An Empirical Artificial Population and Sampling Design for Small-Area Model Evaluation *

Jerzy Wieczorek† Carolina Franco‡

Abstract

When competing small area models are proposed for a particular set of estimates, they should be evaluated and compared not only by large-sample theory but also by their small-sample properties on real or realistic data. As an alternative to evaluations on data simulated directly from simple parametric models, we evaluate models by a simulation study tailored more closely to the source data that will be used in production. We use the 2007-2011 American Community Survey (ACS) 5-year unit-level sample data as a universe, which is likely to account for relationships among variables and other complexities that may not be reflected in a purely model-based artificial population. We then sample the population repeatedly with a design that mimics the ACS sampling design. In that sense this is a "design-based" simulation, although the simulation's response mode and unit nonresponse behavior are model-based, and item nonresponse is not yet implemented. This simulation framework allows for comparison of different statistical inference approaches, with no method being inherently favored over others. Possible future improvements and potential drawbacks of this approach are also discussed.

Key Words: design based simulation; small area estimation; American Community Survey; artificial population

1. Introduction

1.1 Motivation

The U.S. Census Bureau is exploring expanding the use of Small Area Estimation (SAE) models to supplement survey samples. One particular goal is to use SAE models for improving poverty estimates at the census tract level, using the American Community Survey (ACS) as the primary data source. There is also interest in expanding this research to other ACS variables besides poverty.

The Small Area Income and Poverty Estimates (SAIPE) program uses SAE models to produce poverty estimates for states and counties, but not for geographies as small as census tracts. (The SAIPE program also produces estimates for school districts, although using a county-shares method.) Census tracts are

> small, relatively permanent statistical subdivision of a county or equivalent entity ... Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people (U.S. Census Bureau, 2012).

This raises the general problem of evaluating and comparing SAE models. When a statistical agency chooses a particular small area model for producing a particular set of estimates, the agency should ensure that

*This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

†Contact author: jerzy@cmu.edu, Department of Statistics, Baker Hall, Carnegie Mellon University, Pittsburgh, PA 15213

‡Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

- the proposed model has good statistical properties for the problem at hand, and
- its predictions perform at least as well as those of other proposed models.

We have several models under consideration, and we would like to

- understand each models' strengths and limitations, and
- compare the models to each other in order to choose one for use in production.

However, this is difficult, as the proposed models come from a variety of approaches to statistical inference: frequentist vs. Bayesian, design-based vs. model-assisted vs. model-based, etc. It can be challenging to directly compare models inspired by one inference paradigm against models from another paradigm.

Ideally, if we had a "truth deck" of the poverty status of the entire U.S. population, as well as several iterations of the ACS from the same snapshot of that population, we could evaluate and compare SAE models by seeing which model tends to produce estimates closest to the truth across the various samples. Unfortunately, we have no such truth deck: the Decennial Census does not collect poverty information for the entire population, besides suffering from coverage and other nonsampling error. Also, we do not have multiple iterations of the ACS: it is only run once per year. Finally, such a truth deck for our problem would also have to contain the outcomes of all potential respondents' self-selection decisions regarding ACS response mode and item nonresponse, which is clearly not attainable in practice.

Instead, we mimic this approach by generating a realistic artificial population, then drawing repeated samples from this population using an imitation of the ACS sampling process. The following excerpt provides a clear summary of this idea:

> The ACS micro-data could be used to construct a test population for comparison, either by using the entire sample directly, or more likely by using the empirical distribution of the response population to produce a larger artificial population for particular domains and area types. A challenge of this approach would be to reproduce the complex sample design to provide a test ACS-like sample for use by the candidate SAE approach. Non-response behavior and imputation error would also be included (Bell et al., 2011).

