College of Liberal Arts and Sciences
Bureau of Economic and Business Research

Special Population Reports

Number 12, December 2021

An Evaluation of Population Estimates in Florida: April 1, 2020

Stefan Rayer, Population Program Director
Richard Doty, GIS Coordinator/Research Demographer
Suzanne Roulston-Doty, GIS Project Manager/Research Demographer
Ying Wang, Research Demographer

The Bureau of Economic and Business Research (BEBR) has made population estimates for all cities and counties in Florida each year since 1972. These estimates are used for a wide variety of purposes. Businesses use them to develop customer profiles, identify market clusters, and determine optimal site locations. Research analysts use them to study urban sprawl, environmental conditions, and social trends. State and local governments use them to monitor the impact of public policies and to estimate the need for schools, roads, parks, public transportation, fire protection, and other goods and services. Furthermore, they are used for allocating more than one billion dollars each year to city and county governments through Florida's revenue-sharing programs.

Given their many uses, it is essential to evaluate the accuracy of these estimates. In this report, we describe the methodology used for making state and local population estimates in Florida and evaluate the accuracy of the 2020 estimates by comparing them with the results of the 2020 census. We also evaluate the accuracy of previous BEBR estimates and estimates produced by the U.S. Census Bureau. We close with several observations regarding the production of population estimates in Florida.

METHODOLOGY

BEBR population estimates are constructed using the housing unit (HU) method, in which estimates of population change are derived from estimates of changes in occupied housing units. This is the most commonly used method for making small-area population estimates in the United States because it is conceptually simple, can utilize a wide variety of data sources, can be applied at any level of geography, and produces reliable estimates (Siegel 2002). We use this method to construct population estimates for each county and subcounty area in Florida, with subcounty areas defined as incorporated cities, towns, and villages, and the unincorporated balance of each county. The state estimate is calculated as the sum of the county estimates. The estimates refer solely to permanent residents of Florida; they do not include seasonal or other types of temporary residents.

The foundation of the HU method is the fact that almost everyone lives in some type of housing structure, whether a traditional single-family unit, an apartment, a mobile home, a college dormitory, or a state prison. The population of any geographic area can be calculated as the number of occupied housing units (households) times the average number of persons per household (PPH), plus the number of persons living in group quarters such as college dormitories, military barracks, nursing homes, and prisons:

$$P_t = (H_t \times PPH_t) + GQ_t$$

where P_t is the population at time t, H_t is the number of occupied housing units at time t, PPH_t is the average number of persons per household at time t, and GQ_t is the group quarters population at time t. Estimates of the group quarters population typically include persons without permanent living quarters (e.g., the homeless population).

This is an identity, not an estimate. If these three components were known exactly, the total population would also be known. The problem, of course, is that these components are almost never known exactly. Rather, they must be estimated from various data sources, using one or more of several possible techniques. In this section, we provide a brief description of the data and techniques used to estimate these three components for counties and subcounty areas. Other descriptions of the HU method can be found in Murdock and Ellis (1991), Smith (1986), Siegel (2002), and Tayman and Swanson (2012).

Households

Census definitions require a person to be counted as an inhabitant of his/her usual place of residence, which is generally construed to mean the place where he/she lives and sleeps most of the time. This place is not necessarily the same as one's legal or voting residence. A household is the person or group of people occupying a housing unit; by definition, the number of occupied housing units is the same as the number of households. Households refer solely to permanent residents and a housing unit is classified as vacant even when it is continuously occupied, if all the occupants are temporary residents staying only for a few days, weeks, or months.

We use three different data sources to estimate the number of households in Florida. Our primary data source is active residential electric customers. We collect these data from each of the state's 53 electric utility companies. Households can be estimated by constructing a ratio of households to active residential electric customers using data from the most recent census year (e.g., 2010) and multiplying that ratio times the number of active residential customers in some later year (e.g., 2020). This procedure assumes that no changes have occurred in electric company bookkeeping practices or in the proportion of customers who are permanent residents. Although changes do occur, they are generally fairly small. In some places we adjust the household/electric customer ratio to account for likely changes in the proportion of housing units occupied by permanent residents.

We sometimes filter electric customer data to exclude minimum use customers. Minimum use customers are those using less than 200 kilowatt (kWh) hours per month. We believe these customers represent seasonal or other part-time residents or vacant units and excluding them may give a more accurate measure of permanent residents. Because we estimate the change in population since the 2010 Census, excluding minimum use customers can capture changes in unit occupancy over that period. These data are not available for all areas of the state, but in places in which the data are available and appear to be reliable, we often use them in conjunction with other data sources.

