# An Evaluation of Hispanic Population Estimates*

Stanley K. Smith, *University of Florida*

June M. Nogle, *University of Florida*

*Objective.* Estimates of the Hispanic population have traditionally been based on historical trends, ratios, or some variant of the cohort-component method. In this article, we describe and test a methodology in which estimates of the Hispanic population are based on symptomatic indicators of population change such as births, deaths, and school enrollments. *Methods.* Using a variety of techniques, we develop Hispanic population estimates for counties in Florida. We evaluate the accuracy of those estimates by comparing them with 2000 Census counts. *Results.* Hispanic population estimates have larger errors than estimates of total population; errors vary considerably by population size and growth rate; some techniques perform better than others in places with particular population characteristics; and averages often perform better than individual techniques. *Conclusions.* In many circumstances, symptomatic data series can provide more accurate estimates of the Hispanic population than more commonly used techniques.

Information on the Hispanic population is used for analyzing socioeconomic and demographic trends, making business decisions, targeting government programs, allocating public funds, and many other purposes. For example, recent studies have used Hispanic data to evaluate radio and television markets (Raymond, 2002), assess political representation (Meier, Polinard, and Wrinkle, 2000), analyze mortality risks (Hummer et al., 1999), measure income inequality (e.g., Cotter, Hermsen, and Vanneman, 1999), and investigate the incidence of racial/ethnic discrimination in the home insurance market (Klein and Grace, 2001). These data frequently come from the decennial census or large sample surveys. In many states and local areas, however, high growth rates cause census data to become quickly outdated and sample surveys are not representative of the relevant population. Consequently, there is a pressing need for accurate, up-to-date subnational estimates of the Hispanic population. How can such estimates be made?

State and local estimates of the Hispanic population are typically based on historical trends, ratios, or some variant of the cohort-component method

*Direct correspondence to Stanley K. Smith, Bureau of Economic and Business Research, 221 Matherly Hall, University of Florida, Gainesville, FL 32611-7145 ⟨sksmith@ufl.edu⟩. The authors will share all data and coding information with anyone wishing to replicate this study.

SOCIAL SCIENCE QUARTERLY, Volume 85, Number 3, September 2004

(e.g., U.S. Census Bureau, 2000; Texas State Data Center, 1996). Although these techniques often produce reasonable estimates, they do not pick up the effects of postcensal changes in location-specific demographic trends. An alternative approach is to develop estimates using variables that are expected to reflect such changes (e.g., Smith and Nogle, 1997; State of California, 2001). To our knowledge, no evaluations of the estimation accuracy of either approach have yet been published.

In this article, we use birth, death, and school enrollment data to develop Hispanic population estimates for counties in Florida on April 1, 2000. We evaluate the accuracy of these estimates by comparing them with 2000 Census counts. To provide a wider perspective, we also evaluate the accuracy of estimates based on several alternative techniques. We close with a discussion of the accuracy of Hispanic population estimates, the value of symptomatic indicators of Hispanic population change, the use of averages, the role of professional judgment, and opportunities for future research.

## Hispanic Population Growth

The Hispanic population of the United States has grown rapidly in recent decades (U.S. Census Bureau, 1993, 2001a). The 2000 Census counted 35.3 million Hispanic residents, compared to 22.4 million in 1990, 14.6 million in 1980, and 9.1 million in 1970. In 2000, Hispanics accounted for 12.5 percent of the national population, compared to only 4.5 percent in 1970. The Hispanic population grew by 61 percent during the 1970s, by 53 percent during the 1980s, and by 58 percent during the 1990s. These rates were much higher than those for the population as a whole.

The Hispanic population is not evenly distributed throughout the nation. In 2000, California had the largest number of Hispanic residents (11.0 million), followed by Texas (6.7 million), New York (2.9 million), Florida (2.7 million), and Illinois (1.5 million). Hispanics accounted for the largest proportions of total population in New Mexico (42.1 percent), California (32.4 percent), Texas (32.0 percent), Arizona (25.3 percent), and Nevada (19.7 percent). At the other end of the spectrum, Maine, North Dakota, and Vermont had fewer than 10,000 Hispanic residents each. The Hispanic population accounted for less than 2 percent of the total population in Alabama, Kentucky, Maine, Mississippi, New Hampshire, North Dakota, Ohio, South Dakota, Vermont, and West Virginia (Guzman, 2001).

The highest Hispanic growth rates during the 1990s occurred in states not usually thought of as "Hispanic." Of the 10 states with the most rapidly growing Hispanic populations, seven were in the South (Alabama, Arkansas, Georgia, Kentucky, North Carolina, South Carolina, and Tennessee), two were in the North Central region (Minnesota and Nebraska), and one was in the West (Nevada). Hispanic growth rates in these states ranged from 155 percent for Nebraska to 394 percent for North Carolina. Of the 10 states with the largest Hispanic populations, none ranked among the 10 with the highest Hispanic growth rates (U.S. Census Bureau, 2001a).

