Statistical Journal of the IAOS 31 (2015) 661–676
DOI 10.3233/SJI-150884
IOS Press

On the interpretation of multi-year estimates of the American Community Survey as period estimates

Chaitra H. Nagaraja[a,*] and Tucker McElroy[b]
[a]*Gabelli School of Business, Fordham University, New York, NY, USA*
[b]*Center for Statistical Research and Methodology, U.S. Census Bureau, Washington, DC, USA*

Abstract. The rolling sample methodology of the American Community Survey leads to Multi-Year Estimates that measure aggregate activity over one, three, or five years. This paper introduces a novel, non-model-based method for quantifying the impact of viewing multi-year estimates as functions of single-year estimates belonging to the same time span. The method is based on examining the changes to confidence interval coverage. The interpretation of a multi-year estimate as the simple average of single-year estimates is a viewpoint that underpins the published estimates of sampling variability. Therefore, it is vital to ascertain the extent to which this viewpoint is valid. We apply our new methodology to data from the U.S. Census Bureau's Multi-Year Estimates Study and demonstrate that viewing a multi-year estimate as the simple average of single-year estimates typically results in substantial distortions to coverage; therefore, multi-year estimates should not be interpreted as averages, but merely as period estimates.

Keywords: Rolling sample, confidence interval, time series

1. Introduction

The American Community Survey (ACS) of the U.S. Census Bureau was designed as a more timely analog of the Census Long Form, with data on social, demographic, economic, and housing variables published annually for a range of geographical regions. In order to counteract perceived high sampling variability in geographical regions of low population, it was deemed prudent to aggregate information temporally using a rolling sample. This means that data collected over long stretches of time would be formed into one "period estimate." These estimate types are becoming more common worldwide. Within the United States, the U.S. Census Bureau uses the terminology Multi-Year Estimates, and such estimates are currently being published for counties, tracts, block groups, and school districts throughout the USA (including territories), over periods of one-, three-, and five-years.

Other countries use the multi-year estimate methodology as well. For example, in France, the Institute of Statistics and Economic Studies has implemented a rolling "census" where every municipality is sampled at least once in a 5-year period. Only those municipalities under 10,000 people have a full census, whereas larger municipalities sample residents. Also, 5-year and 1-year estimates are annually published at the national and regional levels [12]. This type of publishing schedule is similar to that of the American Community Survey and consequently, the methodology we propose in this paper can be applied to such data.

A common misconception, and one which the U.S. Census Bureau has attempted to rectify, is that these multi-year estimates are computed by averaging the one-year estimates. (Citro and Kalton [10] implicitly use this type of averaging in their description of multi-

*Corresponding author: Chaitra H. Nagaraja, Gabelli School of Business, Fordham University; 5 Columbus Circle, Rm 1116, New York, NY 10019, USA. Tel.: +1 212 636 6678; E-mail: cnagaraja@fordham.edu.

year estimates.) The distinction between a multi-year estimate and an average of the corresponding one-year estimates is quite subtle and has to do with the particulars of the weighting methods and the application of controls (more detail provided in the next section). A more flawed misconception is that the period estimate is interpreted as an estimate for the middle or end year.

For example, a 3-year period estimate for the homeowner vacancy rate from 2004–2006 might be conceived of as an estimate for 2005 (mid-year), as an estimate for 2006 (end-year), or as an average rate computed from individual one-year rates from 2004 through 2006. The period estimate should instead be interpreted as an estimate of the population's vacancy rate across the three-year time span, formalized as the ratio of vacancies *over all three years* relative to total home ownership over the same period. Crudely put, the period estimate is a ratio of sums, whereas the average rate estimate is an average of ratios – these are algebraically dissimilar quantities.

Such interpretations may be "justified" in two situations. First, when both 1-year and period estimates are published, it may be tempting to apply the mid- or end-year interpretations to the period estimates because the sampling error is substantially lower than for the corresponding 1-year estimates. Second, when only multi-year estimates exist, they may be interpreted in terms of the alternate estimates (average, mid-, or end-year) instead of as an aggregate estimate covering the entire period for ease of use.

None of these alternative interpretations are mathematically correct, though they may be approximately correct in some statistical sense. In this paper, we propose a method to measure the impact of using such interpretations. Chand and Alexander [9] attempt to compare such estimates using simulation; however, their comparisons require strong assumptions on the underlying data trends. While such studies may be illuminating about a theoretical description of the data, we focus here on the practical impact of such an alternative interpretation upon statistical inference.

The key contribution of this paper is the development of a simple, non-model-based method that is useful for quantifying the extent of the error in the above-mentioned faulty average, mid-year, and end-year alternate interpretations of the multi-year estimates. When examining multi-year estimates with alternate estimates, we cannot simply compare the point estimates because the corresponding standard errors are not accounted for. Instead, we examine the extent of change in the confidence interval when substituting an alternate estimate for a multi-year estimate. If this change is less than a pre-specified threshold τ, the alternate interpretation is deemed "inferentially τ-close" and is considered an acceptable proxy for the period interpretation. Using confidence intervals allows us to incorporate both differences in the point estimate (period versus alternate) and differences in the corresponding standard errors. Both are necessary for a complete comparison.

In this analysis, we examine 95% confidence intervals and let the threshold $\tau = 0.01$. (We also investigate the impacts of choosing different values for τ.) Consequently, an alternate estimate is considered "inferentially τ-close" if the confidence interval coverage changes by less than a full percentage point. Our method can be readily applied in individual cases to directly determine the suitability of any interpretation constructed as a linear combination of 1-year estimates. We also briefly examine the reverse question: whether a multi-year estimate can be a suitable substitute for an alternate estimate.

We apply this method to data from the Multi-Year Estimates Study, a trial run of the American Community Survey implemented on a limited set of counties in the USA from 2000 through 2005. Estimates such as totals, means, and proportions were computed on a wide range of characteristics (e.g. demographic, social). We explicitly show that alternative estimates, while easier to interpret than multi-year estimates, are generally unsuitable even when given a generous value of τ. In addition, one can easily apply the methodology developed in this paper to American Community Survey data. Furthermore, constructing multi-year estimates using rolling samples is used in other countries such as in France by the National Institute of Statistics and Economic Studies [12] and may eventually be used in Israel by the Central Bureau of Statistics [24]. Therefore, the methodology proposed here could be applied to other surveys as well.

Our paper first introduces notation in Section 2 along with a brief discussion of the Multi-Year Estimates Study and the mechanics of multi-year estimate construction. Section 3 outlines our new methodology. This method of evaluation is probabilistic rather than statistical in nature and we justify it on the basis of the usability and the interpretability of multi-year estimates. Our empirical results are found in Section 4, where we apply our methodology to the Multi-Year Estimates Study database. The vast amount of information is summarized graphically and specific relationships of interest are highlighted through a discussion of selected variables. We conclude in Section 5.

2. Background

The American Community Survey is used to collect data on U.S. residents on a variety of subjects, falling into four categories: demographic, social, economic, and housing characteristics. Roughly 3 million households are sampled every year [4, p. 3]. While data is collected annually, the publishing schedule varies by geographic level (state, county, census tract, and so forth) depending on the population of the area. For all areas, 5-year estimates are published. If the population is above 20,000, then 3-year estimates are published as well. Finally, if the population is above 65,000, estimates at the 1-, 3- and 5-year intervals are published. For instance, state-level estimates are published separately for 1-, 3-, and 5-years, whereas a census tract may only have five-year estimates published if the population is below 20,000. Small areas have relatively fewer sampled units; consequently, pooling data across multiple years results in estimates with lower standard errors. See U.S. Census Bureau [21] and Torrieri [19] for more discussion.

Data is collected using a rolling sample throughout the year, a methodology first proposed by Kish [3,15], and was selected to provide more frequent estimates for small areas and capture long term trends [1, p. 644]. However, as it was recognized that this method would generate a lag in underlying trends and temporal dynamics [7,8], the resulting estimates were to be interpreted as period estimates. Thus, a rolling sample estimate over a span of years was not to be viewed as an estimate of a population quantity at the end year, middle year, or average of the years, but simply as an estimate of aggregate activity over the entire temporal span. This approach is mathematically viable and entirely analogous to the concept of flow estimates used in economic time series. For further discussion of these issues, see the following: Kish [14], Alexander [2], Citro and Kalton [10], Beaghen and Weidman [6], McElroy [16], Beaghen et al. [4].

