					
	1,000- 2,499	6	14.60	14.60	13.49
	2,500- 4,999	21	10.41	14.47	13.81
	5,000- 9,999	40	7.24	12.83	12.24
	10,000-24,999	65	-0.16	7.62	7.36
	25,000-49,999	38	-6.39	9.77	9.84
	50,000-99,999	23	-6.51	10.86	10.87
	100,000+	28	-1.95	9.26	7.77
	All Counties	221	0.63	10.32	8.31

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

Table 1. contd….

Method	Population Size 1990	Number of Counties	MALPE	MAPE	MPAD
Panel IV: Average of Component Method II, Ratio-correlation, and Housing Unit Methods*					
	< 1,000	5	-1.93	15.97	9.89
	1,000- 2,499	18	2.44	7.68	7.88
	2,500- 4,999	29	1.54	8.73	8.31
	5,000- 9,999	43	4.27	7.48	7.25
	10,000-24,999	70	0.37	5.00	4.97
	25,000-49,999	38	-1.78	4.75	4.82
	50,000-99,999	23	-1.88	5.06	5.09
	100,000+	28	-0.58	4.86	4.22
	All Counties	254	0.64	6.21	4.49

* For counties of <1,000 for which housing data were not available to implement the Housing Unit Method, only the Component Method II estimate was employed.

ulation greater than 25,000 and less than 50,000. There are 38 counties in this category. The lowest MPAD (3.18 percent) is for counties with a population 100,000 or more. There are 28 counties in this category. Overall smallest counties have the highest MALPE, MAPE, and MPAD compared with largest counties.

For the Ratio-correlation Method, the overall Mean Algebraic Percent Error is -1.25, and Mean Absolute Percent Error is 4.79, and the Mean Percent Absolute Difference is 2.50 (Table **1**, Panel II). The MALPE, MAPE and MPAD for counties with population less than 1,000 are -1.81, 17.74, and 10.72, respectively. The MALPE, MAPE, and MPAD for counties with population more than 1,000 and less than 2,500 are -1.35, 4.70, and 4.65, respectively. The lowest MALPE, MAPE, and MPAD are for counties with population of 100,000 or more and are -0.16, 2.92, and 2.05, respectively.

For the Housing Unit Method, the overall Mean Algebraic Percent Error is 0.63, the Mean Percent Absolute Error is 10.32, and the Mean Percent Absolute Difference is 8.31 (Table **1**, Panel III). The highest MALPE, MAPE and MPAD, 14.60, 14.60, and 13.49 respectively, are for counties with population more than 1,000 and less than 2,500. The MALPE, MAPE, and MPAD for counties with population more than 2,500 and less than 5,000 are 10.41, 14.47, and 13.81, respectively. The lowest MALPE, MAPE, and MPAD are for counties with population more than 10,000 and less than 25,000 are -0.16, 7.62, and 7.36, respectively.

Averaging the Component Method II, Ratio-correlation Method, and Housing Unit Method produced an overall Mean Algebraic Percent Error of 0.64, a Mean Percent Absolute Error of 6.21 percent, and a Mean Percent Absolute Difference of 4.49 percent (Table **1**, Panel IV). The MALPE, MAPE, and MPAD for counties with population less than 1,000 are -1.93, 15.97, and 9.89, respectively. The MALPE, MAPE, and MPAD are higher for small size counties compared with the large size counties. In general, the data in Table **1** suggest the expected patterns, with error measures being larger for counties with smaller population and smaller for counties with larger populations. The data in Table **1** also suggests 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 1980 to 1990. Panel I presents error measures for Component Method II, Panel II presents error measures for the Ratio-correlation Method, Panel III presents error measures for Housing Unit Method, and Panel IV presents error measures for the average of all three methods. As can be seen from Table **2**, MALPE, MAPE, and MPAD was largest for counties with declining population of 10.0 percent or more and also for fastest growing counties with growth rate of 40.0 percent or more for Housing Unit Method. There were 39 counties with a decline rate of 10.0 percent or more and 16 counties with a growth rate of 40.0 percent or more. The error measures are smaller for moderate growing counties (Table **2**). The average of three methods produced similar type of result. Overall, Table **2** suggests that Component Method II and Rati-correlation Method produced better estimates compared with Housing Unit Method for Counties in Texas.