The Hispanic population of the United States is quite diverse. In 2000, 58.5 percent was Mexican, 9.6 percent was Puerto Rican, 3.5 percent was Cuban, and 28.4 percent was of some other origin. These subgroups differ considerably on a variety of cultural, linguistic, socioeconomic, and demographic characteristics. This article's focus on the Hispanic population as a whole is not intended to mask the diversity found within the Hispanic population itself.

**Hispanics in Florida**

The Hispanic population is growing even more rapidly in Florida than in the nation as a whole, with growth rates of 112 percent during the 1970s, 83 percent during the 1980s, and 70 percent during the 1990s (U.S. Census Bureau, 1972, 1982, 1992, 2001b). The Hispanic share of Florida's population nearly tripled between 1970 and 2000, growing from 6.0 percent to 16.8 percent. Florida's 2.7 million Hispanic residents in 2000 constituted the fourth largest Hispanic population of any state.

The composition of Florida's Hispanic population differs dramatically from the national composition. In 2000, 13.4 percent was Mexican, 18.0 percent was Puerto Rican, 31.1 percent was Cuban, and 37.4 percent was of some other origin (Colombia, Nicaragua, and the Dominican Republic were the top three). Although Florida has more than twice as many residents of Cuban origin than all other states combined, their share of Florida's Hispanic population has declined considerably in recent decades.

The Hispanic population increased in all Florida counties during the 1990s, but growth rates differed considerably from one county to another. The Hispanic population grew by less than 100 percent in 20 of the state's 67 counties, by 100–200 percent in 37 counties, and by more than 200 percent in 10 counties. For individual counties, Hispanic growth rates ranged from 31 percent in Monroe County to 434 percent in Wakulla County. Monroe County is a slowly growing county in the Florida Keys in which the total population grew by only 2 percent between 1990 and 2000; its 12,553 Hispanic residents accounted for 16 percent of its total population in 2000. Wakulla County is a rapidly growing coastal county in the Florida Panhandle, but its 443 Hispanic residents accounted for only 2 percent of its total population in 2000.

Despite generally high growth rates, the Hispanic population accounts for a small proportion of the total population in most Florida counties. In 2000, Hispanic residents accounted for less than 5 percent of the total population in 35 counties and less than 10 percent in 52 counties. Nineteen counties had fewer than 1,000 Hispanic residents; 43 had fewer than 10,000. Miami-Dade County—a highly urbanized county in the southern part of the state—

had by far the largest number of Hispanic residents (1,291,737), representing 57 percent of its total population. At the other end of the spectrum, Dixie County—a largely rural county in the northern part of the state—had only 249 Hispanic residents, representing 1.8 percent of its total population.

Substantial county differences in the size, growth rate, and composition of the Hispanic population—along with the availability of the necessary data—make Florida an excellent testing ground for evaluating alternative population estimation techniques.

## Methodology

The Census Bureau uses the cohort-component method to produce state estimates by age, sex, race, and Hispanic origin (U.S. Census Bureau, 2000). Under this method, birth, death, and migration rates are applied to each age/sex/race/ethnicity cohort to produce estimates of the demographic composition of the population. These rates are typically based on historical birth, death, and migration data. However, the lack of location-specific data makes it difficult to apply this method at the county level. As a result, the Census Bureau makes county-level Hispanic estimates using a "top-down" methodology that combines elements from several different methods (U.S. Census Bureau, 2000). First, data from the most recent census are used to calculate the proportion of each county's population falling into each age/sex/race/ethnicity category (e.g., white Hispanic males age 25–29 as a proportion of total county population). Second, these proportions are applied to the current estimate of total population for each county, as determined by the Census Bureau's Population Division. Finally, the resulting estimates are controlled to the state's age/sex/race/ethnicity estimates, with the additional constraint that county totals remain constant.

These estimates do not incorporate any data that directly reflect postcensal changes in the Hispanic population at the county level. The same is true for other estimates based on the extrapolation of historical trends or ratios. Given the highly variable rates of population growth observed at the county level, this poses a serious problem in many counties. To deal with this shortcoming, we have developed a methodology that utilizes symptomatic indicators of postcensal changes in the Hispanic population.

We investigated several potential indicators. Some were found to be incomplete or unreliable in Florida (e.g., driver license records, Hispanic surname indexes), but three appeared to be reliable and were available annually for all counties in the state.

1. Hispanic and total births.
2. Hispanic and total deaths.
3. Hispanic and total school enrollments in grades K–12.

These variables reflect changes in three different segments of the population. Births occur mostly to females aged 15–44. Deaths are distributed throughout the population, but are most heavily concentrated in the older age groups. School enrollments are composed almost entirely of children aged 5–19. Although these variables do not cover every segment of the population, they cover several major segments and are strongly correlated with population size.

We made estimates of the Hispanic population of each county using seven techniques based on changes in birth, death, and school enrollment data between 1990 and 2000. For comparison purposes, we also made estimates using four techniques that extrapolated 1980–1990 population trends; five averages based on various combinations of techniques; one technique that incorporated the application of professional judgment; the Census Bureau's "top-down" ratio technique; and two techniques that hold population values constant at previous levels. These techniques are described in Table 1.