Our second data source is residential building permits, as collected and distributed by the U.S. Department of Commerce. The housing inventory in 2020 for a city or county can be estimated by adding permits issued since 2010 to the units counted in the 2010 census, and subtracting units lost to destruction, demolition, or conversion to other uses. The time lag between the issuance of a permit and the completion of a unit is assumed to be three months for single-family units and fifteen months for multi-family units. Building permits are not issued for mobile homes, but proxies can be derived from records of shipments to mobile home dealers. Creating a housing inventory for an entire county requires complete permit data for every permitting agency within the county. Although such data are not always available, coverage is sufficient in most Florida cities and counties to provide useful information.

There are no readily available data sources providing comprehensive up-to-date information on occupancy rates. Accurate information can be obtained through special censuses or large sample surveys, but in most instances these methods are too expensive to be feasible. A common solution is to use the occupancy rates reported in the most recent census. We base our occupancy estimates on these values, but we may make adjustments to account for factors reflecting changes in occupancy rates over time. Occupancy changes since 2010 may be captured in places where we use electrical customer data and are able to exclude minimum use customers. Additional factors may include data from the U.S. Census Bureau's American Community Survey (ACS), in cases where they show statistically significant trends over time since the last decennial census.

The product of the inventory figure and the occupancy rate (performed separately for each type of housing unit) provides an estimate of the number of households. There are several potential problems with this estimate. Time lags between the issuance of permits and the completion of units may vary from place to place and from year to year. The proportion of permits resulting in completed units is usually unknown. Data on demolitions and conversions are incomplete and data on mobile homes must be estimated indirectly. Reliable estimates of changes in occupancy rates are generally unavailable. Certificate-of-occupancy data can eliminate problems related to completion rates and time lags but not those related to occupancy rates, demolitions, and conversions. Although these problems limit the usefulness of the data in some places, building permit data often provide reasonably accurate estimates of households.

Our third data source is the number of homestead exemptions reported by the Florida Department of Revenue. Households can be estimated by constructing a ratio of households to exemptions using data from the most recent census year (e.g., 2010) and multiplying that ratio times the number of exemptions in some later year (e.g., 2020). An important advantage of these data is that they cover only housing units occupied by permanent residents, thereby excluding the impact of seasonal and other non-permanent residents. The primary disadvantage is that the data do not include households occupied by renters or other non-homeowners, but those households often change at a similar rate to the households with homestead exemptions. Homestead exemption data is also available from each county's property appraiser at the property parcel level, which can be summarized by subcounty areas. We occasionally use these data to inform our decision making in places where our other primary data sources show significantly different results.

Electric customer, building permit, and homestead exemption data all provide useful information regarding changes in households. Electric customer data generally provide good household estimates but do not provide information on changes in the mix of housing units (single-family, multifamily, mobile home). Building permit data provide somewhat less accurate estimates of households, but they provide information on changes in housing mix. Homestead exemption data refer solely to permanent residents but exclude non-homeowners. Previous research on BEBR population estimates has shown that household estimates based on electric customer data are – on average – more accurate than those based on building permit data (Smith and Cody 1994, 2004, 2013). However, we apply our professional judgment to decide which data source(s) to use in each county and subcounty area. In many instances, we use averages of estimates from more than one data source. The benefits of combining estimates or projections are well-known (Armstrong 2001; Hoque 2012; Rayer 2008; Siegel 2002; Smith and Cody 2004). We also sometimes use GIS-based property parcel data (along with year-built information and detailed land use codes from the Florida Department of Revenue) to evaluate which data source is best for a particular place.

Persons per Household

The second component of the housing unit method is the average number of persons per household (PPH). Florida's PPH dropped steadily from 3.22 in 1950 to 2.46 in 1990 but then leveled off, remaining constant between 1990 and 2000 before rising to 2.48 in 2010. The currently available redistricting data from Census 2020 do not include data on PPH. One can, however, derive an approximate PPH value by dividing the population not in group quarters by the number of occupied housing units. Statewide, this resulted in a PPH of 2.47, which is just slightly lower than in 2010. There is a substantial amount of variation among local areas in Florida, with values in 2020 ranging from 1.9 to 3.0 for counties and from less than 1.4 to more than 3.8 for subcounty areas. PPH values have risen over time in some cities and counties and declined in others.

For each county and subcounty area, we base our PPH estimates on the local PPH value in the most recent census (e.g., 2010). In some instances, we estimate changes in PPH since that census using statistically significant trends in data from the American Community Survey or changes in the mix of single-family, multifamily, and mobile home units since the last census. Again, we use our professional judgment to decide which data sources and techniques to use in each county and subcounty area.

