People Might Move but Housing Units Don't: An Evaluation of the State and
County Housing Unit Estimates

by

Jason Devine and Charles Coleman
Population Division
U.S. Census Bureau
Washington, D.C. 20233

CONTENTS

ABSTRACT

Throughout the 1990s the Population Division of the U.S. Census Bureau produced state and county level housing unit estimates. While the state level estimates were released to the public, the county level estimates were considered experimental. The county level estimates were produced using a component method whereby we begin with the 1990 Census housing unit count and update the Census count using administrative records data on building permits, mobile home shipments, and housing unit loss. The county estimates were summed to obtain the state estimates. The availability of Census 2000 data provides the opportunity to evaluate the accuracy of these estimates.

This paper compares Census 2000 results with April 1, 2000 housing unit estimates using a variety of statistical measures including the Mean Absolute Percent Error and the Mean Algebraic Percent Error. The results of this analysis will be used to inform Census Bureau analysts on ways by which the current housing unit estimates can be improved.

BACKGROUND

This report presents an evaluation of estimates of total housing units for the Nation, states, and counties produced by the Population Division of the Census Bureau. At the national and state level these estimates were released to the public every year in the 1990s except 1997 and 1999. While the county level estimates were not released to the public, they were used in an evaluation of Census 2000 housing unit coverage designed to improve the accuracy of census data (Barron, 2001). The comparison of the April 1, 2000 estimates to the April 1, 2000 decennial census counts forms the basis for this report.

The housing unit estimates were produced using an administrative records based method by adding data on building permits, mobile home shipments, and housing unit loss to the 1990 Census unadjusted housing unit count. These data were collected and processed at the subcounty level for areas such as cities, boroughs, villages, towns, and townships and then summed to the county, state, and national level. See the Appendix for a more detailed explanation of the methodology.

MEASURES OF ACCURACY

For the purposes of this evaluation, the differences between the estimates and the census counts are assumed to be due to errors in the estimates. This paper considers two aspects of quality: bias and accuracy, as measured by the Mean Algebraic Percent Error (MALPE), and Mean Absolute Percent Error (MAPE), respectively.   The MALPE is simply the average of all the percent errors, and differs from the MAPE only in that the MAPE involves taking the absolute values of the percent errors. MALPE is a measure of mean bias and can be used as the basis for testing for the presence of significant mean bias (Coleman, 1999 and 2001). MAPE is a measure of accuracy. That is, it provides a measure of how "close" the estimates get to the truth, on average.

NATIONAL AND STATE LEVEL HOUSING UNIT ESTIMATES

The April 1, 2000 estimate for the Nation of 115,547,749 housing units was 0.3 percent lower than the Census 2000 housing unit count of 115,904,641. The MAPE for all states was 1.5 percent. A similar analysis of the housing unit estimates for the 1980's showed a similar error of 0.3 percent for the nation and a MAPE for all states of 1.8 percent (Prevost, 1994). Half of all states had absolute errors less than 1.0 percent (Table 1, Col. 5). Montana had the highest absolute error of 5.9 percent while Minnesota had the lowest at 0.008 percent. The error varied by number of housing units in the state. Generally, the higher the 1990 housing unit count, the more accurate the estimate. Of the ten states with the largest number of housing units in 1990, all but two, New York (-2.0) and California (1.9), had absolute percent errors lower than 1 percent. The MALPE for all states was -0.9 percent. Thirty-three states had estimates that were below the Census 2000 count, indicating negative bias. These results can also be compared with the
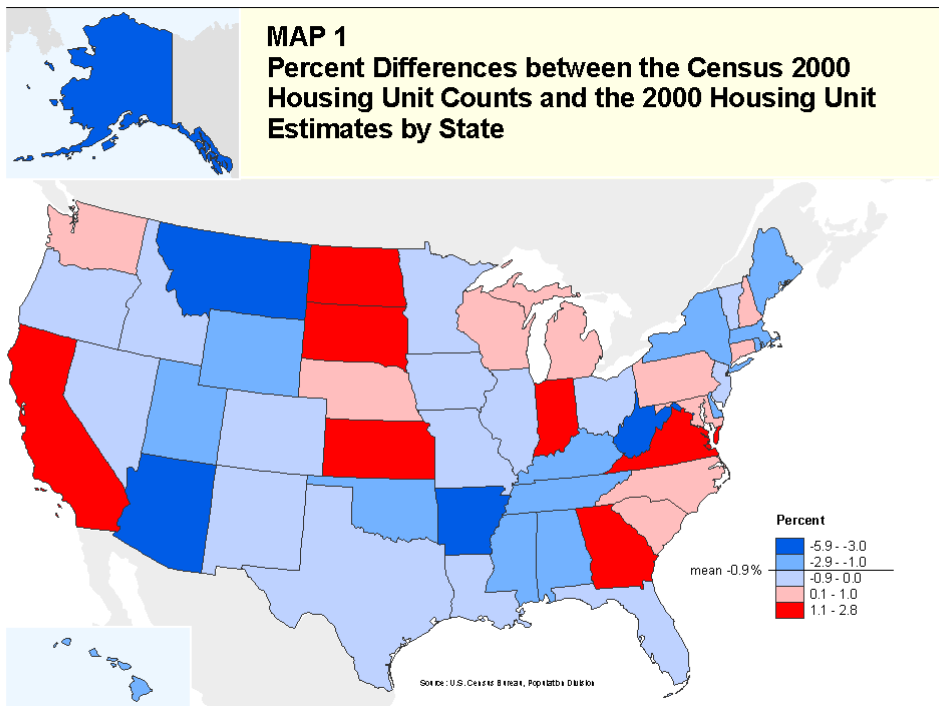
Census Bureau's independently derived 1990 based population estimates which had an error of -2.4 percent for the Nation and a MAPE of 2.6 percent for all states (Davis, 2001).

