# Estimation of Correlation Bias Using Conditional Logistic Regression

Mary Mulry, Kent Wurdeman, and Jay Kim,* U.S. Bureau of the Census,
and Juha Alho, University of Illinois
Mary Mulry, Bureau of the Census, Washington, D.C. 20233

## 1. BACKGROUND

The dual system estimation used for the U.S. Bureau of the Census 1990 Post Enumeration Survey (PES) estimates is based on three independence assumptions: causality, homogeneity, and autonomy. Basically these assumptions say, respectively, that inclusion in the PES sample and the census are independent, that everyone has the same probability of inclusion, and that everyone acts on their own as to whether they are included in the PES sample population or the census. The violation of any of these three assumptions may cause the estimate of the proportion of the population enumerated in the census, and thereby the estimates of the population, to be biased. Such a bias is known as a correlation bias. The focus of this paper is on evaluating whether the homogeneity assumption holds.

## 2. STATISTICAL METHODOLOGY

To discuss the estimation of correlation bias, we need to define the dual system estimator (DSE). The present application of the dual system estimator involves two incomplete lists of the population. The census enumerations of the population not living in institutions or homeless comprise the first list. The second is an implicit list of those persons covered by the sampling frame for the P sample of the PES, which we will call the P-sample population; this list would be obtained if the P sample were conducted for the entire U.S. (instead of a sample) with no measurement errors or missing data.

Whether the i-th individual in the population of size N is in the census or not and in the P sample or not are assumed to be random events with probabilities as shown in Table 2.1. The true population size in each category is also shown in Table 2.1, and $N_{++}=N$ is the total population size. Even if we could observe the $N_{jk}$'s in the first row and first column, the $N_{jk}$'s in parentheses would not be observed directly but would have to be estimated. The estimator, $\tilde{N} = N_{1+}N_{+1}/N_{11}$, is called the *DSE*. The DSE is accurate only to the extent that $N_{11}/N_{+1}$ is an accurate estimate of the proportion of the population enumerated in the census. Accuracy depends on certain independence assumptions being satisfied (Wolter 1986):

Table 2.1. Probabilities of Inclusion and Population Sizes in a Cell

Inclusion Probability | True Population Size

| | | Original Enumeration | | |
|---|---|---|---|---|
| | | In | Out | Total |
| P-sample Pop. | In | $P_{i11}\|N_{11}$ | $P_{i12}\|N_{12}$ | $P_{i1+}\|N_{1+}$ |
| | Out | $P_{i21}\|N_{21}$ | $P_{i22}\|(N_{22})$ | $P_{i1+}\|(N_{2+})$ |
| | Total | $P_{i+1}\|N_{+1}$ | $P_{i+2}\|(N_{+2})$ | $P_{i++}\|(N_{++})$ |

*Causal Independence.* The event of being included in the census is independent of the event of being included in the P-sample population. That is, the cross-product ratio $\theta_i = P_{i11}P_{i22}/P_{i12}P_{i21}$ is equal to 1 for each person i = 1, for i = 1, . . ., N.

*Autonomous Independence.* The two lists, census and the P-sample populations, are formed in N mutually independent trials.

*Heterogeneous Independence.* The covariance between $P_{i1+}$ and $P_{i+1}$ is 0, with covariance defined as $N^{-1}\Sigma(P_{i1+} - \bar{p}_{1+})(P_{i+1} - \bar{p}_{+1})$, with $\bar{p}_{1+} = N^{-1}\Sigma P_{i1+}$ and $\bar{p}_{+1} = N^{-1}\Sigma P_{i+1}$. A sufficient condition for heterogeneous independence is *homogeneity*, i.e., that $P_{i1+} = P_{1+}$ or $P_{i+1} = P_{+1}$ for i = 1, ..., N.

Sekar and Deming (1949) suggested forming poststrata, groupings of the population by demographics (e.g., age, race, sex) and geography, so that the homogeneity assumption holds within each poststratum.

The Census Bureau poststratifies the persons in the PES according to demographic and geographic variables (Alberti et al. 1988). An estimate of the population size in each poststratum is calculated and then the estimates are summed to give an estimate for the total population.

Poststratification reduces but does not eliminate the effect of failure of the heterogenous independence assumption. Having independent field operations avoids failure of the causality assumption. Failure of autonomy tends to increase variance but has only a negligible effect on the bias; see Cowan and Malec (1986) and Wolter (1986).

