CBECS 2018 Misclassification: Issue and Solutions

Frost Hubbard[1], Jim Green[1], David Morganstein[1]
Katie Lewis[2], Michelle Amsbary[1]
[1]Westat, 1600 Research Blvd., Rockville, MD 20850
[2]U.S. Energy Information Administration, 1000 Independence Ave., SW, Washington, DC 20585

Abstract
The 2018 Commercial Buildings Energy Consumption Survey (CBECS) aimed to efficiently oversample buildings greater than 200,000 square feet due to their greater energy consumption. Previous CBECS cycles found building size was correlated with key statistics of interest such as total annual fuel use. In 2018 and previous cycles, CBECS used a multi-frame design (i.e., multiple existing building lists combined with a multi-stage area probability frame) to meet this objective. Thus, the sample design used a building-size categorical variable to help create the relatively homogenous sampling strata for the building-level selection stage. Not unexpectedly, the categorical frame variables used to create the sampling strata had significant classification error. This research explores how misclassification rates can vary by frame in an establishment survey context and the negative effects on precision of the sample. Specifically, we compare the building-size category assigned on the frame to the building-size category reported by the respondent. Based on these results, we suggest two competing design modifications that may generate a more efficient design in future CBECS cycles.

Key Words: sampling frame, stratum misclassification, commercial, buildings, CBECS

1. Introduction

In disproportionate sampling allocations, such as a Neyman allocation (Lohr, 2019), the precision gains from stratification are maximized when all frame units are correctly classified to strata. Stratum misclassification (where the stratum reported by respondents is not the same as the assigned stratum on the sampling frame) leads to increased variation in the base weights (inverse of the probability of selection weights) of cases within each stratum. If there is perfect agreement between the stratum assigned on the sampling frame and the respondent-reported stratum, then all cases within each stratum have the same base weight and there is no variation in base weights among cases within stratum. Stratum misclassification leads to larger variation in the weights, which in turn lowers the precision of the key estimates. A multi-frame sample design can potentially exacerbate the stratum misclassification issue, as different frames can contain different levels of misclassification on the stratification variables and attempting to account for and/or reduce the misclassification across multiple frames can bring increased costs. This paper examines the misclassification rates in each of the six separate sampling frames in one of the stratification variables in the building-level selection process of the 2018 Commercial Buildings Energy Consumption Survey (CBECS), building-size category. It then explores two potential solutions for misclassification in this multi-frame study and addresses the positive and negative aspects of each solution.

2. 2018 CBECS Frame Construction and Sample Design

CBECS is a nationally representative, multi-frame, repeated cross-sectional survey of commercial buildings in the United States sponsored by the Energy Information Administration (EIA). The first CBECS cycle occurred in 1979 and since then new, cross-sectional cycles have been administered every 3 to 6 years. For the last several cycles, the CBECS sample design has used a Neyman allocation with total annual major fuel use as the key statistic. Building size (in square feet) and building use were highly correlated with total annual major fuel consumption in CBECS cycles prior to 2018 (including a correlation of 0.695 in the 2012 CBECS) and thus were used as the building-level stratification variables for the 2018 CBECS. The 2018 CBECS contained 24 building level-sampling strata based on six building-size categories and four building-use categories. This paper focuses solely on misclassification within the six building-size categories for measuring the amount of stratum misclassification in the 2018 CBECS and two potential methods for addressing this misclassification within a future CBECS sample design. The six building-size categories are listed below:

Building-Size Categories (in square feet)

- 1,001 – 10,000
- 10,001 – 25,000
- 25,001 – 50,000
- 50,001 – 100,000
- 100,001 – 200,000
- Greater than 200,000 (> 200,000)

