REVISED 4/24/09

USING A REGRESSION APPROACH TO ESTIMATE PERSONS PER HOUSEHOLD AND VACANCY RATES IN THE PRODUCTION OF HOUSING UNIT-BASED POPULATION ESTIMATES

Kirsten K. West, Bashir Ahmed, Antonio Bruce, and D.H. Judson
Population Division
U.S. Census Bureau

U S C E N S U S B U R E A U

ABSTRACT

In 2007, the Population Division of the U.S. Census Bureau started a project to explore strategies for producing county level housing unit-based population estimates. The Housing Unit Method relies on the number of occupied housing units and the average number of persons per household (PPH) at a particular time. Research by Smith et al. (2002) confirmed that regression models based on symptomatic indicators of PPH change can produce more precise and less biased county PPH estimates than methods that rely on the most recent decennial census for estimation. Following their work, we tested different models and produced county level predicted PPH estimates for 2000 based on the regression results. The regression based PPH estimates were evaluated by comparisons to the PPH values in Census 2000 and PPH estimates based on alternative methods.

The regression approach looks promising for PPH estimation in postcensal years. The selected independent variables explained over 90 percent of the variation in some models for both 1990 and 2000. Measures such as the Mean Absolute Percent Error (MAPE) for the regression based PPH suggest that the error is less than 2 percent in these models. In comparison, were we to keep constant the 1990 PPH values throughout the decade, we would generate an error of 4.2 percent and introduce an upward bias close to 3.8 percent. If we kept constant the change observed in the previous decade, we would find the MAPE to be 3.4 percent and the predicted PPH in 2000 to be lower by 2.5 percent than the actual PPH. Finally, applying the change observed for a state for each county within that state yields a predicted PPH that is 1.4 percent higher than the actual PPH.

We were not successful in using the regression approach to estimate occupancy (percent vacant.)

INTRODUCTION

The research presented in this paper was conducted as part of a collaborative effort by the U.S. Census Bureau, the Federal State Cooperative for Population Estimates (FSCPE), and other experts to examine the use of a housing unit-based method for estimating population. This effort was undertaken in response to input received during a conference sponsored by the Census Bureau on meeting the needs of data users and testimony at a Congressional Oversight Hearing held before the Subcommittee on Federalism and the Census.

The Census Bureau produces annual population estimates for all counties in the United States. In the 1990s, the use of averages, or combinations of estimates from several different methods, was replaced by a single, nationwide series of estimates produced using an administrative records-based method (ADREC).[1, 2] The change to a single method was made after it was decided that the potential gains in accuracy were not large enough to justify the increase in the complexity of production introduced by multiple methods (Davis, 1994). The ADREC method uses administrative records to estimate the components of population change, namely births, deaths, and migration.

Although they are not used to prepare the county population estimates, the Census Bureau produces estimates of housing units for use in the production of subcounty population estimates and for use as statistical controls by the American Community Survey (ACS), the American Housing Survey (AHS), and other Census Bureau surveys. These estimates are released to the public as a separate data product at the state and county level. Population estimates produced using the Housing Unit Method may be accepted when local governments challenge the official population estimates by providing additional housing unit information.

To gather user input on the Population Estimates Program, the Census Bureau sponsored a conference in July 2006 titled "Population Estimates: Meeting User Needs." The conference

[1] ADREC refers to the components of population change method used by the Census Bureau to develop the official county population estimates that uses administrative records to estimate births, deaths, and net domestic migration.
[2] The multiple methods included the Administrative Records (ADREC), the Composite II (COMPII), and the Ratio-Correlation methods.

included presentations by experts from the federal, state, local, and private sectors that use population estimates for a variety of purposes. A common theme heard from many of the representatives at this conference and during the Congressional Oversight Hearing held in September 2006 was the need to explore multiple methods for preparing the official estimates. Recognizing the new opportunities offered by the full implementation of the American Community Survey and the ongoing maintenance of the Master Address File (MAF), participants encouraged the Census Bureau to evaluate the use of a housing unit-based method as part of the program for producing county-level population estimates.

In response to this input, the Census Bureau formed the Housing Unit-Based Estimates Research Team (HUBERT) to conduct research in three areas: the accuracy of the ADREC and housing unit-based population estimates, ways to improve the components currently used in the production of housing unit estimates, and new sources for estimating the components needed to produce population estimates using a housing unit-based method. The results presented in this paper are from the third area of research. The research summarized here focuses on alternative methods for estimating the components, particularly whether a regression approach could be used to estimate the persons per household (PPH) and the current vacancy rates at the county level.

BACKGROUND

The population in a given area equals the sum of the household population and the group quarters (GQ) population. In turn, the household population is the product of the number of occupied housing units (households) and the average household size or number of persons per household.

$$P_t = ((H_t * O_t) * PPH_t) + GQ_t \qquad (1)$$

where P_t = total population at time t, $H_t * O_t$ = occupied housing units (households) at time t, PPH_t = average number of persons per household at time t, and GQ_t = group quarters population at time t.[3]

Estimates of occupied housing units are usually obtained from knowledge of the vacancy rate and the total number of housing units in the most recent census. When the vacancy rate is used to estimate occupancy rates, the formula for occupied housing units is simply:

$$H_t * O_t = HU_c *(1 - VACHU_c), \qquad (2)$$

where HU_c = number of housing units in the census and $VACHU_c$ = the vacancy rate in the census. Throughout the remainder of this paper, the discussion will center on vacancy rates rather than on occupied housing units. The task at hand consists of developing regression models for estimating the PPH and vacancy components in equations 1 and 2.

The regression approach would eliminate the dependency on the decennial census as a data source thus providing timely estimates in the intercensal years. Research conducted on a limited scale for some states and counties suggests that regression models based on symptomatic indicators of change might yield more up-to-date and accurate PPH estimates than approaches using decennial data (National Research Council, 1980; Rives, 1982; Smith and Lewis, 1980 and 1983; Smith et al., 2002).

County-level regression models based on symptomatic indicators have been successfully developed for estimating the PPH. To predict PPH, Smith et al. (2002) chose the following three independent variables on the basis of their availability and prediction power: births per household, school enrollees in grades K-12 per household, and Medicare enrollees aged 65 and over per household. They used data from 462 counties in four states: Florida, Illinois, Texas, and Washington. They were able to develop regression models with a good fit. The models generated predicted PPH values that compared well with the PPH estimated from the census.

[3] The GQ population is out of scope for this analysis.

The vacancy rates have not been similarly studied. Part of this research void might be attributed to the complexity of understanding determinants of vacancy rates. Theoretically, economic indicators or their proxies or indicators reflecting a change in the socioeconomic or demographic make-up of an area should be good predictors of the PPH and vacancy rates. For example, we would expect an influx of population to an area in response to economic growth and an exodus of population in response to economic decline. Population growth could be absorbed by existing housing units or by new housing units being built (higher, stable or decreasing PPH). On the other hand, population decline could take place as within-household attrition from an area, decreasing the PPH, but leaving no housing unit vacant.

To explore the possibilities empirically, we looked at the correlation between change in PPH and change in vacancy at the county level between the 1990 Census and Census 2000. We found only a modest statistical correlation (r=0.2702). Between 1990 and 2000, most counties (n=2,816) experienced a decrease in the PPH, but the response in vacancy rates was not predictable. The vacancy rate decreased in 52 percent of these counties and increased in 48 percent. In the modest number of counties (n=322) where the PPH increased, the vacancy rate decreased in almost 76 percent of the counties (Table 1).

We further examined the county level patterns of relationship between PPH and vacancy rates under different scenarios of housing unit change between 1990 and 2000. (Three counties exhibited no change in the number of housing units over the decade. These counties are included with the counties where the housing unit decreased.) The outcome is shown in Table 2. When the change in housing units is decreasing, and the PPH decreases, 67 percent of the counties experience a decrease in the vacancy rate. Under the same housing unit change scenario, but with an increase in the PPH, the vacancy rate is decreasing in 77 percent of the counties. When the housing unit change is increasing, but the PPH is decreasing, the vacancy rate change is split 50/50 between increase and decrease. When both the housing unit change and the PPH are increasing, the vacancy rate increases in 24 percent of the counties. The change in vacancy rate within a county does not appear to follow a predictable pattern based on knowledge about the PPH in a county.

In summary, previous research has shown that, at least for some states and counties, regression models are capable of producing more precise and less biased PPH estimates than those produced by more common methods. No similar previous research provides further analysis of vacancy rates. In the remaining part of this report, the regression approach is tested for developing estimates of PPH (Part 1) and vacancy rates (Part 2).

PART 1: USING A REGRESSION APPROACH TO ESTIMATE PPH IN THE PRODUCTION OF HOUSING UNIT-BASED POPULATION ESTIMATES

In Part 1, we will focus on persons per household.

INTRODUCTION

This research addresses the question: Can a regression approach be used to estimate current PPH? Nationally, PPH has declined decade by decade but also within each decade, though the speed of the decline was faster in the 1960s, the 1970s, and the 1980s than in the last two decades.

Table 3 shows annual PPH estimates from 1960 to 2006 calculated from Current Population Survey (CPS) data. In 1960, PPH was 3.33. By the end of that decade, it had declined to 3.19. The decline continued in the 1970s and the 1980s, ending in 1989 with a PPH of 2.62. The 1990s showed some fluctuations. In 1993, the survey estimates were revised based on results from the decennial census. This resulted in a small increase between 1993 and 1994. By the end of the 1990s, however, PPH estimates were again lower than the 1990 estimate. Since 2003, the estimates have remained around 2.57.

The decline in PPH over the last decade is also observed in the decennial census data at the national and lower levels of geography. Table 4 shows the percent distribution of counties by PPH calculated from data in the 1980, 1990, and 2000 censuses. In the 1980 census, PPH for most counties (about 94.0 percent of all counties) was in the range of 2.40 to 3.19. In the 1990 census, PPH for most counties shifted to the range of 2.30 to 2.99. Finally, in Census 2000, PPH for most counties fell to the range of 2.20 to 2.89, a downward shift in the distribution. However, it should also be noted that the decline is subtle for many counties and observable only within a relative small range of PPH values. Thus, the research challenge at the county level is to develop a regression model to predict variations of different magnitude and direction for areas with different population sizes and number of housing units.

METHODOLOGY

The methodology section outlines the strategy for developing the regression models, with brief descriptions of the dependent and the independent variables, the models, and data sources. First, we focus on construction of the models. Next, we discuss production of the county-level predicted PPH estimates for 2000 based on the regression results. Finally, we discuss the measures used to evaluate the regression-based PPH estimates compared to Census 2000 PPH.