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. Compared with the actual 1990 Census counts, 52.0 percent of all county estimates produced by Component Method II are within the 5 percent of error range. Another 28.3 percent are in the rage of 5.1 percent to 10.0 percent. The estimates produced by the Ratio-correlation Method, 68.5 percent are within 5 percent

Table 2. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 1990 Census Counts and Estimated Population Produced by Component Method II, Ratio-correlation, and Housing Unit Methods for Counties in Texas by Percent Population Change, 1980-1990

Method	Percent Population Change, 1980-1990	Number of Counties	MALPE	MAPE	MPAD
Panel I: Component Method II					
	<-10.0	39	0.57	7.69	6.11
	-10.1 - 0.0	59	-0.67	6.98	4.11
	0.1 - 10.0	67	-2.98	5.41	3.95
	10.1 - 20.0	39	-2.26	6.53	2.34
	20.1 - 30.0	18	0.08	5.21	5.36
	30.1 - 40.0	16	-2.02	7.48	3.15
	40.1 - 50.0	8	2.59	3.75	3.00
	50.1+	8	-0.88	9.89	10.36
Panel II: Ratio-correlation Method					
	<-10.0	39	1.32	4.19	3.03
	-10.1 - 0.0	59	0.21	4.54	3.72
	0.1 - 10.0	67	-3.00	5.08	3.08
	10.1 - 20.0	39	-1.40	4.96	1.81
	20.1 - 30.0	18	-3.24	4.85	2.41
	30.1 - 40.0	16	-2.36	6.15	3.15
	40.1 - 50.0	8	-2.06	4.19	3.17
	50.1+	8	-1.79	3.96	4.02
Panel III: Housing Unit Method					
	<-10.0	39	20.42	20.98	21.62
	-10.1 - 0.0	59	10.61	11.27	12.94
	0.1 - 10.0	67	2.12	4.18	6.29
	10.1 - 20.0	39	-3.66	6.83	7.13
	20.1 - 30.0	18	-9.25	9.88	3.96
	30.1 - 40.0	16	-12.84	13.57	6.63
	40.1 - 50.0	8	-20.43	20.43	21.11
	50.1+	8	-22.99	22.99	20.82
Panel IV: Average of Component Method II, Ratio-correlation, and Housing Unit Methods*					
	<-10.0	39	10.46	11.02	10.86
	-10.1 - 0.0	59	4.86	5.55	6.63
	0.1 - 10.0	67	-0.49	3.19	3.47
	10.1 - 20.0	39	-2.79	5.10	3.66
	20.1 - 30.0	18	-5.33	5.76	2.59
	30.1 - 40.0	16	-7.67	8.14	3.24
	40.1 - 50.0	8	-10.14	10.14	10.31
	50.1+	8	-11.45	11.45	11.78

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

Table 3. Range of Percent Error for Differences Between1990 Census Counts and Estimated Population Produced by Component Method II, Ratio-correlation, and Housing Unit Methods

	Component Method II		Ration-Correlation Method		Housing Unit Method		Average of CM II, RCM, & HU Methods	
Range	**Number**	**Percent**	**Number**	**Percent**	**Number**	**Percent**	**Number**	**Percent**
0.0- 5.0	132	52.0	174	68.5	86	33.9	135	53.1
5.1-10.0	72	28.3	60	23.6	59	23.2	73	28.7
10.1-15.0	31	12.2	11	4.3	38	15.0	27	10.6
15.1-20.0	13	5.1	5	2.0	31	12.2	13	5.1
20.1-25.0	1	0.4	2	0.8	20	7.9	4	1.6
25.1-30.0	2	0.8	0	0.0	7	2.8	1	0.4
30.1+	3	1.2	2	0.8	13	5.1	1	0.4

Table 4. Number of Counties, Percent of Counties, Mean Algebraic Percent Error (MALPE) and Mean Percent Absolute Difference (MPAD) for Counties with Estimates Above and Below the 1990 Census Counts

	Component Method II		Ration-Correlation Method		Housing Unit Method		Average of CM II, RCM, & HU Methods	
	Below	**Above**	**Below**	**Above**	**Below**	**Above**	**Below**	**Above**
Number*	162	92	152	102	95	158	110	143
Percent	63.78	36.22	59.84	40.16	37.55	62.45	43.48	56.52
MALPE	-6.09	7.22	5.05	2.72	-10.89	10.99	-6.43	6.08
MPAD	3.30	4.09	4.40	2.45	8.43	8.66	4.38	4.61

* One county's 1990 population was exactly estimated by the base estimates and so is not included in the comparisons shown here.

error range and 23.6 percent are with the range of 5.1 to 10.0 percent. For the estimates produced by Housing Unit Method, 33.9 percent are within the range of 5 percent and 23.2 percent are with the rage of 5.1 to 10.0. Using the average of three methods (Component Method II, Ratio-correlation Method, and Housing Unit Method), 53.1 percent of all counties are being estimated within 5 percent of actual counts and 28.7 percent within the range of 5.1 to 10.0. Overall, only 6 of the 254 counties have a 20 percent or more error from the actual 1990 Census counts.