With the highest variance of total annual major fuel use occurring in the largest commercial buildings, the Neyman allocation generates a desired sample size for the largest building-size category that exceeds what a three stage area-probability frame can support. Although the CBECS segments are much larger in geographic size than block-level segments typically generated in area-probability designs for household surveys, the CBECS area segments do not contain enough commercial buildings > 200,000 square feet (sqft) to meet the demands of the highly disproportional Neyman allocation. Thus, in addition to generating an area probability frame of commercial buildings, the CBECS also uses five other list frames to identify commercial buildings > 200,000 sqft within the selected primary sampling units (PSUs). The addition of these five list frames allows CBECS to select the desired number of large commercial buildings identified by the Neyman allocation. The five list frames are:

- Airport (based on publically available campus-level data provided by the Federal Aviation Administration)
- Federal (based on publically available building-level data provided by the General Services Administration)
- College/University (based on campus-level data purchased from a vendor)
- Hospital (based on campus-level data purchased from a vendor)
- Common Premises Location or CPL (based on multiple-business building data purchased from a vendor)

Along with using the area probability frame, Westat used an innovative virtual listing system (VLS[1]) to estimate building square footage for the airport, college, and hospital list frames. The lists for these three frames began at the campus level. From each list, a stratified, systematic sample of campuses was selected. Each selected campus was then virtually listed to identify the commercial buildings > 200,000 sqft. Finally, a set of buildings on the selected campuses believed to be > 200,000 sqft were selected to meet the requirements of the Neyman allocation. The federal and CPL frames were acquired at the building level and included estimates of building square footage. These two frames were restricted to buildings estimated to be > 200,000 sqft, so buildings were selected directly from the lists with no virtual listing required.

Table 1 presents the average final, nonresponse-adjusted weight by sampling frame and frame building-size category. This table demonstrates for the 2018 CBECS how building-size category misclassification affects precision as well as the relationship between the magnitude of the misclassification and the reduction in the precision of key estimates within strata. In order to meet the goals of the Neyman allocation, the buildings > 200,000 sqft were heavily oversampled, which helped produce small, average final weights. The > 200,000 sqft column also shows the difference in final average weights by sampling frame for buildings believed to be > 200,000 sqft. The biggest discrepancy is between the Federal and Area frames, where the average final weight of buildings believed to be > 200,000 sqft on the Area frame is 10 times larger than the average final weight of buildings selected from the Federal frame.

Turning the focus to the "Area" row of Table 1, the average final weight approximately increases by a factor of 2 as one moves right to left from larger to smaller frame building-size categories (i.e., 67 -> 133 -> 302->…2,955). The larger the distance between the building-size category on the sampling frame and the respondent-reported[2] building-size category, the greater the increase in the variation of the weights. For example, for a selected building from the CPL frame, if the respondent reported the building-size category as 100,001 to 200,000 sqft, then the mean final weight of the building and the majority of other buildings in the respondent-reported building-size category differ by a factor of 7 (133/19). However, if the respondent reported a building-size category of 1,001 – 10,000 sqft (as did occur for 11 percent of all selected buildings from the CPL frame), the average final weight of the responding CPL building and the large majority of other buildings that

[1] As discussed in Giangrande et al. (2018), an innovation introduced for the 2018 CBECS was the use of virtual listing to generate a list of commercial buildings within the selected segments, rather than using in-person listers. Westat developed a virtual listing system (VLS) that used satellite imagery to help identify commercial buildings and estimate each commercial building's square footage. Virtual listers estimated a commercial building's size in sqft by outlining the building's footprint on the satellite image and estimating the number of floors in the commercial building. Using geographic information system technology, the VLS estimated the square footage of the building's footprint based on the outline captured by the virtual lister and multiplied the estimated square footage of the footprint by the estimated number of floors to generate a total building square footage.

[2] Throughout this paper, the terms "respondent-reported building sizes" and "respondent-reported building size categories" refer to edited respondent-reported values. Not unexpectedly, there is a fair amount of error each CBECS cycle in the initial respondent-reported building size values. Thus, as a standard quality control procedure during post-data collection processing, EIA reviews the initial respondent-reported building size values and edits them as necessary in an effort to reduce measurement error.

reported being between 1,001 and 10,000 sqft would differ by a factor of 155 (2,955/19), leading to a greater decrease in precision.