Methodology for Constructing Regression Models

The regression approach

Regression analysis is often used in an exploratory fashion to look for empirical relationships between the independent and dependent variables. The goal is to estimate the intercept and the slope of a line described by the equation:

$$y_i = \beta_0 + \beta_1 x_j + \ldots \beta_k x_k + \varepsilon_i, \quad (3)$$

where y_i is the response or dependent variable for the ith case, x_j is the jth regressor or independent variable (j=1,2,…,k), ε_i is the ith error term, and β_0, β_1, …,β_k are unknown parameters to be estimated.

The dependent variable

In this analysis, the dependent variable in each regression model is persons per household (PPH) expressed either in absolute or relative terms or as change, where each county is an observation.

The independent variables

In the initial models, we used exploratory stepwise regression to select the following variables as candidates for independent variables in the models:

1) Births per household (Number of births in the county divided by the number of occupied housing units in the county),[4]
2) Deaths per household (Number of deaths in the county divided by the number of occupied housing units in the county),
3) School enrollees per household (Number of school enrollees in grades kindergarten through 12 in the county divided by the number of occupied housing units in the county),
4) Medicare enrollees per household (Number of Medicare enrollees in the county divided by the number of occupied housing units in the county),
5) Hispanics per household (Number of persons of Hispanic origin in the county divided by the number of occupied housing units in the county),
6) Households in the family formation years (Number of householders aged 25 to 44 divided by the number of occupied housing units in the county),
7) Nonfamily households (Number of nonfamily households[5] divided by the number of occupied housing units in the county), and
8) Tax exemptions claimed per household in the county (Number of tax exemptions claimed by forms filed in the county divided by the number of occupied housing units in the county).

These variables are presumed to capture the socio-economic and demographic dynamics of the counties with regard to household size.

Theoretically, some effects are expected to increase the PPH, others to decrease the PPH. For example, the proportion of infants (measured through births per household) and children (measured through school enrollment) in a county are expected to relate positively to PPH (the more births or children, the higher the PPH). Hispanics, on average, have higher fertility rates than non-Hispanics. Thus, the higher the proportion of Hispanics in a county, the higher would be the PPH. Likewise, the higher the proportion of households in the family formation years (householder in age group 25 to 44), the higher the expected fertility and hence the PPH.

[4] We use the term “household” and “occupied housing unit” interchangeably.
[5] A nonfamily household can be either a person living alone or a householder who shares the housing unit only with his or her nonrelatives – for example, boarders or roommates. The nonrelatives of the householder may be related to each other (that is, they could be part of an unrelated subfamily).

In contrast, the proportion of nonfamily households in a county would be negatively related to PPH because this variable reflects lower fertility rates. The proportion of older people in a county (people aged 65 and older and therefore eligible for Medicare) is also expected to show a negative relationship (the older the population, the lower the PPH), as older people often live in one- or two-person households. Mortality (deaths per household) represents attrition and thus the higher the death rate, the lower the PPH.

We also included the PPH from the previous census (PPH_{CEN} lagged) as a predictor of the current PPH. According to Smith et al. (2002), this variable is expected to have a negative impact on changes in PPH: when overall PPH is falling, for example, the declines are expected to be larger for counties with higher PPH levels than for counties with lower PPH levels.

To account for differences in the sizes of county populations, independent variables were expressed as ratios by dividing by the total number of occupied housing units (households) obtained from Census 2000. To account for differences in the number of county households, ratios of county-to-state values are created for use in ratio models. Change models (basic and ratio) are created by subtracting the earlier-year values from the later-year values.

After examining the direction of the effects and their statistical significance, we decided to drop the following three independent variables from our exploratory models: Medicare enrollees, tax exemptions per household, and the lagged PPH.[6] Thus, the initial analyses include the variables school enrollees, births, deaths, Hispanic origin, householder aged 25 to 44, and nonfamily households. However, we put the variables Medicare enrollees and tax exemptions back into an expanded analysis that also includes the variables school enrollees and births. Thus, the expanded model includes only the independent variables – school enrollees, births, tax exemptions, and Medicare enrollees – that would be readily available or can be estimated in postcensal years, mostly from administrative record files – a requirement for selection of the expanded model determinants.[7]

[6] The lagged PPH variable did not exhibit the impact consistently; therefore, we did not include it in any of the final models.

[7] The variable 'deaths' could be obtained from administrative records, however, we did not include it in the expanded model because it is highly correlated with the variable Medicare enrollees.

Regression models

We initially tested four ordinary least squares regression models: a basic model (Model 1), a ratio model (Model 2), a change model (Model 3), and a ratio change model (Model 4).

The basic model uses PPH from the 1990 census as the dependent variable, and the six independent variables from around the same time.[8] This model can be defined by the equation:

$$\text{(Model 1):}\quad PPH_t = ß_0 + ß_1X_{1t} + ß_2X_{2t} + ß_3X_{3\,t} + ß_4X_{4t} + ß_5X_{5t} + ß_6X_{6t} + \varepsilon, \qquad (2)$$

where,

t=time period for which the estimate is being constructed,

X_1= Ratio of school enrollment in grades kindergarten through 12 to households,

X_2 = Births per household,

X_3 = Deaths per household,

X_4 = Hispanic population per household,

X_5 = Proportion of households with householder aged 25 to 44, and

X_6 = Proportion of households that are nonfamily.

The ratio model is a variant of the ratio-correlation method. In this model, the variables are defined as ratios for one point in time (e.g., 1990), rather than as ratios from two points in time (Namboodiri, 1972). All variables are expressed as ratios of county-to-state values. This model can be defined by the equation:

[8] As stated later in this section, the variables births, tax exemptions, and Medicare enrollees come from administrative record files. The variable school enrollees is taken from the decennial censuses, though it could be estimated using the Common Core of Data (CCD) and the Private School Survey (PSS).

(Model 2): $(PPH_t / PPH^s_t) = ß_0 + ß_1 (X_{1t} / X_1{}^s{}_t) + ß_2 (X_{2t} / X_2{}^s{}_t) + ß_3 (X_{3t} / X_3{}^s{}_t) +$
$ß_4 (X_{4t} / X_4{}^s{}_t) + ß_5 (X_{5t} / X_5{}^s{}_t) + ß_6 (X_{6t} / X_6{}^s{}_t) + \varepsilon,$ (3)

where the independent variables are defined as in (2) above and the superscripts (s) refer to state values.

The third model is a change model. It models the change between two census years. In this analysis, the change between 1980 and 1990 is being modeled in order to predict the change between 1990 and 2000. This model can be defined by the equation:

(Model 3): $(PPH_t – PPH_c) = ß_0 + ß_1 (X_{1t} -X_{1c}) + ß_2 (X_{2t} - X_{2c}) + ß_3 (X_{3t} - X_{3c}) +$
$ß_4 (X_{4t} - X_{4c}) + ß_5 (X_{5t} - X_{5c}) + ß_6 (X_{6t} - X_{6c}) + \varepsilon,$ (4)

where the independent variables are defined as in (2) above.

The last model is a ratio change model, which models the ratio of the change between two census years. This is a variant of the difference-correlation method of population estimation (O'Hare, 1976), and the variables used are based on county and state level data (ratios of county to state). Once again, the change between 1980 and 1990 is being modeled, so that this model can be used to predict the change between 1990 and 2000. This model can be defined by the equation:

(Model 4): $[(PPH_t / PPH^s_t) - (PPH_c / PPH^s_c)] = ß_0 + ß_1 [(X_{1t} / X_1{}^s{}_t) - (X_{1c} / X_1{}^s{}_c)] +$
$ß_2 [(X_{2t} / X_2{}^s{}_t) – (X_{2c} / X_2{}^s{}_c)] + ß_3 [(X_{3t} / X_3{}^s{}_t) - (X_{3c} / X_3{}^s{}_c)] + ß_4 [(X_{4t} / X_4{}^s{}_t) - (X_{4c} / X_4{}^s{}_c)] +$
$ß_5 [(X_{5t} / X_5{}^s{}_t) – (X_{5c} / X_5{}^s{}_c)] + ß_6 [(X_{6t} / X_6{}^s{}_t) - (X_{6c} / X_6{}^s{}_c)] + \varepsilon,$ (5)

where the variables are defined as in (2) above.

The expanded analysis presents an alternative to the basic model, which can be defined by the equation:

$$(\text{Alternative Model 1}):\; PPH_t = ß_0 + ß_1X_{1t} + ß_2X_{2t} + ß_3Z_{1t} + ß_4Z_{2t} + \varepsilon, \qquad (6)$$

where,

X_1 = Ratio of school enrollment in grades kindergarten through 12 to households,

X_2 = Births per household,

Z_1 = Tax exemptions per household, and

Z_2 = Medicare enrollees per household.

and the number of households are derived from 1990-based housing unit estimates and vacancy rates rather than based on counts from the decennial census.

Data sources

The variables births and deaths were derived from administrative records obtained from the National Center for Health Statistics. The variables Medicare enrollees and tax exemptions come from the Centers for Medicare and Medicaid Services and the Internal Revenue Service, respectively. The variables school enrollees, nonfamily households, Hispanic, and householders aged 25 to 44 come from the decennial censuses. Note that, of the variables that come from censuses: 1) the variable school enrollees is from sample data in 1980, 1990, and 2000; 2) the variable Hispanic origin and the variables nonfamily households and householders aged 25 to 44 are taken from 100-percent data in 1980, 1990, and 2000. Also, note that we use data from all counties that provided comparable information for 1980, 1990, and 2000.[9]

Methodology for Producing Predicted PPH Estimates

Each of the four models, basic, change, ratio, and ratio-change, has somewhat unique operational requirements. These requirements may affect the feasibility of using the model(s) in practice.

[9] In order to keep the data sets comparable between 1980, 1990, and 2000, one county in Arizona (La Paz), one county in New Mexico (Cibola) and seven counties in Alaska were omitted from the analysis.
These counties did not exist in either 1980 or 1990. Loving County, Texas and Kalawao County, Hawaii were omitted due to small population size. Furthermore, Washington, DC is excluded from the ratio/ratio change models (dividing by itself). Note that, in practice, if these models were used in non-census years (e.g., 2006), the independent variables would need to be obtained from administrative data and/or ACS data—and the latter would, of course, introduce additional sampling variability.

The ultimate purpose of the model, of course, is to predict PPH at the county level. In this section, we discuss how that is achieved.

For the basic model, implementation is simple. Merely fit the model:

$$P\hat{P}H = X\hat{\beta},$$

where,

X is the matrix of predictor variables, $\hat{\beta}$ is the vector of estimated regression coefficients, and

$P\hat{P}H$ refers to the within-sample predicted value of PPH under the model with estimated coefficients $\hat{\beta}$. [10]

To apply this model to some future time t, define:

$P\hat{\bar{P}}H_t$ = the predicted value of PPH based on the estimated model and predictor values at time t, and

X_t = the values of the predictor variables at time t.