Let $\theta = N_{11}N_{22}/(N_{12}N_{21})$ be the overall cross-product ratio and let $\tau = \theta - 1$. We will refer to $\tau$

as the *correlation bias factor* that reflects failure of the independence assumptions. If the independence assumptions hold then $1 = \theta = \theta_i$ for i = 1,...,N. The correlation bias may be expressed as follows:

$$\tilde{N} - N = -\tau N_{12} N_{21} / N_{11} + O_p(\tilde{N}^{-1/2})$$

with the $O_p$ term the random component of correlation bias that is negligible in this application (Wolter 1986).

Our goal is to estimate the correlation bias factor. A conditional logistic estimation procedure (Alho, 1990) is used to estimate the probabilities of inclusion in the census and the P sample, $P_{i1+}$ and $P_{i+1}$. This method allows analysis of dual system data using individual level covariate information. No grouping of the data is required as in the method of Sekar and Deming, and no completely independent source of information, such as demographic analysis, is needed. Having estimated the inclusion probabilities, we can estimate the correlation and obtain an estimate of $\tau$, the *correlation bias factor.*

### 2.1 Calculation of Inclusion Probabilities

For ease of notation, let $P_{1i} = P_{i+1}$ be the probability of the i-th individual being included in the census, and let $P_{2i} = P_{i1+}$ be probability of the i-th individual being included in the P-Sample population.

Conditional logistic regression requires assuming that we have vectors $X_{1i}$ and $X_{2i}$ of "explanatory" variables giving the characteristics (e.g. age, sex, tenure) of individuals correlated with inclusion. The inclusion probabilities, $P_{1i}$ and $P_{2i}$, can be modeled as follows:

$$\log\left(\frac{P_{1i}}{(1 - P_{1i})}\right) = X_{1i}^T a_1$$

$$\log\left(\frac{P_{2i}}{(1 - P_{2i})}\right) = X_{2i}^T a_2$$

where $a_1$ and $a_2$ are vectors of parameters, which are estimated. Newton's method is used to estimate the parameters, $a_1$ and $a_2$, iteratively.

### 2.2 Estimation

A conditional logistic regression model produces an E-sample inclusion probability and a P-sample inclusion probability for each person. These probabilities may be used to calculate a correlation bias factor using Spencer's estimator.

### 2.2.2 Spencer's Method of Calculating $\tau$ Using Only Resolved Cases

Spencer has developed an estimator of $\tau$ using the covariance of $P_{1i}$ and $P_{2i}$ (1991). When only resolved cases are used, the estimator has the following form.

$$\tau = \frac{Cov(P_{1i}, P_{2i})}{(\bar{P}_1 - \bar{P}_x)(\bar{P}_2 - \bar{P}_x)}$$

*where*

$$Cov(P_{1i}, P_{2i}) = \bar{P}_x - \bar{P}_1\bar{P}_2,$$

$$\bar{P}_x = \frac{\sum_{i=1}^{n} \frac{W_i P_{1i} P_{2i}}{\phi_i}}{\sum_{i=1}^{n} \frac{W_i}{\phi_i}}, \quad \bar{P}_1 = \frac{\sum_{i=1}^{n} \frac{W_i P_{1i}}{\phi_i}}{\sum_{i=1}^{n} \frac{W_i}{\phi_i}},$$

$$\bar{P}_2 = \frac{\sum_{i=1}^{n} \frac{W_i P_{2i}}{\phi_i}}{\sum_{i=1}^{n} \frac{W_i}{\phi_i}}$$

$W_i$ = stratum weight for the i-th individual.
$\phi_i = P_{1i} + P_{2i} - P_{1i} * P_{2i}$ for i = 1,..,n.
n = number of resolved cases.

