The TACPA Program: A Review of Current Structure and Potential Alternatives
Technical Appendix

This appendix provides a detailed overview of the analysis conducted by the California Research Bureau (CRB). It begins by summarizing the impediments to unbiased standard error estimation with U.S. Census Bureau data and the equations CRB used to identify upper and lower bounds of the standard errors for indicators of the eight Target Area Contract Preference Act (TACPA) criteria. We then review the structure of the Monte Carlo simulation analysis used to estimate misclassification rates and present a more detailed discussion of its results.

Standard Error Estimation

The Census Bureau provides methods to estimate the standard errors associated with counts and proportions calculated from data included in Summary Files. These approximations are known to be biased whenever multiple cells are combined to create aggregate counts. The process for generating measures of the eight TACPA conditions often requires some degree of aggregation. As a result, the standard error estimates surrounding the TACPA indicators produced by the approximation methods outlined by the Census Bureau can be either too small or too large.

For example, the Census Bureau reported the unemployed population separately for males and females in 2000. To calculate the unemployment rate, analysts need to combine the contents of two cells, adding the number of unemployed females to the number of unemployed males. In 2010, the Census Bureau reported unemployment numbers by gender and age, meaning analysts have to add up 14 separate cells to measure the unemployment rate. Summing multiple cells does not affect the validity of the generated totals, but it does lead to bias in the standard error approximations provided by the Census Bureau.

The Census Bureau suggests that these issues can be avoided "by creating estimates and standard errors using the Public Use Microdata sample (PUMS) or by requesting a custom tabulation, a fee-based service offered under certain conditions by the Census Bureau."[3] The PUMS data do not include geographic identifiers for any region that contains fewer than 100,000 people, meaning that such data could not provide the necessary information at the block-group level.

Without a custom table from the Census Bureau, then, any attempts to estimate the degree of sampling error in the data or identify its potential effect are limited. To address this issue, CRB estimated the standard errors associated with the seven percentage-based TACPA indicators in two ways. The first method relies on information from the aggregated percentage, ignoring the fact that multiple cells were summed:

$$SE(\hat{p}) = \sqrt{\left(\frac{k}{\widehat{denominator}}\right)\hat{p}(100-\hat{p})} \qquad (1)$$

where $\hat{p}$ is the estimated percentage, k is the inverse of the general sampling rate minus one, and $\widehat{denominator}$ is the base or denominator of the estimated percentage.[1] Equation (1) likely underestimates the standard error of given percentages because it ignores information on the standard errors of individual components used to construct the estimated percentage and because it assumes a constant sampling rate that is not actually found in the data.[*]

The second approach uses information on the standard errors of each component of the percentage.[3] First, the standard errors associated with the aggregated numerator and denominator are estimated:

$$SE(\widehat{numerator}) = \sqrt{\sum SE(\widehat{components}_{num})^2}$$
$$SE(\widehat{denominator}) = \sqrt{\sum SE(\widehat{components}_{den})^2} \qquad (2)$$

where $SE(\widehat{numerator})$ and $SE(\widehat{denominator})$ are the standard errors of the numerator and denominator used in the percentage and $SE(\widehat{components}_{num})$ and $SE(\widehat{components}_{den})$ are the standard errors of each component in the numerator and denominator. These estimates are then used to calculate the standard error of the percentage:

$$SE(\hat{p}) = \frac{\sqrt{SE(\widehat{numerator})^2 - \left(\hat{p}^2 \times SE(\widehat{denominator})^2\right)}}{\widehat{denominator}} \qquad (3)$$

The approximation method outlined in Equations (2) and (3) is biased in two ways. First, Equation (2) will overestimate the standard error of aggregate counts if the estimated counts being summed contain multiple zeros.[3†] Second, Equation (2) will underestimate or overestimate the standard error of the aggregate count if the component estimates are positively or negatively correlated.[3] We treat the estimated standard errors from this approach as an upper bound, but note that these figures could still underestimate the degree of sampling error, depending on the unknown relationships between the components of our aggregated counts.