Then,

$$P\hat{\bar{P}}H_t = X_t\hat{\beta}.$$

For the change model, we fit:

$PPH_t - P\hat{P}H_{t-10} = X\hat{\beta},$ and, solving for PPH_t, we obtain:

$P\hat{\bar{P}}H_t = X_t\hat{\beta} + PPH_{t-10}$ (we used the previous value of PPH for the term PPH_{t-10}, where t-10 refers to the point in time ten years previous to time t).

For the ratio[11] model, we fit:

[10] We use the “hat” notation to denote within-sample estimates of coefficients and predicted values of the response variable; we use the “bar hat” notation to denote the out-of-sample predictions of PPH.

[11] A reviewer correctly points out that, by transforming the response terms of the ratio models, we are disadvantaging them; a regression model is never as good at predicting functions of the response modeled. However, from the point of view of implementation, our ultimate goal is the construction of a predicted PPH at some future time, for use in a population estimates system. Thus, we are required to make this transformation in order to use these models in practice.

$\left(\frac{P\hat{P}H_{c,t}}{PPH_{s,t}}\right) = X\hat{\beta}$, where c = county-level and s = state-level values.

We again solve for PPH_t, and obtain:

$P\hat{\hat{P}}H_{c,t} = X_t\hat{\beta} \cdot P\hat{\hat{P}}H_{s,t}$. (We used the estimated value of PPH for the term $P\hat{\hat{P}}H_{s,t}$ on the right hand side.)

The ratio-change would be the most complicated. We fit:

$$\frac{PPH_{c,t}}{PPH_{s,t}} - \frac{P\hat{P}H_{c,t-10}}{PPH_{s,t-10}} = X\hat{\beta}.$$

Solving for $PPH_{c,t}$ yet again, we obtain:

$$P\hat{\hat{P}}H_{c,t} = \left[X_t\hat{\beta} + \frac{PPH_{c,t-10}}{PPH_{s,t-10}}\right]P\hat{P}H_{s,t}.$$

In this case, we used the previous county- and state-level PPH values for the term $\left[\frac{PPH_{c,t-10}}{PPH_{s,t-10}}\right]$ and the estimated value of PPH for the term $P\hat{\hat{P}}H_{s,t}$, to derive the final county PPH estimates.

Methodology for Evaluating the Predicted PPH

To evaluate the predicted PPH estimates, we compare them to the actual estimates obtained from the 2000 decennial census. The difference is considered an error. Two error measures are used in the evaluation: the mean absolute percent error (MAPE) and the mean algebraic percent error (MALPE).

The MAPE is defined as: $\frac{\sum_{\text{ith county}} \frac{|Estimate - Criterion|}{Criterion}}{\text{Number of counties}}$, while the MALPE is defined as: $\frac{\sum_{\text{ith county}} \frac{Estimate - Criterion}{Criterion}}{\text{Number of counties}}$. The MAPE is the average error with the direction of error ignored.

This is a measure of how close the predicted estimates are to the census values, regardless of whether they are too high or too low. The MALPE is the average error with the direction of error included. This is a measure of the bias, or the tendency of the estimates to be too high or too low. The evaluations are done at the national and state levels.

Furthermore, as a standard of comparison, we also compare the results obtained from the regression approach against three other simple but widely used traditional methods (Starsinic and Zitter, 1968). As defined below, we will refer to these traditional methods as Method A, Method B, and Method C.

Method A: The county PPH in the most recent census is used as an estimate of the county PPH during the following decade, i.e., assume no change in PPH. For example, take the 1990 Census PPH for PPH estimates in 2000, assuming no change from 1990 to 2000.

Method B: The percentage change in each county's PPH during the previous decade is used as an estimate of the county PPH change during the following decade, i.e., assume that the change remains constant. In our case, the county percentage change from 1980-1990 was applied to the 1990 county PPH to predict the county PPH in 2000.

Method C: The percentage change in a state's PPH since the most recent census is used as an estimate of PPH change for each of the state's counties, i.e., assume that the change observed for the state applies to each county and that the change is constant. In this report, the percentage change in each state's PPH from 1990-2000 was applied to its county PPH in 1990 to predict the county PPH in 2000.

These approaches are often preferred because they are straightforward and require simple extrapolations of trends generated from decennial data.

FINDINGS

Regression Results

Our initial analysis was exploratory in nature. We tested four outcome variables (basic, ratio, change, and ratio change) and different independent variables. Next, for the model that appears to have most promise for future applications, we show an alternative model based only on administrative records. Were we to adopt the regression approach, for the purpose of producing annual county level population estimates, we would prefer the data to be available annually at the county level. The regression results show findings from both the initial and the expanded analysis.

Results from initial analysis[12]

For Models 1 and 2, we ran separate regressions for 1990 and 2000. For Models 3 and 4, we ran separate regressions for 1980-1990 and 1990-2000. The results for Models 1 and 2 are shown in Table 5 and those for Models 3 and 4 in Table 6.[13]

The coefficients that are statistically significant at $p < 0.05$ or better are noted and the models shown are the models with the best fit. Note that we use the same variable names for regression results in the tables, though the variables are defined somewhat differently in each model. For example, in Model 1 (basic) the variable births is defined as the number of births per household at the county level, whereas in Model 2 (ratio) it is the ratio of the county-to-state level births per household.

Examining Model 1 for 1990 and 2000 (Columns 1 and 2 in Table 5), we find that the independent variables explain a relatively high proportion of variation in the dependent variable in both years: 93 percent in 1990 ($R^2 = 0.9297$) and 90 percent in 2000 ($R^2 = 0.9013$). The signs

[12] We focus on regression R^2 values, but it should be noted for the record that because the response variables are in different forms, these values for models 1-4 are, strictly speaking, not comparable.

of the coefficients in Model 1 are consistent with expectations. A higher number of school enrollees per household, births, Hispanics, and a higher proportion of householders 25-44 years of age produce higher values of PPH, while deaths per household and the proportion of nonfamily householders are inversely related to PPH. The regression coefficients are also consistent between 1990 and 2000.

In Model 2 (Columns 3 and 4 in Table 5), we find that the independent variables together explain a relatively small proportion of variation in the dependent variable in both 1990 and 2000: 19 percent in both years (1990 $R^2 = 0.1882$; 2000 $R^2 = 0.1896$). Moreover, Model 2 shows counter-intuitive negative coefficients for the variables Hispanics (defined as the county-to-state ratio of Hispanics per household) and Householder (defined as the county-to-state ratio of the proportion of households with a householder aged 25 to 44).[14] Other coefficients are in the expected directions and statistically significant at $p<0.05$ or better.

Models 3 and 4 show the regression results when the variables are expressed as changes over time (see Table 6). The coefficient of determination or R^2 is 77 percent for the 1980-1990 change model and 69 percent for the 1990-2000 change model. All the effects in Model 3 are in the expected direction for the 1980-1990 change model, but the change in the mortality level between 1990 and 2000 shows a small and ambiguous effect on household size.

Model 4 is the counterpart to Model 3 and, similar to Model 2, it measures changes in the county-state relationship of the independent and dependent variables. As shown in Columns 3 and 4 in Table 6, the effects of all the independent variables except the county-to-state change in mortality are significant at $p< 0.05$ or better. The ratio change in the proportion of householders in the age group 25 to 44 is now negatively related to the ratio change in PPH. Overall, the

[13] Note that the ratio-change models have correlated error terms within states; therefore for these models, standard errors (and hence p-values) are not correct. A future improvement is to apply the Huber (1967) robust standard error method (see also Statacorp, 2005: 275-280 for further discussion of robust standard errors).
[14] Two alternative explanations for counter-intuitive results include: 1) Omission of interaction terms that would render the joint result interpretable; and 2) Presence of high multi-collinearity. In this exploratory model, we did not attempt to fit multiple interaction terms, focusing instead on the simpler to explain first-order models. We performed tests for multicollinearity, and found that it was not sufficiently high in models presented here to explain the existence of counter-intuitive results.

model explains only 12 percent of the variation in PPH change in either the 1980-1990 or the 1990-2000 period.

Results from expanded analysis

Based on the results of exploratory analyses, we repeated the initial analysis with Model 1 (basic) including only variables that would be available in intercensal years from administrative data sources. School enrollment, births, tax return data, and Medicare enrollment data meet this criterion. The results are shown in Table 7.

The independent variables in Alternative Model 1 explain a relatively high proportion of variation in the dependent variable in both 1990 and 2000, though less than Model 1. In 1990, the explained variation in Alternative Model 1 is 89 percent compared with 93 percent in Model 1. In 2000, the explained variation in Alternative Model 1 is 86 percent compared to 90 percent in Model 1. The signs of the coefficients of the independent variables are in the expected direction: positive for school enrollment, births, and tax exemptions, and negative for Medicare enrollment. Like births and school enrollment, the higher the average number of exemptions per federal income tax return the higher the expected PPH. The signs are in the same direction in 1990 and 2000, and all the coefficients are significant at $p<0.01$ in both years.

Evaluation of the Regression Models

The results obtained from the different regression models are evaluated by comparing the predicted PPH values for 2000 with the actual PPH values derived from Census 2000. We also compare the regression models against three traditional estimation approaches, as outlined in the methodology section. In addition, we compare the PPH estimates obtained by averaging the results from the individual regression models.[15] Averages are expected to capture more information than can be incorporated in a single model and hence reduce errors in the estimates.

[15] In theory averaging of models does not necessarily improve sample predictability; however, in practice, averaging often in fact does better (see, e.g., Madigan and Raftery, 1994, for a statistical justification, or Bryan, 2004, for a demographic justification). When averaging, for simplicity, we did not try averaging all possible combinations of models.

Evaluation of the predicted PPH at the national level: Predictive accuracy and bias

Table 8 presents the MAPE and MALPE values at the national level. The MAPEs for all the regression models (including Alternative Model 1) fall within a relatively small range. Except for Model 2, the ratio model, MAPEs for all models range from 1.52 percent to 2.41 percent (Column 1 in Table 8). Model 2 exhibits a MAPE value of 4.48 percent. Alternative Model 1 has a higher MAPE than Model 1--2.41 percent compared with 1.74 percent.[16] As expected, the MAPEs for the Average models are within the range of the MAPEs for the individual regression models. Finally, the regression models (except for Model 2) generally have lower MAPEs than any of the traditional methods.