### 2.2.3 Spencer's Method of Calculating $\tau$ Using Resolved and Unresolved Cases

The difference between the estimator using unresolved E-sample cases and the estimator using unresolved P-sample cases is in the calculation of $\bar{P}_1$, $\bar{P}_2$, and $\bar{P}_x$. Using unresolved E-sample cases, they are calculated as follows:

$$\bar{P}_x = \frac{\sum_{i=1}^{n} \frac{W_i P_{1i} P_{2i}}{\phi_i} + \sum_{i=1}^{U_E} W_i \gamma_{1i} P_{2i}}{\tilde{N}_E},$$

$$\bar{P}_1 = \frac{\sum_{i=1}^{n} \frac{W_i P_{1i}}{\phi_i} + \sum_{i=1}^{U_E} W_i \gamma_{1i}}{\tilde{N}_E},$$

$$\bar{P}_2 = \frac{\sum_{i=1}^{n} \frac{W_i P_{2i}}{\phi_i} + \sum_{i=1}^{U_E} \frac{W_i \gamma_{1i} P_{2i}}{P_{1i}}}{\tilde{N}_E},$$

*where* $\tilde{N}_E = \sum_{i=1}^{n} \frac{W_i}{\phi_i} + \sum_{i=1}^{U_E} \frac{W_i \gamma_{1i}}{P_{1i}}$

$U_E$ = number of unresolved E-sample cases.
$\gamma_{i1}$ = the probability of being correctly enumerated, which is estimated for each unresolved E-sample case.

Using P-sample unresolved cases, $\bar{P}_1$, $\bar{P}_2$, and $\bar{P}_z$ are calculated as follows:

$$\bar{P}_z = \frac{\sum_{i=1}^{n} \frac{W_i P_{1i} P_{2i}}{\phi_i} + \sum_{i=1}^{U_P} W_i P_{1i}}{\tilde{N}_P},$$

$$\bar{P}_1 = \frac{\sum_{i=1}^{n} \frac{W_i P_{1i}}{\phi_i} + \sum_{i=1}^{U_P} \frac{W_i P_{1i}}{P_{2i}}}{\tilde{N}_P},$$

$$\bar{P}_2 = \frac{\sum_{i=1}^{n} \frac{W_i P_{2i}}{\phi} + \sum_{i=1}^{U_P} W_i}{\tilde{N}_P},$$

$$\text{where } \tilde{N}_P = \sum_{i=1}^{n} \frac{W_i}{\phi_i} + \sum_{i=1}^{U_P} \frac{W_i}{P_{2i}}$$

$U_P$ = number of unresolved P-sample cases.

## 3. IMPLEMENTATION

### 3.1 Adjustment of P-sample Capture Probabilities for Migration

Applying the conditional logistic modeling in the PES setting is complicated because of migration. Some people move in to PES sample blocks, and others move out. Only those individuals who were present in the PES sample area at census time should be included in the estimation of correlation bias. Therefore the inclusion probability for each person in the P-sample is multiplied by an estimate of the probability that the person was in the PES sample area at census time. The interpretation of the E-sample probabilities is in no way confounded by migration.

### 3.2 Accounting for Data Errors

The application of the conditional logistic model also is complicated by errors during the PES data collection. These data errors create issues of which cases to include in model fitting and which to include in estimation. One issue of this is the cases which remain unresolved at the end of the matching operation. These cases are excluded from the model fitting but are included in the two of three estimators in Section 2.2. The underlying assumption is that the unresolved individuals have the same capture characteristics as those individuals that were resolved, given equal covariates.

Another issue is caused by geocoding error. Some P-sample people match to census people in the search area of the PES block. The search area is a ring of blocks surrounding the PES block. Such a match is allowed to compensate for minor geocoding errors in the census or PES. Such cases are included in the model fitting as a match. This formulation has the effect of adding the census people who match P-sample people to the E-sample.

## 4. DATA ANALYSIS

### 4.1 Model Building

Models were built for the following four minority evaluation poststrata which are aggregates of PES poststrata: Northeast central cities, South central cities, West central cities, and Midwest central cities. Models were built for E- and P-sample separately in each evaluation poststratum. Table 4.1 displays the regression coefficients for the E-sample and P-sample models.

Sex, race, tenure, place type 0, relationship, marital status, and census division (CD) are indicator variables, i.e. they receive a value of 0 or 1. The values for the indicator variables are assigned as follows: sex is '1' if female, race is '1' if black, tenure is '1' if renter, place type 0 is '1' if living in central cities in primary metropolitan statistical areas (the most densely populated areas), relationship is '1' if not related to the person who completed the questionnaire, marital status is '1' if married, and census division is '1' if living in the particular CD.

The 'rate' variables, such as renter rate, are block level variables. The variables age, household size, and the block level variables are standardized to give them a level of magnitude equivalent to the indicator variables.