Group Quarters Population

The household population is calculated as the product of households and PPH. To obtain an estimate of the total population, we must add an estimate of the group quarters population. In most places, we estimate the group quarters population by assuming that it accounts for the same proportion of total population in 2020 as it did in 2010. For example, if the group quarters population accounted for 2% of the total population in 2010, we assume that it accounted for 2% in 2020. In places where there are large group quarters facilities, we collect data directly from the administrators of those facilities and add those estimates to the other group quarters population. Inmates in state and federal institutions are accounted for separately in all local areas; these data are available from the Federal Bureau of Prisons, the Florida Department of Corrections, the Florida Department of Veteran Affairs, the Florida Agency for Persons with Disabilities, the Florida Department of Health, the Florida Department of Juvenile Justice, and the Florida Department of Children and Families. The total population estimate is made by adding the estimate of the group quarters population to the estimate of the household population.

EVALUATING PRECISION AND BIAS

We constructed population estimates for April 1, 2020 for each incorporated city, town, and village; each county; and the unincorporated balance of each county in Florida. We evaluated these estimates by comparing them with census counts for the same date. Although census counts contain errors, they are quite accurate in most places and provide a widely used standard for evaluating population estimates. We refer to differences between estimates and census counts as estimation errors, but they may have been caused partly by enumeration errors.

We use five measures for evaluating the accuracy of the population estimates. The mean absolute percent error (MAPE) is the average error when the direction of the error is ignored. The proportions of errors less than 5% and greater than 10% indicate the frequency of relatively small and large errors, respectively. These are measures of precision, or how close the estimates were to census counts, regardless of whether they were too high or too low. The mean algebraic percent error (MALPE) is the average error when the direction of error is included. This is a measure of bias: a positive error indicates a tendency for estimates to be too high and a negative error indicates a tendency for estimates to be too low. Since a few extreme errors in one direction can strongly influence the MAPE and MALPE, we also calculate the proportion of estimates above the census count

(%POS), which is another measure of bias. We considered the median absolute percent error (MedAPE) and the median algebraic percent error (MedALPE) as well, which are sometimes used as more robust alternatives to the MAPE and the MALPE. The results from these two alternative accuracy measures were generally similar to the MAPE and MALPE, and they are not shown in this report.

State Estimates

BEBR's state population estimate for April 1, 2020 was 21,596,068, less than 0.3% above the census count of 21,538,187. This error is very small for a state that grew by almost 15% during the decade; had large numbers of interstate migrants, seasonal residents, and foreign immigrants; was struck by several devastating hurricanes; and recovered from the Great Recession of 2007 and 2009.

Table 1. Errors in State Population Estimates, Florida, 1980–2020

Year	Percent Error
1980	-2.7
1990	1.6
2000	-1.8
2010	-0.2
2020	0.3

Table 1 shows errors for the state population estimates for each census year since 1980. Errors were below census counts in three years (1980, 2000, and 2010) and above the count in two years (1990 and 2020). Although there was not a perfectly monotonic relationship, errors have generally fallen over time. The very small positive error in 2020 is almost the exact mirror image of the tiny negative error in 2010, and the small negative error in 2000 also balanced the small positive error in 1990. These results demonstrate that the BEBR population estimates for the state exhibit no consistent downward or upward bias, that they are very accurate, and that accuracy has increased over time.

County Estimates

Table 2 summarizes the errors for the 2020 county population estimates. The MAPE for all counties was 2.8%. Most counties had errors of less than 5% and only one county had an error greater than 10%. The county estimates displayed very little bias, as the MALPE was 1.0% and errors were about evenly split between those that were too high and those that were too low.

MAPEs were larger for small counties than large counties, but there was little change in the relationship between accuracy and population size for counties with more than 100,000 residents. Estimates for counties with populations below 100,000 were somewhat too high, estimates for counties with populations between 100,000 and 300,000 were slightly too low, while estimates for the largest counties with populations of 300,000 or more had essentially no bias.

There was little indication of any relationship between errors and population growth rates, except for counties with declining populations, for which accuracy was lowest. MALPEs showed a U-shape pattern: estimates for counties with declining populations were too high and estimates for counties with fast-growing populations were too low; the error for the counties with declining populations was much higher than the one for the fast-growing counties. These relationships are mirrored in the proportion of positive errors, which show an upward bias for the declining counties and a downward bias for the most rapidly growing counties.