Next, we evaluate the bias (MALPE) in the predicted PPH (Column 2 in Table 8). At the national level, the lowest positive MALPE is observed to be 0.49 percent for Model 2 (ratio). The MALPEs for the other three models are in the opposite direction: Model 1 (basic) at –0.55 percent, Model 4 (ratio change) at –0.65 percent, and Model 3 (change) at –0.68 percent. The value of MALPE for Alternative Model 1 is identical to Model 1 (basic), but in the opposite direction. MALPEs for the Average models are the lowest of all: -0.35 percent for Average 1 (average of the four initial models) and –0.07 percent for Average 2 (average of Alternative Model 1 plus the ratio, change, and ratio-change models). Note that all MALPEs here are quite similar and close to zero. As with predictive accuracy, all regression models show less bias than the three traditional methods. The MALPEs for the traditional methods were 1.38 percent for the state percentage approach, –2.49 percent for the county percentage change approach, and 3.79 percent for the no-change approach.

---

[16] Note that in all the regression models tested, each independent variable was divided by the county-level number of households obtained from the decennial census. In order to test our ability to estimate change using a model with the data that will be available to us in postcensal years, we created a new household variable for the denominator. The numbers of households were obtained from a 1990 Census-based estimate of housing units for April 1, 2000 and the occupancy rate from the 1990 Census. We multiplied the coefficients generated from the 1990 Alternative Model 1 with the independent variables thus derived to obtain a new predicted PPH value for 2000. We compared the predicted values to the observed values for Census 2000 and obtained a MAPE of 2.95 percent and a MALPE of 1.34 percent.

Evaluation of the predicted PPH at the state level: Predictive accuracy and bias

The counties within the nation were grouped into their respective states. The predicted state level PPH was derived from averaging the predicted county level PPH values for all the counties within the state. The MAPE values are shown in Table 9 and MALPE values in Table 10.

For predictive accuracy (Table 9), Model 1 (basic) generally performs very well compared with the traditional methods for all states with a few exceptions, including California (notable because of its size). The MAPE for California under Model 1 is 4.18, compared with values under 3 for all of the traditional methods. The traditional methods also fare considerably better than Model 1 for Hawaii and Iowa. Model 2 (ratio) fares less well than the traditional approaches, and generally less well than Model 1, especially in Hawaii, where it reaches a value of 26.15 percent. Model 3 (change) generally fares the best of the four regression models and produces lower MAPE values than the traditional approaches for most states, including California. The exceptions are Hawaii and Iowa.

In terms of bias (Table 10), the change model (Model 3) shows the smallest magnitudes of MALPEs among the regression models. The range is from –8.96 percent for Iowa to +0.79 percent for South Dakota. The range is rather similar for Model 1, but there are larger numbers of MALPE values in Model 1 in excess of one percent, positive or negative.

Between the two models that predict on the basis of ratios, Models 2 and 4, Model 2 generally shows less bias, with the notable exception of Hawaii, with a mean overstatement of 21.10 percent, the largest magnitude bias in the table.

Among the traditional methods used for comparison, the “no change” method (Method A) almost consistently overestimates PPH, while the “county change” method (Method B) underestimates it with the same consistency. Method C, the “state change” method, produces levels comparable with the regressions, but the overstatement of the PPH is far more consistent among states.

Evaluation of the predicted PPH at the county level: Predictive accuracy and bias

Next, we examine the county-level MAPE and MALPE values based on the five regression models, including Alternative Model 1. The counties are grouped first by county population size in Census 2000 and then by county population change between 1990 and 2000. The results by population size are shown in Tables 11 and 12 and by population change in Tables 13 and 14.

Predictive accuracy and bias by county population size. Except for Model 3 (change), the mean absolute percent error (MAPE) values are generally higher for counties with fewer than 25,000 people, lower for counties with 25,000 to 99,999 people and higher again for counties with 100,000 or more people (see Table 11 and Figure 1). Model 3 (change) shows the smallest MAPE among all models when compared by size of the county, except for counties with 10,000 to 24,999 people. The next smallest MAPEs by size of county are shown by Model 1 (Column 2 in Table 11).

Alternative Model 1 and the ratio model show MALPE values of 2.08 and 2.98, respectively, for counties with fewer than 10,000 people (see Table 12). Other models do better for these small counties. Model 1 understates PPH for all size categories except the category with fewer than 10,000 people. Alternative Model 1, on the contrary, overstates PPH for all size categories except the ones with 250,000 to 499,999 people and 500,000 or more people. Model 3 (change) understates PPH in all size categories. Model 2 (ratio) understates PPH for all size categories except the ones with fewer than 25,000 people whereas Model 4 (ratio change) understates PPH for all size categories except the ones with fewer than 10,000 people.

Predictive accuracy and bias by county population change. Table 13 and Figure 2 exhibit the MAPE values by percent change in county population between 1990 and 2000. Unlike county population size, county growth does not show any particular pattern for MAPEs. However, the group of counties that grew between 15.0 percent to 25.0 percent in 1990-2000 shows the smallest MAPEs produced by three regression models: Model 1 (1.52 percent), Alternative Model 1 (1.85 percent), and Model 3 (1.20 percent). Similarly, the group that grew 40 percent or more commonly shares one of the largest MAPEs produced by all five regression models.

Finally, we present the MALPE values by percent change in county population between 1990 and 2000 (Table 14). Both Model 1 (basic) and Model 4 (ratio change) show underestimation of PPH at all levels of county change except less than –5.0 percent. Alternative Model 1 (basic) and Model 2 (ratio), on the contrary, show overestimation of PPH at lower-to-mid levels of county change (from –5.0 percent to +15.0 percent) and underestimation of PPH at higher levels of county change (25.0 percent or more). Model 3 (change) is the only model that consistently underestimates PPH irrespective of the magnitude of the percent change in county population between 1990 and 2000. This also holds true when examined by county population size (Column 5 in Table 12).

CONCLUSION

Previous research has shown that regression models are capable of producing more precise and less biased PPH estimates than those produced by more common methods. Our research supports these findings. Results from the basic regression models (both Model 1 and Alternative Model 1) as well as the change model (Model 3) indicate that the regression approach may be an effective methodology for estimating the county-level PPH in postcensal years. It is worth noting that the change model had the lowest overall MAPE compared to other models, but it consistently underestimated the PPH regardless of the county population size in 2000 or population growth between 1990 and 2000. Nevertheless, the selected independent variables explained most of the variance in the basic and change models.

The regression based MAPE is 1.74 percent for Model 1 (basic) and 2.41 percent for Alternative Model 1 (basic). Thus, the precision of the regression based PPH appears to be good. However, the regression based predicted PPH from Model 1 (basic) appears to be slightly lower on average than the actual value, i.e., we observe a small downward bias (MALPE = -0.55 percent). On the contrary, the regression based predicted PPH from Alternative Model 1 (basic) appears to be slightly higher on average than the actual value (MALPE = 0.55 percent).

In comparison, were we to keep constant the 1990 PPH values throughout the decade (Method A), we would generate an error of 4.2 percent and introduce an upward bias close to 3.8 percent. If we kept constant the change observed in the previous decade (Method B), we would find the MAPE to be 3.4 percent and the predicted PPH in 2000 to be lower by 2.5 percent than the actual PPH. Finally, using the change observed for a state for each county within that state (Method C), yields a predicted PPH that is 1.4 percent higher than the actual PPH.

The initial models we tested rely on Census 2000 data for predicted 2000 PPH, which would not be available in postcensal years. Thus, we refined Model 1 (basic) to explore if we could create an acceptable model based only on administrative records. We were able to fit a model using school enrollment data, birth records, tax records (IRS exemptions per household), and enrollment in Medicare per household. Although information on school enrollment was taken from censuses, it could be estimated using the Common Core of Data (CCD) or extrapolation of values from earlier censuses. Assuming that the county-level data on the explanatory variables would be available (or estimated) annually from administrative records files, they can be applied to Alternative Model 1 (basic) for 2000 to develop the county-level PPH estimates in postcensal years. The next step in the research process is to produce postcensal PPH estimates using this model and examine how they compare with survey based PPH estimates.

PART 2: USING A REGRESSION APPROACH TO ESTIMATE VACANCY RATES IN THE PRODUCTION OF HOUSING UNIT-BASED POPULATION ESTIMATES

The research in the second part of the report focuses on the vacancy component. Specifically, the research addresses the question: Can a regression approach be used to estimate vacancy rates?

The vacancy rate is one of the components of the Housing Unit Method for estimating population. In the Housing Unit Method, population is calculated as the number of households times the average number of persons per household (PPH) plus the population residing in group-quarters facilities. A household is conceptualized as an occupied housing unit and occupancy, in turn, is derived from knowledge of vacancy. Typically, the data come from the decennial census. This data source works best when the period of estimation is close to the census year or when the vacancy rates remain constant or follow stable trends. It produces less accurate estimates when vacancy rates or trends change rapidly over a decade. Thus, the regression approach might serve as an alternative methodology for deriving estimates of the vacancy rates at the county level.

Vacancy rates do not necessarily follow the same path as the PPH. Nationally, the vacancy rate changed by more than 1 percentage point between the last two censuses, declining from around 10 percent in the 1990 Census to around 9 percent in Census 2000. At lower levels of geography such as state and county levels, it is difficult to detect a pattern of change. For example, at the state level, between 1990 and 2000, the percent change was as high as 70.8 percent in Hawaii and as low as 0.3 percent in Virginia (Woodward and Damon, 2001). At the county level, where the number of housing units varies considerably and the population size can be very small, the change in the vacancy rates also varies accordingly. Illustratively, the change was more than 300 percent in a Census Area in Alaska, but only 0.01 percent in one county in Nebraska. Thus, as was observed for the PPH component, the challenge is to develop a regression model to predict variations of different magnitude and direction for areas with different population sizes and numbers of housing units.

Methodology for Constructing Regression Models for Vacancy Rates

The methodological steps for constructing regression models for the vacancy rates are identical to the steps outlined in Part 1 of this paper with the dependent variables being the vacancy rates. Lacking a theoretical framework, we explored if variables identified as successful when predicting PPH at the county level also predict vacancy. In the county level regression models, the presence of infants (births per household) and children (school enrollment) in an area were positively related to PPH, while the presence of older people (those aged 65 and older) was negatively related because older people often live in one- or two-person households (Smith et al., 2002). Theoretically, it is not clear if these variables would have the same impact on vacancy rates as they have on the PPH, but they do seem to reflect the demographic composition in an area that should affect the vacancy rate of that area.

In addition to these variables, we tested variables reflecting the population available to participate in the labor force (age of householder), number of family households, percent Hispanic, and whether or not the county is urban or rural in composition. County in- or out-migration might also be related to vacancy. The annual change in the number of tax exemptions in a county was used to measure the net change in population due to migration.