For the actual construction of multi-year estimates, see the discussion in Fay [13], Starsinic and Tersine [17], Tersine and Asiala [18], McElroy [16], and Beaghen et al. [4]. The procedure is complicated, involving sampling weights, nonresponse adjustment, and population controls; the result is a highly nonlinear operation on the samples, further interfering with the linear picture adopted in Citro and Kalton [10]. In particular, so-called base weights – defined as the inverse of sampling probabilities – are utilized to weight pooled results from both housing units and group quarters, the pooling an assembly of responses over the entire time period (either one year, three years, or five years) being considered. This is followed by nonresponse adjustment, taking into account three modes of interviews – mail, phone, and on-site. All of these facets serve to generate a complex, and somewhat bewildering, definition of multi-year estimates, which defies simplistic portraits.

We will illustrate our methodology on the Multi-Year Estimates Study data. This study was a trial for the American Community Survey from 1999 through 2005 and restricted to 34 counties. However, multi-year estimates were actually published starting in 2000. The data from the Multi-Year Estimates Study and the current American Community Survey are extremely similar; there are only a few minor differences in methodology (e.g., group quarters were not included in the Multi-Year Estimates Study). Furthermore, like the American Community Survey, the Multi-Year Estimates Study controls population totals using external information. The advantage in using the Multi-Year Estimates Study is that it has been studied by researchers for several years now, so our work can be compared to other existing literature on the same database.

In our analysis, we restrict our sample to 19 counties, which are administrative subdivisions of states, where 1-year estimates for the year 2000 are available, and 1-, 3-, and 5-year estimates exist for each county. These counties are listed in Table 1 with the Federal Information Processing Standard (FIPS) county codes (the FIPS codes are used to designate geographic areas within the United States and associated localities).

Ideally the time series, at least for these 19 counties, could be concatenated with current American Community Survey estimates to form a longer sample, but unfortunately this is impossible due to the publication schedule. That is, the current American Community Survey estimates (for all of the USA) began publication with 1-year estimates in 2006. This leaves a gap in the time series data, which in McElroy [16] was resolved through the crude device of forecasting. In this paper we focus instead on the Multi-Year Estimates Study, with the assumption that results for the trial period should also hold valid for the nation at large (counties for the Multi-Year Estimates Study were selected by experts with the belief that they constituted a diverse and representative picture of the entire USA).

Each of the 19 counties have populations above 65,000 for all years of data. Therefore, 1-, 3-, and 5-year estimates are available for each county. Further-

Table 1
County, state, and FIPS county codes

County	State (postal code)	FIPS county code
Black Hawk County	Iowa (IA)	19013
Bronx County	NewYork (NY)	36005
Broward County	Florida (FL)	12011
Calvert County	Maryland (MD)	24009
Douglas County	Nebraska (NE)	31055
Flathead County	Montana (MT)	30029
Franklin County	Ohio (OH)	39049
Hampden County	Massachusetts (MA)	25013
Jefferson County	Arkansas (AR)	05069
Lake County	Illinois (IL)	17097
Madison County	Mississippi (MS)	28089
Multnomah County	Oregon (OR)	41051
Pima County	Arizona (AZ)	04019
Rockland County	NewYork (NY)	36087
San Francisco	California (CA)	06075
Schuylkill County	Pennsylvania (PA)	42107
Sevier County	Tennessee (TN)	47155
Tulare County	California (CA)	06107
Yakima County	Washington (WA)	53077

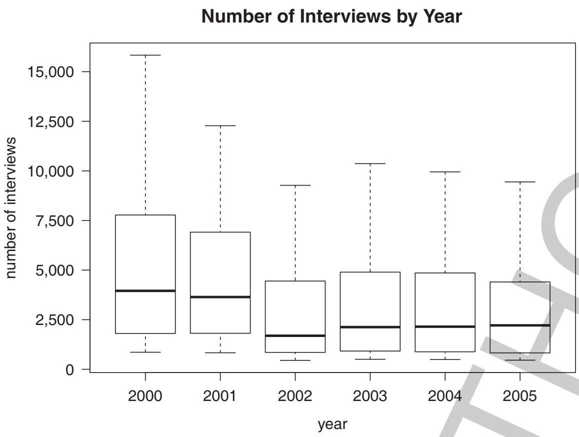

Fig. 1. Number of interviews across 19 counties for each year in the Multi-Year Estimates Study.

more, the total number of interviews conducted annually for each county ranges from 449 to 15,830 and increases as the population of the county increases. Box plots showing the distribution of sample sizes are provided in Fig. 1 and are calculated from the Multi-Year Estimates Study Quality Measures information [23]. These sample sizes are high enough to implement our methodology. There are hundreds of variables available, which the U.S. Census Bureau classifies into four tables along with their corresponding table numbers: Demographic (DP01), Social (DP02), Economic (DP03), and Housing (DP04). We retain this classification scheme in our analyses. Some of these figures are numeric (e.g. totals, rates, averages), whereas others are percents; in each case sampling variances are published.

As we are interested in making comparisons between multi-year estimates of different period lengths, it makes sense to consider only those pairings drawn from the same underlying spans of time. Let $\hat{x}_t^{(k)}$ denote a period estimate, where k is the period length (1-, 3- or 5-years) and t denotes the final year of data used to compute the estimate. For example, if $k = 3$ and $t = 2002$, $\hat{x}_{2002}^{(3)}$ specifies the 3-year estimate combining data from 2000, 2001, and 2002. Using this notation and given the published multi-year estimates available, it follows that we have four possible comparisons of 1-year to 3-year multi-year estimates: (a) $\left\{\hat{x}_{2000}^{(1)}, \hat{x}_{2001}^{(1)}, \hat{x}_{2002}^{(1)}\right\}$ with $\hat{x}_{2002}^{(3)}$, (b) $\left\{\hat{x}_{2001}^{(1)}, \hat{x}_{2002}^{(1)}, \hat{x}_{2003}^{(1)}\right\}$ with $\hat{x}_{2003}^{(3)}$ (c) $\left\{\hat{x}_{2002}^{(1)}, \hat{x}_{2003}^{(1)}, \hat{x}_{2004}^{(1)}\right\}$ with $\hat{x}_{2004}^{(3)}$, and (d) $\left\{\hat{x}_{2003}^{(1)}, \hat{x}_{2004}^{(1)}, \hat{x}_{2005}^{(1)}\right\}$ with $\hat{x}_{2005}^{(3)}$. There are also two possible comparisons of 1-year to 5-year multi-year estimates: (a) $\left\{\hat{x}_{2000}^{(1)}, \hat{x}_{2001}^{(1)}, \hat{x}_{2002}^{(1)}, \hat{x}_{2003}^{(1)}, \hat{x}_{2004}^{(1)}\right\}$ with $\hat{x}_{2004}^{(5)}$ and (b) $\left\{\hat{x}_{2001}^{(1)}, \hat{x}_{2002}^{(1)}, \hat{x}_{2003}^{(1)}, \hat{x}_{2004}^{(1)}, \hat{x}_{2005}^{(1)}\right\}$ with $\hat{x}_{2005}^{(5)}$.

An important note is needed here: the literature on this topic typically assumes that all statistical variation is due to sampling mechanisms. This approach ignores variation in the underlying population quantities, which would typically be modeled with a stochastic process. The published standard errors of the American Community Survey focus entirely on sampling variability, which amounts to viewing population quantities as deterministic. We maintain this viewpoint throughout the paper, treating underlying population estimands as deterministic for the purposes of calculating variances. Details on estimating sampling error variances can be found in U.S. Census Bureau [21,22].

3. Methodology

Consider a population quantity x that represents a variable of interest measured over a time span (e.g., several years), computed exactly like a multi-year estimate, but instead based upon the full population in each year. The population quantity x will be called a multi-year value, in contrast to the multi-year estimate $\hat{x}$, which is its estimate. Any discounting for inflation, or other standardization techniques that are used for multi-year estimates, also occur in the definition of the multi-year value. For example, suppose that $\hat{x}$ represents a 3-year multi-year estimate for median commuting time to work in Bronx County, NY from 2001–2003. Then, x is defined by taking the entire population of Bronx County, NY, year by year, pooling all commuting times for all persons and all years (consequently many people would enter the calculation three times, if they did not move to a different home), and computing the median.