Subcounty Estimates

Table 3 shows errors for subcounty areas (i.e., incorporated cities, towns, and villages, and the unincorporated balance of each county). The MAPE for all subcounty areas was 6.8%, about two and a half times larger than the MAPE for counties. This is not surprising, given the large number of subcounty areas with very small populations. Over half of the errors were less than 5%, and slightly under one-fifth

Table 2. Population Estimation Errors by Population Size and Growth Rate: Florida Counties, 2020

Size (2010) and Growth Rate (2010–2020)	N	MAPE	MALPE	% POS	Percent of Absolute Errors < 5%	Percent of Absolute Errors > 10%
< 25,000	15	4.2	2.8	73.3	66.7	6.7
25,000 – 99,999	19	3.6	2.2	68.4	68.4	0.0
100,000 – 299,999	16	1.9	-1.1	31.3	100.0	0.0
≥ 300,000	17	1.6	0.0	41.2	100.0	0.0
< 0.0%	17	4.9	4.6	88.2	47.1	5.9
0.0 – 9.9%	19	2.4	0.7	68.4	100.0	0.0
10.0 – 19.9%	18	1.7	-0.8	38.9	94.4	0.0
≥ 20.0%	13	2.3	-1.0	7.7	92.3	0.0
Total	67	2.8	1.0	53.7	83.6	1.5

were greater than 10%. There was a slight upward bias in the sub-county estimates, as indicated by a MALPE of 1.7% and 59% positive errors.

Differences in population size and growth rate had a much greater impact on estimation errors for subcounty areas than for counties. This occurred because the number of observations was much greater for subcounty areas, and there was much more variation in population size and growth-rate characteristics. The MAPE was 15.7% for places with fewer than 500 residents and declined as population size increased, reaching 2.6% for places with 100,000 or more residents. The results are mirrored by the percent of absolute errors below 5%, which go up with increasing population size, and by the percent of absolute errors above 10%, which go down with increasing population size. There was no consistent relationship between population size and the tendency for estimates to be too low or high; estimates for subcounty areas of all sizes but the largest size category showed a positive MALPE, which was quite small for most size categories. The percent positive errors mirror the MALPE results.

Differences in population growth rates had a strong effect on errors for subcounty areas. Similar to the results for counties, there was a U-shaped relationship between MAPEs and growth rates. MAPEs were smallest in places with moderate positive growth rates but increased as growth rates deviated in either direction from those levels. MAPEs were less than 5% for places growing by less than 20% during the decade; in contrast, they were 21.9% for places losing more than 10% of their residents, and 16.2% for places where the population increased by more than 50%. The percent of absolute errors below 5% and above 10% show a similar relationship to growth rates; places with moderate positive growth rates mostly had very small errors, while those with rapidly growing or declining population often exhibited errors above 10%.

There was a strong tendency for estimates to be too high for places losing population and too low for rapidly growing places. Places losing more than 10% of their residents were overestimated by 14.9%, on average; estimates were too high in 88% of those places. At the other end of the spectrum, places with growth rates above 50% were underestimated by

Table 3. Population Estimation Errors by Population Size and Growth Rate: Florida Subcounty Areas, 2020

Size (2010) and Growth Rate (2010–2020)	N	MAPE	MALPE	% POS	Percent of Absolute Errors < 5%	Percent of Absolute Errors > 10%
< 500	46	15.7	0.4	56.5	32.6	52.2
500 – 999	39	9.6	7.6	76.9	38.5	35.9
1,000 – 2,499	61	8.8	2.2	55.7	52.5	27.9
2,500 – 4,999	53	7.2	1.2	49.1	52.8	26.4
5,000 – 9,999	56	6.6	3.2	62.5	46.4	21.4
10,000 – 14,999	47	5.0	1.1	63.8	59.6	12.8
15,000 – 24,999	42	4.5	1.4	71.4	61.9	7.1
25,000 – 49,999	48	3.8	0.3	58.3	68.8	2.1
50,000 – 99,999	42	3.1	0.3	50.0	78.6	2.4
≥ 100,000	45	2.6	0.0	48.9	91.1	2.2
< -10.0%	41	21.9	14.9	87.8	9.8	78.1
-10.0 – -5.1%	33	10.0	9.0	84.9	30.3	42.4
-5.0 – 0.0%	57	5.5	3.9	86.0	49.1	14.0
0.0 – 4.9%	69	3.6	2.0	71.0	76.8	5.8
5.0 – 9.9%	76	3.6	1.0	63.2	80.3	5.3
10.0 – 14.9%	74	3.7	-0.4	48.7	75.7	6.8
15.0 – 19.9%	42	4.4	-2.5	28.6	61.9	4.8
20.0 – 29.9%	38	5.7	-2.5	26.3	57.9	18.4
30.0 – 49.9%	36	8.3	-4.6	30.6	41.7	22.2
≥ 50.0%	13	16.2	-9.4	23.1	15.4	69.2
Total	482	6.8	1.7	58.9	57.9	19.5

9.4%, on average; estimates were too low in 77% of those places. The percent positive errors mirror the MALPE results.