[*] The 2000 Census long form survey was designed to sample approximately 17% of the population. The mean sampling rate across California block groups was actually 13%, with a standard deviation of 5%. The 5-year estimates included in the 2006-2010 ACS were intended to include about 13% of the population, but the mean sampling rate across California block groups is closer to 7%, with a standard deviation of 2%. Design factors are often used to address this issue, but they are unavailable for the 2010 data. Because our goal is to compare sampling error in 2000 to 2010, CRB relied on the unadjusted standard errors, without the use of design factors.

[†] CRB minimizes this issue by following the Census Bureau's suggestion to only sum one of the standard errors associated with the zero estimates when multiple zeros are present, though, as they note, some bias still remains.

Standard Errors

Table 1 presents the mean estimated standard errors for each of the eight criteria using data from 2000 and 2010. Estimates in the "Low" columns are based on Equation (1) while estimates in the "High" column are based on Equations (2) and (3).

A few of features of this table are worth noting. First, there is considerable variation between the lower and upper bound estimates, particularly in 2010. Without a custom table created by the Census Bureau, it is impossible to know which estimates are a better reflection of the true sampling error associated with the eight TACPA criteria. Second, there are no standard error estimates for per capita income in 2000 (because the necessary data are unavailable at the block-group level) and the upper and lower estimates are identical in 2010 (because only one approximation method is provided by the Census Bureau). Third, regardless of the approximation method, the standard errors associated with the eight TACPA criteria are larger in 2010 than they were in 2000.

Table 1: Mean Standard Errors Associated with the Eight TACPA Conditions

Block groups:	2000		2010	
	Low	High	Low	High
Percentage of population over 25 with less than high school degree	2.98	7.84	3.04	12.92
Percentage of civilian labor force who are unemployed	2.40	5.79	2.82	7.27
Per capita income	.	.	4,811	4,811
Percentage of families with children, headed by a female, in poverty	2.66	8.06	3.04	23.19
Percentage of population over 65 in poverty	5.52	24.35	.	.
Percentage of households with more than 1.01 persons per room	3.45	9.48	3.09	20.74
Percentage of population under 18 in poverty	4.10	12.60	.	.
Percentage of population who are nonwhite or Hispanic	2.54	4.34	2.83	16.12

Coefficients of Variation

It is difficult to put these values into perspective and to assess their potential effect on TACPA eligibility determination. In the *Briefly Stated*, CRB presents results from simulation analysis designed to estimate the number of block groups that could be misclassified under different levels of error. Another approach might consider the coefficient of variation (CV) associated with the TACPA indicators. CVs are a standardized indicator of reliability based on the ratio of an estimate's standard error to itself, expressed as a percent.[2] They are used by the U.S. Census Bureau to assess the reliability of estimates. While the Census Bureau does not provide an explicit rule for determining if a set of estimates is reliable enough to be used by researchers, it uses at least two thresholds in their publications.

When deciding if data are reliable enough to warrant inclusion in its publicly-available datasets, the Census Bureau calculates the CV for each estimate in a geographic area and publishes those with a median CV below 61%.[4] Alternatively, when providing case studies of ACS applications in their handbook for state and local governments, the Census Bureau suggests that "estimates with CVs of more than 15% are considered cause for caution when interpreting patterns in the data."[3]

Table 2 presents the median CV for the eight TACPA criteria, using the two approximation methods previously outlined. In 2010, as many as five of the available six indicators may fail to reach the thresholds employed by the U.S. Census Bureau. While these results do not indicate that the TACPA indicators are highly reliable, readers should not overreact to Table 2. These CVs are estimates based on approximation methods that are known to be biased. Without standard error estimates from a custom table produced by the Census Bureau, strong conclusions should not be drawn. However, users of these data should be aware of the relatively large standard errors.