This approach has been referred to as *design-based simulation* (see for instance Alfons et al. 2010). In contrast, in a strictly *model-based simulation*, the simulated attributes are often drawn from oversimplified parametric models based on simple distributions such as Binomial, Normal or Poisson, and sometimes drawn i.i.d. across the population without using different parameters for different geographic-by-demographic cells. The latter approach may miss important relationships in the real data and ignore the effects of sampling, nonresponse, etc.

In practice, even so-called design-based simulations often incorporate model-based elements. For instance, our simulation uses estimated response propensities and a Bernoulli model to decide each household's response mode and nonresponse status anew within each simulation run. We have not yet implemented item nonresponse, but it will necessarily be model-based as well. In a setting like this, where response mode self-selection and nonresponse exist, there is no way to escape modeling and perform purely design-based simulation. Still, although the terms design-based and model-based simulation are used loosely in common practice, we find them useful to distinguish drawing complex samples from an empirical distribution vs. drawing i.i.d. samples from, say, a Normal distribution.

1.2 Properties of a desirable simulation design

A useful design-based simulation procedure requires:

- Ways to generate a "universe" that resembles the population of interest.
- Processes for generating samples from this universe, including:
  - Sample selection procedures that can resemble the survey(s) being studied.
  - Production procedures that handle mode selection, unit and item nonresponse, imputation, calibration, etc.

If these components are available, the production and sample selection procedures can be used to reconstruct an existing survey and production or to test the consequences of changes in the design and production. Given a particular combination of universe, sampling design, and set of production mechanisms, the universe can be sampled repeatedly as the basis for comparing inference techniques.

This universe or artificial population should be realistic in accounting for the complex relationships between units and between variables. The sampling and weighting process should mimic that of the target survey. The mechanisms for mode selection, nonresponse, etc. should be as realistic as possible, and should be adjustable (e.g., simulating either ignorable or nonignorable nonresponse) to allow for different test scenarios.

1.3 ACS design overview

The ACS sample design and methodology are documented in detail by U.S. Census Bureau (2009). In summary, the ACS has sampled approximately 3 million addresses each year since its full implementation in 2005. The complex sample design involves stratification, systematic sampling, and clustering of persons within households. Individuals in the sample who do not reply by mail or telephone are subsampled for Computer Assisted Personal Interviewing (CAPI). A non-negligible proportion of addresses are "unmailable" and must be sampled directly into CAPI at a higher rate, skipping the mail and telephone modes entirely. After the data are collected and processed, the estimation procedure begins with basic weights that then undergo several adjustments for post-stratification, nonresponse, population controls, and so on.

The ACS surveys and calculates estimates both for the population in housing units (HUs) and for the group quarters (GQ) population. However, currently our methods are exclusively focused on inference for the HU universe.

2. Related work

Several previous authors have considered the problem of evaluating models by repeated sampling from complex artificial populations. Their work provided us with valuable guidance regarding important details to consider and possible approaches to take.

Schafer and Kang (2008) build an artificial population for model-comparison purposes, avoiding the use of the competing models in the construction of their population. Alfons et al. (2010, 2011) document design decisions in creating a pair of software packages that can generate an empirical artificial population and repeatedly draw complex samples. Franco (2011) gives a concrete example of generating and sampling from an artificial population using a U.S. Census Bureau dataset and sampling design.

2.1 Schafer and Kang (2008)

Schafer and Kang (2008) draw repeated samples from a simulated population of adolescent girls to compare different methods of computing Rubin's Average Causal Effect (ACE). The authors use distributions estimated from the National Longitudinal Study of Adolescent Health ("Add Health") to create a synthetic population of 1 million adolescent girls, and then draw a simple random sample (SRS) of size 6000 repeatedly. They take special efforts to capture nonnormal shapes, nonlinearity, structural missingness, and interactions when constructing the artificial population.

Their estimation methods are based on simpler models that do not precisely agree with those used to create the population, so that they can see how well these models perform when all of them are misspecified. As in our work, Schafer and Kang have the goal of constructing a population in a complex enough manner to be realistic, without favoring any one estimation model over others. However, unlike in our work, the sampling design for their simulations is SRS, since Schafer and Kang are not interested in mimicking any particular survey's design or incorporating complex sampling features into the simulation.