The result in Table **4** shows that the Component Method II and the Ratio-correlation Method tend to be biased downward with 63.78 and 59.84 percent of the counties, respectively being underestimated while the Housing Unit Method produced estimates that tended to overestimate the population of the counties with 62.45 percent of the counties being overestimated. An average of the three methods produced estimates that tended to overestimate the population of the counties with 43.48 percent of the counties being underestimated and 56.52 percent being overestimated.

VII.B. The Results of the Evaluation of Place-Level Estimates

Table **5** presents error measures for place estimates for Component Method II, Housing Unit Method, and the average of Component Method II and the Housing Unit Method. For the Component Method II, the overall Mean Algebraic Percent Error was 9.10, the Mean Percent Absolute Error was 19.79, and the Mean Percent Absolute Difference was 10.81 (Table **5**, Panel I).

For the Housing Unit Method the overall Mean Algebraic Percent Error was 7.83, the Mean Percent Absolute Error was 15.94, and the Mean Percent Absolute Difference was 8.86 (Table **5**, Panel II). Averaging the Component Method II and Housing Unit Method produced an overall Mean Algebraic Percent Error of 7.95, a Mean Percent Absolute Error of 18.14 percent, and a Mean Percent Absolute Difference of 9.49 percent (Table **5**, Panel III). In general, the data on Table **5** suggest the expected patterns, with error measures being larger for places with smaller population and smaller for place with larger populations.

Table **6** presents the same error measures by the rate of population change from 1980 to 1990. Differences in population growth rates had the same impact on errors for places as for counties (i.e., fastest declining and growing places 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

Table 5. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 1990 Census Counts and Estimated Population Produced by Component Method II and Housing Unit Method for Places in Texas

Population Size 1990		Number of Places	MALPE	MAPE	MPAD
Panel I: Component Method II					
	< 1,000	452	15.13	26.29	22.88
	1,000- 2,499	306	6.66	17.33	17.15
	2,500- 4,999	167	8.36	17.30	17.81
	5,000- 9,999	104	2.37	13.38	13.51
	10,000-24,999	107	3.42	15.37	15.63
	25,000-49,999	34	-1.03	15.29	14.89
	50,000-99,999	19	4.74	9.14	9.33
	100,000+	19	2.46	5.42	7.59
	All Places	1208	9.10	19.79	10.81
Panel II: Housing Unit Method					
	< 1,000	111	4.96	24.30	22.94
	1,000- 2,499	206	10.63	19.09	19.38
	2,500- 4,999	144	9.41	14.15	14.14
	5,000- 9,999	96	7.48	13.35	13.12
	10,000-24,999	101	5.49	10.46	10.54
	25,000-49,999	30	3.46	7.96	7.87
	50,000-99,999	19	5.76	8.19	8.11
	100,000+	19	5.51	9.13	7.32
	All Places	726	7.83	15.94	8.86
Panel III: Average of Component Method II and Housing Unit Methods*					
	< 1,000	453	11.47	24.19	21.36
	1,000- 2,499	307	6.23	16.37	16.52
	2,500- 4,999	167	8.52	15.76	16.16
	5,000- 9,999	104	4.43	12.76	12.69
	10,000-24,999	107	4.23	12.30	12.43
	25,000-49,999	34	-1.23	12.88	12.51
	50,000-99,999	19	6.24	8.17	8.12
	100,000+	19	5.00	5.40	6.80
	All Places	1210	7.95	18.14	9.49

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

Table 6. Mean Algebraic Percent Error (MALPE), Mean Absolute Percent Error (MAPE), and Mean Percent Absolute Difference (MPAD) between 1990 Census Counts and Estimated Population Produced by Component Method II and Housing Unit Method for Places by Percent Population Change, 1980-1990