Table 2: Median Coefficient of Variation Associated with TACPA Criteria

Block groups:	2000		2010	
	Low	High	Low	High
Percentage of population over 25 with less than high school degree	16%	36%	22%	75%
Percentage of civilian labor force who are unemployed	35%	61%	33%	57%
Per capita income	.	.	14%	14%
Percentage of families with children, headed by a female, in poverty	90%	100%	100%	100%
Percentage of population over 65 in poverty	77%	100%	.	.
Percentage of households with more than 1.01 persons per room	31%	72%	57%	110%
Percentage of population under 18 in poverty	34%	82%	.	.
Percentage of population who are nonwhite or Hispanic	6%	8%	6%	29%

Estimating Misclassification Rates with Simulations

Another reason that the results in Tables 1 and 2 may exaggerate the issue of sampling error is that, by relying on multiple indicators of economic distress, the current eligibly process tends to minimize the effect any one unreliable measure may have on the process. Assessing the reliability of the indicators individually provides only a limited understanding of the eligibility determination process's reliability as a whole. This is one reason CRB pursued an analytical strategy based on Monte Carlo simulations, which were designed to provide a close approximation of real-world data while simplifying the structure enough so that useful conclusions could be drawn. In that effort, CRB began with the assumption that the 2000 Census long form survey data were accurate and that standard error estimates reflect the true amount of sampling error likely found in the 2000 survey data.

Relationships in Census 2000 Data

Using a pre-existing data source to structure the simulation is a useful approach in this instance because it preserves some of the relationships between the indicators and the error surrounding them. For example, Table 3 presents results that suggest that the TACPA indicators and their standard errors are strongly correlated.

The average correlation between the 8 indicators is .5 and the scale reliability, a measure of the internal consistency of the items, is .89. Factor analysis on the eight indicators yields a single Eigenvalue above 1 and relatively high factor loadings for each criteria (>.5). These results suggest that indicators of the eight TACPA criteria tap into a similar underlying concept. This is good news since TACPA's eligibility

determination process is built on the notion that the eight indicators are related and measure a block group's level of economic distress.

The standard errors associated with the TACPA indicators, which are estimates of the magnitude of sampling error, are correlated as well. The primary determinant of an estimate's standard error is the size of the sample used to construct it. Since estimates of all eight indicators come from the same sample, it is not surprising that the average correlation between the errors is .63 or that the scale reliability is .92.

Table 3: Relationships across Indicators and Their Standard Errors

	Indicators	Standard Errors
Average Inter-item Correlation	0.50	0.63
Scale Reliability	0.89	0.92
Factor Loadings		
Criteria 1	0.91	0.89
Criteria 2	0.70	0.80
Criteria 3	-0.74	.
Criteria 4	0.72	0.84
Criteria 5	0.51	0.50
Criteria 6	0.85	0.85
Criteria 7	0.81	0.56
Criteria 8	0.82	0.88

Note: Estimates of the average inter-item correlation, Cronbach's alpha scale reliability statistic, and factor loadings from principal-component factor analysis for indicators and estimated standard errors of the eight TACPA criteria. The standard error for per capita income, criterion 3, cannot be estimated at the block-group level given the data available for the 2000 Census long form survey.

Because we base the Monte Carlo simulations that follow on the Census 2000 long form data, we are able to maintain these relationships. Specifically, the average inter-item correlation between the eight TACPA indicators is held to the .50 observed in the Census 2000 data. The average inter-item correlation between the standard errors is held around .63, which reflects the fact that block groups with a noisy and imprecise measure on one indicator are likely to have noisy and imprecise measures on the other seven as well.

Not every important relationship can be preserved, however. By using the Census 2000 data to structure the simulation, the relationships between the magnitudes of sampling error across items are maintained, but relationships between the directions of error are not. For example, if Census Bureau estimates of the unemployment rate for a given block group are too low, estimates of poverty are likely low as well. We only have access to estimates of the TACPA criteria so we cannot know the extent of this problem. Because the indicators and their standard errors are strongly correlated and the data used to measure all eight come from the same source, we have every reason to suspect that the errors are correlated across the eight indicators. To capture this issue and its effect on eligibility determination, CRB considered a range of correlations between block group error across items in the simulations presented in the *Briefly Stated* and this appendix.