Table 1: Average Final Weight by Sampling Frame and Frame Building-Size Category

Frame	Frame building size category (in square feet)[3]					
	1,001 – 10,000	10,001 – 25,000	25,001 – 50,000	50,001 – 100,000	100,001 – 200,000	> 200,000
Airport						13
Federal						6
College						9
Hospital						12
CPL						19
Area	2,955	1,553	655	302	133	67

3. Potential Misclassification Solutions and Research Questions

3.1 Misclassification Solutions

Stratum misclassification in disproportionate allocations is not a new phenomenon, and sample design literature discusses two potential solutions to this issue—mathematical programming and two-phase sampling. Mathematical programming (Dantzig 1965; Valliant and Gentle 1997) is defined as the optimum allocation of limited resources among competing priorities, under a set of constraints. In a sample design context, the goal of the mathematical program is either to (a) minimize the overall sample size given the constraints of minimum precision goals by stratum and fixed misclassification rates, or (b) maximize precision by stratum within the constraints of a fixed sample size and misclassification rates. Option (a) will be the mathematical programming objective discussed later in this analysis.

Two-phase sampling was introduced by Jerzy Neyman (1938). In this method, the sample designer obtains auxiliary information from a large sample of units via an inexpensive method. This is called Phase 1. The sampling statistician then uses the auxiliary information to subsample units for more expensive data collection at Phase 2. In the context of CBECS stratum misclassification, the Phase 1 auxiliary information is an indicator of a building > 200,000 sqft and the less expensive Phase 1 method is virtual listing estimates. Given that only the Federal and CPL frames were not virtually listed, two-phase sampling will only be a useful solution if either of those two frames is shown to have relatively high levels of stratum misclassification.

3.2 Research Questions

With these issues in mind, this paper investigates two research questions:

1. Which sampling frame(s) generated the highest building-size misclassification rates?
2. Among the two potential options for addressing the loss in precision within stratum due to misclassification—mathematical programming and two-phase sampling—which method or combination of methods appears most useful for a future CBECS sample design?

[3] All but the > 200,000 sqft frame building size category for the five list frames have been shaded in grey because all buildings on each list frame were believed to be > 200,000 sqft prior to data collection and were assigned to the > 200,000 sqft frame building size category.

4. Results

4.1 Misclassification Rates

To address the first research question, "Which sampling frame(s) generated the highest building-size misclassification rates?," we examined misclassification rates by sampling frame. Table 2 presents the weighted percentage of commercial buildings by respondent-reported building-size category and list sampling frame. In an ideal (and highly unlikely) scenario in the context of CBECS, each respondent-reported building size on each of the five list frames would be > 200,000 sqft. The green-shaded cells are the percentage of responding buildings that were correctly classified on the sampling frame in the > 200,000 sqft building-size category. The cells shaded in yellow are the percentage of responding buildings that were classified on the sampling frame as > 200,000 sqft, but the respondent reported were between 100,001 and 200,000 sqft. The yellow-shaded buildings were only off by one size category, minimizing the decrease in precision due to unequal weighting. Ideally, all respondent-reported building-size category values on the five list frames would be greater than 100,000 sqft and most would be greater than 200,000 sqft. With 11 percent of responding buildings selected from the CPL frame reporting a building size between 1,001 and 10,000 sqft, the CPL frame is the major culprit in the building-size misclassification for the 2018 CBECS.