2.2 Alfons et al. (2010, 2011)

Alfons et al. document a pair of packages, `simPopulation` and `simFrame`, for the statistical software `R`. The package `simPopulation` provides tools for creating a complete artificial population based on a survey sample. Its output can be used in `simFrame` to create a sampling frame, repeatedly draw and analyze samples, and summarize the results.

`simPopulation` (Alfons et al. 2011) expands a weighted-sample dataset to a population sized by the sum of the weights. The package initializes the estimated number of households of each size and age-by-gender structure, then imputes other variables conditionally using models estimated on the original sample.

`simFrame` (Alfons et al. 2010) lets users specify the desired sampling design, sample size, data contamination features, missing values (i.e. unit or item nonresponse patterns), and a function (such as a SAE model estimation procedure) to run on each simulated dataset. The package processes all of the simulations and automatically summarizes the across-simulations results.

Initially, we had hoped to use Alfons et al.'s `R` packages for our research, since they are designed to be extensible by users. Unfortunately, we encountered several obstacles:

- several elements of the particular survey design we require (including sample selection, nonresponse, cell-collapsing, weighting, replicate weights. . .) are beyond the scope of the `simFrame` framework as it now stands
- our simulations must be easily run by other staff at the Census Bureau, where the statistical software `SAS` is the standard rather than `R`, and
- our artificial population is too big to load into our available working memory (RAM), which is less of a limitation in `SAS` than in `R`.

Consequently, we decided to write our own `SAS` program custom-tailored to mimicking the ACS complex sample design. Following Alfons et al.'s approach in spirit, our code is modular so that in future research we can incorporate other assumptions. For example, we could try out several different non-response models, or we could use a sampling design and weighting adjustments from a different survey such as the Current Population Survey (CPS).

This approach also means we do not benefit from the simulation-results summaries built into Alfons et al.'s packages. However, by coordinating with the various small-area researchers on our project, we can ensure that everyone uses the same set of simulations and the same metrics for evaluating their models.

2.3 Franco (2011)

During the 2000 decennial census, approximately one-sixth of households were sampled to receive the “long form” including many detailed questions that were omitted from the “short form” received by other households. The long form has been supplanted by the ACS, so in the 2010 census all households received a short form only. In Franco (2011), a “true population” is constructed from the Census 2000 long form by imputing housing units to all addresses in the Master Frame that were not selected in the 2000 long form sample in a way that accounts for the sampling weights of the original sample. As in the present paper, Franco’s artificial population contains complete data on variables of interest, but not on the nonresponse behavior, which is dealt with via a simplified model.

Franco’s artificial population is sampled 1000 times, using a simplification of the ACS design (not the long form design) that includes stratification, clustering of people within households, and systematic sampling. A logistic regression model based on ACS data predicts which units sampled are likely to become eligible for CAPI, and a subset of these are retained in each iteration of sampling. These respondents’ weights are adjusted accordingly. The resulting artificial population and simulated samples are used to compare the performance of confidence intervals for proportions in small areas.

Imputing a census population using the 2000 long form data was sufficient for the purpose of evaluating confidence intervals for proportions. However, for the present research, our desire to evaluate ACS-specific models means that we cannot rely on long form data that is over a decade old and that asks slightly different questions than the ACS. Instead, we need to build our artificial population from the ACS itself.

3. Methods

3.1 Artificial population

Our current first attempt at constructing an artificial population is to use the complete, unweighted, unit-level 2011 5-year ACS dataset and treat it as the entire universe. (As mentioned above, we omit the group quarters (GQ) population and only use the housing unit (HU) population.) This approach ensures that the data has all the variables we need: demographics, geographic IDs, links between persons and households, auxiliary data, etc. Any administrative records used as SAE model covariates are converted to rates (e.g., county-level food stamp program participation counts are divided by county size and transformed to rates), then multiplied by the artificial population domain sizes to get auxiliary counts for the artificial population.