Based on these results, we can say that precision generally increases as population size increases up to a certain point, but then levels off; that precision declines as growth rates deviate (in either direction) from moderate but positive levels; and that bias is largely unaffected by differences in population size but strongly affected by differences in population growth rates (negative growth rates are associated with a strong tendency to overestimate and high positive growth rates are associated with a strong tendency to underestimate). Similar results have been reported in many other studies (Harper, Devine, and Coleman 2001; Siegel 2002; Smith 1986; Smith and Cody 1994, 2004, 2013). We believe these patterns can be accepted as general characteristics of population estimates.

Errors by Component

Which component of the HU method can be estimated most accurately? Table 4 shows that errors were smallest for PPH and largest for the group

Table 4. Estimation Errors by Component: Florida Counties and Subcounty Areas, 2020

	Component	MAPE	MALPE	% POS	Percent of Absolute Errors < 5%	Percent of Absolute Errors > 10%
Counties	Households	2.1	0.3	52.2	94.0	1.5
	PPH	1.5	0.8	62.7	98.5	0.0
	GQ	14.9	-0.8	38.8	31.3	47.8
Subcounty Areas	Households	6.5	2.4	57.5	64.7	17.6
	PPH	3.4	0.2	50.4	82.2	4.6
	GQ	40.2	6.9	29.0	40.0	54.1

quarters population (GQ). For counties, MAPEs were 1.5% for PPH, 2.1% for households, and 14.9% for GQ. There was a slight tendency for PPH estimates to be too high and GQ estimates to be too low. For subcounty areas, MAPEs were 3.4% for PPH, 6.5% for households, and 40.2% for GQ. The GQ estimates for subcounty areas tended to be too high; there was also a slight tendency for household estimates to be too high, while PPH estimates showed very little bias. Although numeric errors for the GQ population were generally quite small, percent errors were very large because in many places they were based on very small numbers of people.

A number of studies have found errors for households to be greater than errors for PPH (Lowe, Myers, and Weisser 1984; Smith and Cody 1994, 2004, 2013; Starsinic and Zitter 1968). This most likely reflects the fact that growth rates are generally higher and more variable for households than for PPH. Whereas PPH changed by less than 5% between 2010 and 2020 for most counties and subcounty areas in Florida, the number of households often changed by 20%, 30%, 40%, or more. There is simply more potential for error in estimates of households than in estimates of PPH.

For both counties and subcounty areas, errors for GQ were much larger than errors for households and PPH. Does this mean that GQ errors contributed the most to overall estimation error? One way to answer this question is to construct synthetic population estimates using a combination of estimated values and census counts. We made estimates for counties and sub-county areas under three scenarios. The first used estimates of households and census counts for PPH and GQ; the second used estimates of PPH and census counts for households and GQ; and the third used estimates of GQ and census counts for households and PPH. For each scenario, then, errors in the population estimates were due solely to errors in the single estimated component. The results are shown in Table 5.

It is clear that errors in GQ estimates did not contribute the most to overall estimation errors; in fact, they contributed the least. For both counties and subcounty areas, Scenario 1 had the largest MAPE, the most large errors, and the fewest small errors. Even with perfect estimates of PPH and GQ, errors in household estimates would have led to population estimation errors averaging 1.9% for counties and 5.9% for subcounty areas (ignoring the direction of error). With perfect estimates of households and GQ, errors in PPH estimates would have created population estimation errors averaging 1.4% for counties and 3.6% for subcounty areas (ignoring the direction of errors). With perfect estimates of households and PPH, errors in GQ estimates would have created population estimation errors of only 0.5% for counties and 1.4% for sub-county areas (again, ignoring the direction of errors). Although errors were much larger

Table 5. Florida Population Estimation Errors Under Alternate Scenarios

	Scenario	MAPE	MALPE	% POS	Percent of Absolute Errors < 5%	Percent of Absolute Errors > 10%
Counties	1	1.9	0.2	50.7	95.5	1.5
	2	1.4	0.8	61.2	98.5	0.0
	3	0.5	-0.1	38.8	100.0	0.0
Subcounty Areas	1	5.9	2.2	57.5	66.0	16.4
	2	3.6	0.2	50.6	82.8	4.4
	3	1.4	-0.7	33.2	95.2	1.5

for GQ estimates than for household or PPH estimates, those errors contributed relatively little to over-all estimation errors because the GQ population generally accounts for a very small proportion of total population. Similar results were found in evaluations of the 2000 and 2010 estimates in Florida (Smith and Cody 2004, 2013).