Techniques 1–3 apply 1990–2000 growth rates in Hispanic births, deaths, and school enrollments to the 1990 Hispanic population. Techniques 4–6 assume that the Hispanic share of total population changes at the same rate as the Hispanic share of births, deaths, and school enrollments between 1990 and 2000. Technique 7 is based on the assumption that the relationship between the Hispanic population growth rate and the Hispanic school enrollment growth rate calculated for 1980–1990 remains constant over time. Techniques 8–11 extrapolate 1980–1990 Hispanic growth trends forward to 2000 (the SHARE technique was not applied in five counties in which Hispanic and total population growth had opposite signs). Techniques 12–16 are averages of various combinations of estimates from the first 11 techniques.

Technique 17 incorporates the application of professional judgment. We examined each data series for each county and considered how well the estimates fit with each other, with estimates based on extrapolation techniques, and with independently produced estimates of total population. We asked ourselves questions such as: Do birth, death, and school enrollment data provide similar estimates, or does one of the three variables provide an estimate that differs considerably from the estimates provided by the other two? Are the data series based on a reasonably large number of observations? How consistent are the estimates with previous trends and with overall population growth in the county? Based on our answers to these questions, we chose the estimate we believed would be the most accurate for each county. Sometimes this estimate was based on birth, death, or school enrollment data; sometimes it was based on one of the extrapolation techniques; and sometimes it was based on one of the averages. These judgmental estimates were made without reference to the 2000 Census results.

Technique 18 holds the number of Hispanic residents constant at 1990 levels. Technique 19 holds Hispanic residents as a share of the total population constant at 1990 levels. Technique 20 is based on the Census Bu-

TABLE 1

Description of Estimation Techniques

| | |
|---|---|
| 1. SCHOOL-%CHANGE | The percent change in Hispanic school enrollment from 1990 to 2000 is applied to the 1990 Hispanic population. |
| 2. BIRTH-%CHANGE | The percent change in Hispanic births from 1990 to 2000 is applied to the 1990 Hispanic population. |
| 3. DEATH-%CHANGE | The percent change in Hispanic deaths from 1990 to 2000 is applied to the 1990 Hispanic population. |
| 4. SCHOOL-SHARE | The percent change in the Hispanic share of total school enrollments between 1990 and 2000 is applied to the Hispanic share of total population in 1990 and multiplied by the 2000 estimate of total population. |
| 5. BIRTH-SHARE | The percent change in the Hispanic share of total births between 1990 and 2000 is applied to the Hispanic share of total population in 1990 and multiplied by the 2000 estimate of total population. |
| 6. DEATH-SHARE | The percent change in the Hispanic share of total deaths between 1990 and 2000 is applied to the Hispanic share of total population in 1990 and multiplied by the 2000 estimate of total population. |
| 7. SCHOOL-RATIO | The ratio of the Hispanic population growth rate 1980–1990/Hispanic school enrollment growth rate 1980–1990 is applied to the school enrollment growth rate from 1990 to 2000 and multiplied by the 1990 Hispanic population. This ratio was truncated at 0.1 and 2.0 to reduce the impact of outliers. |
| 8. LINEAR | The average annual numerical change in the Hispanic population 1980–1990 is extrapolated to 2000. |
| 9. EXPONENTIAL | The average annual percentage change in the Hispanic population 1980–1990 is extrapolated to 2000. |
| 10. SHARE | Hispanic population change 1980–1990 as a share of total population change 1980–1990 is applied to total population change estimated for 1990–2000 and added to the 1990 Hispanic population. |
| 11. SHIFT | The average annual change in the Hispanic share of total population between 1980 and 1990 is extrapolated to 2000 and applied to the estimate of total population in 2000. |
| 12. AVE-1 | The average of estimates from Techniques 1–3. |
| 13. AVE-2 | The average of estimates from Techniques 4–6. |
| 14. AVE-3 | The average of estimates from Techniques 1–7. |
| 15. AVE-4 | The average of estimates from Techniques 8–11. |
| 16. AVE-5 | The average of estimates from Techniques 1–11. |
| 17. JUDGMENT | The estimate judged most likely to provide an accurate prediction of the Hispanic population in 2000. |
| 18. CONSTANT-POP | The Hispanic population is assumed to be the same in 2000 as in 1990. |
| 19. CONSTANT-SHARE | The Hispanic population is assumed to constitute the same proportion of total population in 2000 as in 1990. |
| 20. USCB | Census Bureau estimates of the Hispanic population as a proportion of the total population in 1999 are projected forward to 2000, and multiplied by the Census Bureau estimates of the total population in 2000. |

reau's 1999 estimates for Florida counties (U.S. Census Bureau, 2000). Since the Census Bureau did not release a set of county Hispanic estimates for 2000, we developed a set by calculating the number of Hispanics as a proportion of total population in 1999 for each county, extrapolating those proportions forward to 2000, and multiplying them by the Census Bureau's 2000 county estimates of total population. For all other techniques requiring postcensal county estimates of total population (Techniques 4, 5, 6, 10, 11, and 19), we used the estimates produced by the Bureau of Economic and Business Research at the University of Florida.