Currently, the artificial population domains are determined by the U.S. county where their ACS data was sampled. However, the artificial population sizes within each domain are much smaller than typical county populations. Instead, they tend to be closer to U.S. tract sizes, which are the real population domains we want to imitate in our current SAE model evaluations. Future versions of the software will allow the user to select other geographies for defining small area domains or to specify their own domains directly.

This simplistic approach has been quick to implement, while preserving essential data features (such as within-household clustering) that would be lost under a more traditional model-based simulation approach. It also requires no modeling to create the artificial population. However, we recognize that this is a very simple “placeholder” artificial population, while our main focus so far has been on improving the other steps of the process: sampling, weighting, nonresponse, etc. Naturally, there are many weaknesses with this approach.

First, the domains have incorrect sizes, being reduced to a fraction of their full size approximately equivalent to the overall ACS sampling fraction (roughly $1/40$). The small areas of interest (census tracts) become too small to work with when they are sampled. One possible way to alleviate this is by joining census tracts together to create "artificial census tracts" whose artificial-population sizes sum up to the size of a real census tract. However, this can lead to increased heterogeneity compared to the real census tracts, which are drawn to be as homogenous as possible. Also, it leads to having a smaller number of domains, even though each domain would have an appropriate size.

Next, since we ignore the informative weights in the ACS data, our resulting artificial population is not representative of the U.S. population. This leads to altering the relationship between the response and the administrative records used as covariates. These covariates may not have as strong a relationship with the response in the reduced population. Also, the relationship between variables within the reduced sample may not be the same as the relationships between the variables in the real U.S. population.

To address these weaknesses, we intend to build a more complex artificial population of the same size as the real U.S. population it mimics. We are conducting preparatory research to determine how best to accomplish this, drawing on the methods of Schafer and Kang (2008), Alfons et al. (2011), and Franco (2011).

3.2 Sampling and nonresponse

Our sampling process mimics the annual ACS sampling procedure. The ACS performs sampling independently within each U.S. county or county equivalent, except for the very smallest counties, which are combined with neighbors into county-groups. Since tracts are sub-county geographies, this process alone does not guarantee that every tract will be sampled. However, the sampling within each county is a systematic sample in geographic sort order, with a skip much smaller than the smallest tract size, which in practice ensures that every tract is sampled.

We ignore several features of the HU sampling design that were deemed unnecessarily complicated for our current simulations. First, the ACS actually separates the sampling frame into 5 subframes and samples from one per year, to ensure that no housing unit (HU) is sampled twice within any 5-year period. For simplicity, we ignore the 5 subframes and just sample from the whole artificial population at once.

Second, we also ignore the ACS stratification of HUs by Measure Of Size (MOS). The real ACS determines the smallest sampling entity (school district, county, etc.) to which each block belongs, then assigns all HUs in that block to one of five strata based on the MOS of that smallest sampling entity. This is meant to allow for higher sampling rates in small-MOS entities, ensuring sufficiently-large sample sizes. However, determining the MOS would be overcomplicated for our current simplistic artificial population, so we ignore this stratification step, although we intend to use it in the future with improved artificial populations.

Third, we ignore the fact that the ACS assigns each sampled HU to a sampling month. This will be incorporated in the future once we have a nonresponse model that accounts for sampling month. Since our current nonresponse rate is constant, the sampling month would have no effect.

Finally, we ignore several practical constraints on sampling that the ACS has to manage. To begin with, some real HU addresses are unmailable and the ACS must send these straight to in-person (CAPI) followup instead of starting with mail and phone contact. We simplify by assuming our entire sampling frame is mailable. Furthermore, some HU addresses in the sampling frame are actually vacant or fall into other special categories such as temporarily occupied, not residences but businesses, etc. Our simulation combines all of these special cases into one category, “vacant,” but again this is a considerable simplification.