Percent Population Change, 1980-1990	Number of Places	MALPE	MAPE	MPAD
Panel I: Component Method II				
<-10.0	267	28.12	29.65	21.39
-10.1- 0.0	281	12.58	15.24	10.01
0.1-10.0	212	8.44	13.02	13.18
10.1-20.0	122	7.67	14.18	6.22
20.1-30.0	61	1.58	13.11	5.93
30.1-40.0	44	0.50	15.57	9.71
40.1-50.0	32	0.14	15.97	9.38
50.1+	133	-21.78	25.12	18.39
Panel II: Housing Unit Method				
<-10.0	127	28.30	28.46	26.27
-10.1- 0.0	183	13.22	14.49	15.50
0.1-10.0	148	7.34	8.02	8.76
10.1-20.0	77	4.24	8.79	6.99
20.1-30.0	43	-0.37	7.48	4.96
30.1-40.0	28	-3.52	8.46	2.66
40.1-50.0	24	0.01	9.22	7.73
50.1+	84	-7.49	19.85	10.85
Panel III: Average of Component Method II and Housing Unit Methods				
<-10.0	267	30.35	30.56	24.32
-10.1- 0.0	281	12.48	12.75	11.54
0.1-10.0	212	6.97	9.06	10.59
10.1-20.0	122	3.54	9.37	6.41
20.1-30.0	61	-1.61	9.59	5.17
30.1-40.0	44	-4.60	11.36	5.36
40.1-50.0	32	-2.90	12.05	5.67
50.1+	133	-22.29	25.72	14.36

and underestimate the fastest growing places. The Mean Algebraic Percent Error was 28.12 for Component Method II and 28.30 for Housing Unit Method for places with declining population of 10.0 percent or more. The Mean Percent Absolute Difference was 21.39 for Component Method II and 26.27 for Housing Unit Method with declining population of 10.0 percent or more. The MALPE, MAPE, and MPAD were higher for the fastest growing places as well. This was evident from both Component Method II and Housing Unit Method.

Table **7** shows results of the evaluation of estimates in terms of range of errors. Overall (average of the Component Method II and the Housing Unit Method) 20.9 percent of the places (compared to 53.1 percent of the counties) were estimated within 5 percent, 21.7 percent within 5.1 to 10.0 percent, and another 16.0 percent within 10.1 to 15.0 percent of the 1990 population count. For the Component Method II, 42.9 percent of the places were estimated within 10.00 percent while for the Housing Unit Method 40.9 percent of the places were estimated within the 10.00 percent of the actual census count.

Table 7. Range of Percent Error for Differences between Census Counts and Estimated Population Produced by Component Method II and Housing Unit Method for Places in Texas for 1990

Absolute Percent Error	Component Method II		Housing Unit Method		Average of CM II and HU Methods	
	Number	Percent	Number	Percent	Number	Percent
0.0 - 5.0	298	24.7	150	20.7	253	20.9
5.1 - 10.0	220	18.2	147	20.2	262	21.7
10.1 - 15.0	177	14.7	152	20.9	194	16.0
15.1 - 20.0	111	9.2	97	13.4	153	12.6
20.1 - 25.0	87	7.2	74	10.2	89	7.4
25.1 - 30.0	80	6.6	31	4.3	74	6.1
30.1+	235	19.5	75	10.3	185	15.3

Table 8. Number of Places, Percent of Places, and Mean Algebraic Percent Error (MALPE) and Mean Percent Absolute Differences (MPAD) for Places with Estimates Above and Below the 1990 Census Counts

	Component Method II		Housing Unit Method		Average of CM II and HU Methods	
	Below	Above	Below	Above	Below	Above
Number	443	761	163	562	362	846
Percent	36.79	63.21	22.48	77.52	29.97	70.03
MALPE	-14.58	22.97	18.06	15.36	-17.04	14.38
MPAD	11.90	10.48	12.32	8.39	18.66	8.70

The results in Table **8** point to a tendency for the population estimates of places to be overestimated compared with county estimates both for Component Method II and Housing Unit Method. This underestimation may be due to the undercount in the 1980 census, creating a base population that was lower than it should have been. Murdock and Hoque [5] evaluated the impact of undercount on the accuracy of small-area population estimates and found significant differences in using adjusted and non-adjusted population for 1980 (for detailed discussion see [5]). Another possible explanation is that there might have been an overestimation of population in small places in 1990. To my knowledge, there is no study that evaluates the impact of adjustment to the 1990 census.

The data in Tables **6** through 8 also suggest that in nearly all cases, the use of the Housing Unit Method improves the accuracy of the estimates obtained. The results of the Housing Unit Method averaged with Component Method II appear to lead to a reduction in the error of estimates for places in Texas.

In review, the data for places suggest that place level estimates are less accurate than those for counties. They suggest that the use of the Housing Unit Method averaged with estimates from the Component Method II may provide the means of moving toward improvement of the estimates.