Variables are selected based on significance tests, multicollinearity, and assumed sociological importance. There is a group of "core variables" which are common to each of the eight models. Not all the "core variables" have an impact on each of the four models, however, each of these variables does have an impact for some of the evaluation poststrata and was included for each poststrata in order to make comparisons and to simplify the model building process. The remaining variables are place type, place type x race, census division, and Hispanic origin. Place type and census division are geographic variables which vary according to evaluation poststrata. The Hispanic indicator and Hispanic rate, a block level variable, are used in the

Table 4.1 Regression Coefficients (and Standard Errors) for the Minority, Central City Evaluation Poststrata

| | E Sample | | | | P Sample | | | |
|---|---|---|---|---|---|---|---|---|
| | Northeast | South | Midwest | West | Northeast | South | Midwest | West |
| Intercept | 1.863 (0.119) | 2.757 (0.116) | 2.422 (0.109) | 2.853 (0.115) | 1.190 (0.109) | 2.387 (0.096) | 2.291 (0.100) | 2.208 (0.126) |
| Age | -0.054 (0.064) | 0.326 (0.112) | -0.112 (0.089) | 0.143 (0.069) | -0.032 (0.058) | 0.353 (0.108) | 0.014 (0.010) | 0.105 (0.056) |
| (Age)2 | 0.088 (0.029) | 0.764 (0.110) | 0.132 (0.036) | 0.059 (0.040) | 0.144 (0.028) | 0.618 (0.081) | 0.110 (0.032) | 0.141 (0.035) |
| (Age)3 | -0.021 (0.018) | -0.357 (0.125) | -0.012 (0.024) | -0.022 (0.023) | -0.058 (0.015) | -0.889 (0.149) | -0.072 (0.017) | -0.055 (0.020) |
| Sex | 0.199 (0.043) | 0.281 (0.041) | 0.256 (0.049) | 0.111 (0.058) | 0.212 (0.040) | 0.187 (0.034) | 0.262 (0.044) | 0.140 (0.049) |
| Race (Black) | -0.166 (0.129) | -0.315 (0.144) | -0.187 (0.071) | -0.589 (0.155) | 0.351 (0.118) | -0.464 (0.119) | -0.288 (0.068) | -0.143 (0.137) |
| Hispanic | | | | 0.077 (0.128) | | | | 0.006 (0.101) |
| Tenure | -0.657 (0.108) | -0.603 (0.086) | -0.715 (0.063) | -0.755 (0.089) | -0.446 (0.101) | -0.773 (0.074) | -0.578 (0.056) | -0.357 (0.071) |
| HH Size | | -0.163 (0.032) | -0.155 (0.037) | -0.338 (0.042) | | -0.146 (0.030) | -0.165 (0.031) | -0.308 (0.039) |
| Renter Rate | 0.103 (0.037) | -0.041 (0.034) | 0.136 (0.042) | -0.080 (0.052) | -0.299 (0.033) | -0.424 (0.028) | 0.128 (0.038) | -0.380 (0.045) |
| Black Rate | -0.012 (0.037) | 0.026 (0.057) | | -0.337 (0.066) | -0.283 (0.036) | 0.041 (0.046) | | 0.000 (0.054) |
| Hispanic Rate | | | | -0.169 (0.070) | | | | -0.008 (0.054) |
| Multiunit Rate | -0.124 (0.034) | 0.023 (0.029) | -0.260 (0.039) | 0.067 (0.045) | 0.155 (0.030) | 0.009 (0.022) | -0.003 (0.035) | 0.108 (0.037) |
| Vacancy Rate | -0.047 (0.021) | -0.196 (0.019) | -0.109 (0.024) | -0.018 (0.031) | 0.017 (0.021) | -0.108 (0.016) | -0.156 (0.022) | -0.066 (0.025) |
| Place Type 0 | 0.045 (0.087) | -0.340 (0.087) | -0.275 (0.071) | | 0.017 (0.081) | -0.532 (0.069) | -0.401 (0.066) | |
| Relationship | -1.020 (0.090) | -0.962 (0.085) | -0.797 (0.099) | -1.141 (0.096) | -0.878 (0.081) | -1.054 (0.071) | -0.884 (0.089) | -1.180 (0.082) |
| Marital Status | 0.180 (0.610) | | | | 0.326 (0.057) | | | |
| Tenure*Race | -0.035 (0.117) | 0.167 (0.094) | | 0.752 (0.127) | 0.089 (0.109) | 0.487 (0.081) | | 0.037 (0.120) |
| Tenure*HH Size | | 0.194 (0.038) | 0.240 (0.048) | 0.163 (0.053) | | 0.149 (0.035) | -0.010 (0.041) | 0.164 (0.048) |
| Age*Race | 0.139 (0.054) | 0.202 (0.115) | 0.270 (0.078) | 0.157 (0.065) | 0.171 (0.049) | 0.263 (0.088) | 0.128 (0.072) | 0.163 (0.060) |
| Race*Pl. Type 0 | 0.065 (0.100) | 0.021 (0.115) | | | -0.285 (0.092) | 0.256 (0.093) | | |
| Age*HH Size | 0.043 (0.024) | 0.313 (0.055) | 0.075 (0.030) | 0.044 (0.098) | 0.053 (0.023) | 0.004 (0.041) | 0.006 (0.024) | -0.355 (0.084) |
| Sex*Age | 0.094 (0.045) | 0.440 (0.096) | 0.119 (0.054) | -0.027 (0.062) | 0.020 (0.042) | 0.204 (0.076) | 0.132 (0.045) | -0.025 (0.054) |
| South Atlantic Census Division | | -0.366 (0.065) | | | | -0.117 (0.052) | | |
| East South Central Census Division | | -0.287 (0.072) | | | | -0.001 (0.059) | | |
| East North Central Census Division | | | 0.114 (0.097) | | | | 0.099 (0.089) | |
| Pacific Census Division | | | | -0.344 (0.031) | | | | -0.031 (0.027) |