## Empirical Results

We compared the estimates from each of these 20 techniques with 2000 Census counts for the state and each county. We refer to the resulting differences as estimation errors, although they may have been caused partly by errors in the census counts themselves. Nationally, the Hispanic undercount was estimated as 2.9 percent in 2000, down from 5.0 percent in 1990 (Guzman and McConnell, 2002). We did not attempt to account for changes in census coverage because direct estimates of census coverage are not available for counties; rather, they must be estimated indirectly from survey data for larger places. Although changes in census coverage may have contributed to the estimation errors calculated for individual counties, we believe the net effect of those changes was small when aggregated over all counties and—compared to other causes of error—had relatively little impact on the results reported here.

The errors are summarized in Table 2. The mean absolute percentage error (MAPE) is the average when the direction of error is ignored. It is a measure of precision, or how close the estimates were to census counts regardless of whether they were too high or too low. The mean algebraic percentage error (MALPE) is the average when the direction of error is accounted for. It is a measure of bias, or the tendency for estimates to be too high or too low. These measures have been widely used for evaluating the precision and bias of population estimates (e.g., Davis, 2001; Shahidullah and Flotow, 2001; Smith and Cody, 1994; Smith and Mandell, 1984).

At the state level, estimates were about evenly split between those that were too low and those that were too high, with errors ranging from −41.3 to 27.8 percent. The smallest errors were for DEATH-%CHANGE (−0.3 percent), BIRTH-SHARE (−0.8 percent), and JUDGMENT (−0.8 percent). The three averages based on symptomatic indicators of population change had errors ranging from 4.0 to 6.4 percent, whereas the average based on extrapolation techniques had an error of −9.2 percent. The average of all 11 techniques, however, had a substantially smaller error than any of the other averages and most of the individual techniques (1.7 percent). The USCB estimate had an error of –9.8 percent. This was the only

TABLE 2

Percentage Errors for Hispanic Estimates, by Technique

| Technique | State % Error | County MAPE | County MALPE |
|---|---|---|---|
| SCHOOL-%CHANGE | 27.8 | 58.6 | 49.4 |
| BIRTH-%CHANGE | −15.6 | 33.4 | 0.1 |
| DEATH-%CHANGE | −0.3 | 41.7 | 3.6 |
| SCHOOL-SHARE | 11.3 | 64.2 | 57.8 |
| BIRTH-SHARE | −0.8 | 39.5 | 22.0 |
| DEATH-SHARE | 3.9 | 38.3 | −1.4 |
| SCHOOL-RATIO | 18.4 | 46.6 | 11.1 |
| LINEAR | −14.6 | 35.3 | −34.8 |
| EXPONENTIAL | 18.2 | 38.9 | 3.5 |
| SHARE | −19.8 | 42.7 | −17.6 |
| SHIFT | −9.3 | 29.4 | −26.3 |
| AVE-1 | 4.0 | 42.2 | 24.2 |
| AVE-2 | 4.8 | 47.5 | 35.4 |
| AVE-3 | 6.4 | 40.1 | 24.8 |
| AVE-4 | −9.2 | 28.4 | −21.7 |
| AVE-5 | 1.7 | 26.4 | 6.4 |
| JUDGMENT | −0.8 | 21.2 | −8.3 |
| CONSTANT-POP | −41.3 | 55.4 | −55.4 |
| CONSTANT-SHARE | −28.8 | 43.3 | −43.3 |
| USCB | −9.8 | 26.3 | −11.0 |

estimate based on a state-level model; for the other 19 techniques, state-level estimates were calculated as the sum of county-level estimates. For most techniques, errors were considerably larger than the 1 to 3 percent errors normally found for estimates of the total population of states (e.g., Long, 1993).

We also evaluated the 1999 USCB Hispanic estimates for all states in the United States, extrapolated forward to 2000 (not shown here). The MAPE was 16.9 percent and the MALPE was −15.9 percent. We believe these errors were so large primarily because so many states had small Hispanic populations and high Hispanic growth rates. Undocumented international migration during the 1990s (especially from Mexico) undoubtedly played a role as well, contributing both to the large absolute percentage errors and the strong downward bias of the estimates.

Not surprisingly, errors for Florida counties were larger than errors for the state as a whole. MAPEs ranged from 21.2 to 64.2 percent. The two individual techniques that performed best at the state level (DEATH-%CHANGE and BIRTH-SHARE) did not perform particularly well at the county level; it appears that accurate state estimates for these techniques were the fortuitous result of large positive errors for some counties offsetting large negative errors for others. AVE-4, based on four simple extrapolation techniques, had a smaller MAPE than AVE-1, AVE-2, and AVE-3, which used various combinations of estimates based on birth, death, and school enrollment data. AVE-5—based on all 11 of the individual techniques making up the first four averages—had a smaller MAPE (26.4 percent) than any of the other averages and every individual technique except JUDGMENT (21.2 percent) and USCB (26.3 percent). The two techniques based on assumptions of no change (CONSTANT-POP and CONSTANT-SHARE) performed very poorly, with MAPEs that were larger than all of the averages and all but two of the individual techniques. To put these errors in perspective, MAPEs for county estimates of total population typically fall between 3 and 6 percent (e.g., Davis, 2001; Long, 1993; Shahidullah and Flotow, 2001; Smith and Cody, 1994, 2002).