Table 1. Measures of Error in the April 1, 2000 Housing Unit Estimates by State

State Name	1990 to 2000 Percent Change ((4)-(2))/(2) (1)	Census 1990 Housing Unit Count (2)	April 1, 2000 Housing Unit Estimate (3)	Census 2000 Housing Unit Count (4)	Percent Difference ((3)-(4))/(4) (5)
United States	13.3	102262196	115547749	115904641	-0.3
Alabama	17.6	1670258	1915989	1963711	-2.4
Alaska	12.2	232608	251878	260978	-3.5
Arizona	31.9	1659439	2111282	2189189	-3.6
Arkansas	17.2	1000579	1121396	1173043	-4.4
California	9.2	11182503	12448944	12214549	1.9
Colorado	22.4	1477348	1804536	1808037	-0.2
Connecticut	4.9	1320851	1396296	1385975	0.7
Delaware	18.3	289919	335201	343072	-2.3
District of Columbia	-1.3	278489	262386	274845	-4.5
Florida	19.7	6100248	7262782	7302947	-0.5
Georgia	24.4	2637906	3336316	3281737	1.7
Hawaii	18.1	389811	447488	460542	-2.8
Idaho	27.7	413322	523021	527824	-0.9
Illinois	8.4	4506275	4847481	4885615	-0.8
Indiana	12.7	2246044	2573171	2532319	1.6
Iowa	7.8	1143666	1226337	1232511	-0.5
Kansas	8.3	1044111	1154001	1131200	2.0
Kentucky	16.2	1506930	1707718	1750927	-2.5
Louisiana	7.6	1716229	1835845	1847181	-0.6
Maine	11.0	587045	635294	651901	-2.5
Maryland	13.4	1891665	2146667	2145283	0.1
Massachusetts	6.0	2472710	2594259	2621989	-1.1
Michigan	10.0	3847940	4260370	4234279	0.6
Minnesota	11.8	1848567	2066101	2065946	0.0
Mississippi	15.0	1010421	1136163	1161953	-2.2
Missouri	11.0	2199084	2439544	2442017	-0.1
Montana	14.3	361155	388380	412633	-5.9
Nebraska	9.4	660634	723446	722668	0.1
Nevada	59.5	518778	825925	827457	-0.2
New Hampshire	8.6	503904	549664	547024	0.5
New Jersey	7.6	3075310	3286619	3310275	-0.7
New Mexico	23.5	632058	773301	780579	-0.9
New York	6.3	7227059	7525613	7679307	-2.0
North Carolina	25.0	2818073	3525871	3523944	0.1
North Dakota	4.8	276340	297404	289677	2.7
Ohio	9.4	4371944	4765853	4783051	-0.4
Oklahoma	7.7	1406495	1481939	1514400	-2.1
Oregon	21.7	1193574	1446582	1452709	-0.4
Pennsylvania	6.3	4938225	5296608	5249750	0.9
Rhode Island	6.1	414572	434637	439837	-1.2
South Carolina	23.2	1423771	1756098	1753670	0.1
South Dakota	10.5	292436	327816	323208	1.4
Tennessee	20.4	2026066	2392755	2439443	-1.9
Texas	16.4	7008887	8109855	8157575	-0.6
Utah	28.4	598388	759971	768594	-1.1
Vermont	8.5	271214	292473	294382	-0.6
Virginia	16.3	2496519	2936497	2904192	1.1
Washington	20.6	2032344	2462487	2451075	0.5
West Virginia	8.1	781295	798643	844623	-5.4
Wisconsin	12.9	2055774	2331401	2321144	0.4
Wyoming	10.0	203413	217445	223854	-2.9

The relationship between growth and accuracy was less clear. Some states, such as Nevada, had high growth but low absolute percent errors while other states such as Arizona had high growth and high absolute percent errors. Table 1 shows the percent change in housing units between 1990 and 2000, the 1990 Census count, the April 1, 2000 housing unit estimate and Census 2000 count, and the percent error for each state. Map 1 shows the percent error for each state.

Map 1. Percent Error in the April 1, 2000 Housing Unit Estimates by State

MAP 1
Percent Differences between the Census 2000 Housing Unit Counts and the 2000 Housing Unit Estimates by State

Percent

■	-5.9 - -3.0
■	-2.9 - -1.0
■	-0.9 - 0.0
■	0.1 - 1.0
■	1.1 - 2.8

mean -0.9%

Source: U.S. Census Bureau, Population Division

COUNTY LEVEL HOUSING UNIT ESTIMATES

A comparison of the Census 2000 housing unit counts with the April 1, 2000 housing unit estimates shows that the overall MAPE for all 3,141 counties is 4.6 percent. The overall MALPE is -1.5 percent, which indicates that, on average, the estimates were below the Census 2000 housing unit counts. On average, the counties experienced a 12.0 percent change in housing units between the 1990 Census and Census 2000. For comparison, the county level population estimates had a MAPE of 3.3 percent and a MALPE of -1.6 percent (Davis, 2001).