Table 2: Weighted Percentage of Responding Commercial Buildings by List Frame and Respondent-Reported Building-Size Category

List frame	**Respondent-reported building size category (in square feet)**					
	1,001 – 10,000	**10,001- 25,000**	**25,001 – 50,000**	**50,001 – 100,000**	**100,001 – 200,000**	**> 200,000**
Airport	0%	0%	0%	0%	0%	100%
Federal	0%	0%	1%	1%	2%	96%
College	0%	0%	0%	0%	24%	76%
Hospital	0%	0%	0%	12%	4%	84%
CPL	11%	3%	8%	8%	19%	51%

Although the CPL frame was the outlier in terms of misclassification, there were other frames that contained misclassification, but to a lesser degree. Within the hospital list frame, 12 percent of the selected buildings had respondent-reported building sizes between 50,001 and 100,000 sqft. The virtual listing team explained this may have occurred because it was difficult to estimate the number of floors within some oddly shaped hospital buildings. Often, these hospital buildings had different sections that varied widely in the number of floors, and the number of floors in each section was difficult to determine from bird's-eye view or street-view images of the building.

Table 3 presents building-size misclassification rates for the area probability frame. Unlike the list frames, the buildings on the area probability frame could be any size greater than 1,000 sqft to be eligible for CBECS sample selection. Thus, Table 3 contains a row for each of the six frame building-size categories.

In this table, the numbers off of the diagonal represent the weighted misclassification rates by different frame building-size and respondent-reported building-size combinations. The pattern follows the pattern of the virtually listed list frames in Table 1 (i.e., the airport, college, and hospital frames). The cell shading in Table 3 follows the same rules as those in Table 2: green-shaded cells contain the percent of responding buildings assigned to the correct building-size category on the area sampling frame, while the yellow-shaded cells are responding buildings that were off by one building-size category on the area frame. The large majority of the frame building-size categories are correct (cells shaded in green on the diagonal) or off by one size category (cells shaded in yellow), which minimizes the decrease in precision due to misclassification. Interestingly, the misclassification percentages are higher below the diagonal than above the diagonal. This means that a higher percentage of the misclassification occurred due to the virtual listers overestimating the building size than underestimating it.

Table 3: Weighted Percentage of Responding Commercial Buildings by Area Frame Building-Size Category and Respondent-Reported Building-Size Category

Area frame building size category	**Respondent-reported building size category (in square feet)**					
	1,001 – 10,000	**10,001 – 25,000**	**25,001 – 50,000**	**50,001 – 100,000**	**100,001 – 200,000**	**> 200,000**
1,001 – 10,000	93%	7%	<1%	0%	0%	<1%
10,001 – 25,000	26%	60%	14%	<1%	0%	0%
25,001 – 50,000	3%	25%	58%	14%	0%	0%
50,001 – 100,000	1%	3%	24%	62%	10%	0%
100,001 – 200,000	<1%	4%	3%	32%	56%	5%
> 200,000	0%	<1%	1%	4%	32%	63%

4.2 Mathematical Programming Solution

To address the final research question, "Which method or combination of methods appears most useful for a future CBECS sample design?," we began by developing a mathematical programming approach that would minimize the overall sample size while accounting for building-size category misclassification rates and achieve the effective sample size within each building-size category indicated by the original Neyman allocation. As described in Green (2000), mathematical programming approaches require the following elements in order to obtain an optimal result (i.e., a desired maximum or minimum): an objective function; a set of decision variables; and other parameters and constraints. We defined each of these elements as:

Objective Function: Minimize the overall actual sample size, after accounting for misclassification (n'')

Decision Variables: Actual sample size by frame building-size category (n_h'')

Other Parameters: Frame building-size population counts (N_h); misclassification rates

Constraints: The effective sample size post-misclassification (n_{eff}'') cannot be less than the effective sample size from the original Neyman allocation (n_{eff}'), both overall and by building size class *h*.

We used Kish's (1992) formula to calculate the design effect due to unequal weighting by building size class h due to implementing a disproportionate allocation: $d_{h\_eff} = (\sum(\frac{N_{hd}^2}{n_{hd}})) * \frac{n}{N^2}$

Constraints in formulas:

$$n_{eff}'' >= n_{eff}', \text{ which is equivalent to } n''/d_{eff}'' >= n'/ d_{eff}'$$

$$n_{h\_eff}'' >= n_{h\_eff}', \text{ which is equivalent to } n_h''/ d_{h\_eff}'' >= n_h'/ d_{h\_eff}'$$

Table 4 contains the desired number of completed interviews, the design effect due to unequal weighting, and the effective number of completed interviews, both for the original Neyman allocation and the mathematical programming solution after accounting for the building-size misclassification.