An alternative to a multi-year estimate will be denoted by $\tilde{x}$. Our chief interest will focus on the estimate generated by averaging 1-year estimates, but we also consider the mid-year or end-year estimates. All three will be denoted as alternate estimates.

The sampling error ε is defined as the difference between the multi-year estimate and the target multi-year value: $\hat{x} = x + \varepsilon$. These sampling errors tend to be heteroscedastic and correlated across region and time due to many factors, including variation in sample sizes over years, weighting patterns, non-response adjustments, and the application of population controls. However, some authors have assumed these sampling errors to be uncorrelated in the case of a 1-year estimate, which is really more of a working assumption than a verifiable reality (cf. Beaghen and Weidman [6]).

The variances of the sampling errors will be denoted by $\hat{\sigma}^2_{\hat{x}}$, and are random variables like the multi-year estimates themselves. For the alternate estimates of interest, we can also generate variance estimates, denoted by $\hat{\sigma}^2_{\tilde{x}}$, using naïve formulas. For the *average* alternate estimates, constructed by averaging the 1-year estimates, we assume that the 1-year estimates are approximately independent of each other as they are constructed from non-overlapping data. We follow Citro and Kalton [10] and Beaghen and Weidman [6, p. 23] here for computing variances for temporal differences of multi-year estimates. Consequently, the variance estimate is simply the sum of the individual variances divided by the square of the multi-year estimate period length. For the mid- and end-year alternate estimates, we use the variance value for the corresponding 1-year estimate.

To summarize the notation, x is the population multi-year quantity, which is estimated by $\hat{x}$, and $\tilde{x}$ is the alternate estimate; similarly $\hat{\sigma}^2_{\hat{x}}$ and $\hat{\sigma}^2_{\tilde{x}}$ are the estimated variances for $\hat{x}$ and $\tilde{x}$ respectively. The relative difference of a multi-year estimate and an alternate estimate is an important quantity in our analysis, and is defined by

$$D = \frac{\tilde{x} - \hat{x}}{\hat{x}}. \tag{1}$$

Now, we define the intervals $I_{\hat{x}} = \hat{x} \pm z_{\alpha/2}\hat{\sigma}_{\hat{x}}$ and $I_{\tilde{x}} = \tilde{x} \pm z_{\alpha/2}\hat{\sigma}_{\tilde{x}}$ where we let $z_{\alpha/2}$ be the upper $\alpha/2$ normal quantile for a symmetric 2-sided confidence interval with coverage $(1-\alpha)$ of the population quantity x. The first confidence interval is appropriate for the multi-year estimate, and by presupposition (a working assumption) we suppose that it is a correct interval for x. That is, $\mathbb{P}[x \in I_{\hat{x}}] = 1 - \alpha$. Use of the normal quantile presupposes the validity of an underlying central limit theorem, each sampling unit being viewed as identically distributed and roughly independent. (This is another common assumption in the sampling literature.) The interval $I_{\hat{x}}$ constructed using normal quantiles is published by the U.S. Census Bureau. Assuming the interval is correct, we are chiefly concerned with the question of whether $I_{\tilde{x}}$ is a valid confidence interval for x.

One might imagine substituting an alternate estimate for the multi-year estimate for purposes of analysis. There are many reasons this might be desirable: end-year and mid-year estimates involve less smoothing than multi-year estimates, and hence will tend to respond more rapidly to turning points or other changes in trend. On the other hand, one might utilize an average estimate in lieu of a multi-year estimate in order to calculate standard errors for temporal differences as was developed in Beaghen et al. [4]. For example, the standard error of the consecutive difference in time of two average estimates can be calculated (assuming that 1-year sampling errors are independent), whereas this is not possible to do for multi-year estimates. Given that a practitioner wishes to make such a substitution, it is important to assess the impact. As confidence intervals for population quantities x are the most fundamental application of the multi-year estimates, we ask: How is inference for x altered when an alternate estimate is swapped for a multi-year estimate?

To quantify this question, we seek to know the coverage of $I_{\tilde{x}}$, the confidence interval for x based on using an alternate estimate in lieu of a multi-year estimate. When $D = 0$, there is no difference between $\hat{x}$ and $\tilde{x}$, so that if $\hat{\sigma}^2_{\hat{x}} = \hat{\sigma}^2_{\tilde{x}}$ as well, then $I_{\hat{x}} = I_{\tilde{x}}$ and there is no loss in making the substitution. In practice, estimates of D are nonzero and the variances can be quite different too, so the confidence intervals come out differently. However, how much is the inference really altered? We aim to provide a simple model-free analysis of the question, which has the advantage of being quickly applicable and replicable to any data aggregation method, not just the American Community Survey.

We proceed as follows. Let

$$I(s,r) = \left[\hat{x} + (s - \sqrt{r} z_{\alpha/2})\hat{\sigma}_{\hat{x}},\ \ \hat{x} + (s + \sqrt{r} z_{\alpha/2})\hat{\sigma}_{\hat{x}}\right]. \quad (2)$$

Here, s is a real number and $r \geqslant 0$. As these numbers vary, they distort the usual confidence interval in various ways. Note that $I(0,1) = I_{\hat{x}}$ and $I\left(D\hat{x}/\hat{\sigma}_{\hat{x}},\ \hat{\sigma}^2_{\tilde{x}}/\hat{\sigma}^2_{\hat{x}}\right) = I_{\tilde{x}}$. Hence we can examine the coverage of $I(s,r)$ for various values of s and r. Of principal interest is the case when $|s| \geqslant D\hat{x}/\hat{\sigma}_x$ and/or r differs from $\hat{\sigma}^2_{\tilde{x}}/\hat{\sigma}^2_{\hat{x}}$. This suggests the designations:

$$S = \frac{D\hat{x}}{\hat{\sigma}_{\hat{x}}} = \frac{\tilde{x} - \hat{x}}{\hat{\sigma}_{\hat{x}}} \qquad R = \frac{\hat{\sigma}^2_{\tilde{x}}}{\hat{\sigma}^2_{\hat{x}}}. \quad (3)$$

These quantities can be computed from the data, and will be viewed as realizations of the corresponding random variables s and r. The coverage of $I(s,r)$ is given by

$$\mathbb{P}[x \in I(s,r)] = \Phi(-s + \sqrt{r} z_{\alpha/2}) - \Phi(-s - \sqrt{r} z_{\alpha/2}), \quad (4)$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function. Let $\mathbb{P}[x \in I(s,r)]$ be denoted as $C(s,r)$ for short. Now if $1-\alpha$ is the coverage of the interval $I_{\hat{x}}$, then $C(0,1) = 1-\alpha$. We are interested in determining values of s and r such that $C(s,r)$ is suitably close to $1-\alpha$. Defining a value $\tau > 0$ to be our threshold of coverage error, let $N_\tau = \{(s,r) : |C(s,r) - (1-\alpha)| < \tau\}$ be the neighborhood of such values. If the computed $(S,R) \in N_\tau$, then we know that $I_{\tilde{x}}$ has close to correct coverage, because for any (s,r) such that $|s| \leqslant |S|$ and $|r-1| \leqslant |R-1|$, the coverage $C(s,r)$ of the interval $I(s,r)$ differs from $1-\alpha$ by at most τ. We then say the the corresponding alternate estimate is "inferentially τ-close" to the multi-year estimate.

The neighborhoods N_τ can be computed as unions of level curves of the function $C(s,r)$. That is, for any δ such that $|\delta| \leqslant \tau$, we can compute the level curve $L_{1-\alpha+\delta} = \{(s,r) : C(s,r) = 1-\alpha+\delta\}$, so that $N_\tau = \bigcup_{-\tau \leqslant \delta \leqslant \tau} L_{1-\alpha+\delta}$.