Of all the counties, 643 had estimates within one percent of the census count. Only 359 counties had estimates that were more than 10 percent different than the Census 2000 count. 1,813 counties had housing unit estimates below the Census 2000 count, 1,327 had estimates that were higher, and one county (Potter County, South Dakota) had an estimate exactly the same as the Census 2000 count. 420

6

counties had housing unit loss between 1990 and 2000. For 374 (89.0 percent) of these counties, the estimates correctly indicated that the county had housing unit loss.

ERRORS BY SIZE AND GROWTH

When considered by population size (Table 2 and Figure 1), the larger counties were found to have smaller MAPEs. This is a pattern found also in prior analyses of county (Davis, 1994) and subcounty population estimates (Galdi, 1985), (Harper, Devine and Coleman, 2001).

Table 2. MAPE of April 1, 2000 County Housing Unit Estimates by Size

1990 Housing Unit Count	Number of Areas	Mean Absolute Percent Error
All	3141	4.6
0-2,499	336	7.3
2,500-4,999	518	5.7
5,000-9,999	749	5.0
10,000-19,999	653	4.4
20,000-49,999	507	3.1
50,000-99,999	178	2.2
≥100,000	200	1.9

Figure 1. MAPE of April 1, 2000 County Housing Unit Estimates by Size Class

7
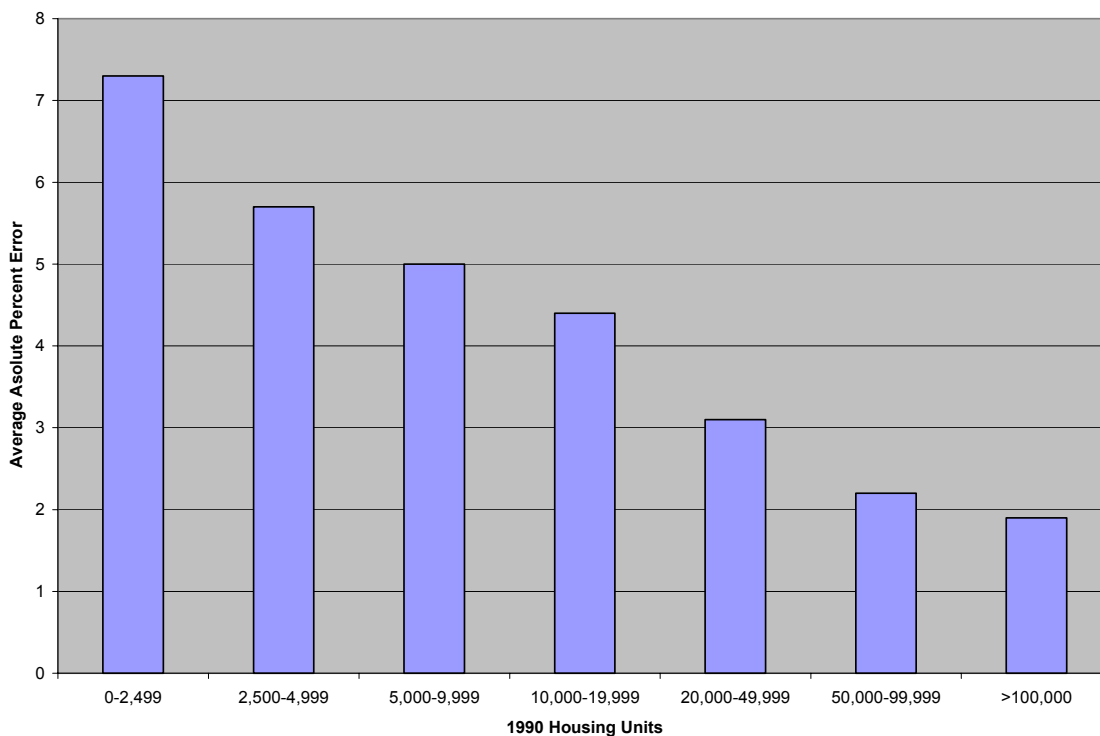
When classified by change in the number of housing units between the 1990 Census and Census 2000 (Table 3 and Figure 2), counties with little change have lower MAPEs than counties with large changes in housing unit counts. This is a common pattern found also in prior analyses of subcounty (Galdi, 1985), (Harper et al., 2001) and county population estimates (Davis, 1994) .

Table 3. MAPE of County Housing Unit Estimates by Change in Housing Units: 1990 to 2000

Percent Change	Number of Areas	Mean Absolute Percent Error
All	3141	4.6
≤ -5	185	8.1
-4.9 to 0.1	231	3.1
0 to 4.9	456	2.7
5 to 9.9	565	2.9
10 to 14.9	500	3.4
15 to 24.9	712	5.5
≥ 25	492	7.0