Table 4: Sample Size, Design Effect Due to Disproportionate Allocation, and Effective Sample Size by Frame Building-size Category for Original Neyman Allocation and Mathematical Program Solution Accounting for Building-Size Misclassification

	Building size category						Overall
	1,001 – 10,000	**10,001-25,000**	**25,001 – 50,000**	**50,001 – 100,000**	**100,001 – 200,000**	**>200,000**	
Neyman allocation							
n’	1,550	964	561	557	531	2,297	6,500
d_{eff}’	1.0	1.0	1.0	1.0	1.0	1.6	2.3
n_{eff}’	1,590	964	561	557	531	1,471	2,826
Mathematical program solution for misclassification							
n′	2,160	1,549	565	546	1,110	2,716	8,636
% moved out of n’	7%	40%	42%	38%	18%	40%	N/A
% moved in to n”	21%	29%	58%	64%	56%	4%	N/A
n”	2,537	1,301	785	935	1,378	1,700	8,636
d_{eff}”	1.1	1.3	1.2	1.7	2.1	1.2	2.2
n_{eff}”	2,254	964	656	557	662	1,471	3,857

The upper portion of Table 4 contains the following three statistics, both overall and by building-size category: (1) the baseline Neyman allocation of 6,500 completed interviews prior to accounting for any building-size misclassification; (2) the Kish design effect due to unequal weighting (d_{eff}'); and (3) the effective sample size (n_{eff}'). The d_{eff}' of 1.6 for the >200,000 sqft building-size category, as opposed to the design effect of 1.0 for the other five categories, is driven by the fact that buildings in the largest building-size category are selected from six separate sampling frames that were sampled at much different rates (see Table 1). The buildings in the other frame building-size categories are only selected from the area frame and have the same sampling rate within the category, hence producing a d_{eff}' of 1. The overall d_{eff}' of 2.3 is due to the fact that the goal of the Neyman allocation is not to generate a proportionate allocation, but instead to minimize variance of an estimator for a fixed sample size, assuming the costs per unit are the same across all building-size categories. The Neyman allocation will deviate from proportional allocation when the estimated population standard deviations (S_h) are different across the strata. This deviation from proportionate allocation is the case in CBECS, where the S_h for the key statistic, total annual major fuel use, is much greater for the larger buildings than the smaller ones.

The lower portion of Table 4 displays six statistics, by building size: (1) the allocation generated by the mathematical program according to the frame building-size category (n'); (2) the percent of responding buildings that were in that building-size category on the frame and "moved out" to a different building-size category based on the respondent-reported building size; (3) the percent of responding buildings in that building-size category based on the respondent-reported building size that "moved in" from a different frame building-size category; (4) the allocation generated by the mathematical program according to the respondent-reported building-size category (n''; i.e., after accounting for misclassification); (5) the design effect due to unequal weighting after accounting for misclassification (d_{eff}''); and (6) the effective sample size after accounting for misclassification (n_{eff}'').

The mathematical program solution generates an overall sample size (n = 8,636) that is 33 percent larger than the initial sample size of 6,500 completed interviews.[4] One way to explain what is driving the mathematical program solution is to measure the amount of movement within each building-size category based on classification based on the sampling frame values versus the respondent-reported values. As shown in Table 1, the cases moving into a building-size category have a different final weight than those originally selected in the category, while the cases moving out reduce the sample size within the category and remove cases with similar weights. Forty percent of the 2,716