To determine whether an alternate estimate is "inferentially τ-close" to the multi-year estimate, and therefore is an acceptable proxy for it, we implement the following procedure:

1. Set α to specify the confidence level and choose a value τ which delineates the neighborhood N_τ around the confidence interval $I_{\hat{x}}$ that is considered acceptable.
2. Compute $R = \hat{\sigma}^2_{\tilde{x}}/\hat{\sigma}^2_{\hat{x}}$ and $S = (\tilde{x} - \hat{x})/\sqrt{\hat{\sigma}^2_{\hat{x}}}$.
3. Check if $|C(S,R) - (1-\alpha)| < \tau$; if so, the alternate estimate is considered "inferentially τ-close" and is a reasonable proxy for the corresponding multi-year estimate.

Note that R can vary widely; it can be large or even close to zero. Since $C(S,\infty) = 1$, we have $N_\tau = \emptyset$ so long as $\tau < \alpha$. (Consequently, we should certainly set τ much smaller than α in practice.) However, if $R = 0$, then $C(S,0) = 0$ and likewise $N_\tau = \emptyset$ if $\tau < (1-\alpha)$. With $\alpha = 0.05$, we recommend $\tau = 0.01$ so that the effective coverage is altered by less than a full percentage point. The above procedure can be easily applied to all of the variables in the Multi-Year Estimates Study (and for American Community Survey estimates for areas with more than 20,000 people) and we do so in the next section.

4. Results

We now apply the methodology to compare multi-year estimates and alternative estimates for each possible combination of county, time span, and variable. For each case, we use our technique to determine whether the alternative estimate is an acceptable interpretation for the multi-year estimate. Given all of the possible combinations, the results are too innumerable to display individually. To provide sufficient detail, however, we proceed as follows. In Section 4.1, we focus on the results for one specific comparison: the number of persons over 16 in the labor force in Franklin County, OH. This example illustrates the mechanics of the methodology. In Section 4.2, we provide a graphical summary of a selection of the results to identify overall patterns. We then examine a set of variables in Section 4.3 which are of particular interest, chosen to represent a wide range of types of variables found in the data including totals for subpopulations, means, and income. We examine 95% confidence intervals ($\alpha = 0.05$) although any confidence level could be used. In Sections 4.1 through 4.3, we set $\tau = 0.01$ as justified in Section 3. However, we do study the effects on the results if we vary τ in Section 4.4. Finally, in Section 4.5, we examine the reverse question: whether the multi-year estimate is a suitable proxy for the average estimate. We have chosen a representative sample of results to present here; please contact the authors for a complete set.

Table 2
Number of persons over 16 in the labor force in Franklin County, OH where $\alpha = 0.05$, $\tau = 0.01$ (note: decimals have been truncated for the display); values in bold indicate that the alternate estimate is inferentially τ-close to the multi-year estimate

	Time span			
	2000–2002	2001–2003	2002–2004	2003–2005
3-year estimate: $\hat{x}$	577,738	577,958	575,307	575,189
Variance of 3-year estimates	3,168,400	3,791,871	3,385,600	3,751,852
S_{mid}	6.274	−3.687	−0.158	1.354
R_{mid}	2.797	2.992	4.431	2.479
$\lvert C(S_{mid}, R_{mid}) - (1-\alpha)\rvert$	0.9486	0.5667	0.0499	**0.0083**
S_{end}	−3.910	−1.510	1.361	−0.362
R_{end}	3.581	3.956	2.747	2.873
$\lvert C(S_{end}, R_{end}) - (1-\alpha)\rvert$	0.5297	0.0415	0.0204	0.04834
S_{avg}	0.584	0.142	−0.419	0.301
R_{avg}	0.953	1.032	1.170	1.039
$\lvert C(S_{avg}, R_{avg}) - (1-\alpha)\rvert$	0.0482	**0.0013**	**0.00004**	**0.0056**

4.1. Comparing interpretations for Franklin County, OH

In this section we examine the results for an individual case: the number of persons over 16 in the labor force in Franklin County, OH. We show both numerically and graphically how the procedure introduced in Section 3 is implemented and interpreted. The 3-year estimates will be compared with the average, mid-year, and end-year interpretations constructed using the corresponding 1-year estimates. In Table 2, the 3-year estimates (and variances) are provided along with the average, middle, and end interpretation values of S, R, and $|C(S,R) - (1-\alpha)|$. Finally, we compare $|C(S,R) - (1-\alpha)|$ with $\tau = 0.01$. If the value is less than τ, then the alternate estimate is inferentially τ-close and considered an acceptable interpretation of the multi-year estimate. Such values are shown in bold in Table 2. Otherwise, the alternate estimate is not a suitable interpretation. Only four of the twelve comparisons are considered inferentially τ-close in this example: the middle interpretation for 2003–2005 and the average interpretation for 2001–2003, 2002–2004, and 2003–2005.

We can see this even more clearly in the graphs in Fig. 2. In each plot, the 3-year multi-year estimate is shown with a solid line and the specified alternate estimate is shown with a dashed line. The 95% confidence bands for the respective estimators are plotted as well. (Note that the listed year on the x-axis represents the *final* year in the 3-year period; for example, the 3-year estimates containing data from 2000–2002 would be be plotted at year 2002.) The confidence intervals for the middle and end year interpretations are wider than the 3-year multi-year estimate and the average interpretation since they incorporate fewer observations. The years for which the alternate estimate is inferentially τ-close, from the calculations in Table 2 are highlighted with arrows.

One important feature to note is that some of the alternate estimates fall within the pointwise confidence bands of the 3-year multi-year estimate but are *not* considered inferentially τ-close when we set $\tau = 0.01$. The second plot in Fig. 2 for the 2005 end interpretation is one such example. The methodology proposed in this paper incorporates the variances of the estimates into the methodology, which is vital to make a like for like comparison.

4.2. Summary of results

We have mentioned that many comparisons are possible once we account for all combinations of alternate estimate type, time period, county, and variable. We will show results by variable type (demographic, social, economic, and housing), and also separately display results by time period. In order to make a comparison, an estimate and its variance for a variable must exist for both the multi-year estimate and the relevant 1-year estimates used to calculate the alternate estimate. This restriction means that for each county, the number of variables to which we apply the methodology varies considerably. Tables 3 and 4 show typical variable counts for a 3-year (2000–2002) and a 5-year (2001–2005) comparison respectively.

In Figs 3 and 4, we plot a representative sample of results (the authors can supply the remaining results for interested readers). Each figure displays the results from a specific time span (i.e., 2000–2002 and 2001–2005). The top plot represents results from numeric variables and the bottom plot for percentage variables. We separate out results for numerics (e.g. totals, means, and medians) and percents because the U.S. Census Bureau makes this distinction in the Multi-

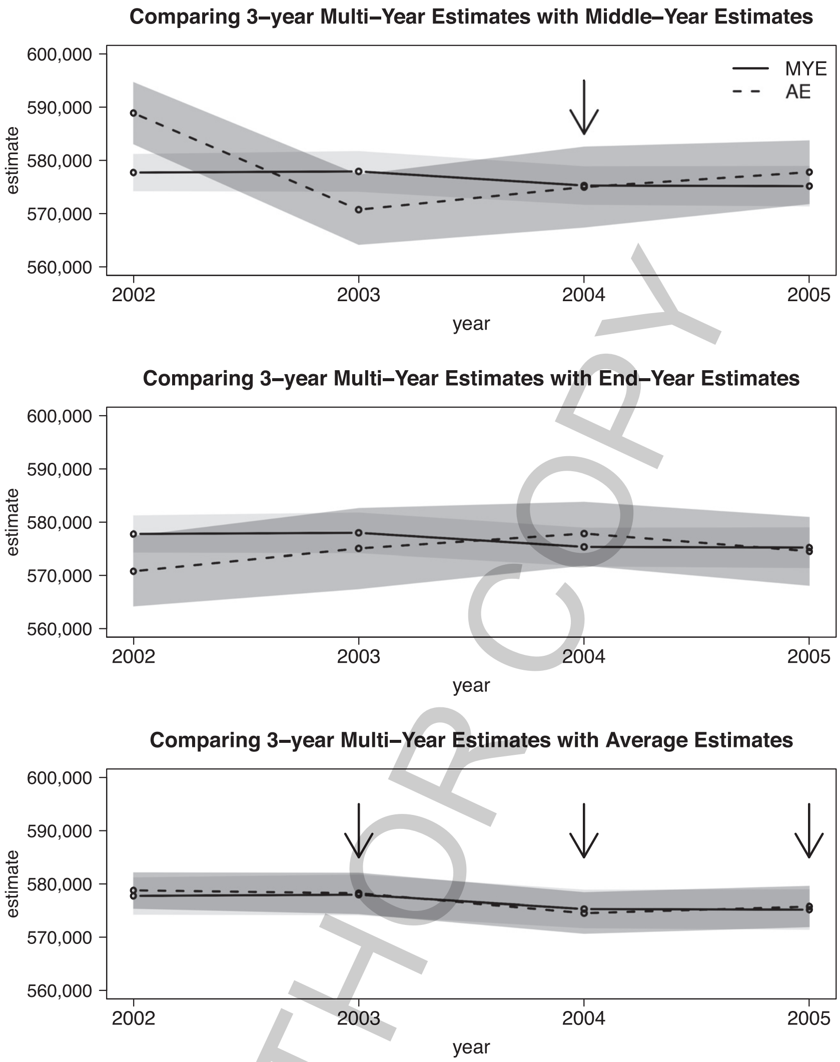

Fig. 2. Comparisons for number of persons over 16 in the labor force in Franklin County, OH (arrows represent alternate estimates which are inferentially τ-close to the corresponding multi-year estimate).