Figure 2. MAPE of April 1, 2000 County Housing Unit Estimates by Change in Housing Units
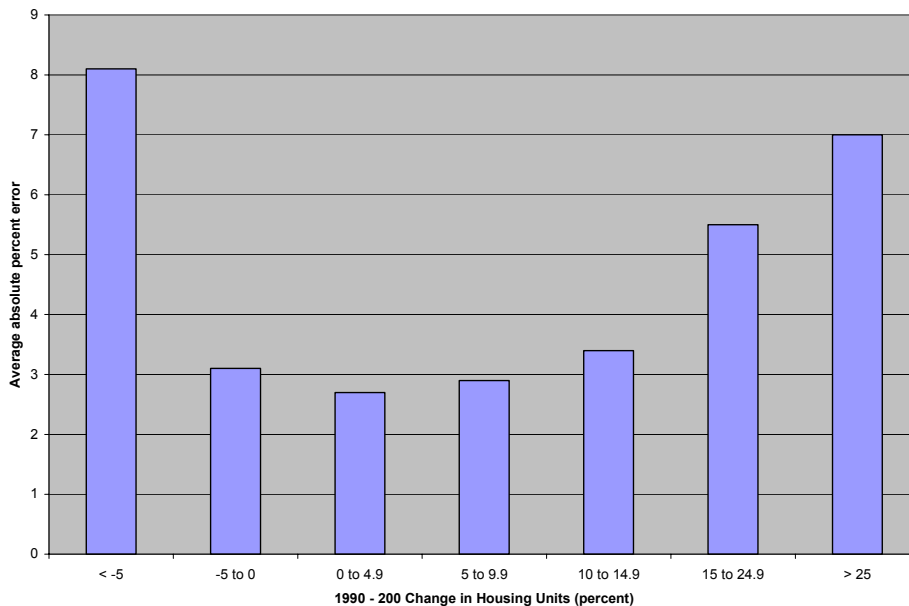

Figure 3 shows a 3-dimensional graph of MAPEs by size and growth categories, the same categories that are used in Figures 1 and 2. Figure 3 shows that the U-shaped curves vary dramatically by size class, achieving troughs at differing growth classes. Moreover, the 100,000+ housing units size class has two troughs: at the 5-9.9 percent and 15-24.9 percent growth classes. The trend for increased accuracy as the number of housing units increases remains generally true for growing counties, but not for declining counties.

Figure 3: MAPE of April 1, 2000 County Housing Unit Estimates by Size and Growth Class
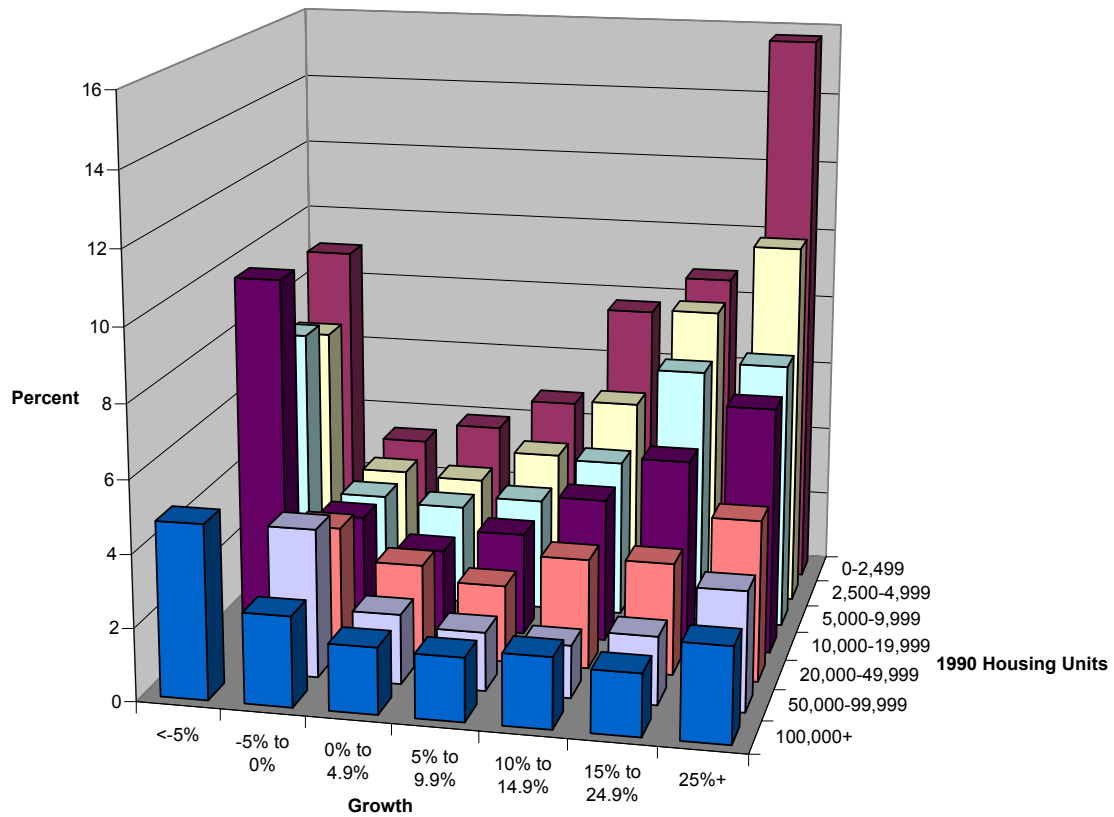

Figure 4 shows MALPE by size and growth class. Again, these classes are the same as in the preceding Figures. Figure 4 shows the origin of the U-shaped curves: growing areas were systematically underestimated while declining areas were systematically overestimated. In the growth classes of 5 percent and more, the magnitude of MALPE systematically declines in size as the 1990 housing unit count increases, which leads to MAPE's general declines in these classes. The growth classes under 5 percent do not show this systematic decline, leading one to suspect that they were generated by different processes. These differences account for the breakdown between size and accuracy in these classes. An important interaction effect shows up dramatically: small size interacts with large growth rates to decrease MALPE.