West poststratum because this poststratum includes a substantial number of Asians.

### 4.2 Analysis of Odds Ratios

If the sex variable is coded as 0 for male and 1 for female, then the odds of being captured, or enumerated, for females is defined as P(1)/[1-P(1)], where P(1) is the capture probability for females. Similarly, the odds of being captured for males is defined as P(0)/[1-P(0)], where P(0) is the capture probability of male. The odds ratio, denoted by $\Psi$, is defined as the ratio of odds for females to the ratio of odds for males. Thus

$$\Psi = \frac{P(1)/[1-P(1)]}{P(0)/[1-P(0)]}.$$

The odds ratios for both the E and P sample for the Northeast and Midwest minority/central city evaluation poststrata are given in tables 4.2.1 and 4.2.2. The odds ratios for the South and West are similar to those for the Midwest.

Among the five effects considered, three effects, renter among black females in place type 0, non-relative, and black renter among females in place type 0, have an odds ratio consistently less than one for both the E and P sample. This implies that each of the three groups has a lower inclusion probability than its respective counterpart. The odds ratio for female among black renters in place type 0 is greater than one for both the E and P samples in both poststrata. Except for the P-sample for the Northeast, the odds ratio for black among female renters in place type 0 is less than one. Non-relatives and black renters have the consistently lowest odds ratios.

### 4.3 Inclusion Probabilities

Table 4.3 shows the average and range of inclusion probabilities for the four minority, central city poststrata. The South minority/central city poststratum had the highest average inclusion probability for both the E-sample and P-sample, .920 and .813 respectively, and the Northeast had the lowest for both the E and P sample, at .816 and .751 respectively. For each of the four poststrata, the average E-sample inclusion probability is higher than the average P-sample inclusion probability.

Table 4.2.1 Estimated Odds Ratios for Northeast, Minority, Central City Evaluation Poststratum

| Effect | Among | Odds Ratio E | Odds Ratio P |
|---|---|---|---|
| Female | Black renter place type 0 | 1.220 | 1.236 |
| Black | Female renter place type 0 | .847 | 1.421 |
| Renter | Black female place type 0 | .519 | .640 |
| Non-relative | All | .503 | .416 |
| Black Renter | Female place type 0 | .398 | .748 |

Table 4.2.2 Estimated Odds Ratios for Midwest, Minority, Central City Evaluation Poststratum

| Effect | Among | Odds Ratio E | Odds Ratio P |
|---|---|---|---|
| Female | Black renter place type 0 | 1.292 | 1.298 |
| Black | Female renter place type 0 | .829 | .750 |
| Renter | Black female place type 0 | .489 | .561 |
| Non-relative | All | .451 | .413 |
| Black Renter | Female place type 0 | .406 | .421 |