Year Estimates Study (and in current American Community Survey estimates). The box plots are divided into four groups, each representing one of the variable types (demographic, social, economic, and housing). There are three box plots for each group representing results from the mid-year, end-year, and average ("avg") alternative estimates. For each county, we apply the methodology to all of the variables and compute the fraction of variables which are *not* inferentially τ-close; that is, the alternate estimate is consequentially different from the multi-year estimate and is therefore *not* a suitable substitute. Each box plot represents the distribution of these fractions across the 19 counties (see Table 1).

We have two main results: (i) nearly all of the multi-year estimates are not substitutes for the corresponding alternate estimate and (ii) in general, this is more so the case for the middle and end interpretations than for the average interpretation. For middle and end interpretations, at least 70–80% of the variables checked in each county are not inferentially τ-close. For average interpretations, this percentage is a bit lower, at approximately 60%. Most likely, this is because the average alternate estimate is the only one of the three which incorporates data from all of the years included in the multi-year estimate. Finally, the spread of the fraction of consequential differences is much less for numeric variables corresponding to middle and end alternate estimates than for the average estimate or percentage

Table 3
Number of confidence intervals checked (3-year estimates: 2000–2002)

	Numeric (average year)				Percent (mid-year)			
	Demographic	Social	Economic	Housing	Demographic	Social	Economic	Housing
Black Hawk County, IA	70	92	112	125	71	81	66	105
Bronx County, NY	86	111	112	125	80	87	66	105
Broward County, FL	86	111	112	125	83	87	66	105
Calvert County, MD	74	86	85	113	67	70	66	98
Douglas County, NE	84	103	112	125	82	87	66	105
Flathead County, MT	72	86	86	125	64	70	66	105
Franklin County, OH	78	109	112	125	77	87	66	105
Hampden County, MA	85	111	112	125	79	87	66	105
Jefferson County, AR	69	59	112	125	68	66	66	105
Lake County, IL	86	111	112	125	79	87	66	105
Madison County, MS	51	85	99	116	54	66	66	96
Multnomah County, OR	89	109	112	125	86	87	66	105
Pima County, AZ	90	111	112	125	87	87	66	105
Rockland County, NY	83	111	112	125	76	87	66	105
San Francisco County, CA	86	111	112	125	83	87	66	105
Schuylkill County, PA	72	92	112	125	64	70	66	105
Sevier County, TN	49	51	86	113	63	66	66	105
Tulare County, CA	83	103	112	125	78	81	66	105
Yakima County, WA	74	104	112	125	73	81	66	105

Table 4
Number of confidence intervals checked (5-year estimates: 2001–2005)

	Numeric (end-year)				Percent (average year)			
	Demographic	Social	Economic	Housing	Demographic	Social	Economic	Housing
Black Hawk County, IA	81	92	112	122	63	70	66	102
Bronx County, NY	86	111	112	125	78	87	66	105
Broward County, FL	86	111	112	125	78	87	66	105
Calvert County, MD	70	92	87	103	63	66	57	87
Douglas County, NE	86	104	112	125	77	81	66	105
Flathead County, MT	71	88	90	125	45	65	51	102
Franklin County, OH	85	111	112	125	78	87	66	105
Hampden County, MA	86	111	112	125	71	87	66	105
Jefferson County, AR	69	59	105	116	43	39	60	96
Lake County, IL	85	111	112	125	78	87	66	105
Madison County, MS	71	92	95	107	44	66	51	89
Multnomah County, OR	89	109	112	125	82	87	66	105
Pima County, AZ	90	111	112	125	83	87	66	105
Rockland County, NY	80	110	112	125	74	87	66	105
San Francisco, CA	88	111	112	125	77	87	66	105
Schuylkill County, PA	71	92	112	125	63	70	66	102
Sevier County, TN	49	86	82	125	42	38	45	95
Tulare County, CA	82	104	112	125	73	81	66	105
Yakima County, WA	76	92	112	125	67	70	66	105

variables, possibly because percentages may be more stable across time than counts given shifts in population controls. However, according to our methodology, we see that, overall, it is not appropriate to interpret the multi-year estimate as *any* of the alternate estimates.

4.3. *Examining specific variables*

Next we examine thirteen specific variables of interest, chosen to represent a variety of "types" found in the Multi-Year Estimates Study. These variables are listed in Table 5 along with the table and profile line numbers (line in the original data file for that variable). We refer to each variable as follows: (table number/profile line). For example, median age is represented as DP01/17.

1. Stable variables: There are a few variables where we would not expect large differences in a three- or five-year period. For such variables, we would expect some, if not all, of the three interpretations of the period estimate to be reasonable. We examine three such variables here: median age in years

Table 5
Variables selected for further examination

Profile line	Variable name
Table DP01: demographic variables	
17	Median age in years
42	Number of Filipino residents
103	Average family size
Table DP02: social variables	
31	Fertility per 1,000 unmarried women
37	Number of grandparents who are responsible for their grandchildren
41	Veteran status/civilian veterans
76	Language other than English spoken at home
Table DP03: economic variables	
62	Mean household income in dollars
94	% of all families in poverty
Table DP04: housing variables	
19	House built between 1950 to 1959
59	House heating fuel is solar energy
72	Value of housing unit between $150,000–$199,000
99	Selected monthly house owner costs as a percentage of household income, 35% or more

Table 6
Proportions of non-inferentially τ-close alternative estimates for selected variables

Profile line		3-year (mid)	3-year (end)	3-year (avg)	5-year (mid)	5-year (end)	5-year (avg)
Table DP01: demographic variables							
17	Numeric	0.921	0.973	0.947	1	1	0.868
	Percent	–	–	–	–	–	–
42	Numeric	0.976	0.931	0.619	0.954	1	0.650
	Percent	1	0.954	0.714	0.954	0.909	0.700
103	Numeric	0.947	0.947	0.710	0.921	1	0.842
	Percent	–	–	–	–	–	–
Table DP02: social variables							
31	Numeric	0.960	0.960	0.671	0.945	0.973	0.666
	Percent	–	–	–	–	–	–
37	Numeric	1	0.972	0.643	0.972	0.945	0.666
	Percent	0.986	0.932	0.712	0.972	0.918	0.722
41	Numeric	0.902	1	0.742	0.970	0.973	0.687
	Percent	–	–	–	–	–	–
76	Numeric	0.941	0.959	0.723	0.961	0.913	0.913
	Percent	0.941	0.979	0.765	0.923	0.913	0.826
Table DP03: economic variables							
62	Numeric	0.947	1	0.986	0.921	1	1
	Percent	–	–	–	–	–	–
94	Numeric	0.955	0.984	0.716	0.968	1	0.800
	Percent	–	–	–	–	–	–
Table DP04: housing variables							
19	Numeric	0.960	0.973	0.723	0.973	0.947	0.631
	Percent	0.934	0.986	0.644	0.973	1	0.631
59	Numeric	1	1	0.676	1	1	0.885
	Percent	0.444	0.478	0.929	0.444	0.542	1
72	Numeric	0.959	0.986	0.630	1	0.972	0.722
	Percent	0.972	0.972	0.780	0.945	1	0.833
99	Numeric	0.921	0.934	0.750	1	0.947	0.657
	Percent	0.934	0.947	0.723	1	1	0.710

(DP01/17), average family size (DP01/103), and fertility per 1,000 unmarried women (DP02/31).