Figure 4: MALPE of April 1, 2000 County Housing Unit Estimates by Size and Growth Class
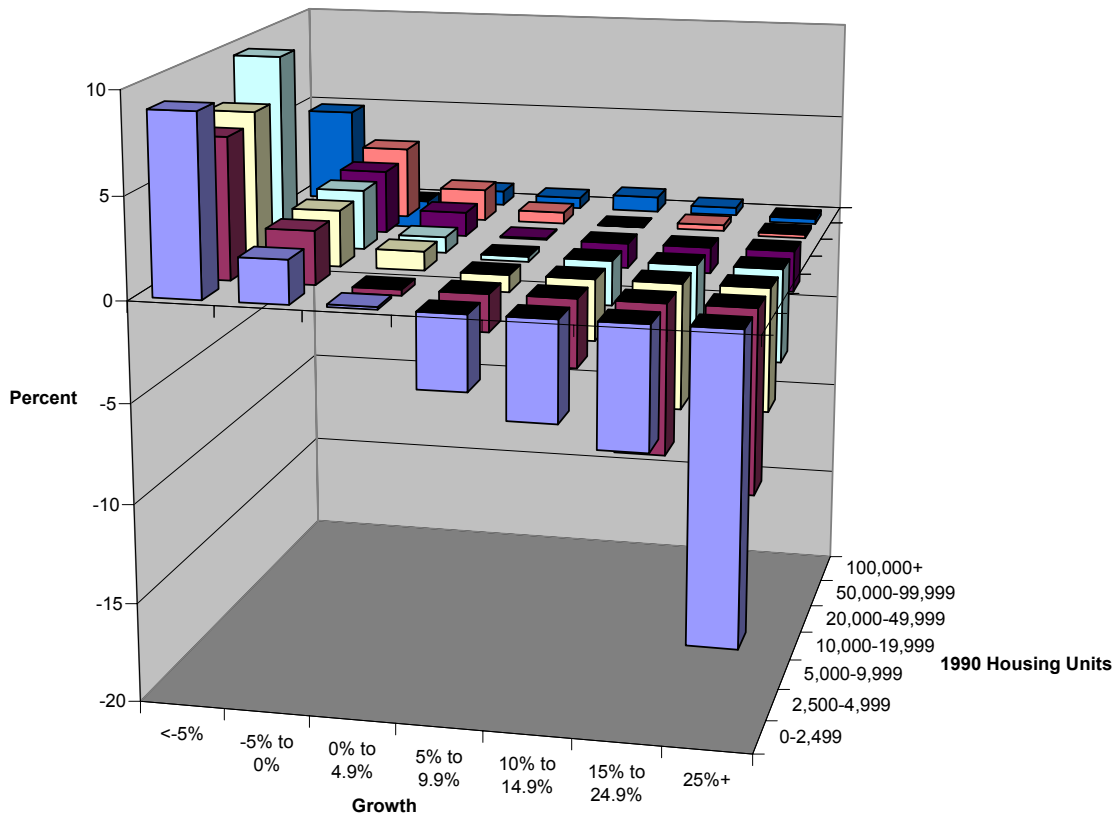
GEOGRAPHIC DIFFERENCES

Map 2. MAPE of April 1, 2000 County Housing Unit Estimates by State

MAP 2
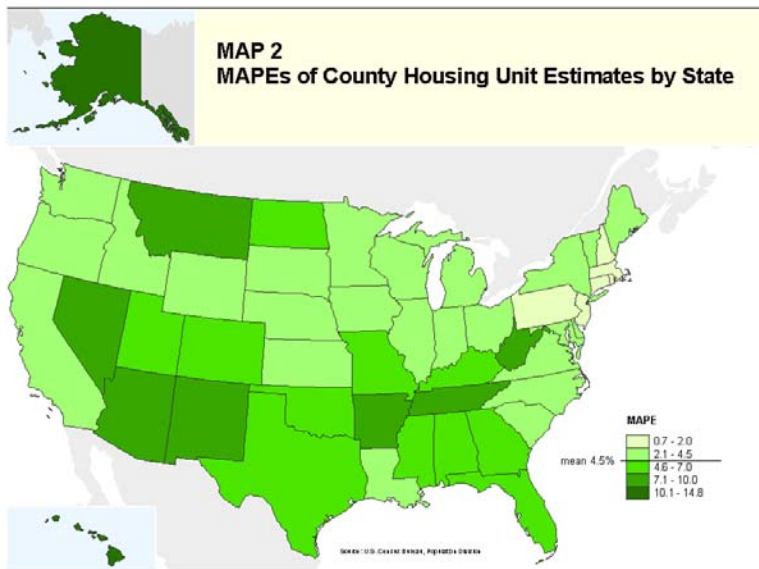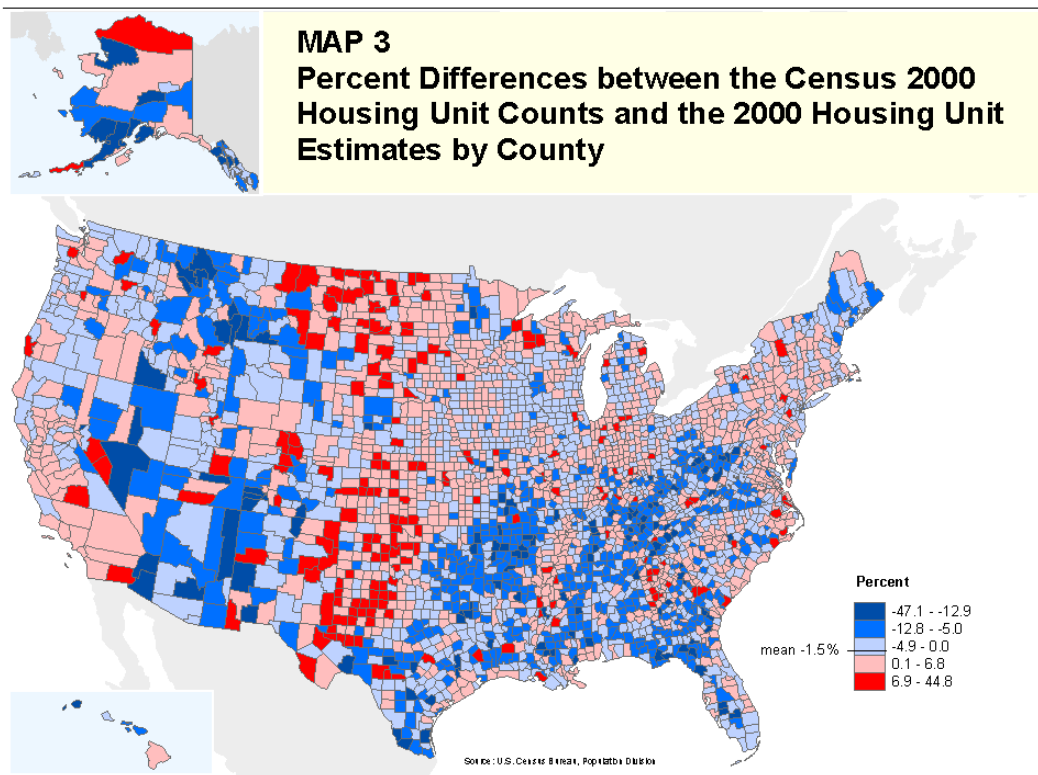MAPEs of County Housing Unit Estimates by State