VIII. CONCLUSION

Accurate estimates are difficult for small areas and for areas showing inconsistency in the direction of change during the estimation period. Most of the places in Texas experienced rather rapid change making it more difficult to do accurate estimates of population. Texas is one of the few states that has produced population estimates for the counties and places since the mid-1980s. Texas Population Estimates and Projections Program's population estimates are calculated using an average of the Component Method II, the Ratio-correlation Method, and the Housing Unit Method. For this paper, population estimates calculated using the Component Method II, the Ratio-correlation Method, and the Housing Unit Method separately as well as estimates calculated using an average of the three methods were evaluated against actual 1990 census counts.

Three error measures are used to assess the accuracy of population estimates of Texas for 1990. They are the Mean Algebraic Percent Error, the Mean Absolute Percent Error, and the Mean Percent Absolute Difference. At the county level, Component Method II did better than the Housing Unit

Method and Ratio-correlation Method. At the place level, the Housing Unit Method did better than the Component Method II. However, as mentioned earlier, some counties and places in Texas neither issue building permits nor provide data to the U.S. Department of Commerce or Texas State Data Center to be used in the Housing Unit Method. Recent data on vacancy rate is also unavailable for most areas.

The evaluation of the Texas Population Estimates and Projections Program's population estimates 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 error rate results because of 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 Ratio-correlation method and compare the estimated population produced by Ratio-correlation to the actual census counts and also taking the average of all three methods to compare with the census counts for 2000. For county estimates, I would recommend adding the administrative record method provided that input data are available to do so and again compare with the census counts.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Bryan T. Population Estimates. In: Siegel J. Swanson DA, Eds. The methods and materials of demography, 2nd Edition. San Diego, CA: Elsevier Academic Press 2004; pp. 523-60.

[2] Hoque N. An Evaluation of population estimates for counties and places in Texas for 2000. In: Murdock SH, Swanson DA, Eds. Applied Demography in the 21st Century. Springer Science and Business Media B.V.: Dordrecht, The Netherlands 2008; pp. 125-48.

[3] Murdock SH, Ellis D. Applied demography: an introduction to basic concepts, methods, and data. Boulder, Colorado: Westview Press Inc 1991.

[4] Murdock SH, Hoque MN. Evaluation of Texas Population Estimates and Projections Program's Population Estimates and Projections for 1990. Department of Rural Sociology, Texas A&M University, College Station, Texas 1992, Departmental Miscellaneous Report No 92-1.

[5] Murdock SH, Hoque MN. The Effect of Undercount on the Accuracy of Small-Area Population Estimates: Implications for the Use of Administrative Data for Improving Population Enumeration. Popul Res Pol Rev 1995; 14(2): 251-71.

[6] Smith SK, Cody S. An evaluation of population estimates in Florida: April 1, 2000, Popul Res Policy Rev 2004; 23: 1-24.

[7] Siegel JS. Applied Demography: Applications to Business, Government, Law and Public Policy. San Diego: Academic Press 2002.

[8] Rowland DT. Demographic methods and concepts. New York: Oxford University Press 2003.

[9] Serow JW, Terrie B, Waller, Wichmann RW. The use of intercensal population estimates in political redistricting. In Kinter JH, Merrick TW, Morrison PA, Voss PR. Eds. Demographics: A Casebook for Business and Government. Rand Publication 1997; pp. 33-54.

[10] Bolton N. The use of intercensal population estimates and projections in the court-mandated reapportionment of Los Angeles County. In Kinter JH, Merrick TW, Morrison PA, Voss PR. Eds. Demographics: A Casebook for Business and Government. Rand Publication 1997; pp. 55-69.

[11] Raymondo JC. Population estimation and projection: methods for marketing, demographic, and planning personnel. One Madison Avenue, N.Y.: Quorum Books 1992.

[12] Smith SK, Lewis B. Some new techniques for applying the housing unit method of local population estimation. Demography 1980; 17: 323-39.

[13] U.S. Census Bureau. Methodology for estimates of state and county total population. [Online] 2001. Available from: http://www.census.gov/popest/topics/methodology/state.php

[14] Shryock H, Siegel JS. The Methods and materials of demography. Washington, D.C. Government Printing Office 1980.

[15] Smith SK, Tayman J, Swanson DA. State and local population projections: methodology and analysis. New York: Kluwer Academic/Plenum Publishers 2001.

[16] Murdock SH, Kelley C, Jordan J, Pecotte B, Luedke A. Demographics: a guide to methods and sources of data for demographic analysis in the media, business, and government. Boulder, Colorado: Paradigm Publishers 2006.

Received: April 19, 2010 Revised: May 18, 2010 Accepted: May 19, 2010