2. Subpopulations: This is a key situation wherein estimates can differ substantially from year to year, because very few people may be sampled who exhibit a given characteristic (see Beaghen and Stern [5] for a specific example in a related context). Therefore, interpreting a multi-year estimate to be an estimate of the middle or end of the period may not be appropriate for such variables. The variables selected specifically for this characteristic are: number of Filipino res-

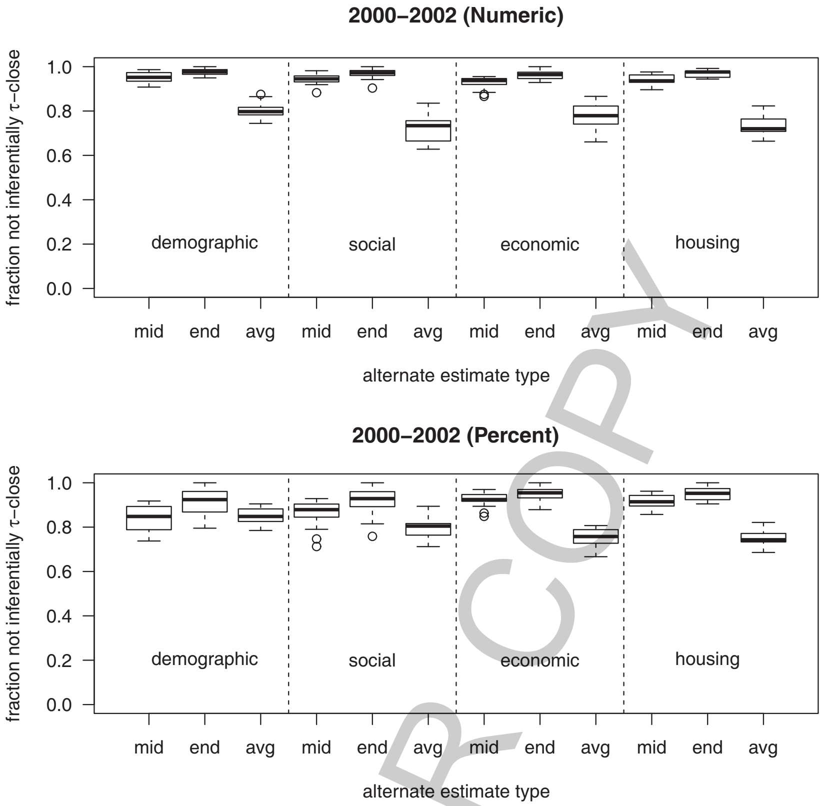

Fig. 3. Fraction of alternate estimates not inferentially τ-close to multi-year estimate (3-year estimates: 2000–2002) by variable (demographic, social, economic, housing) and estimate (middle, end, average) types.

idents (DP01/42), veteran status/civilian veterans (DP02/41), number of unpaid family workers (DP03/49), and number of people whose house heating fuel is solar energy (DP04/59).

3. Variables with high item allocation rates: There are some questions in the American Community Survey that respondents may not know an answer to, or feel hesitant to respond to. Consequently these questions may be left blank, or are answered with the respondent's best guess, which may not be accurate [20, p. 860] Many of these survey items would be imputed resulting in higher item allocation rates (i.e., high item nonresponse rates), possibly affecting the results of our study. Some variables computed from survey questions with high item allocation rates are: number of grandparents who are responsible for their grandchildren (DP02/37), language other than English spoken at home (DP02/76), and house built between 1950 and 1959 (DP04/19) [23]. (Note that money-related questions also have high allocation rates and we discuss those variables next.)
4. Money-related variables: Questions regarding money are in a category of their own, not only because they are sensitive questions but also because they may be hard to answer on behalf of other household members [11, p. 292]. Furthermore, the survey answers are adjusted for inflation and are used in computing poverty estimates (Beaghen et al. [4, p. 15]). All of these factors may affect our results. Such variables are based on the Multi-Year Estimate Study survey questions with high allocation rates [23]. Here we examine: mean household income (DP03/62), percent of all families in poverty (DP03/94), value of housing unit between \$150,000–\$199,000 (DP04 /72), and selected monthly house owner costs as a percentage of household income is 35% or higher (DP04/99).

In Table 6, proportions of confidence interval comparisons that were found to be consequential, or non-inferentially τ-close, are listed for each of the thirteen selected variables by time period (3- or 5-year) and alternate estimate (middle, end, average). Overall, the re-

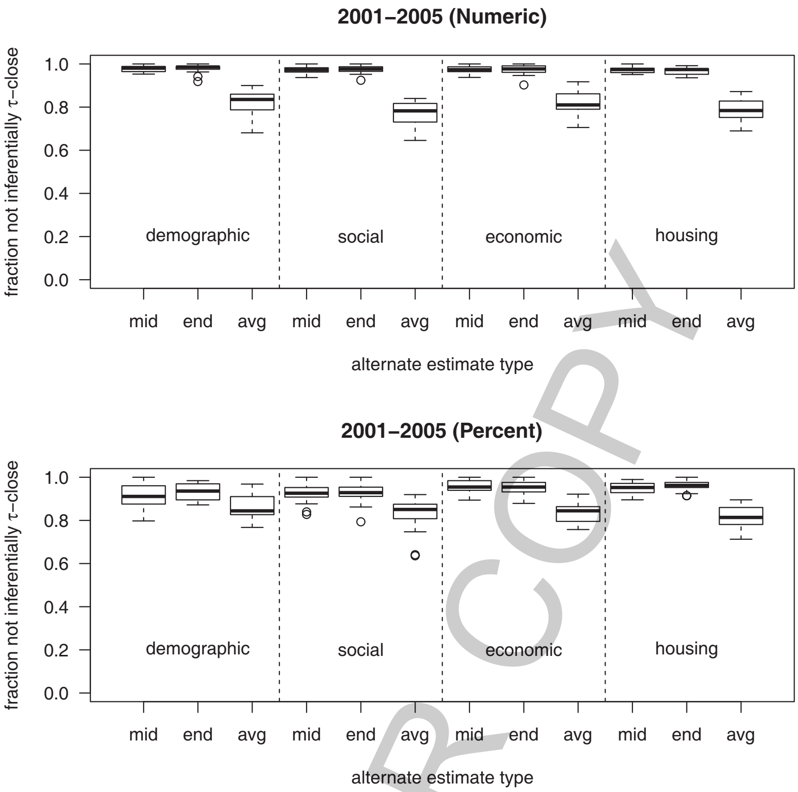

Fig. 4. Fraction of alternate estimates not inferentially τ-close to multi-year estimate (5-year estimates: 2001–2005) by variable (demographic, social, economic, housing) and estimate (middle, end, average) types.

sults from each variable tend to be similar to the overall results. Most often, the middle and end interpretations of the period estimate are the least acceptable, followed by the average interpretation. One notable exception is for the percentage variable "house heating fuel is solar energy" (DP04/59). Between 44% and 55% of the comparisons are consequentially different for the middle and end interpretations, whereas the nearly all of the average interpretations are. It is also interesting to note that variables categorized as stable, such as fertility rates, violate the linearity assumptions at nearly the same rates as other variables.

4.4. Varying τ

Until now, we have fixed the maximum discrepancy (τ) to be 0.01 when $\alpha = 0.05$. One could argue that this threshold is too strict and that it is unsurprising most alternate estimates are considered consequentially different from their corresponding multi-year estimates. If we increase τ, the fraction of variables which are not inferentially τ-close will decrease for any of the three interpretations. In this section, we examine how much our results actually change if we choose alternative values of τ (keeping α still fixed at 0.05).

To answer this question, we perform an analysis similar to the one in Section 4.2 for four counties, but vary τ. The four counties, Bronx County, NY, Yakima County, WA, Schuylkill County, PA and Flathead County, MT were chosen to represent the wide range of county population sizes (listed in descending order here). In Fig. 5, we plot the fraction of variables which are not inferentially τ-close for different combinations of time span and type of variable (numeric versus percent). As a benchmark, $\tau = 0.01$, which we used for the results in Sections 4.1–4.3, is highlighted on each plot.