Table 4 and Map 2 show the MAPEs of the county housing unit estimates by state. These MAPEs range from a low of 0.7 percent for Connecticut to a high of 14.8 percent for Hawaii. Connecticut, Rhode Island, New Hampshire, New Jersey, Massachusetts, and Pennsylvania have county MAPEs lower than 2 percent. Montana, Nevada, Tennessee, Arkansas, Alaska, and Hawaii have county MAPEs higher than 7 percent.

Table 4. **County Measures of Error in the April 1, 2000 Housing Unit Estimates by State**

State Name	April 1, 2000 Housing Unit Estimate (2)	Census 2000 Housing Unit Count (3)	County MALPE (4)	County MAPE (5)
United States	115547749	115904641	-1.5	4.6
Alabama	1915989	1963711	-5.0	5.6
Alaska	251878	260978	-8.0	10.1
Arizona	2111282	2189189	-9.3	9.3
Arkansas	1121396	1173043	-5.6	7.6
California	12448944	12214549	0.5	2.4
Colorado	1804536	1808037	-0.8	5.7
Connecticut	1396296	1385975	0.5	0.7
Delaware	335201	343072	-1.9	2.6
District of Columbia	262386	274845	-4.5	4.5
Florida	7262782	7302947	-4.8	5.7
Georgia	3336316	3281737	-2.1	5.6
Hawaii	447488	460542	-14.4	14.8
Idaho	523021	527824	-0.7	4.1
Illinois	4847481	4885615	-0.4	2.8
Indiana	2573171	2532319	2.0	3.6
Iowa	1226337	1232511	-0.7	2.0
Kansas	1154001	1131200	2.7	3.7
Kentucky	1707718	1750927	-5.0	6.1
Louisiana	1835845	1847181	-2.2	3.9
Maine	635294	651901	-3.4	4.1
Maryland	2146667	2145283	0.0	2.5
Massachusetts	2594259	2621989	-1.5	1.8
Michigan	4260370	4234279	0.2	2.5
Minnesota	2066101	2065946	-0.2	2.7
Mississippi	1136163	1161953	-4.3	6.1
Missouri	2439544	2442017	-3.0	5.4
Montana	388380	412633	-3.4	7.3
Nebraska	723446	722668	0.0	2.9
Nevada	825925	827457	-2.0	7.4
New Hampshire	549664	547024	0.6	1.6
New Jersey	3286619	3310275	-0.2	1.6
New Mexico	773301	780579	2.2	8.5
New York	7525613	7679307	0.0	2.3
North Carolina	3525871	3523944	-0.9	3.4
North Dakota	297404	289677	4.9	5.2
Ohio	4765853	4783051	-1.6	2.9
Oklahoma	1481939	1514400	-2.5	5.6
Oregon	1446582	1452709	0.1	2.8
Pennsylvania	5296608	5249750	0.9	1.8
Rhode Island	434637	439837	-1.0	1.1
South Carolina	1756098	1753670	-1.3	3.2
South Dakota	327816	323208	2.0	4.4
Tennessee	2392755	2439443	-5.7	7.5
Texas	8109855	8157575	-0.8	6.3
Utah	759971	768594	-2.0	5.0
Vermont	292473	294382	-1.3	2.4
Virginia	2936497	2904192	-0.2	3.5
Washington	2462487	2451075	-0.9	3.1
West Virginia	798643	844623	-7.5	8.2
Wisconsin	2331401	2321144	0.3	3.0
Wyoming	217445	223854	-3.4	4.5

11

Map 3 shows the percent error for each county. Various geographic patterns may be observed. The South contains many counties with underestimates of housing units. The northeastern end of their range is in West Virginia. This range extends westward through Kentucky into Arkansas and eastern Oklahoma and southwestward to the southern tip of Texas, interrupted by a set of overestimates in the Mississippi River valley. The eastern boundary of this region is an arc through Kentucky, Tennessee and Alabama into the Florida Panhandle. Underestimates also predominate in eastern Mississippi and parts of southern Louisiana. Another region of underestimates occurs in the Mountain West, particularly in New Mexico and western Montana. Other areas of this region contain clusters of under- and overestimates. Alaska contains similar clusters of under- and overestimates. The western Great Plains generally contain overestimates. All of these areas reflect various problems with the housing unit estimation process, generally the lack or poor quality of building permit data. The overestimates in the Great Plains may reflect underestimation of demolitions.