DISCUSSION

Comparison with Previous BEBR Estimates

The BEBR estimate for the state of Florida was 2.7% below the census count in 1980, 1.6% above the count in 1990, 1.8% below the count in 2000, 0.2% below the count in 2010, and 0.3% above the count in 2020. The change in errors from negative in 1980 to positive in 1990 and back to negative in 2000 was most likely caused – at least in part – by changes in census under-count. Nationally, census undercount declined between 1970 and 1980, rose between 1980 and 1990, and declined again between 1990 and 2000; the 2010 census had essentially no under-over overcount. Because each set of estimates is based on the most recent census, errors in census counts are built into subsequent estimates and changes in undercount (or overcount) from one census to another influence the size and direction of estimation errors. Detailed data on the undercount (or overcount) for the 2020 census are not yet available.

Table 6 compares errors for 2020 with errors for 1980, 1990, 2000, and 2010 for counties and subcounty areas in Florida. With respect to precision, the 2020 estimates for counties were similar in accuracy to the ones for 2010, while the 2020 estimates for subcounty areas were more accurate than those produced in any previous year. For counties, the proportion of large errors was identical to 2010, while the proportion of small errors was slightly lower; for subcounty areas, the proportion of large errors was smaller than ever before, and the proportion of small errors was larger.

MALPEs for counties in 2020 were slightly higher than in 2000 and 2010, but the proportion of positive errors was still close to a 50/50 split. For subcounty areas, the 2020 MALPE was the lowest of any year; the proportion of positive errors was slightly higher than in 2000 and 2010. Whereas the 1980 estimates had a tendency to be too low and the 1990 estimates had a tendency to be too high, the 2000, 2010, and 2020 estimates displayed very little bias. Viewed as a whole, these results suggest that the methodology employed by BEBR has no systematic bias toward either overestimation or underestimation.

Why were the 2020 estimates so accurate? There are several possible explanations. Population sizes were generally larger and population growth rates slower than in previous decades; both of these factors lead to greater accuracy, on average. The insights gained

Table 6. Errors in County and Subcounty Estimates, 1980–2020

	Year	MAPE	MALPE	% POS	Percent of Absolute Errors < 5%	Percent of Absolute Errors > 10%
Counties	1980	5.4	-2.9	34.3	53.7	10.4
	1990	5.4	3.3	74.6	58.2	16.4
	2000	4.2	0.8	50.7	73.1	10.4
	2010	2.7	0.5	49.3	88.1	1.5
	2020	2.8	1.0	53.7	83.6	1.5
Subcounty Areas	1980	14.4	3.5	46.7	33.6	42.4
	1990	11.9	6.0	68.4	36.5	40.5
	2000	10.4	2.3	51.2	46.6	32.3
	2010	9.2	2.0	55.1	49.1	26.1
	2020	6.8	1.7	58.9	57.9	19.5

through an additional ten years of studying estimation methods, sources of data, and the dynamics of population change in Florida most likely contributed to better estimates as well. Perhaps luck played a role. Whatever the causes, the 2020 subcounty estimates were the most accurate ever produced by BEBR, and the 2020 state and county estimates were similar in accuracy to the 2010 estimates, which had been the most accurate up to that point.

Comparison with Other Estimates

How do the BEBR estimates stack up against those produced by other agencies? The only other agency making independent population estimates for all cities and counties in Florida is the U.S. Census Bureau. Although several private data companies produce small-area population estimates for Florida, they base them on estimates produced by the Census Bureau or by state demographic agencies such as BEBR. Some local governments make estimates for places in their own jurisdictions, but not for other places throughout the state.