The ACS samples HUs directly and then collects information on each resident of the responding HUs. Therefore, our simulation design needs to mimic the sampling process and the nonresponse or mode selection among HUs.

In our simulations, independently by county:

- We ignore the MOS strata, sort the artificial population HUs in geographic order, and perform a simple systematic sample of HUs at a constant rate of 1/40 or 2.25%, close to the ACS target base rate which was 2.23% in 2007 (U.S. Census Bureau, 2009). In our artificial population, each of these sampled HUs is categorized either as eligible for the ACS or as vacant (though in the data used to generate our universe, some of these “vacant” HUs were actually in other categories, as mentioned above).

- Of the sampled nonvacant HUs, we use Bernoulli sampling at a rate of 67% to indicate the sampled HUs that respond by mail or phone, requiring no further followup. This corresponds roughly to the national unweighted HU interview rate, which has been between 0.650 and 0.701, or approximately 2/3. This rough estimate is based on the ratios of Final Interviews to Initial Addresses Selected for 2000 through 2011, according to U.S. Census Bureau (2013).

- Of the remaining HUs in the sample (either vacant or nonresponding to mail & phone), we take a simple random sample (SRS) of 33% to be selected for in-person (CAPI) followup. This is the default 1/3 rate for mailable addresses; U.S. Census Bureau (2009) also lists several address and tract categories with higher CAPI subsampling rates, but we ignore this complication for our simulations.

- Among the CAPI followup HUs, we find any vacant HUs and use Bernoulli sampling to choose 67% of the nonvacant HUs as responding to CAPI. Again, the 67% was chosen to correspond to the national unweighted HU interview rate, although that rate is not in fact closely related to the CAPI followup success rate, and better estimates will be chosen in future work.

- The remaining HUs are treated as nonrespondents. Again, we ignore here the real ACS’ followup for detecting nonresponding addresses that turn out to be ineligible such as businesses, vacant units, etc.

All persons in responding HUs become the person sample. All items from the responding HUs and persons are treated as known in our samples: we do not perform any item nonresponse. We use an imputed version of the ACS dataset in constructing the artificial population, in order to have complete responses for every unit in the universe.

There are currently two response rates, both set to a constant 67%: one for HU propensity to respond at the mail & phone stage, and another for propensity to respond at the CAPI stage. In the near future, we plan to replace the flat 67% rates with nonresponse propensity models. Further along, we hope to include item nonresponse as well. Of course, the simulation procedure will then also require an imputation module, so that demographic variables can be imputed for weighting and other variables can be imputed for aggregation and further analysis.

3.3 Weighting

First, note that we ignore the original weights in the ACS data used to create the artificial population. All weights in our simulated datasets are created through the simulation's own sampling and weighting process.

Weighting is performed independently within each county, just as sampling is. The real ACS process pools several of the smallest counties into combined county-groups for sampling and weighting purposes, but our current simulation ignores this step. For any weighting adjustments that involve raking to population totals, we currently assume that these totals are known from the artificial population. Future versions may allow users to mimic the real-world uncertainty about population totals.

As the ACS does, we begin by assigning each HU a base weight equal to the inverse of its base sampling rate; in our case the base weights are all 40. For all HUs that responded in CAPI, we upweight them by a factor of 3 (to 120) to account for the followup subsampling rate of 1/3. We then rake all HU weights to the population total numbers of HUs by vacancy status (vacant vs. nonvacant), and we assign each responding nonvacant HU's weight to all of its residents as their initial person weights.

Next, we use iterative proportional fitting (IPF) to rake these person weights by two factors: by simplified demographic cells, with collapsing of cells if needed, and by householder versus all others. We alternate raking by demographics and by householder status repeatedly until the raking factors converge to $1 \pm .001$.