The second important relationship not accounted for by relying on the Census 2000 data is the relationship between sampling error and missing data. In some cases, the Census Bureau will fail to interview members from all relevant subcategories in a block group when collecting its survey data. In the simulation that follows, CRB combines information on each block group's total population, the subpopulation associated with each of the TACPA criteria, and the observed sampling rate to simulate data that reflect the potential for missing data.

Monte Carlo Simulation Setup

The analysis proceeds by (1) generating simulated indicators for each of the eight criteria based on the block group's true value as well as information about the likely level of sampling error, (2) determining eligibility for the TACPA program using these new estimates for each of the eight criteria, and (3) comparing these results to the eligibility as determined using the Census 2000 long form survey data. The first step assigns simulated values to each of the eight criteria:

$$\widehat{item}_{ij} = item_{ij} + \alpha\left[error_{ij} + \beta\left(bias_j\right)\right] \qquad (4)$$

where:

- i indexes items and j indexes block groups
- $\widehat{item}_{ij}$ is the simulated value assigned to the i^{th} item for the j^{th} block group
- $item_{ij}$ is the observed value in the 2000 Census long form survey data
- $error_{ij}$ is distributed $N(0, SE_{ij})$ such that SE_{ij} is the standard error approximated using the Census 2000 data, reflecting randomly distributed sampling error for each item/block group
- $bias_j$ is distributed $N(0, 1)$ such that it varies across block groups but is constant across items, reflecting the possibility that the sampling error found on the eight indicators are related
- α is a parameter, varied by CRB, that determines the magnitude of error found in the simulated values
- β is a parameter, also varied by CRB, that shapes the degree to which sampling error is correlated across items

As the α parameter increases, the reliability of the simulated data ($\widehat{item}_{ij}$) decreases, allowing CRB to estimate the misclassification rate under different levels of sampling error. As the β parameter increases, the correlation of errors across the eight items increases, meaning, for example, that a block group with underestimated unemployment is likely to also have underestimated poverty.[*]

The next step of the simulation is to assign missing data to some block groups in a manner that reflects real world issues associated with survey implementation. Imagine a block group with a population of 1000, 10 of whom are over the age of 65. If the Census Bureau takes a 10% sample of all individuals on that block group, it is possible that no one over the age of 65 will be included. Specifically, there is a 37% chance that no one over the age of 65 would be sampled in this example block group and it would be

[*] To ensure that increasing the β parameter does not also increase the amount of sampling error in the simulated dataset, the entire $error_{ij} + \beta\left(bias_j\right)$ portion of the equation is standardized to have a mean of zero and standard deviation of one before multiplying by α. Also, to ensure that the β parameter has an equivalent effect across all eight of the criteria, per capita income is recoded to range from high to low values, making it positively correlated with the other seven items.

impossible to estimate the poverty rate of individuals over 65. CRB can estimate the likelihood that certain subpopulations will be included in the Census sample by using data provided by the Census Bureau on the subgroup's prevalence in the block group's overall population and the number of total individuals sampled. The probability with which a given $\widehat{item}_{ij}$ is assigned a value of missing is determined as follows:

$$p(\widehat{item}_{ij} = NA) = \left(1 - \frac{\widehat{denominator}_j}{population_j}\right)^{\gamma(n_j)} \qquad (5)$$

where $\widehat{denominator}_j$ is the denominator used to construct the given indicator, $population_j$ is the 100% count of the block group's true population or number of households, n_j is the number of individuals/ households sampled in 2000, and γ is a parameter that CRB uses to decrease the sample size to reflect the ACS sampling rate.[*]

Before turning to the results, we summarize the steps involved in the simulation:

1. Generate values of $\widehat{item}_{ij}$ for each block group/item using Equation (4).
2. Assign block groups values of "missing" on some items using probabilities outlined in Equation (5).
3. Calculate eligibility using the current rule and the alternatives summarized in the *Briefly Stated*.
4. Save misclassification rates as well as information on the relationships between indicators and error.
5. Repeat steps 1-4 100 times.
6. Repeat steps 1-5, varying α and β to range from 0 to 10 and γ to range between .5 and 1.[†]