MALPEs ranged from –55.4 to 57.8 percent. Four techniques had MALPEs of less than 5 percent (ignoring the direction of error). JUDGMENT, USCB, and AVE-5—the techniques with the smallest MAPEs—also had relatively small MALPEs. Except for EXPONENTIAL, all the extrapolation techniques (including those based on no-change assumptions) had a downward bias. Except for DEATH-SHARE, all the techniques based on symptomatic indicators had an upward bias. Given that bias tends to vary from one set of estimates to another (e.g., Davis, 2001; Smith and Cody, 1994, 2002), we believe these patterns reflect the characteristics of this particular data set rather than characteristics of Hispanic population estimates generally.

Based solely on this evidence, it would not appear that techniques incorporating symptomatic indicators of Hispanic population change can—by themselves—improve on the performance of trend or ratio techniques. Estimates based on births, deaths, and school enrollments were not consistently more precise or less biased than estimates based on the extrapolation of historical trends or ratios. However, techniques based on symptomatic indicators did improve the precision and reduce the bias of trend and ratio extrapolation estimates when averaged together with those estimates (AVE-5). In addition, estimates based on birth, death, and school enrollment data provided critical inputs into the judgmental estimates, which had the best overall performance of any of the 20 techniques. In combination with other techniques, then, symptomatic indicators *did* lead to more accurate population estimates.

Before we can draw any firm conclusions, however, it is important to consider the potential impact of differences in population size and growth rate. As shown in Table 3, MAPEs generally declined as population size increased. Often, the differences were very large. For SCHOOL-SHARE, for example, the MAPE for the smallest size category was more than 10 times larger than the MAPE for the largest category. For most techniques, MAPEs were greater than 40 percent for counties with fewer than 1,000 Hispanic residents in 1990 and less than 20 percent for counties with 25,000 or more. Indeed, JUDGMENT and three of the averages had errors of less than 10 percent for counties in the top size category. A negative

TABLE 3

Mean Absolute Percentage Errors by Number of Hispanic Residents in 1990

| Technique | Number of Hispanic Residents | | | | | |
|---|---|---|---|---|---|---|
| | <1,000 | 1,000–4,999 | 5,000–9,999 | 10,000–24,999 | 25,000+ | Total |
| SCHOOL-%CHANGE | 90.3 | 38.0 | 41.6 | 38.9 | 33.8 | 58.6 |
| BIRTH-%CHANGE | 48.9 | 31.3 | 22.3 | 20.7 | 18.2 | 33.4 |
| DEATH-%CHANGE | 68.4 | 33.9 | 59.9 | 13.7 | 12.5 | 41.7 |
| SCHOOL-SHARE | 116.7 | 33.6 | 36.4 | 31.3 | 11.0 | 64.2 |
| BIRTH-SHARE | 53.0 | 40.0 | 29.4 | 30.8 | 15.3 | 39.5 |
| DEATH-SHARE | 62.8 | 30.4 | 53.2 | 15.1 | 12.9 | 38.3 |
| SCHOOL-RATIO | 68.0 | 43.0 | 33.3 | 21.0 | 17.8 | 46.6 |
| LINEAR | 46.7 | 27.1 | 30.1 | 30.9 | 22.9 | 35.3 |
| EXPONENTIAL | 45.9 | 43.1 | 13.9 | 49.5 | 13.4 | 38.9 |
| SHARE | 63.1 | 29.6 | 34.0 | 36.4 | 25.0 | 42.7 |
| SHIFT | 39.2 | 24.0 | 25.5 | 22.7 | 17.0 | 29.4 |
| AVE-1 | 70.5 | 28.6 | 29.9 | 18.1 | 9.1 | 42.2 |
| AVE-2 | 81.9 | 29.9 | 31.1 | 20.3 | 10.2 | 47.5 |
| AVE-3 | 66.0 | 29.1 | 29.3 | 16.5 | 7.4 | 40.1 |
| AVE-4 | 40.3 | 22.6 | 22.8 | 18.6 | 15.1 | 28.4 |
| AVE-5 | 42.3 | 19.9 | 17.8 | 13.3 | 5.7 | 26.4 |
| JUDGMENT | 30.0 | 20.5 | 13.8 | 11.8 | 7.6 | 21.2 |
| CONSTANT-POP | 59.8 | 52.3 | 51.5 | 56.3 | 48.2 | 55.4 |
| CONSTANT-SHARE | 47.6 | 38.3 | 41.7 | 45.5 | 37.8 | 43.3 |
| USCB | 34.6 | 23.4 | 21.5 | 19.9 | 13.3 | 26.3 |
| Number of Counties | 26 | 18 | 9 | 9 | 5 | 67 |

relationship between MAPEs and population size has been reported in many studies of population estimation errors (e.g., Davis, 2001; Galdi, 1985; Harper, Devine, and Coleman, 2001; Smith and Cody, 1994, 2002).