[4] Note that due to the constraint that the effective number of completed interviews cannot be less after accounting for building-size misclassification and design effect due to unequal weighting in the mathematical programming approach than the effective number of completed interviews from the original Neyman allocation (n_{h_eff}'' >= n_{h_eff}'), the overall number of completed interviews generated by the mathematical program had be larger than that of the Neyman allocation. To help verify that the allocation of 8,636 completed interviews generated by the mathematical program was the minimum value that could be achieved given the constraints stated above, we set extra constraints on the n'' values to see if a lower minimum could be achieved. For example, we added a constraint of n'' for the 1,001 – 10,000 sqft category of 2,400. However, this resulted in generating a new allocation of 8,713, with the decrease of 137 completed interviews allocated to the smallest building size category (2,537 – 2,400) being outnumbered by increases to the number of completed interviews allocated to the 10,001 – 25,000 and 25,001 – 50,000 building size in order to meet the n_{eff}'' of 964 for the 10,001 – 25,000 building size category.

completed interviews assigned as > 200,000 sqft on the sampling frame moved to a smaller category when reclassified with the respondent-reported values. On the other hand, only four percent of the 1,700 completed interviews were cases that moved into the > 200,000 sqft category from any smaller category on the sampling frame. As Tables 2 and 3 show, the largest proportion of cases that moved out were from the Area and CPL frames, the frames with the largest average final weights for cases > 200,000 sqft. This combination of few cases moving in and many cases moving out among cases with the largest weights drove the d_{eff}' down from 1.6 in the Neyman allocation to 1.2 in the mathematical program allocation.

While movement helped reduce the d_{eff}'' for the largest building-size category, the deleterious effects of the movement, especially the movement of buildings from the largest size category to smaller categories, can be seen in the weighting effects greater than one for all other building-size categories, most notably the 100,001 – 200,000 sqft category with a d_{eff}'' = 2.1. As seen in Tables 2 and 3, when buildings in the largest building-size category on the frame are misclassified, they most often move to the adjoining building-size category of 100,001 to 200,000 sqft. These buildings "moving in" to the 100,001 to 200,000 sqft category from the largest building-size category have different sampling rates because they were selected from different sampling frames, adding to the increase in variation of the weights. In addition, as seen in the 50,001 – 100,000 row of Table 3, buildings from this building-size category on the frame are also moving into the 100,001 to 200,000 sqft category, generating even more variation in the weights and producing a weighting effect of 2.1.

4.3 Two-Phase Sampling Solution

Another method to address the building-level misclassification would be to implement a two-phase sample (Neyman, 1934) on the frame with the highest misclassification rates, the CPL frame.[5] While the mathematical programming approach starts from a place of assuming the misclassification rates are fixed, two-phase sampling attempts to reduce the misclassification rates via less expensive data collection methods in the first phase. With only 50 percent of all responding buildings selected from the CPL frame found to be > 200,000 sqft, a two-phase design would select twice as many buildings from the CPL frame as initially believed to be needed after applying the estimated eligibility rates and response rates to the original Neyman allocation solution. Virtual listers would determine which of the selected CPL listings were > 200,000 sqft. Among the CPL selections believed to be > 200,000 sqft, the sampling statisticians would sub-select to get to the desired number of sample cases to complete the full interview.

5. Discussion

A Neyman allocation, like the one used for the 2018 CBECS, does not take into account stratum misclassification. However, misclassification of the frame building-size category, one of the two variables used to generate CBECS building-level sampling strata, is an issue for the multi-frame CBECS sample design. Of the six CBECS sampling frames, the CPL frame is the biggest offender, both in terms of the misclassification rate and the increase in

[5] The Federal list frame could also implement a two-phase sample. However, given that 96 percent of all responding buildings from the Federal list frame were correctly classified on the frame as > 200,000 sqft, the small amount of reduction in misclassification generated by a two-phase design would not likely be worth the extra costs of implementing the first phase of the two-phase design.

relative difference in the weights generated by responding cases “moving” to strata with much different sampling rates. Forty-nine percent of all responding buildings selected from the CPL frame were reported to be less than 200,001 sqft by respondents. In addition, 11 percent of responding buildings from the CPL frame were reported to be in the 1,001 to 10,000 sqft category. Buildings in this smallest building-size category that were correctly assigned to this category on the sampling frame had an average weight that is greater than buildings selected from the CPL frame by a factor of 155. The area frame also experienced a fair amount of misclassification; however, the misclassification was predominantly contained to responding buildings “moving” to contiguous building-size categories whose sampling rates are more similar than non-contiguous building-size categories, helping limit the loss in precision due to misclassification.