The Census Bureau provides a good standard for comparison because it is the nation's premier demographic agency. It has been producing state and local population estimates for many years and has pioneered in the development of several estimation techniques and data sources. At the county level, the Census Bureau uses an administrative record-based (AR) component of change method which updates the latest census population using the following inputs: vital statistics (using birth and death data from the National Center for Health Statistics and the Federal-State Cooperative for Population Estimates); net domestic migration (using Internal Revenue Service tax return data for ages 0–64, Medicare enrollment data for ages 65+, Social Security Administration's Numerical Identification File (NUMIDENT) data for all ages, and data on changes in the group quarters population); and net international migration (using data on foreign-born immigration, foreign-born emigration, migration between the United States and Puerto Rico, native-born migration, and the movement of the Armed Forces population to and from overseas) (U.S. Census Bureau 2021a). County estimates are controlled to add to the Cen-sus Bureau's national population estimate and state estimates are calculated as the sum of each state's county estimates. Subcounty estimates are developed separately for the household and group quarters populations, which are then combined to calculate the resident population. The household population is

Table 7. Comparison of Population Estimation Errors, BEBR and U.S. Census Bureau, 1980–2020

	Year	Percent Error BEBR	Percent Error USCB
State	1980	-2.7	-5.6
	1990	1.6	0.3
	2000	-1.8	-4.4
	2010	-0.2	-0.9
	2020	0.3	0.7

	Year	MAPE BEBR	MAPE USCB	MALPE BEBR	MALPE USCB
Counties	1980	5.4	5.7	-2.9	-5.1
	1990	5.4	4.9	3.3	2.7
	2000	4.2	5.5	0.8	-5.1
	2010	2.7	3.2	0.5	-1.8
	2020	2.8	2.9	1.0	1.6

	Year	MAPE BEBR	MAPE USCB	MALPE BEBR	MALPE USCB
Subcounty Areas	1980	14.4	15.7	3.5	—
	1990	11.9	—	6.0	—
	2000	10.4	16.1	2.3	4.2
	2010	9.2	9.6	2.0	2.1
	2020	6.8	8.3	1.7	3.3

estimated by applying a "Distributive Housing Unit Method" to the county-level household population to distribute it to each subcounty area. The subcounty estimates are controlled to the county estimates (U.S. Census Bureau 2021b).

The Census Bureau estimate for Florida on April 1, 2020 was 21,688,239, which was 150,052 above the census count of 21,538,187. This estimate was quite accurate by most standards, but the error was about two-and-a-half times larger than BEBR's error of 57,881.

Table 7 provides a summary of BEBR and Census Bureau estimation errors from 1980 to 2020. At the state level, the BEBR estimates were more accurate in four of the five years; the greater accuracy of the BEBR estimates was particularly notable in 1980 and 2000. At the county level, the BEBR estimates were more precise and less biased than the Census Bureau estimates in every year except 1990. The Census Bureau did not release data on subcounty estimates in 1990, but the BEBR estimates had smaller MAPEs and MALPEs in all years for which comparable data were available.

Why were BEBR estimates more accurate than those produced by the Census Bureau? There are several possible explanations. First, the Census Bureau's state and local estimates are controlled to its national population estimate; as a result, errors at the national level carry over to state and local levels. Second, the AR method used by the Census Bureau for county estimates may not be as accurate as the HU method (at least, in Florida); several studies have reported smaller errors for estimates based on the HU method than for estimates based on the AR method (Smith 1986; Smith & Mandell 1984). Third, the Census Bureau's application of the HU method relies solely on building permit data, whereas BEBR's relies primarily on electric customer data. Several studies have found that electric customer data generally provide more accurate estimates of households than do building permit data (Smith and Cody 1994, 2004, 2013). Fourth, the Census Bureau is restricted to using data sources that are available every-where because it makes population estimates for all cities and counties in the United States. BEBR, on the other hand, makes estimates only for Florida and can use any type of data it chooses. This greater flexibility allows BEBR to draw on a greater variety of data sources than the Census Bureau. Finally, the application of professional judgment based on BEBR's knowledge of local population dynamics and data idiosyncrasies may have improved the accuracy of its estimates. Any (or all) of these factors may have played a role in the greater accuracy of the BEBR estimates.

We close with the observation that bias in the BEBR and Census Bureau population estimates for 1980–2020 was almost always in the same direction. At the state and subcounty levels, estimate bias was in the same direction in all five years; years in which the BEBR estimates tended to be too high were also years in which the Census Bureau estimates were too high, and vice versa. Bias in the county estimates was also in the same direction in 1980, 1990, and 2020, but differed in 2000 and 2010 when the Census Bureau estimates had a negative bias while the BEBR estimates had a small positive bias. Overall, though, the similarities in the direction of bias between the two sets of estimates stand out. Despite using different estimation methodologies, bias tended to be in the same direction in all five decades. This is an interesting finding because, in contrast to the precision of population estimates, which is somewhat predictable and is associated with area characteristics such as population size and growth rate, estimation bias is more unpredictable and generally unknowable ahead of time.