Smith et al. (2002) noted that a lagged PPH variable is a good predictor of the current PPH. The variable has a negative impact on changes in PPH. When overall PPH levels are falling, declines are expected to be larger for counties with high PPH values than for counties with low PPH values. Replicating this approach, we also included the vacancy rate from the previous census as a predictor of the current vacancy rate.

Results of Regression Analyses for Vacancy Rates

The basic model and the ratio model were the models with the best fit. The change and the ratio change models introduced for the prediction of PPH values were not pursued for this analysis. Examining Model 1 (basic) for 1990 and 2000, the independent variables explain a relatively modest proportion of variation in the dependent variable in both years: about 53 percent in 1990 and about 65 percent in 2000. The statistical significance of the coefficients changes from 1990 to 2000, and some of the coefficients change sign (school enrollees, change in tax exemptions, householder aged 35-44). In 1990, the variable ‘births’ makes a statistically significant contribution, but this determinant is not a significant factor in 2000. Similarly, the variable ‘deaths’ contributes to explaining the variation in 2000 but not in 1990. Population growth is expected to be captured through the number of births per household in the county and to be negatively correlated with vacancy rates. The variable representing the deaths per household is thought to capture the aging of the population in the county and to be positively correlated with the vacancy rate (Table 15).

In Model 2 (ratio), the proportion of older people, the proportion of births, and the proportion nonfamily households in the county are no longer relevant in the model. The variables representing the number of deaths and the number of householders aged 35-44 are present in the model for 2000 but not 1990 (Table 15).

In summary, the ability of the selected variables to explain variation is inconsistent from one period to the next. For the variables common to all models, such as proportion Hispanic and classification as rural, the sign changes across models. This suggests that the identified variables are not reliable as predictors of vacancy.

Evaluation of the Regression Models for Predicting Vacancy Rates

Given the instability of the models and the relatively modest ability to explain variation in county-level vacancy rates compared with PPH, the evaluation of the regression results was limited to the basic regression model. The results from the basic regression model for 1990 were used to generate predicted vacancy rates in 2000. The predicted values were compared with the observed Census 2000 values. The regression model produced an estimate that was about 12 percent lower than the observed national vacancy rate (MALPE = 12.22 percent). When the direction of the differences is ignored, the mean absolute percent error (MAPE) at the national level was more than 50 percent. Inspection of the data showed that the results were particularly erroneous for small counties. The regression approach generated predicted vacancy rates that were very different from the observed vacancy rate in 2000.

Conclusion for Using the Regression Approach to Estimate Vacancy Rates

Regardless of the indicators identified, we surmise that it is difficult to predict the county-level effect. What holds true for vacancy rates in one part of the county may be offset by what happens in another part. Thus, the selection of predictor variables is not straightforward. Additionally, for our purposes, it is desirable that the determinants or predictor variables, once identified through the regression approach, be available at the county level and on a regular and frequent basis for use in the estimation. The present research was not successful in identifying appropriate regression models. Further exploratory research is needed before it can be determined if the regression approach is a proper methodology for vacancy rate estimation.

OVERALL CONCLUSION

The purpose of this research was to report on the regression approach as a method to estimate PPH values and the vacancy rates at the county level. It was the intent to learn if a regression approach might yield PPH values and vacancy rates that are more accurate than those based on previous decennial values or historical trends.

Results from the basic regression models as well as the change model indicate that the regression approach may be an effective methodology for estimating the county-level PPH in postcensal years. The selected independent variables explained most of the variance in the estimates, especially in the basic models. The amount of variation explained in the change model was less than the basic models suggesting that it is more difficult to explain change in PPH over time than to predict the current PPH, or that variables that are able to predict the current PPH do not necessarily explain all the change over time. The amounts of variation explained in both the ratio and the ratio-change models were even lower suggesting that these models do not look promising for predicting the PPH.

The mean absolute percentage errors generated by comparing the regression-based predictions (basic models) with the observed values in the decennial census suggest that the precision of the regression based PPH appears to be good (less than 2 percent). The regression based PPH for Model 1 (basic) appears to be slightly lower on average than the actual value (MALPE = -0.55 percent) whereas for Alternative Model 1 (basic) the regression based PPH appears to be slightly higher than the actual PPH (MALPE = 0.55 percent).

In comparison, were we to keep constant the 1990 PPH values throughout the decade, we would generate an error of 4.2 percent and introduce an upward bias close to 3.8 percent. If we kept constant the change observed in the previous decade, we would find the MAPE to be 3.4 percent and the predicted PPH in 2000 to be lower by 2.5 percent than the actual PPH.

Our research was inconclusive in regard to the vacancy rates. We were not successful in finding good predictor variables. Future research might explore if variables other than the ones tested in this report could be used as predictors of vacancy. Furthermore, in the analysis conducted here, the counties were combined into a single sample. It is possible that a “one size fits all” approach does not work as well as an approach for groups of counties. (For example, separate regression models might be constructed for urban counties, rural counties, counties with large populations, and counties with small populations.) Furthermore, in practice, if these models were used in non-census years (e.g., 2006), the independent variables would need to be obtained from administrative data and/or ACS data—and the latter would, of course, introduce additional sampling variability, a topic that would itself require research.

REFERENCES

Bryan, T., 2004, “Population Estimates,” pp. 523-560 in J.S. Siegel and D.A. Swanson (Eds.) The Methods and Materials of Demography, Second Edition. San Diego, CA: Elsevier.

Davis, S. T, 1994, “Evaluation of Postcensal County Estimates for the 1980s,” Population Division Working Paper No. 5, Issued in March.

Huber, P.J., 1967, “The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions,” Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1:221-233.

Madigan D., and Raftery A.E., 1994, “Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window,” Journal of the American Statistical Association, 89: 1535-1546.

Namboodiri, N.K, 1972, “On the Ratio-Correlation and Related Methods of Subnational Population Estimation,” Demography 9:443-53.

National Research Council, 1980, Estimating Population and Income of Small Areas, Washington, DC: National Academy Press.

O’Hare, W., 1976, “Report on a Multiple Regression Method for Making Population Estimates,” Demography 13:369-79.

Rives, N., 1982, “Assessment of a Survey Approach,” pp. 79-96 in E. Lee and H. Goldsmith (Eds.) Population Estimates: Methods for Small Area Analysis, Beverly Hills, CA: Sage Publications.

Smith, S. and B. Lewis, 1980, “Some New Techniques for Applying the Housing Unit Method of Local Population Estimation,” Demography 17:323-339.

_____1983, “Some New Techniques for Applying the Housing Unit Method of Local Population Estimation: Further Evidence,” Demography 20:407-413.

Smith, S., J. Nogle, and S. Cody, 2002, “A Regression Approach to Estimating the Average Number of Persons Per Household,” Demography 39:697-712.

Starsinic D. and Meyer Zitter, 1968, “Accuracy of the Housing Unit Method in Preparing Estimates for Cities,” Demography 5: 475-484.

StataCorp, 2005, Stata Statistical Software: Release 9, College Station, TX: StataCorpLP.

Woodward, J. and B. Damon, 2001, “Housing Characteristics: 2000,” Census 2000 Brief; Issued in October.

Table 1. Number of Counties with Change in Persons per Household (PPH) and Vacancy Rates: 1990-2000

	Persons Per Household					
	Total		Decrease		Increase	
Vacancy Rates	Number	Percent	Number	Percent	Number	Percent
Total	3,138[1]	100.0	2,816	100.0	322	100.0
Decrease	1,718	54.8	1,474	52.3	244	75.8
Increase	1,420	45.2	1,342	47.7	78	24.2

[1] In order to keep the data sets comparable between 1990 and 2000, only counties that existed in both years were kept in the analysis
(see http://www.census.gov/geo/www.tiger/ctychng.html)
Source: U.S.Census Bureau, 1990 Census and Census 2000.

Table 2. Number of Counties with Change in the Number of Housing Units, Persons per Household (PPH) and Vacancy Rates: 1990-2000

	Persons Per Household					
	Total		Decrease		Increase	
	Number	Percent	Number	Percent	Number	Percent
Total	3,138[1]	100.0	2,816	89.7	322	10.3
Counties where the number of housing units decreased from 1990 to 2000						
Vacancy Rates						
Total	418	100.0	383	91.6	35	8.4
Decrease	285	68.2	258	67.4	27	77.1
Increase	133	31.8	125	32.6	8	22.9
Total		100.0		100.0		100.0
Counties where the number of housing units increased from 1990 to 2000						
Vacancy Rates						
Total	2,720	100.0	2,433	89.4	287	10.6
Decrease	1,433	52.7	1,216	50.0	217	75.6
Increase	1,287	47.3	1,217	50.0	70	24.4
Total		100.0		100.0		100.0

[1] In order to keep the data sets comparable between 1990 and 2000, only counties that existed in both years were kept in the analysis
(see http://www.census.gov/geo/www.tiger/ctychng.html).
Source: U.S.Census Bureau, 1990 Census and Census 2000.

Table 3. National Level Average Persons per Household: 1960-2006

Decade	Years									
	00	01	02	03	04	05	06	07	08	09
1960s	3.33	3.36	3.31	3.33	3.33	3.31	3.30	3.28	3.23	3.19
1970s	3.14	3.11	3.06	3.01	2.97	2.94	2.89	2.86	2.81	2.78
1980s	2.76	2.73	2.72	2.73	2.71	2.69	2.67	2.66	2.64	2.62
1990s	2.63	2.63	2.62	2.63	[1]2.67	2.65	2.65	2.64	2.62	2.61
2000s	2.62	2.58	2.58	2.57	2.57	2.57	2.57	----	----	----

[1] Revised PPH based on population from the decennial census for that year = 2.66.
Source: U.S. Census Bureau, Current Population Survey, March and Annual Social and Economic Supplements, 2006 and earlier. Internet Release Date: March 27, 2007.

Table 4. Distribution of Counties by PPH level: 1980, 1990, and 2000

PPH	1980		1990		2000	
	Number	Percent	Number	Percent	Number	Percent
Total	3,135	100.0	3,139	100.0	3,141	100.0
Less than 2.20	6	0.2	14	0.4	35	1.1
2.20 - 2.29	12	0.4	36	1.2	142	4.5
2.30 - 2.39	30	1.0	180	5.7	500	15.9
2.40 - 2.49	152	4.9	515	16.4	859	27.4
2.50 - 2.59	307	9.8	819	26.1	712	22.7
2.60 - 2.69	522	16.7	668	21.3	422	13.4
2.70 - 2.79	672	21.4	431	13.7	205	6.5
2.80 - 2.89	561	17.9	218	6.9	114	3.6
2.90 - 2.99	375	12.0	119	3.8	66	2.1
3.00 - 3.09	226	7.2	61	1.9	23	0.7
3.10 - 3.19	132	4.2	22	0.7	23	0.7
3.20 - 3.29	68	2.2	19	0.6	11	0.4
3.30 - 3.39	26	0.8	8	0.3	8	0.3
3.40 - 3.49	11	0.4	8	0.3	8	0.3
3.50 - 3.59	10	0.3	4	0.1	2	0.1
3.60 or more	25	0.8	17	0.5	11	0.3

Source: U.S. Census Bureau, 1980 census, 1990 census, and Census 2000.