Map 3. Percent Error in the April 1, 2000 Housing Unit Estimates by State

MAP 3
Percent Differences between the Census 2000 Housing Unit Counts and the 2000 Housing Unit Estimates by County

Percent
-47.1 - -12.9
-12.8 - -5.0
-4.9 - 0.0
0.1 - 6.8
6.9 - 44.8

mean -1.5%

Source: U.S. Census Bureau, Population Division

ALTERNATIVE ESTIMATES

Because of the effort required to produce housing unit estimates based on administrative records, mainly building permits, it is worth asking whether these estimates offer any improvement over using an easier estimation method. This section compares the housing unit estimates developed from administrative records (building permit method) to two alternative sets of estimates. The first set was produced by applying the vacancy and persons per household rates from the 1990 Census to the April 1, 2000 county population estimates developed using the tax return method (population estimate method). The second set simply used the 1990 Census housing unit count as the estimate. The results of this comparison appear in Table 5 and Figure 5. For each size class, the housing unit method is preferable to the other estimates. This demonstrates the value of using a building permit based approach to estimate housing units.

Table 5. MAPE of April 1, 2000 County Housing Unit Estimates by size

1990 Census Housing Unit Count	Number of Areas	MAPE – Building Permit Method	MAPE - Population Estimate Method	MAPE - 1990 Census Housing Unit Count
All	3141	4.6	5.8	12.0
0-2,499	336	7.3	8.1	9.8
2,500-4,999	518	5.7	5.9	10.2
5,000-9,999	749	5.0	6.0	11.7
10,000-19,999	653	4.4	5.5	13.1
20,000-49,999	507	3.1	5.4	14.0
50,000-99,999	178	2.2	4.6	14.0
>100,000	200	1.9	3.9	11.0

Figure 5. Accuracy of April 1, 2000 County Estimates versus Alternative Estimates

13
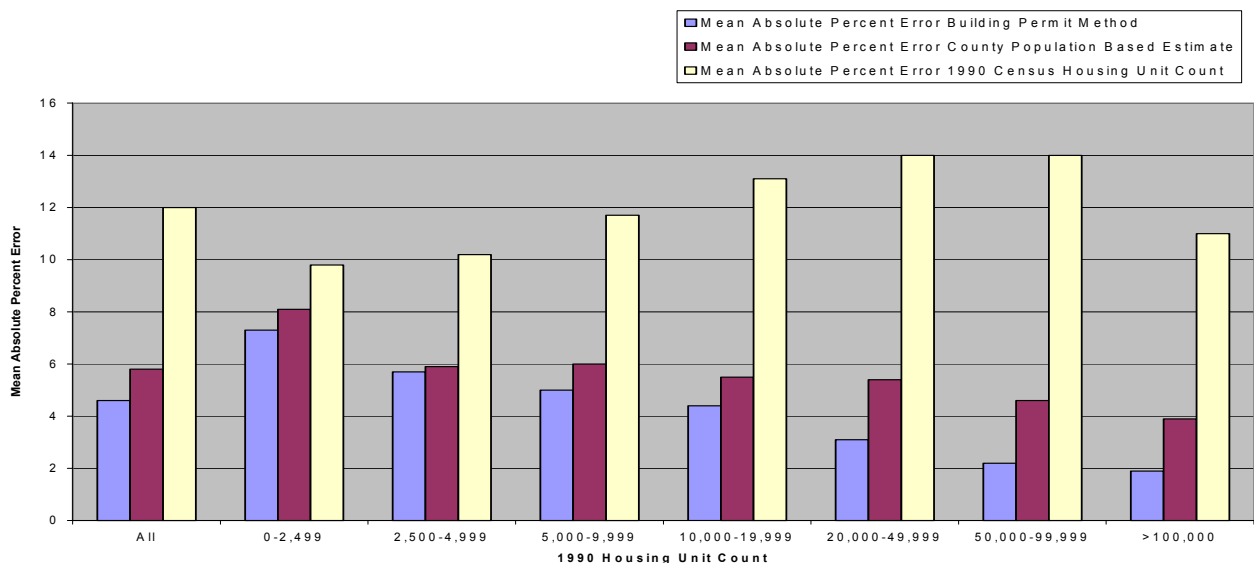
CONCLUSION

This evaluation has found that the 2000 state and county level housing unit estimates developed from building permit, mobile home shipment, and demolition data performed with a degree of accuracy similar to the state and county April 1, 2000 population estimates produced by the Population Division of the Census Bureau. The housing unit estimates follow a pattern similar to other estimates produced by Population Division in that they tended to be more accurate for larger states and counties. The county level housing unit estimates follow a pattern similar to other estimates in that they are more accurate for areas that experienced the smallest amount of housing unit change throughout the decade. At the state level the relationship between housing unit change and accuracy is less clear.