More critical to our analysis is the finding that the precision of estimates based on birth, death, and school enrollment data—relative to that of trend and ratio techniques—increased dramatically with population size. For counties with fewer than 1,000 Hispanic residents, the MAPE for AVE-3 (based on symptomatic indicators) was larger than the MAPEs for AVE-4 (based on extrapolation techniques), CONSTANT-POP, CONSTANT-SHARE, and USCB. For counties with 1,000–9,999 Hispanic residents, the MAPE for AVE-3 was larger than the MAPEs for AVE-4 and USCB, but smaller than the MAPEs for CONSTANT-POP and CONSTANT-SHARE. For counties with 10,000 or more Hispanic residents, the MAPE for AVE-3 was smaller than the MAPEs for all the other techniques. It appears that a certain population size must be reached before variables such as births, deaths, and school enrollments can provide more precise estimates than can be obtained by holding ratios constant or extrapolating past trends.

It should be noted that AVE-5—based on the seven symptomatic indicator estimates *and* the four trend extrapolation estimates—produced lower MAPEs than any of the other averages for all size categories except the smallest. JUDGMENT also performed consistently well, with the smallest MAPE of any technique in three categories, the second smallest in one category, and the third smallest in one category. When used in combination with other techniques, then, the symptomatic indicators improved the precision of the estimates even in the smaller size categories.

As shown in Table 4, there was no clear relationship between MALPEs and population size. For some techniques, MALPEs changed from positive to negative as population size increased; for others, they changed from negative to positive. Sometimes they had the same sign in all size categories and sometimes they fluctuated between positive and negative signs. These results are consistent with those reported in several other studies (e.g., Shahidullah and Flotow, 2001; Smith and Cody, 1994, 2002). We do not believe that differences in population size have a consistent impact on the tendency for Hispanic population estimates to be too high or too low.

We also evaluated the impact of differences in Hispanic population growth rates on estimation errors (not shown here). Many studies have found MAPEs to increase as growth rates deviate in either direction from moderate but positive levels (e.g., Davis, 2001; Galdi, 1985; Harper, Devine, and Coleman, 2001; Smith and Cody, 1994, 2002). In this study, however, we found the expected relationship only for some of the estimation techniques. For other techniques, MAPEs fluctuated from one growth-rate

TABLE 4

Mean Algebraic Percentage Errors by Number of Hispanic Residents in 1990

| | Number of Hispanic Residents | | | | | |
|---|---|---|---|---|---|---|
| Technique | <1,000 | 1,000–4,999 | 5,000–9,999 | 10,000–24,999 | 25,000+ | Total |
| SCHOOL-%CHANGE | 69.5 | 34.0 | 41.6 | 38.9 | 33.8 | 49.4 |
| BIRTH-%CHANGE | 1.7 | 2.4 | – 4.7 | 5.6 | – 16.5 | 0.1 |
| DEATH-%CHANGE | – 23.1 | 17.3 | 29.2 | – 4.3 | – 2.6 | 3.6 |
| SCHOOL-SHARE | 104.2 | 30.4 | 34.6 | 28.1 | 11.0 | 57.8 |
| BIRTH-SHARE | 23.9 | 26.9 | 17.5 | 27.1 | – 4.5 | 22.0 |
| DEATH-SHARE | – 25.7 | 9.5 | 20.7 | – 8.3 | – 0.2 | – 1.4 |
| SCHOOL-RATIO | 5.2 | 11.7 | 25.4 | 10.4 | 15.4 | 11.1 |
| LINEAR | – 45.8 | – 26.7 | – 30.1 | – 30.9 | – 21.8 | – 34.8 |
| EXPONENTIAL | – 21.0 | 23.5 | – 1.9 | 34.1 | 13.0 | 3.5 |
| SHARE | 7.7 | – 29.6 | – 34.0 | – 36.4 | – 25.0 | – 17.6 |
| SHIFT | – 35.4 | – 18.4 | – 25.5 | – 22.7 | – 15.0 | – 26.3 |
| AVE-1 | 36.7 | 17.9 | 22.0 | 13.4 | 4.9 | 24.2 |
| AVE-2 | 61.6 | 22.2 | 24.3 | 15.6 | 2.1 | 35.4 |
| AVE-3 | 36.9 | 18.9 | 23.5 | 13.9 | 5.2 | 24.8 |
| AVE-4 | – 28.6 | – 15.0 | – 22.8 | – 18.0 | – 14.5 | – 21.7 |
| AVE-5 | 8.1 | 7.4 | 6.6 | 3.8 | – 1.1 | 6.4 |
| JUDGMENT | – 17.6 | – 0.3 | – 3.6 | – 6.6 | – 0.5 | – 8.3 |
| CONSTANT-POP | – 59.8 | – 52.3 | – 51.5 | – 56.3 | – 48.2 | – 55.4 |
| CONSTANT-SHARE | – 47.6 | – 38.3 | – 41.7 | – 45.5 | – 37.8 | – 43.3 |
| USCB | – 10.2 | – 5.8 | – 14.8 | – 18.8 | – 13.2 | – 11.0 |
| Number of Counties | 26 | 18 | 9 | 9 | 5 | 67 |

category to another, following no consistent pattern. This unexpected result was most likely caused by the small number of Hispanic residents in many Florida counties and by the paucity of counties with slow or moderate Hispanic growth rates in the data set.