Table 5. Ordinary Least Squares (OLS) Regression Coefficients and Standard Errors for Models 1 and 2: 1990 and 2000

Variable	Model 1 (Basic) 1990 (1)	Model 1 (Basic) 2000 (2)	Model 2 (Ratio) 1990 (3)	Model 2 (Ratio) 2000 (4)
PPH Intercept	2.244 ** (0.019)	2.245 ** (0.020)	1.005 ** (0.001)	0.991 ** (0.001)
School Enrollees	0.824 ** (0.015)	0.658 ** (0.016)	3.956 ** (0.580)	3.749 ** (0.552)
Births	5.773 ** (0.142)	6.356 ** (0.184)	3.389 ** (0.362)	3.126 ** (0.413)
Deaths	-1.352 ** (0.251)	-2.633 ** (0.224)	-2.518 ** (0.323)	-2.349 ** (0.303)
Hispanic Origin	0.021 ** (0.003)	0.035 ** (0.004)	-0.120 * (0.054)	-0.037 (0.057)
Householder Aged 25-44	0.225 ** (0.031)	0.451 ** (0.036)	-3.631 ** (0.651)	-2.809 ** (0.621)
Nonfamily Household	-1.299 ** (0.026)	-1.195 ** (0.029)	-0.944 * (0.448)	-1.535** (0.408)
R^2	93.0 percent	90.1 percent	18.8 percent	19.0 percent

Note: Values in parentheses are the standard error (s.e.) of the coefficients. The s.e. allows for the computation of a confidence interval for a parameter estimate.

** indicates that the parameter estimate is significant at the 99 percent confidence level.

* indicates that the parameter estimate is significant at the 95 percent confidence level.

Source: Administrative Records and Decennial Census Data.

Table 6. Ordinary Least Squares (OLS) Regression Coefficients and Standard Errors for Models 3 and 4: 1980-1990 and 1990-2000

Variable	Model 3 (Change) 1980 - 1990 (1)	Model 3 (Change) 1990 - 2000 (2)	Model 4 (Ratio Change) 1980 - 1990 (3)	Model 4 (Ratio Change) 1990 – 2000 (4)
PPH Intercept	-0.020 ** (0.003)	-0.014 ** (0.002)	-0.008 ** (0.000)	-0.013 ** (0.000)
School Enrollees	0.805 ** (0.014)	0.359 ** (0.013)	2.715 ** (0.247)	4.003 ** (0.329)
Births	3.542 ** (0.107)	1.178 ** (0.138)	1.060 ** (0.161)	1.068 ** (0.230)
Deaths	-0.112 (0.214)	0.029 (0.195)	0.215 (0.198)	-0.445 (0.248)
Hispanic Origin	0.306 ** (0.016)	0.323 ** (0.012)	0.225 ** (0.065)	0.096 * (0.048)
Householder Aged 25-44	0.161 ** (0.036)	0.254 ** (0.031)	-0.966 ** (0.374)	-1.009 ** (0.385)
Nonfamily Household	-1.581 ** (0.052)	-2.172 ** (0.053)	-2.182 ** (0.237)	-2.627 ** (0.291)
R^2	76.9 percent	68.9 percent	11.6 percent	12.3 percent

Note: Values in parentheses are the standard error (s.e.) of the coefficients. The s.e. allows for the computation of a confidence interval for a parameter estimate. The ratio-change model has correlated error-terms within states. The standard errors (and hence the p-values) are not exact.

** indicates that the parameter estimate is significant at the 99 percent confidence level.

***** indicates that the parameter estimate is significant at the 95 percent confidence level.

Source: Administrative Records and Decennial Census Data.

Table 7. Ordinary Least Squares (OLS) Regression Coefficients and Standard Errors for Alternative Model 1: 1990 and 2000

	Alternative Model 1 (Basic)	
Variable	1990 (1)	2000 (2)
PPH Intercept	1.818 ** (0.011)	1.736 ** (0.012)
School Enrollees	1.168 ** (0.015)	0.858 ** (0.018)
Births	4.139 ** (0.168)	6.941 ** (0.195)
Tax Exemptions	0.040 ** (0.002)	0.110 ** (0.005)
Medicare Enrollees	-0.245 ** (0.014)	-0.275 ** (0.0149)
R^2	89.3 percent	85.7 percent

Note: Values in parentheses are the standard error (s.e.) of the coefficients. The s.e. allows for the computation of a confidence interval for a parameter estimate.

** indicates that the parameter estimate is significant at the 99 percent confidence level.

* indicates that the parameter estimate is significant at the 95 percent confidence level.

Source: Administrative Records and Decennial Census Data.

Table 8. Mean Absolute Percentage Errors (MAPEs) and Mean Algebraic Percentage Errors (MALPEs) for County PPH Estimates at the National Level: All Models, 2000

Model or Approach	Evaluation: MAPE (1)	Evaluation: MALPE (2)
Regression Models		
Model 1 (Basic)	1.74	-0.55
Model 2 (Ratio)	4.48	0.49
Model 3 (Change)	1.52	-0.68
Model 4 (Ratio change)	2.39	-0.65
Alternative Model 1 (Basic)	2.41	0.55
Average 1 (Average of Models 1-4)	1.82	-0.35
Average 2 (Average of Models 2-4 and Alternative Model 1)	1.91	-0.07
Traditional Approaches		
A (No change)	4.20	3.79
B (County percentage change)	3.37	-2.49
C (State percentage change)	2.41	1.38

Source: Administrative Records and Decennial Census Data.

Table 9. Mean Absolute Percentage Errors (MAPEs) for County PPH Estimates by State: All Models, 2000

State	Regression Models					Traditional Methods		
	Model 1 (Basic) (1)	Model 2 (Ratio) (2)	Model 3 (Change) (3)	Model 4 (Ratio-Change) (4)	Alternative Model 1 (Basic) (5)	Method A: No Change (6)	Method B: 1980-1990 County Change (7)	Method C: 1990-2000 State Change (8)
Alabama	1.08	2.98	1.01	2.42	1.79	6.49	2.87	2.15
Alaska	2.88	8.63	1.43	3.39	3.81	4.85	4.46	3.54
Arizona	2.05	7.10	1.50	3.30	1.92	4.46	4.01	5.23
Arkansas	0.91	3.31	1.07	2.01	1.57	4.08	3.10	2.07
California	4.18	8.49	1.78	6.82	3.51	2.90	2.24	2.99
Colorado	2.34	5.14	2.28	3.29	3.11	3.84	5.45	4.34
Connecticut	1.84	4.69	0.41	3.13	1.14	3.15	3.36	1.46
Delaware	1.13	8.02	0.68	2.29	0.84	3.18	3.47	0.91
Florida	1.78	4.96	3.81	2.13	2.17	4.96	5.25	4.97
Georgia	1.77	3.74	1.41	2.33	2.40	4.26	4.40	3.69
Hawaii	8.66	26.15	5.43	6.12	12.37	4.90	2.33	2.18
Idaho	2.01	6.70	1.63	2.51	2.23	3.96	4.00	3.20
Illinois	1.30	4.60	1.07	2.44	1.89	2.94	2.45	2.37
Indiana	1.01	3.09	0.92	1.33	1.43	3.57	2.47	1.35
Iowa	7.82	2.62	8.96	2.47	9.23	2.68	3.87	1.43
Kansas	1.59	6.23	1.30	1.77	3.17	2.43	2.71	1.98
Kentucky	1.13	2.84	0.96	2.06	1.42	5.60	2.98	2.20
Louisiana	1.33	3.37	1.23	1.07	2.03	5.83	1.95	1.48
Maine	0.44	2.05	0.70	3.57	2.53	7.58	0.91	1.15
Maryland	1.69	5.46	0.70	3.57	1.81	3.46	3.60	1.84
Massachusetts	3.10	2.49	1.02	5.60	1.81	3.39	2.14	1.59
Michigan	1.17	4.60	1.23	2.43	1.87	5.04	2.56	1.68
Minnesota	0.98	3.67	0.84	2.20	2.20	4.13	3.23	2.06
Mississippi	1.16	3.18	1.10	1.67	1.56	5.30	3.40	1.80
Missouri	1.02	3.06	1.15	2.10	1.82	2.57	2.85	1.65
Montana	1.51	3.89	2.19	2.58	3.49	4.22	4.43	2.34
Nebraska	1.60	4.33	1.34	2.27	3.29	3.44	3.89	2.34
Nevada	2.96	4.78	1.61	3.09	4.56	3.71	3.64	5.63
New Hampshire	1.65	4.97	0.89	5.08	1.15	4.50	1.66	1.15

Table 9. Mean Absolute Percentage Errors (MAPEs) for County PPH Estimates by State: All Models, 2000--Continued

	Regression Models					Traditional Methods		
State	Model 1 (Basic) (1)	Model 2 (Ratio) (2)	Model 3 (Change) (3)	Model 4 (Ratio-Change) (4)	Alternative Model 1 (Basic) (5)	Method A: No Change (6)	Method B: 1980-1990 County Change (7)	Method C: 1990-2000 State Change (8)
New Jersey	3.94	4.10	0.51	5.29	3.36	2.08	3.89	1.76
New Mexico	2.26	6.34	2.46	2.48	2.87	5.84	11.55	2.91
New York	2.26	3.71	0.57	2.53	2.39	4.59	1.97	3.92
North Carolina	1.16	3.69	0.89	2.15	1.44	3.88	5.70	2.41
North Dakota	0.93	3.89	1.30	2.52	3.45	6.75	3.23	2.13
Ohio	1.01	3.80	0.85	1.59	1.29	4.74	1.81	1.21
Oklahoma	1.16	3.52	1.09	1.68	2.04	2.34	2.51	1.47
Oregon	1.81	4.69	1.68	2.48	1.84	2.75	3.43	2.56
Pennsylvania	1.81	2.83	0.53	3.01	1.44	4.44	2.77	1.55
Rhode Island	2.51	3.76	0.70	4.06	1.48	4.78	2.63	1.95
South Carolina	0.93	2.12	1.47	1.93	1.60	7.29	2.82	2.36
South Dakota	1.86	6.64	1.49	2.16	3.22	3.99	3.69	2.24
Tennessee	0.87	3.09	0.86	1.88	1.30	4.43	4.12	2.02
Texas	1.69	7.63	1.39	2.25	2.35	2.95	3.13	3.21
Utah	2.51	8.70	2.28	3.71	2.92	4.66	3.82	4.08
Vermont	1.26	3.59	1.20	4.52	1.94	5.46	2.09	0.90
Virginia	1.77	5.49	1.24	2.20	2.33	5.39	3.77	3.13
Washington	1.41	5.88	2.00	3.45	1.67	2.59	3.06	2.60
West Virginia	1.11	1.91	0.83	1.27	1.11	6.79	2.64	1.51
Wisconsin	1.17	3.67	0.68	3.03	1.62	5.14	2.05	1.17
Wyoming	1.74	2.60	2.09	2.20	3.28	6.33	3.11	1.79

Source: Administrative Records and Decennial Census Data.