The housing unit estimates show clear geographic variations. The number of housing units in counties were generally underestimated in large parts of the South and Mountain West, and generally overestimated in the western Great Plains. Some areas, such as Alaska contain mixtures of under- and overestimates. These areas are the most difficult to estimate. We hypothesize that the problems are due to input data deficiencies.

The comparison with estimates produced using alternative methods indicates that the building permit method performed better than the alternative methods for all size and growth categories. While the building permit based estimates are more accurate than the county population based estimates, the accuracy of the county population based estimates relies heavily on the accuracy of the vacancy and persons per household rates. The county population based method used persons per household and vacancy rates from the 1990 Census. Improvements in our ability to estimate these rates would improve the county population based estimates.

Through this analysis we have begun to look at the discrepancies between our housing unit estimates and Census 2000. Future research should focus on identifying the components of the estimates that contributed the most to these discrepancies.

REFERENCES

Barron, W. J., Jr. 2001. "Recommendation on Adjustment of Census Counts." Memorandum to Donald L. Evans, Secretary of Commerce, March 1.

Breiman, Leo, 1999. "Random Forests-Random Features," Technical Report No. 567, Statistics Department, University of California-Berkeley.

Coleman, Charles D. 1999, "Nonparametric Tests for Bias in Estimates and Forecasts," in *American Statistical Association: 1999 Proceedings of the Business and Economic Statistics Section*, 251-256.

Coleman, Charles D. 2001, "Non-i.i.d. Generalizations of the Matched-Pairs *t*-Test," in *American Statistical Association: 2001 Proceedings of the Business and Economic Statistics Section*.

Davis, ST 1994. "Evaluation of Postcensal County Estimates for the 1980s", Technical Working Paper #5.

Davis, Sam T., Josephine D. Baker, Marc J. Perry, Signe Wetrogan, and Carolette Norwood, 2001. "An Early Comparison of Postcensal County Population Estimates with Results from the 2000 Census." Paper Presented at the Annual Meetings of the American Statistical Association, Atlanta, GA, August 2001.

Friedman, Jerome, 2001. "A Statistical View of Boosting," presentation made to the Joint Statistical Meetings 2001, Atlanta, GA.

Galdi, David 1985. "Evaluation of 1980 Subcounty Population Estimates, U.S. Census Bureau." Current Population Reports, Series P-25, No. 963, U.S. Government Printing Office.

Harper, Greg, Jason Devine and Charles Coleman, 2001. "Evaluation of 2000 Subcounty Population Estimates." Paper Presented at the Annual Meetings of the Southern Demographic Association, Miami, FL, September 2001.

Prevost, Ron 1994. "State Housing Unit and Household Estimates: April 1, 1980, to July 1, 1993, U.S. Census Bureau." *Current Population Reports*, Series P-25, No. 1123, U.S. Government Printing Office.

APPENDIX: HOUSING UNIT ESTIMATES METHODOLOGY

State and County Level Housing Unit Estimates Methodology

The Population Estimates Branch produces the housing unit estimates in the following steps:

1. Obtain unadjusted total housing units from the 1990 Census.

Note: The April 1, 1990 census housing unit count used for these estimates is a count of the number of housing units in an area as reported in the 1990 census of population or recalculated to match legal boundary changes and geocoding enhancements in the Census Bureau's TIGER system. The closing date to include these revisions in this set of estimates was January 1, 1998.

2. Estimate residential construction from building permits compiled from internal data files developed by the Manufacturing and Construction Statistics Division.

Note: Estimates of nonpermitted residential construction are calculated using the following steps at the county level:

 a) Calculate the number of 1990 county housing units located in non-permit-issuing jurisdictions within the county by subtracting the number of units located in permit-issuing jurisdictions from the total units.

 b) Sum the housing units located in non-permit issuing jurisdictions to the national level.

 c) Calculate the county's share of units located in jurisdictions not issuing permits by dividing the county's units by the nation's units (a)/(b).

 d) Multiply the total number of units reported in the Survey of Construction as constructed without building permits by the county's share of units located in jurisdictions not issuing building permits.

3. Estimate new mobile home placements from Series C-40 reports. Ninety-eight percent of all mobile homes shipped to states are used for residential purposes. State mobile home information is distributed to counties based upon a county's proportion of the state's mobile homes as of the 1990 census.

4. Estimate housing loss from demolition permits from internal data files developed by the Manufacturing and Construction Statistics Division (MCSD). These files include imputed permits where a permit-issuing jurisdiction did not report permit issuance for the entire year. No lag time is assumed for demolition permits.

Note: MCSD stopped collecting data on demolition permits in 1995. After 1995, all data on housing unit loss is estimated. Estimates of nonpermitted housing loss are calculated from the county's share of structures in region at risk of loss. The risk is based on the county's share of the following types of structures: (1) Mobile homes and other; (2) Older units (pre-1939 construction); (3) Vacant for Seasonal and Recreational Use; and (4) Boarded up units.

5. Adjust estimates. After the housing unit estimates are produced, adjustments are made to ensure that the estimates are consistent with the county population estimates produced using the Tax Return method. County household estimates are produced using tax return data, which are controlled to state household estimates. The state household estimates are based on household formation rates from the Current Population Survey (CPS). Controlled county household estimates and county housing unit estimates are used to derive a vacancy rate. The county household population estimates from the Tax Return method and the controlled county household estimates are used to estimate persons per household (PPH). Adjustments are made to the housing estimates to ensure that these vacancy and PPH values fall within certain tolerances.

6. State Estimates. Sum the county level estimates to get the state level estimates