Simulation Results

The results of these simulations are presented in Figures 1, 2, and 3 of this appendix and in Table 3 of the *Briefly Stated*. Figure 1 plots the misclassification rate associated with the current "*5/8 at the Block-Group Level*" rule against the average reliability of the eight TACPA indicators, measured as the squared correlation between $\widehat{item}_{ij}$ and $item_{ij}$. Moving from left to right, the reliability of the indicators increases as the amount of sampling error decreases. When the reliability equals one, there is no sampling error and the simulated measures of the TACPA criteria are identical to those observed in the 2000 data. The blue bar reflects a range of potential misclassification rates, which depend on the assumed relationship between item-specific errors. The bottom of the bar, associated with the lowest misclassification rate, is based on the assumption that there is no relationship, that the errors are not

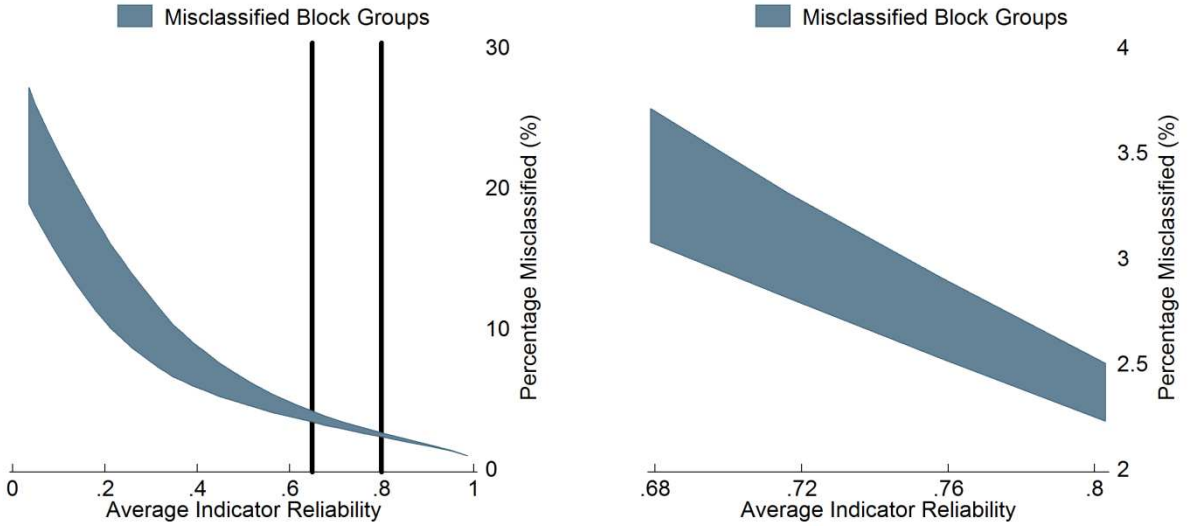
[*] This is simply the probability density function for a binomial distribution where the number of successes is zero, reflecting the probability that zero members of a subgroup are sampled from a given population.

[†] By varying α from 0 to 10, we alter the level of overall random sampling error. When α equals 0, there is no error, the eight TACPA criteria are measured perfectly, and the average indicator reliability is 1.0. As α approaches 10, the degree of error increases and the average reliability of the indicators approaches 0. The prospect of correlated errors is accounted for by varying the β parameter. When $\beta$is 0, the average correlation across iteration-specific errors is 0, reflecting the unlikely scenario that error on one item is not related to error on others. As β reaches a value of 10, the average correlation across errors is .6. Because we cannot know with certainty how correlated the errors are in the 2000 or 2010 data, CRB chose .6 as a reasonable middle ground between the average correlation between the indicators (.5) and standard errors (.63). To the extent that the correlation between the errors is higher than .6, the analysis presented here underestimates the misclassification rate. Finally, we vary γ from .5 to 1 to reflect the fact that the observed sample size in 2010 is about half that observed in 2000.

correlated across items. The top of the bar, where the misclassification rate is highest, represents an assumption that the errors are moderately correlated, with an average interitem correlation of .6.