Our demographic cells include all crossings of four race/ethnicity categories, two sexes, and two age groups. If any of the demographic cell in the sample contains fewer than 10 people, we collapse it with other demographic cells until the combined total count is over 10. This prevents small cells' weights from fluctuating too much from sample to sample. The cell-collapse order follows the current ACS procedure as documented in U.S. Census Bureau (2009).

Our demographic cells are a simplification from the real ACS, which considers six race/ethnicity categories, two sexes, and 13 age groups. We felt that the complications surrounding use of the complete demographic cell breakdown, especially the cell-collapse rules, were not worth the time it would require to program and debug. The ACS also rakes by a third factor as well, for marital status: its categories are householders in two-person-relationship households, their spouses or partners, and everyone else. We have not included this third raking factor yet, but it should not be difficult to incorporate in the near future.

Once the IPF procedure has converged, we replace each HU's weight with its householder's new weight. These new HU weights still sum up to the population total of nonvacant HUs, since the number of householders and the number of occupied HUs are the same.

This concludes the process of creating primary weights for each HU and person in the sample. We also create 80 replicate weights for each HU and person, in order to mimic the Successive Differences Replication Method used to compute sampling variances in the ACS. To generate these 80 sets of replicate weights, we return to the initial weights of 40 or 120; multiply them by an 80-column matrix of replicate-weight-factors computed from an 80 by 80 Hadamard matrix; and run the resulting perturbed weights through the same raking and IPF weighting process as above. U.S. Census Bureau (2009) documents a simple example and the rationale behind the use of the Hadamard matrix, replicate-weight-factors, etc.

3.4 Aggregation

Since many of the proposed SAE models for evaluation are area-level rather than unit-level models, we include a module for aggregating the samples into area-level estimates. Of course, users can always perform this aggregation themselves from the weighted unit-level data if they need any other statistics that we do not calculate.

We aggregate independently in each small domain, which by default uses the same county ID that is currently used for sampling and weighting. However, users can currently calculate estimates at alternative geographic levels by switching out the county-level geographic ID variable with, e.g., census tract IDs. We aim to make aggregation domains much simpler for users to define in future versions.

Our aggregation procedure computes weighted point estimates of poverty count, rate, and log-count. These are the three most common point estimate types needed for our proposed SAE models of interest. We also use the 80 replicate weights to compute the variance estimates for each of the three point estimate types, since their sampling variances are often used as inputs to SAE models as well. Again, users can manually substitute the default poverty variable with another target variable of interest, but we plan to make the choice of aggregation variables easier in the future.

We also compute several other weighted statistics used for variance modeling, estimating design effects, etc. Finally, the aggregation module also computes and records the true artificial-population values for poverty count, rate, and log-count, so that we can use them as benchmarks for our SAE models.

One concern is that when estimated counts are 0, the standard replicate-weight-based method gives incorrect variance estimates of 0. However, good variance estimates are necessary as inputs for some SAE models. We intend to implement the ACS' approach in the near future and, eventually, to supplement it with better variance estimates for this kind of situation: see Gilary, Maples and Slud (2012) and Slud (2012).

3.5 Evaluation

Once an artificial population is generated and many samples are drawn from it, researchers can apply their SAE models to each sample and calculate various estimates from each sample, such as point estimates, standard error estimates, and confidence or credible intervals (CIs).

One useful evaluation is to consider each small area separately and examine the properties of these estimates across all the samples: What are the approximate bias, variance, and MSE of the model's point estimates? What is the bias of the model's standard error estimates? What is the coverage of the model's CIs? Then these properties can be compared across the various proposed models. For example, perhaps Model 1 tends to have less-biased point estimates than Model 2, but Model 2's CI coverage is closer to nominal. Comparison of such properties can help a statistical agency decide which SAE model to use in production.