Table 10. Mean Algebraic Percentage Errors (MALPEs) for County PPH Estimates by State: All Models, 2000

State	Regression Models					Traditional Methods		
	Model 1 (Basic) (1)	Model 2 (Ratio) (2)	Model 3 (Change) (3)	Model 4 (Ratio-Change) (4)	Alternative Model 1 (Basic) (5)	Method A: No Change (6)	Method B: 1980-1990 County Change (7)	Method C: 1990-2000 State Change (8)
Alabama	0.67	1.68	0.65	2.17	1.25	6.49	-2.48	1.36
Alaska	-1.86	-4.51	-0.89	-0.98	2.16	4.42	-3.51	2.26
Arizona	-1.52	-3.04	-1.09	2.82	0.80	4.04	-2.75	4.91
Arkansas	0.54	2.21	-0.24	1.29	0.78	3.77	-2.57	0.57
California	-3.93	-0.94	-1.60	-6.75	-1.92	-0.78	-0.17	2.03
Colorado	-0.11	0.87	-1.95	0.26	2.31	2.48	-4.21	3.50
Connecticut	-1.84	-0.96	0.11	-3.13	0.10	3.10	-3.36	0.88
Delaware	-1.13	3.30	-0.68	-1.74	-0.26	3.18	-3.47	0.43
Florida	-1.09	-1.93	-0.66	-0.36	-0.34	2.76	-3.03	2.78
Georgia	-0.14	-0.14	0.32	0.98	0.06	3.78	-3.94	3.06
Hawaii	-8.66	21.10	-5.43	-6.12	0.56	4.90	-0.47	1.78
Idaho	-0.81	0.65	-1.10	0.16	-0.05	2.58	-2.31	0.96
Illinois	-0.65	3.17	-0.96	-2.37	0.93	2.81	-2.35	2.14
Indiana	-0.52	-0.92	-0.53	-0.64	0.00	3.56	-2.30	0.40
Iowa	-7.82	-0.46	-8.96	-2.34	-9.23	2.48	-3.73	0.04
Kansas	0.98	3.60	-0.44	0.32	2.94	1.64	-1.79	0.69
Kentucky	0.53	0.22	-0.09	0.32	0.42	5.60	-2.87	0.36
Louisiana	0.01	-0.98	0.39	0.45	1.00	5.83	-0.84	1.07
Maine	-0.33	-1.57	-0.58	0.15	2.53	7.58	-0.11	0.59
Maryland	-1.67	-1.62	-0.11	-3.30	-0.44	3.44	-3.35	1.01
Massachusetts	-3.10	-1.32	-0.59	-3.57	0.64	2.89	-1.79	-0.07
Michigan	-0.54	0.98	-0.89	-5.60	1.23	5.04	-2.34	1.20
Minnesota	-0.25	-1.25	-0.36	-2.26	1.98	4.11	-3.04	1.72
Mississippi	-0.31	0.45	0.03	-1.94	0.57	5.30	-2.82	0.71
Missouri	0.05	0.21	-0.81	0.30	1.23	2.17	-2.58	-0.22
Montana	0.46	-0.57	-1.66	-1.69	3.29	4.13	-3.55	0.59
Nebraska	1.03	1.97	-0.31	-1.47	3.04	2.73	-2.58	0.75
Nevada	0.59	1.08	0.47	-0.42	3.82	1.90	-0.83	5.53
New Hampshire	-1.65	-0.90	-0.89	0.44	0.96	4.50	-0.36	0.89

Table 10. Mean Algebraic Percentage Errors (MALPEs) for County PPH Estimates by State: All Models, 2000--Continued

State	Regression Models					Traditional Methods		
	Model 1 (Basic) (1)	Model 2 (Ratio) (2)	Model 3 (Change) (3)	Model 4 (Ratio-Change) (4)	Alternative Model 1 (Basic) (5)	Method A: No Change (6)	Method B: 1980-1990 County Change (7)	Method C: 1990-2000 State Change (8)
New Jersey	-3.94	-3.56	-0.21	-5.29	-3.27	1.37	-3.69	0.52
New Mexico	1.15	1.82	-1.11	-0.43	2.11	5.79	0.93	1.58
New York	-2.19	-1.55	-0.31	-2.42	0.41	4.33	-1.42	3.51
North Carolina	0.34	1.96	-0.01	1.61	0.59	3.76	-5.70	1.51
North Dakota	0.34	-0.74	-0.21	-1.67	3.42	6.74	-1.86	0.63
Ohio	-0.80	-3.35	-0.72	-1.46	0.14	4.74	-1.62	0.56
Oklahoma	0.69	2.19	-0.82	0.99	1.67	1.94	-1.69	0.19
Oregon	-1.14	-0.95	-1.68	-1.94	0.31	1.59	-3.02	1.22
Pennsylvania	-1.75	-2.31	-0.03	-3.00	-0.37	4.39	-2.67	0.93
Rhode Island	-2.51	-1.50	0.63	-4.06	-0.17	4.78	-2.63	1.45
South Carolina	0.20	-0.51	0.79	0.69	1.20	7.29	-2.13	1.23
South Dakota	-0.08	-0.96	-0.61	-0.11	2.34	3.78	-2.65	0.40
Tennessee	0.26	1.08	-0.24	1.02	0.14	4.37	-4.12	1.12
Texas	0.59	3.84	0.21	0.63	0.51	2.35	-1.15	2.77
Utah	-2.26	4.80	-2.26	3.11	-2.41	4.14	1.19	3.29
Vermont	-1.25	-3.37	-1.20	-4.52	1.77	5.46	-1.31	0.33
Virginia	-1.08	2.00	0.13	-0.26	0.05	5.22	-3.18	2.32
Washington	-1.18	-2.45	-1.72	-3.24	0.71	0.67	-2.73	0.68
West Virginia	-0.91	-0.61	-0.25	-0.36	-0.36	6.79	-2.42	0.33
Wisconsin	-0.93	-2.09	-0.01	-3.00	1.20	5.14	-1.89	0.58
Wyoming	0.37	0.41	-1.72	-1.44	3.13	6.33	0.31	0.33

Source: Administrative Records and Decennial Census Data.

Table 11. Mean Absolute Percent Errors (MAPEs) by County Population Size in Census 2000

County Population Size, Census 2000	Number of Counties (1)	Model 1 (Basic) (2)	Model 2 (Ratio) (3)	Model 3 (Change) (4)	Model 4 (Ratio-Change) (5)	Alt. Model 1 (Basic) (6)
All	3,141	1.74	4.48	1.52	2.39	2.41
<10,000	696	2.05	6.04	1.88	2.80	3.26
10,000 - 24,999	886	1.62	4.12	1.72	2.29	2.33
25,000 - 39,999	446	1.30	3.77	1.13	1.97	1.93
40,000 - 59,999	309	1.50	4.07	1.17	2.06	1.85
60,000 - 99,999	280	1.58	3.68	1.12	2.00	1.84
100,000 - 249,999	293	1.81	3.64	1.19	2.31	2.06
250,000 - 499,999	119	2.16	3.96	1.12	2.83	2.55
500,000 or more	112	3.21	5.75	1.12	4.17	3.36

Source: Administrative Records and Decennial Census Data.

Table 12. Mean Algebraic Percent Errors (MALPEs) by County Population Size in Census 2000

County Population Size, Census 2000	Number of Counties (1)	Model 1 (Basic) (2)	Model 2 (Ratio) (3)	Model 3 (Change) (4)	Model 4 (Ratio-Change) (5)	Alt. Model 1 (Basic) (6)
All	3,141	-0.55	0.49	-0.68	-0.65	0.55
<10,000	696	0.33	2.98	-0.39	0.36	2.08
10,000 - 24,999	886	-0.53	0.69	-0.67	-0.10	0.13
25,000 - 39,999	446	-0.38	-0.28	-0.61	-0.65	0.31
40,000 - 59,999	309	-0.66	-0.88	-0.72	-1.02	0.14
60,000 - 99,999	280	-0.88	-0.88	-0.84	-1.16	0.07
100,000 - 249,999	293	-1.23	-0.39	-1.04	-1.91	0.20
250,000 - 499,999	119	-1.87	-0.69	-1.00	-2.55	-0.18
500,000 or more	112	-3.07	-3.12	-0.89	-3.57	-0.63

Source: Administrative Records and Decennial Census Data.

Table 13. Mean Absolute Percent Errors (MAPEs) by Percent Population Change: 1990-2000

Percent Change in County Population 1990-2000	Number of Counties (1)	Model 1 (Basic) (2)	Model 2 (Ratio) (3)	Model 3 (Change) (4)	Model 4 (Ratio-Change) (5)	Alt. Model 1 (Basic) (6)
All	3,141	1.74	4.48	1.52	2.39	2.41
Less than -5.0	295	2.01	5.48	1.65	2.82	3.57
-5.0 to less than 0.0	391	2.00	3.84	1.99	2.23	3.14
0.0 to 4.99	510	1.78	3.57	1.58	2.16	2.45
5.0 to 9.99	555	1.57	3.96	1.26	2.27	2.06
10.0 to 14.99	427	1.76	4.44	1.32	2.61	2.09
15.0 to 24.99	503	1.52	4.65	1.20	2.47	1.85
25.0 to 39.99	311	1.70	5.55	1.38	2.30	2.12
40.0 or more	149	2.02	6.11	1.71	2.60	2.71

Source: Administrative Records and Decennial Census Data.