We did find a strong, consistent relationship between MALPEs and growth rates. For every technique, MALPEs declined as growth rates increased. For some techniques, large positive errors for the most slowly growing counties became small positive errors for the most rapidly growing counties; for other techniques, small negative errors for the most slowly growing counties became large negative errors for the most rapidly growing counties; and for many techniques, positive errors for the most slowly growing counties became negative errors for the most rapidly growing counties. For every technique, then, higher growth rates were associated with a greater tendency to underestimate population growth. This finding is consistent with the results reported in a number of other studies (e.g., Galdi, 1985; Harper, Devine, and Coleman, 2001; Smith and Cody, 1994, 2002).

**Summary and Conclusions**

The Hispanic population is a large and rapidly growing segment of the U.S. population, but no standard methodology for estimating that population has been developed and few (if any) studies have attempted to evaluate the accuracy of Hispanic population estimates. Most estimation techniques have been based on historical trends and ratios rather than on postcensal data specific to the Hispanic population. In this article, we developed a series of estimates based on symptomatic indicators of population change; specifically, births, deaths, and school enrollments. For comparison purposes, we also developed estimates based on a variety of other techniques. We evaluated the precision and bias of these estimates using 2000 Census data for counties in Florida and—based on those evaluations and the literature on population estimation errors—have drawn a number of preliminary conclusions.

1. Errors for estimates of the Hispanic population will generally be larger (sometimes much larger) than errors for estimates of the total population. This is not surprising, given that many places have small numbers of Hispanic residents and high rates of Hispanic population growth. We believe that MAPEs for Hispanic estimates are likely to decline as Hispanic population size increases and as Hispanic growth rates decline from their currently high levels.
2. Estimates of the Hispanic population based on symptomatic indicators of population change will generally be more precise than estimates based on the extrapolation of historical trends and ratios, but only when the number of Hispanic residents is above some threshold level. For small places, estimates based on the extrapolation of historical

trends and ratios may perform at least as well as those based on symptomatic indicators.
3. Averages will often provide more accurate estimates of the Hispanic population than individual techniques. Not only do averages reduce the odds of making large errors, but they typically outperform most individual techniques and sometimes produce smaller errors than *any* individual technique. In this study, averages consistently produced smaller MAPEs than most (or all) of the techniques making up those averages. Similar results have been reported in several other studies (e.g., Armstrong, 2001; Smith and Mandell, 1984).
4. Hispanic population estimates that incorporate professional judgment will often be more precise and less biased than estimates based on the mechanical application of specific estimation techniques. In every size category, we found JUDGMENT to have a smaller MAPE than all (or almost all) of the other techniques. It also performed consistently well with respect to the absolute values of the MALPEs. Several other studies have reported similar results for estimates of total population (e.g., Smith and Cody, 1994, 2002).

   However, it is difficult to draw general conclusions regarding the role of professional judgment. Whose judgment should be used? How should it be applied? Will it always improve the results? Do the benefits of incorporating judgment into the estimation process outweigh the costs? Clearly, many questions remain to be answered.
5. A "composite" approach may prove to be particularly useful for the production of Hispanic population estimates. A composite approach is one in which different techniques are used for estimating (or projecting) the populations of places with differing characteristics (e.g., Isserman, 1977; Smith and Shahidullah, 1995). This approach is based on the assumption that some techniques tend to perform consistently better than others for places with particular characteristics. The data presented in this article suggest that techniques based on symptomatic indicators of population change may provide the best estimates for counties with large numbers of Hispanic residents, whereas techniques based on trend or ratio extrapolation may provide the best estimates for counties with small numbers of Hispanic residents. This approach could be extended to account for a variety of other characteristics as well (e.g., differences in the composition of the Hispanic population). We believe the composite approach holds a great deal of promise for improving the quality of Hispanic (and other) population estimates.

## Future Research

In this article, we described and evaluated a variety of techniques for making Hispanic population estimates. Our conclusions, however, must be viewed as preliminary. Would similar results be found in other states? What other estimation techniques might be developed (e.g., regression models that incorporate the influence of several variables simultaneously)? Which techniques work best in specific circumstances? What other symptomatic indicators of Hispanic population change might be used (e.g., lists of Hispanic surnames)? Can data from the American Community Survey be used to improve the estimation models? Further research is needed before we can answer these questions conclusively.

It may be particularly useful to explore the impact of differences in Hispanic ancestry (e.g., Mexican, Puerto Rican, and Cuban) on population estimates. The Hispanic population is not monolithic; rather, it is an amalgam of subgroups with widely varying characteristics (e.g., age structures, income levels, and mortality, fertility, and migration patterns). Do differences in the composition of the Hispanic population—or changes in composition over time—affect the accuracy of the estimates? Can techniques be developed that account for these differences, or that provide separate estimates for each Hispanic subgroup? The techniques described in this article are promising, but they are merely the first step in the use of symptomatic indicators for the production of state and local Hispanic population estimates.