4. Conclusion

Our objective has been to create a simple-to-use, well-documented simulation module that can easily be updated from year to year with new ACS data. Already, we have successfully used the design-based simulation procedure described in this paper to generate the inputs to research by Basel and Taciak (forthcoming). An earlier version of our procedure was also used by Wieczorek et al. (2012) to evaluate a particular SAE model.

As future versions of the artificial population will be designed to resemble the current population of the U.S. in demographic makeup and survey responses, the module could serve to evaluate statistical and methodological questions related to other U.S. household surveys as well. Different assumptions could be implemented regarding the creation of the artificial population. Likewise, different unit nonresponse and mode selection mechanisms could be implemented and their impact could be assessed. Moreover, the sensitivity to item nonresponse and imputation methods could also be assessed.

In particular, we are currently pursuing an improved artificial population which resembles that of the U.S. by attempting to:

- augment the survey data to cover all HUs in the master frame by intelligently imputing the nonsampled HUs with data from sampled ones
- respect true sizes of domains
- represent marginal distributions and interactions between variables
- retain homogeneities within small domains, and
- avoid imputing implausible zero counts to the population merely because the underlying sample count was zero.

Although our artificial population may reuse some of the same variables of interest for different surveys, response mode and nonresponse will depend on the survey to be studied. We are also developing an improved ACS unit nonresponse model that does not assume missing completely at random (MCAR). Logistic regression models for response mode and nonresponse propensities in the ACS are being estimated from the unit-level ACS data.

Other Census Bureau researchers have plans to use the simulation procedure as we continue to improve it. Apart from testing SAE models, it can also be used for projects such as evaluating confidence intervals, as in Franco (2011), or examining the effects of the ACS weighting adjustments.

REFERENCES

Alfons, A., Templ, M., and Filzmoser, P. (2010), "An Object-Oriented Framework for Statistical Simulation: The `R` package `simFrame`," *Journal of Statistical Software*, 37(3).

Alfons, A., Kraft, S., Templ, M., and Filzmoser, P. (2011), "Simulation of close-to-reality population data for household surveys with application to EU-SILC," *Statistical Methods & Applications*, 20(3), 383-407.

Basel, W., and Taciak, J. (forthcoming), "Small-Domain Estimation with Limitations on the Direct Estimate," U.S. Census Bureau working paper (forthcoming).

Bell, B., Jarmin, R., Maples, J., Navarro, A., Basel, W., McHugh, J., Brault, M., and Bauder, M. (2011), "Business Case: Use of Small Area Methods for More Data Concepts and/or Surveys," U.S. Census Bureau internal document.

Franco, C. (2011), "Report on Confidence Intervals for Proportions in Small Areas," DSSD Results Seminar Series, U.S. Census Bureau.

Gilary, A., Maples, J., and Slud, E. (2012), "Small-Area Confidence Bounds on Small Cell Proportions in Survey Populations," *Proceedings, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.

Schaefer, J., and Kang, J. (2008), "Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example," *Psychological Methods*, 13(4), 279-313.

Schirm, A., and Zaslavsky, A. (1997), "Reweighting Households to Develop Microsimulation Estimates for States," *Proceedings, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.

Slud, E. (2012), "Assessment of Zeroes in Survey-Estimated Tables via Small-Area Confidence Bounds," *Journal of the Indian Society of Agricultural Statistics*, 66(1), 157-169.

U.S. Census Bureau (2009), "Design and Methodology: American Community Survey," Washington, DC: U.S. Government Printing Office.

U.S. Census Bureau (2012), "Geographic Terms and Concepts - Census Tract," http://www.census.gov/geo/reference/gtc/gtc_ct.html

U.S. Census Bureau (2013), "Initial Addresses and Sample Selected and Final Interviews," http://www.census.gov/acs/www/methodology/sample_size_data/

Wieczorek, J., Nugent, C., and Hawala, S. (2012), "A Bayesian Zero-One Inflated Beta Model for Small Area Shrinkage Estimation," *Proceedings, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.