Conclusion

The BEBR population estimates for April 1, 2020 were very accurate at the state level, for counties, and for subcounty areas in Florida. The BEBR estimate for the state came within 0.3% of the Census 2020 count, the county estimates had a MALPE of 1.0% and a MAPE of 2.8%, and the subcounty estimates had a MALPE of 1.7% and a MAPE of 6.8%. These levels of accuracy are as good as – and in many cases better than – those achieved in the past. Estimation errors did vary by population size and growth rate; the relationships were generally similar to those found in previous evaluations of BEBR estimates. The accuracy of the BEBR estimates for April 1, 2020 also compared favorably to estimates produced by the U.S. Census Bureau, with the BEBR estimates having lower MAPEs and MALPEs at all three levels of geography.

BEBR has developed annual population estimates for Florida, its counties, and subcounty areas for almost fifty years. Evaluated against the decennial census, the BEBR estimates for the state, counties, and subcounty areas in Florida have proven more precise and less biased than the population estimates produced by the U.S. Census Bureau in four of the past five decades. The results of our accuracy evaluations demonstrate that the BEBR estimation methodology is well suited to providing high quality population estimates for Florida. Over the past decade, we have investigated improvements to our estimation models and incorporated new data items such as minimum use electric customer data, trends in occupancy and average household size from the American Community Survey, and GIS-based property parcel data. We will continue to look for improvements and potential new data sources over the coming decade.

ACKNOWLEDGEMENT

Funding for this report was provided by the Florida Legislature.

REFERENCES

Armstrong, J.S. 2001. Combining Forecasts, pp. 417–439 in J.S. Armstrong (ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Harper, G., Devine, J. and Coleman, C. 2001. Evaluation of 2000 Subcounty Population Estimates. Paper presented at the annual meeting of the Southern Demographic Association, Miami Beach.

Hoque, N. 2012. Evaluation of Small Area Population Estimates Produced by Housing Unit, Ratio-Correlation, and Component Method II Compared to 2000 Census Counts. *Canadian Studies in Population* 39 (1-2), 91–108.

Lowe, T.J., Myers, W.R. and Weisser, L.M. 1984. A Special Consideration in Improving Housing Unit Estimates: The Interaction Effect. Paper presented at the annual meeting of the Population Association of America, Minneapolis, MN.

Murdock, S. H. and Ellis, D. R. 1991. *Applied Demography*. Boulder: Westview Press.

Rayer, S. 2008. Population Forecast Errors: A Primer for Planners. *Journal of Planning Education and Research* 27, 417–430.

Siegel, J. S. 2002. *Applied Demography*. San Diego: Academic Press.

Smith, S. K. 1986. A Review and Evaluation of the Housing Unit Method of Population Estimation. *Journal of the American Statistical Association* 81: 287–296.

Smith, S.K. and Cody, S. 1994. Evaluating the Housing Unit Method: A Case Study of 1990 Population Estimates in Florida, *Journal of the American Planning Association* 60: 209–221.

Smith, S.K. and Cody, S. 2004. An Evaluation of Population Estimates in Florida: April 1, 2000, *Population Research and Policy Review* 23: 1–24.

Smith, S.K. and Cody, S. 2013. Making the Housing Unit Method Work: An Evaluation of 2010 Population Estimates in Florida, *Population Research and Policy Review* 32: 221–242.

Smith, S.K. and Mandell, M. 1984. A Comparison of Population Estimation Methods: Housing Unit versus Component II, Ratio Correlation, and Administrative Records, *Journal of the American Statistical Association* 79: 282–289.

Starsinic, D.E. and Zitter, M. 1968. Accuracy of the Housing Unit Method in Preparing Population Estimates for Cities, *Demography* 5: 475-484.

Tayman, J. and Swanson, D.E. 2012. *Subnational Population Estimates*. The Springer Series on Demographic Methods and Population Analysis. Dordrecht: Springer.

U.S. Census Bureau. 2021a. Methodology for the United States Population Estimates: Vintage 2020. Nation, States, Counties, and Puerto Rico – April 1, 2010 to July 1, 2020. (https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2020/methods-statement-v2020-final.pdf).

U.S. Census Bureau. 2021b. Methodology for Subcounty Total Resident Population Estimates (Vintage 2020): April 1, 2010 to July 1, 2020. (https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2020/2020-subco-meth_final.pdf).

Bureau of Economic and Business Research
College of Liberal Arts and Sciences
UNIVERSITY *of* FLORIDA

Ayers Technology Plaza
720 SW 2nd Ave, Suite 150
P.O. Box 117148
Gainesville, Florida 32611-7148
Phone: 352.392.0171
Fax: 888.534.2404
www.bebr.ufl.edu