## REFERENCES

Armstrong, J. Scott. 2001. "Combining Forecasts." Pp. 417–39 in J. Scott Armstrong, ed., *Principles of Forecasting*. Boston, MA: Kluwer Academic Publishers.

Cotter, David A., Joan M. Hermsen, and Reeve Vanneman. 1999. "Systems of Gender, Race, and Class Inequality: Multilevel Analyses." *Social Forces* 78:433–60.

Davis, Sam T. 2001. "Evaluating County Population Estimates: The Big Picture, Hard to Enumerate Counties, and Counties with Special Treatment." Paper presented at the Annual Meeting of the Southern Demographic Association. Miami Beach, FL.

Galdi, David. 1985. *Evaluation of 1980 Subcounty Population Estimates*. Current Population Reports, Series P-25, No. 963. Washington, DC: Government Printing Office.

Guzman, Betsy. 2001. *The Hispanic Population*. U.S. Census Bureau, Census 2000 Brief, C2KBR/01-3. Washington, DC: U.S. Census Bureau.

Guzman, Betsy, and Eileen Diaz McConnell. 2002. "The Hispanic Population: 1990–2000 Growth and Change." *Population Research and Policy Review* 21:109–28.

Harper, Greg, Jason Devine, and Chuck Coleman. 2001. "Evaluation of 2000 Subcounty Population Estimates." Paper presented at the Annual Meeting of the Southern Demographic Association. Miami Beach, FL.

Hummer, Robert A., Richard G. Rogers, Charles B. Nam, and Felicia B. LeClere. 1999. "Race/Ethnicity, Nativity, and U.S. Adult Mortality." *Social Science Quarterly* 80:136–53.

Isserman, Andrew. 1977. "The Accuracy of Population Projections for Subcounty Areas." *Journal of the American Institute of Planners* 43:247–59.

Klein, Robert W., and Martin F. Grace. 2001. "Urban Homeowners Insurance Markets in Texas: A Search for Redlining." *Journal of Risk and Insurance* 68:581–614.

Long, John F. 1993. "Postcensal Population Estimates: States, Counties, and Places." Paper presented at the Annual Meeting of the American Statistical Association. San Francisco, CA.

Meier, Kenneth J., J. L. Polinard, and Robert D. Wrinkle. 2000. "Micheal Giles and Mancur Olson Meet Vincent Ostrom: Jurisdiction Size and Latino Representation." *Social Science Quarterly* 81:123–35.

Raymond, Joan. 2002. "Tienen Numeros?" *American Demographics* 24:22–25.

Shahidullah, Mohammed, and Mark Flotow. 2001. "An Evaluation of the Accuracy of 2000 Population Estimates for Counties in Illinois." Paper presented at the Annual Meeting of the Southern Demographic Association. Miami Beach, FL.

Smith, Stanley K., and Scott Cody. 1994. "Evaluating the Housing Unit Method: A Case Study of 1990 Population Estimates in Florida." *Journal of the American Planning Association* 60:209–21.

———. 2002. "An Evaluation of Population Estimates in Florida: April 1, 2000." Paper presented at the Annual Meeting of the Southern Demographic Association. Austin, TX.

Smith, Stanley K., and Marylou Mandell. 1984. "A Comparison of Population Estimation Methods: Housing Unit Versus Component II, Ratio Correlation, and Administrative Records." *Journal of the American Statistical Association* 79:282–89.

Smith, Stanley K., and June M. Nogle. 1997. "An Experimental Methodology for Estimating Hispanic Residents for States and Counties." *Journal of Economic and Social Measurement* 23:263–75.

Smith, Stanley K., and Mohammed Shahidullah. 1995. "An Evaluation of Population Projection Errors for Census Tracts." *Journal of the American Statistical Association* 90:64–71.

State of California. 2001. *Race/Ethnic Population Estimates: Components of Change for California Counties April 1990 to July 1999*. Sacramento, CA: Department of Finance.

Texas State Data Center. 1996. *Estimates of the Total Population of Counties in Texas by Age, Sex and Race/Ethnicity for July 1, 1995*. College Station, TX: Texas A & M University.

U.S. Census Bureau. 1972. *1970 Census of Population*. PC(1)-D11. Washington, DC: Government Printing Office.

———. 1982. *1980 Census of Population*. PC80-1-B11. Washington, DC: Government Printing Office.

———. 1992. *1990 Census of Population*. CP-1-11. Washington, DC: Government Printing Office.

———. 1993. "Hispanic Americans Today." In *Current Population Reports*. P23-183. Washington, DC: Government Printing Office.

———. 2000. *Estimates of the Population of Counties by Age, Sex, Race and Hispanic Origin: 1990 to 1999*. Internet report. Available at ⟨http://eire.census.gov/popest/archives/county/casrh_doc.tx⟩.

———. 2001a. *Census 2000: Hispanics in the U.S.A.*. Internet report. Available at ⟨http://www.census.gov/mso/www/pres_lib/hisorig/sld001.ht⟩.

———. 2001b. *2000 Census of Population*, Summary File 1. Washington, DC: U.S. Census Bureau.