Table 4.3 Average and Range of Inclusion Probabilities for Minority, Central City Evaluation Poststrata

| | E-sample | | | P-sample | | |
|---|---|---|---|---|---|---|
| | Ave | Max | Min | Ave. | Max | Min |
| NE | .816 | .944 | .496 | .751 | .948 | .350 |
| SO | .920 | .998 | .507 | .813 | .979 | .429 |
| MW | .864 | .991 | .373 | .794 | .963 | .368 |
| WE | .879 | .981 | .308 | .785 | .974 | .249 |

### 4.4 Correlation Bias

The correlation bias factors are estimated using the capture probabilities for E- and P-sample based on Spencer's method for these four evaluation poststrata. Table 4.4 shows the correlation bias factors, their conditional standard errors, the undercount rates calculated using the usual DSEs, and undercount rates calculated using DSEs adjusted for correlation bias. The inclusion of unresolved cases has little impact on the estimates probably because the number of unresolved cases is small relative to the number of resolved cases. Thus, estimates which include unresolved cases are given for the Midwest only. MW(ue) includes unresolved E-sample cases, and MW(up) includes unresolved P-sample cases. T-test values are 2.022 for MW and MW(ue) and 2.454 for MW and MW(up). The correlation bias factor estimate for the South using unresolved P-sample cases was the only other estimate using unresolved cases which differed significantly, at the .05 level, from the corresponding estimate using only resolved cases.

Table 4.4 Correlation Bias Factors and the Effect of Correlation Bias on Undercount Rates for Minority, Central City Evaluation Poststrata

| | Corr. Bias Factor | Std.* Error | Undct Rate | Adj. Undct Rate |
|---|---|---|---|---|
| NE | 0.14 | 0.02 | 6.83% | 7.31% |
| SO | 0.34 | 0.04 | 5.68% | 6.13% |
| WE | 0.42 | 0.03 | 6.14% | 7.27% |
| MW | 0.25 | 0.03 | 3.97% | 4.12% |
| MW(ue) | 0.26 | 0.03 | 3.97% | 4.12% |
| MW(up) | 0.27 | 0.03 | 3.97% | 4.12% |

* Standard errors are conditional on the models.

We conclude by noting that the undercount estimates based on conditional logistic regression in table 4.4 are all higher than the ones based on the usual DSEs. This suggests that there has been some residual heterogeneity in the inclusion probabilities that the logistic regression has revealed. The E and P samples appear to have had a higher positive correlation than the one expected based on the usual stratified analysis.

## 5. REFERENCES


Alberti,N., Diffendal, G., Hogan, H., Isaki, C., Monsour, N., Passel, J., Robinson, G., Schenker, N., Thompson, J., Wolter, K., and Woltman, H. (1988) "Preliminary Poststratification Schemes for the 1990 Census Coverage by Measurement Programs," unpublished manuscript August 3, 1988. Bureau of the Census, Washington, D.C.

Alho, J. M. (1990) "Logistic Regression in Capture-Recapture Models," Biometrics, 46, 623-635.

Cowan C.D. and Male, D.J. (1986) "Capture-Recapture Models When Both Sources Have Clustered Observation," Journal of the American Statistical Association, 81, 347-353.

Ericksen, E.P. and Kadane, J.B. (1985) "Estimating the Population in a Census Year: 1980 and Beyond," Journal of the American Statistical Association, 80, 98-108, 129-131.

Sekar, C. C. and Deming, W. E. (1949) "On a Method of Estimating Birth and Death Rates and the Extent of Registration," Journal of the American Statistical Association, 44, 101-115.

Spencer, B.D. (1991) Derivation of $\tau$, March 4, 1991, unpublished manuscript.

Wolter, K. M. (1986) "Some Coverage Error Models for Census Data," Journal of the American Statistical Association, 81, 338-346.

Wolter, K. M. (1986) "A Combined Coverage Error Model for Individuals and Housing Units," SRD Research Report Number Census/SRD/RR-86/27, Statistical Research Division Report Series, U.S. Bureau of the Census, Washington, D.C.



Acknowledgements

The authors wish to thank Ven Sathyamoorthy, David Carr, and Robert Fay for their contributions.

Juha Alho wishes to thank the Census Bureau for supporting this research through a joint statistical agreement with the NORC.


* This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.