For each county, as the maximum permitted discrepancy (τ) increases, the proportion of consequential differences drops off much faster for the average interpretation than for the middle and end interpretations. We expect more alternate estimates to be acceptable as τ increases because we are weakening the restrictions on what is considered inferentially τ-close. Given the results in Section 4.2, we also expect that the average interpretation has lower rates of consequential differ-

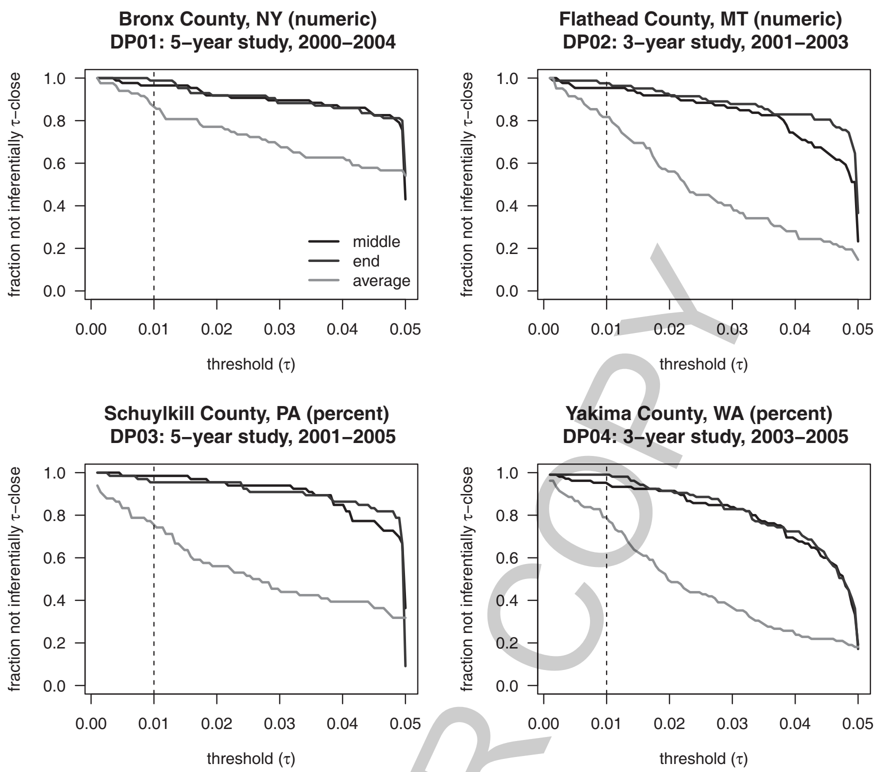

Fig. 5. Fraction of consequential differences (i.e., not inferentially τ-close) by threshold τ ($\alpha = 0.05$) for four counties.

ences than the other two interpretations since the same data is used for both the average estimate and the corresponding multi-year estimate. An important feature is that these characteristics are consistent across all combinations of county and period.

For the middle and end interpretations, even at large values of τ, we find that a majority of the variables are consequentially different for each county. For the average interpretation, as τ increases, the fraction of inferentially τ-close variables increases more rapidly for 3-year multi-year estimates than for 5-year multi-year estimates and soon become the majority. The same conclusions hold if we examine the corresponding plots for other county, variable type, and time period combinations.

4.5. *Multi-year estimates as proxies for average estimates*

Suppose now that the population value of interest is the true average over multiple years. Here, the alternative estimate, the average, $\tilde{x}$, is the estimate of this population value. We ask the following question: Is the corresponding multi-year estimate a suitable proxy for the average estimate? We can modify the methodology outlined in Section 3 to answer this question. As before, $\tilde{x}$ is the average estimate and $\hat{x}$ is the multi-year estimate. However, now the confidence interval is written as $\tilde{I}(s,r) = \left[\tilde{x} + (s - \sqrt{r}z_{\alpha/2})\hat{\sigma}_{\tilde{x}},\ \tilde{x} + (s - \sqrt{r}z_{\alpha/2})\hat{\sigma}_{\tilde{x}}\right]$. Now, if we let $\tilde{D} = (\hat{x} - \tilde{x})/\tilde{x}$, $\tilde{S} = (\hat{x} - \tilde{x})/\hat{\sigma}_{\tilde{x}}$, and $\tilde{R} = \hat{\sigma}^2_{\hat{x}}/\hat{\sigma}^2_{\tilde{x}}$, we can use

$$\left|\left[\Phi\left(-\tilde{S} + \sqrt{\tilde{R}}z_{\alpha/2}\right) - \Phi(-\tilde{S} - \sqrt{\tilde{R}}z_{\alpha/2})\right] - (1-\alpha)\right| < \tau \tag{5}$$

to determine whether the multi-year estimate is inferentially-τ close to the average estimate.

While the fraction of confidence intervals not inferentially-τ close is lower than in the original setting (Section 4.2) for the different variable categories and time periods, they are almost always above 60%. Consequently, the multi-year estimate is generally not a suitable proxy for the average estimate. In Fig. 6, we display the results for three typical time periods: 2000–2002, 2002–2004, and 2001–2005. Each box plot represents the fraction of not inferentially-τ close variables for each variable category across the 19 counties. Similarly, we can also apply this methodology to the mid-year and end-year estimates as well; we obtain

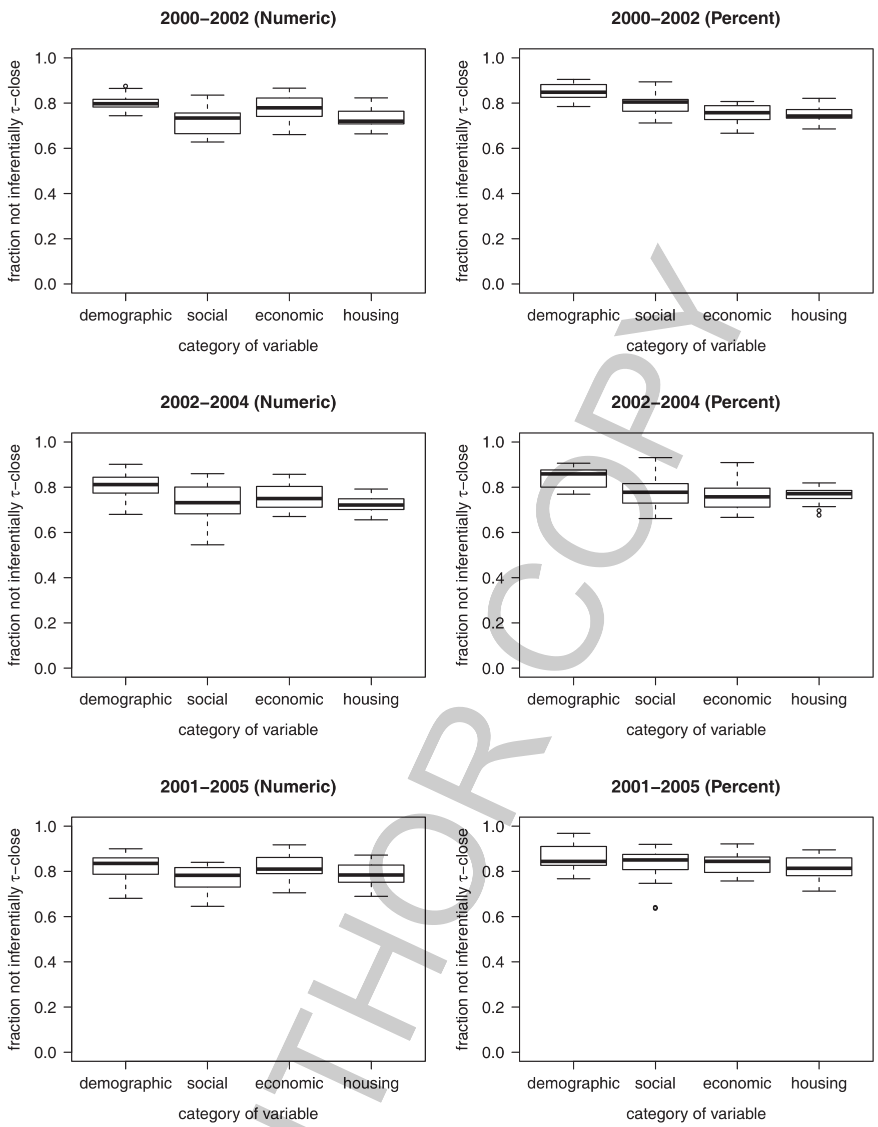

Fig. 6. Fraction of variables which are not inferentially τ-close when using a multi-year estimate as an interpretation for the average estimate.

similar results for the Multi-Year Estimates Study data when we do so.