Table 14. Mean Algebraic Percent Errors (MALPEs) by Percent Population Change: 1990-2000

Percent Change in County Population 1990-2000	Number of Counties (1)	Model 1 (Basic) (2)	Model 2 (Ratio) (3)	Model 3 (Change) (4)	Model 4 (Ratio-Change) (5)	Alt. Model 1 (Basic) (6)
All	3,141	-0.55	0.49	-0.68	-0.65	0.55
Less than -5.0	295	0.49	3.83	-0.10	1.42	2.72
-5.0 to less than 0.0	391	-0.41	2.09	-1.01	-0.39	1.15
0.0 to 4.99	510	-0.62	0.70	-0.82	-0.92	0.69
5.0 to 9.99	555	-0.55	0.30	-0.56	-0.89	0.67
10.0 to 14.99	427	-0.68	0.19	-0.53	-0.65	0.41
15.0 to 24.99	503	-0.70	-0.91	-0.55	-1.10	0.08
25.0 to 39.99	311	-0.93	-1.03	-0.81	-0.97	-0.65
40.0 or more	149	-1.47	-2.14	-1.44	-1.26	-1.73

Source: Administrative Records and Decennial Census Data.

Table 15. Ordinary Least Squares (OLS) Regression Coefficients and Standard Errors for Vacancy Rates: 1990 and 2000

Variable	Model 1 (Basic)		Model 2 (Ratio)	
	1990	2000	1990	2000
Intercept	-13.445 ** (3.537)	41.731 ** (4.110)	1.784 ** (0.252)	6.759 ** (0.334)
Births	-4.226 * (1.839)	------	------	-------
Deaths	------	-176.610 ** (20.767)	------	-0.317 ** (0.053)
School Enrollees	-25.498 ** (1.871)	15.213 ** (0.285)	-0.962 ** (0.079)	0.436 **
Householder Aged 25-34	-18.707 (5.647)	-61.456 ** (6.215)	-0.770 ** (0.292)	-1.510 ** (0.105)
Householder Aged 35-44	25.625 ** (6.329)	-34.022 ** (6.488)	------	-0.942 ** (0.168)
Population Aged 65+	-4.282 ** (0.532)	-7.145 * (2.586)	------	------
Hispanic Origin	4.626 ** (1.840)	5.244 ** (0.323)	-0.033 * (0.012)	0.081 ** (0.017)
Rural	15.077 ** (0.291)	15.852 ** (0.285)	0.236 ** (0.006)	0.279 ** (0.006)
Nonfamily Household	24.224 ** (2.923)	29.961 ** (3.175)	-------	------
Tax Exemptions	9.559 ** (1.339)	-14.762 ** (1.181)	0.771 * (0.292)	-3.904 ** (0.351)
R^2	53.1 percent	65.4 percent	41.7 percent	58.5 percent

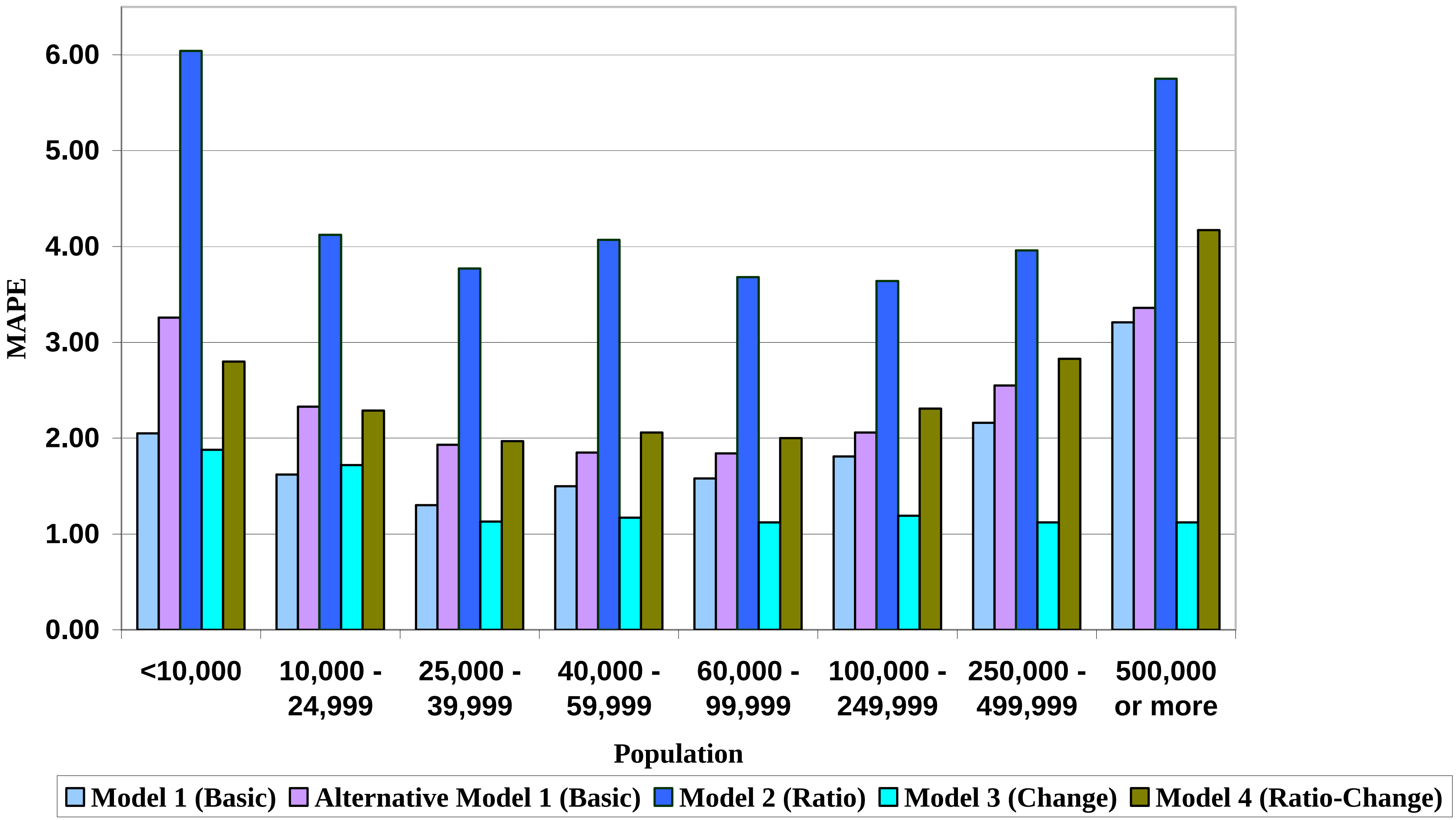
Note: Values in parentheses are the standard error (s.e.) of the coefficients. The s.e. allows for the computation of a confidence interval for a parameter estimate.

** indicates that the parameter estimate is significant at the 99 percent confidence level.

* indicates that the parameter estimate is significant at the 95 percent confidence level.

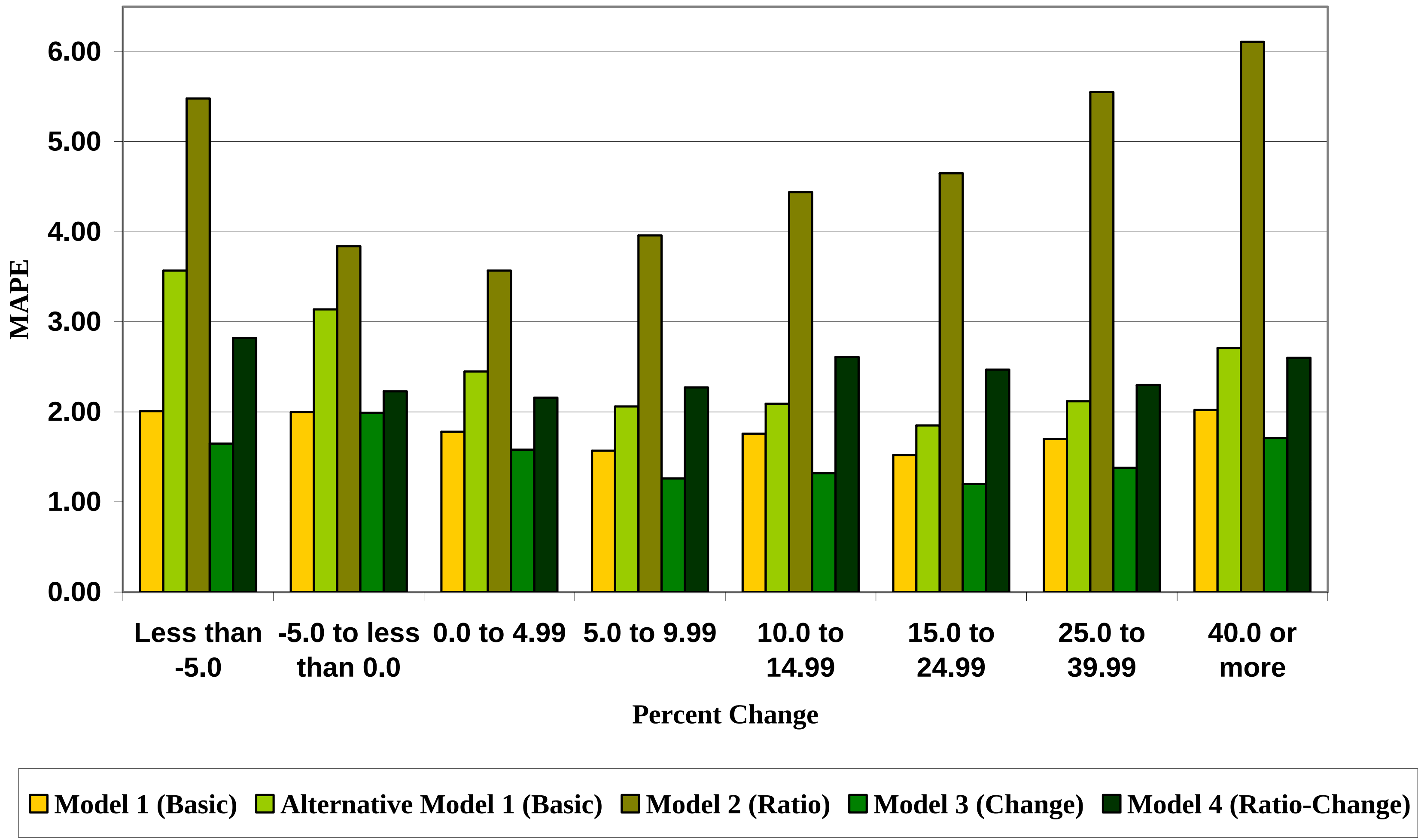
Source: Administrative Records and Decennial Census Data.

Figure 1. Mean Absolute Percent Errors (MAPEs) by County Population Size in Census 2000



Source: Administrative Records and Decennial Census Data.

Figure 2. Mean Absolute Percent Errors (MAPEs) by Percent Change in County Population: 1990-2000



Source: Administrative Records and Decennial Census