A useful finding for future CBECS cycles is that when virtual listers made building-size category misclassifications, they tended to be errors of overestimation. This was not unexpected as virtual listers were told to overestimate the building size when in doubt, to avoid undercovering and potentially missing large buildings. However, the concerns about undercoverage of large buildings did not materialize—only 5 percent of responding buildings selected from the area frame categorized as 100,001 to 200,000 sqft by virtual listers were reported by respondents to be > 200,000 sqft and no responding area frame buildings categorized as less than 100,001 sqft by virtual listers were reported by respondents to be > 200,000 sqft. Given this, the next CBECS sample design could be improved by removing any instructions to overestimate building-size estimates when in doubt.

We discussed two potential solutions to the misclassification issue: mathematical programming and two-phase sampling. The mathematical program assumes that the misclassification is fixed and generates a minimum allocation that meets specified precision levels while accounting for the misclassification and its negative effects on precision. The mathematical program produced an allocation (n = 8,636) that was 33 percent larger than the original Neyman allocation (n = 6,500). On the other hand, the two-phase sample design attempts to reduce the misclassification, while minimizing the increase in costs by using a less expensive data collection method in the first phase. For CBECS, the two-phase sample design would be focused on reducing the misclassification rate in buildings selected from the CPL frame, given that the CPL frame is the key driver of building-size misclassification in CBECS.

Each method has its advantages and disadvantages. The advantage of the mathematical program is it accounts for the misclassification observed across all frames. However, this advantage comes with a 33 percent increase in allocation, a substantial cost that few projects would be eager to absorb. For the two-phase design, the advantages are that (a) it would not require an increase in the allocation of 6,500 completed interviews; (b) it can target the main driver of the CBECS building-size misclassification, the CPL frame; and (c) it is easy to implement at a relatively low cost. The downside is that would only address the misclassification of the CPL frame and none of the other frames.

Given these trade-offs, future CBECS sample designs may want to consider one of the following two approaches. A lower cost approach would be to focus all efforts on reducing misclassification. This would attempt to reduce misclassification by implementing the two-phase design on the CPL frame while also doing further research on the buildings that were misclassified by virtual listers and devising methods to increase accuracy in the building-size estimates of the virtual listing team. A more expensive approach would implement everything from the lower cost approach as well as a mathematical program to help account for the remaining misclassification after reasonable reduction efforts have been exhausted.

Acknowledgements

The authors would like to thank Kelly Daley for her review and the Energy Information Administration for their support of this research.

References

Giangrande, M., Brick, J.M. Morganstein, D., Lewis, K. (2018), “Virtual Listing: GIS Approaches to Improve Survey Listing Efficiency,” In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1554-1560

Dantzig, G. B. (1965), *Linear Programming and Extension*, 48, Princeton University Press.

Green, J. L. (2000), “Mathematical Programming for Sample Design and Allocation Problems,” In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 688-692.

Kish, L. (1992) “Weighting for Unequal Pi.” *Journal of Official Statistics*, 8-2, 83-2000.

Lohr, S. L. (2019), *Sampling: Design and Analysis*, CRC Press.

Neyman, J. (1938), “Contribution to the Theory of Sampling Human Populations,” *Journal of the American Statistical Association*, 33(201), 101-116.

Valliant, R., and Gentle, J. E. (1997), “An Application of Mathematical Programming to Sample Allocation,” *Computational Statistics & Data Analysis*, 25(3), 337-360.