5. Discussion

This paper addresses a long-standing issue of concern in the American Community Survey: can multi-year estimates be viewed as linear functions of one-year estimates? Although official U.S. Census Bureau policy answers this question negatively, an empirical exploration of the issue was lacking until now. We propose a simple, replicable methodology that is directly matched to the typical user's concerns via the novel concept of "inferentially τ-close," and apply this technique to the Multi-Year Estimates Study database. Our conclusion is that the official policy is correct. Note that fundamental reasons can be given for why multi-year estimates should not be viewed as linear functions of one-year estimates, but even so it is not *a priori* clear that ignoring this injunction will have negative repercussions in terms of data analytical conclusions. However, the results given in Section 4 demonstrate that there are indeed negative consequences in practice.

The method of consequential differences is directly linked to underlying assumptions and usages of the American Community Survey already in place, and adds no further axiomatic burden. There are few "tuning parameters" to the method; just the choice of τ and the default values used in the Multi-Year Estimates Study evaluation are sensibly motivated such that no adjustment is typically required. The actual application of the method in practice is quite simple: any survey with multi-year estimates and standard errors along with the potential to compute alternative estimates and their standard errors is sufficient. This allows one to compute R and S from step 2 of the algorithm given at the end of Section 3, and the question of whether or not the estimates are inferentially τ-close can then be answered.

The vast Multi-Year Estimates Study serves as a proxy for the still larger American Community Survey; hence our conclusions here should be expected to extend to the full survey. We argue for the validity of this extrapolation on the grounds of the initial design of the Multi-Year Estimates Study – it was constructed to be just such an antecedent for the more comprehensive American Community Survey. The actual verification of this extrapolation requires one to apply our methodology to the publicly available data.

Overall, we find solid support that substitution is unwise, though the exact consequence depends on many factors: period length, variable, region (county in our study), and alternate interpretation type. We conclude by reinforcing the U.S. Census Bureau cautions against making such *ad hoc* comparisons.

Disclaimer

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Acknowledgements

We want to thank Krizstina Marton and Mike Cohen from the Committee on National Statistics at the National Academies in the U.S. for their assistance on identifying surveys around the world which use multi-year estimates.

References

[1] C. Alexander, Impact of multiyear averaging of data from the American Community Survey, *Proceedings of the Survey Research Methods Section, ASA*, (1996), 644–649.

[2] C. Alexander, Recent developments in the American Community Survey, *Proceedings of the Survey Research Methods Section, ASA*, (1998), 92–100.

[3] C. Alexander, Still rolling: Leslie Kish's "rolling samples" and the American Community Survey. *Proceedings of Statistics Canada Symposium, Achieving Data Quality in a Statistical Agency: A Methodological Perspective*, (2001). http://www5.statcan.gc.ca/olc-cel/olc.action?ObjId=12-001-X20020016413&ObjType=47&lang=en.

[4] M. Beaghen, T. McElroy, L. Weidman, M. Asiala and A. Navarro, Interpretation and Use of American Community Survey Multi-Year Estimates. *U.S. Census Bureau Research Report, 2012–03*, (2012).

[5] M. Beaghen and S. Stern, Usability of the American Community Survey Estimates of the Group Quarters Population for Substate Geographies, *Proceedings of the 2009 Joint Statistical Meetings on CD-ROM* American Statistical Association, 2009, pp. 2123–2137.

[6] M. Beaghen and L. Weidman, Statistical Issues of Interpretation of the American Community Survey's One-, Three-, and Five-Year Period Estimates, U.S. Census Bureau, 2008 American Community Survey Research Memorandum Series #ACS08-R-4, 2008.

[7] W. Bell, Borrowing Information Over Time in Small Area Estimation: Thoughts with Reference to the American Community Survey, unpublished paper presented at the American Community Survey workshop, September 13, 1998, Michael Cohen (workshop study director), Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council, Washington, DC, 1998.

[8] F. Breidt, Alternatives to the Multi-Year Period Estimation Strategy for the American Community Survey, Appendix C in *Using the American Community Survey: Benefits and Challenges, National Research Council, Panel on the Functionality and Usability of Data from the American Community Survey*, Constance F. Citro and Graham Kalton, editors, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press, 2007.

[9] N. Chand and C. Alexander, Multi-year averages from a rolling sample survey. *Proceedings of the Survey Research Methods Section, ASA*, 2000, 301–306.

[10] C. Citro and G. Kalton, *Using the American Community Survey: Benefits and Challenges, National Research Council, Panel on the Functionality and Usability of Data from the American Community Survey*, Constance F. Citro and Graham Kalton, editors, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press, 2007.

[11] D.A. Dillman, M.D. Sinclair and J.R. Clark, Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys, *The Public Opinion Quarterly* **57** (1993), 289–304.

[12] J. Durr, (National Institute of Statistics and Economic Studies (INSEE, France)) (2004). The new French rolling census. Statistical Commission and UN Economic Commission for Europe, Conference of European Statisticians, UNECE Seminar on New Methods for Population Censuses, Geneva (Working paper). http://www.unece.org/fileadmin/DAM/stats/documents/2004/11/censussem/wp.2.e.pdf.

[13] R. Fay, Imbedding model-assisted estimation into ACS estimation, *Proceedings of the 2007 Joint Statistical Meetings on*

CD-ROM, American Statistical Association, 2007, pp. 2946–2953.

[14] L. Kish, Using cumulated rolling samples to integrate census and survey operations of the Census Bureau, Washington, D.C., U.S. Government Printing Office, 1981.

[15] L. Kish, olling samples and censuses, *Survey Methodology* **25** (1999), 129–138.

[16] T. McElroy, Incompatibility of Trends in Multi-Year Estimates from the American Community Survey, *The Annals of Applied Statistics* **3** (2009), 1493–1504.

[17] M. Starsinic and A. Tersine, Analysis of variance estimates from American Community Survey multi-year estimates, *Proceedings of the 2007 Joint Statistical Meetings on CD-ROM,* American Statistical Association, 2007, pp. 3011–3017.

[18] A. Tersine and M. Asiala, Methodology for the production of American Community Survey multi-year estimates, *Proceedings of the 2007 Joint Statistical Meetings on CD-ROM,* American Statistical Association, 2007, pp. 3018–3023.

[19] N. Torrieri, America is Changing, and so is the Census: The American Community Survey, *The American Statistician* **61** (2007), 16–21.

[20] R. Tourangeau and T. Yan, Sensitive questions in surveys, *Psychological Bulletin* **133** (2007), 859–883.

[21] U.S. Census Bureau (2006) Technical Paper 67. *Design and Methodology, American Community Survey.* http://www.census.gov/history/pdf/ACSHistory.pdf.

[22] U.S. Census Bureau (2007a) Multi-Year Estimates Study, Accuracy of the Data. http://www.census.gov/acs/www/Downloads/methodology/special_data_studies/multiyear_estimates/Accuracy_of_the_Data_MYE.pdf.

[23] U.S. Census Bureau (2007b) Multi-Year Estimates Study, Multi-Year Estimate Study Quality Measures Definitions. ftp://ftp.census.gov/acs/MultiYearEstimates/Quality_Measures/myeqmdocumentation.doc.

[24] P. Zadka and Y. Fienstien, (Central Bureau of Statistics, Israel) (2012). The 2008 integrated census in Israel and future censuses. United Nations International Seminar on Population and Housing Censuses: Beyond the 2010 Round, 27–29 November 2012, Seoul, Republic of Korea. http://unstats.un.org/unsd/demographic/meetings/Conferences/Korea/2012/docs/s04-2-1-Israel.pdf.