Figure 1(a) presents the misclassification rate across a wide range of reliabilities, ranging almost from 0 to 1. The width of the blue bar declines as the reliability increases because CRB's assumption regarding the correlation of errors across items means less as the total amount of error decreases. Overall, the misclassification rate ranges from 0% to about 27%. Using the low and high estimates of the standard errors associated with the TACPA indicators, we can focus on a range of reliabilities likely to be found in the 2000 Census long form survey data. The vertical black lines in Figure 1(a) identify this range and Figure 1(b) narrows in on this region of the plot. Figure 1(b) is identical to Figure 1(a), except for the scales of the y- and x-axis. CRB estimates that the overall misclassification rate due to sampling error in 2000 is about 2-4%.

Figure 1: Estimated Misclassification Rate Associated with the Current TACPA Rule
Based on Simulation Analyses



(a) Full Reliability Range

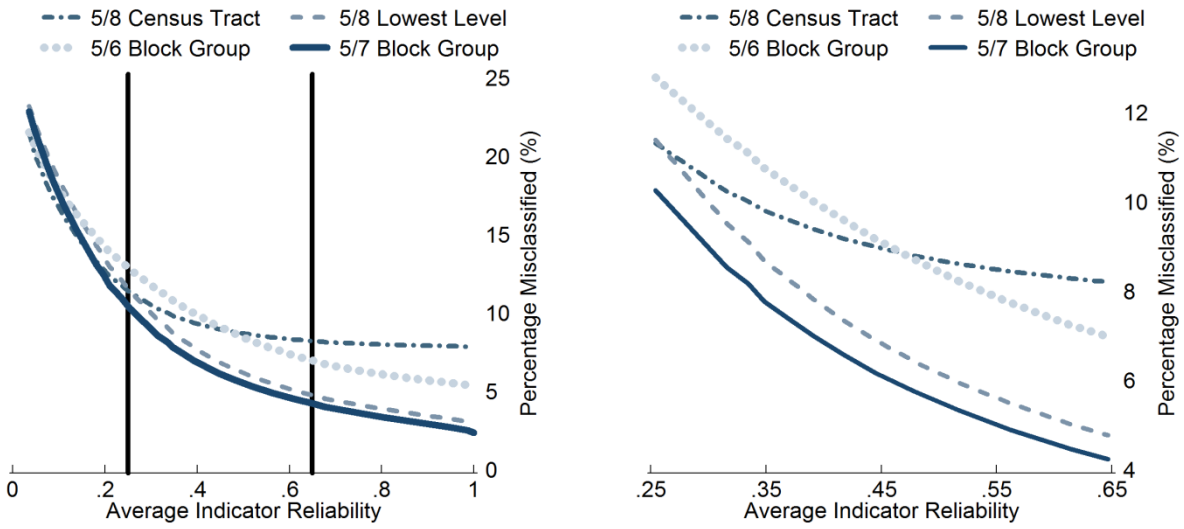
(b) Likely Reliability Range in 2000

Figure 2 plots the misclassification rates associated with the four alternative rules against the average reliability of the TACPA indicators. Instead of plotting a range associated with each level of reliability as in Figure 1, we include a single estimate based on the assumption that the average correlation between errors among the TACPA criteria is .5. The patterns observed in Figure 2 do not change when other assumptions are made.

Figure 2(a) includes a wide reliability range on the x-axis, while Figure 2(b) narrows in on the likely reliability found in the 2010 ACS data. Note that the range of likely reliability is lower (reflecting more sampling error) and wider (reflecting more uncertainty in the level of sampling error) in Figure 2 than in Figure 1. As is mentioned in the *Briefly Stated*, the "*5/7 at Block-Group Level*" has the lowest misclassification rate under most reliability levels, certainly among those expected in the 2010 ACS data. At the very lowest reliability, however, the "*5/8 at Census-Tract Level*" misclassifies the fewest block

groups. Any determination process conducted at the census-tract level will be less sensitive to changes in reliability than those conducted at the block-group level because census-tract estimates are based on a larger number of survey respondents.

Figure 2: Estimated Misclassification Rate Associated with Alternative TACPA Rules Based on Simulation Analyses



(a) Full Reliability Range

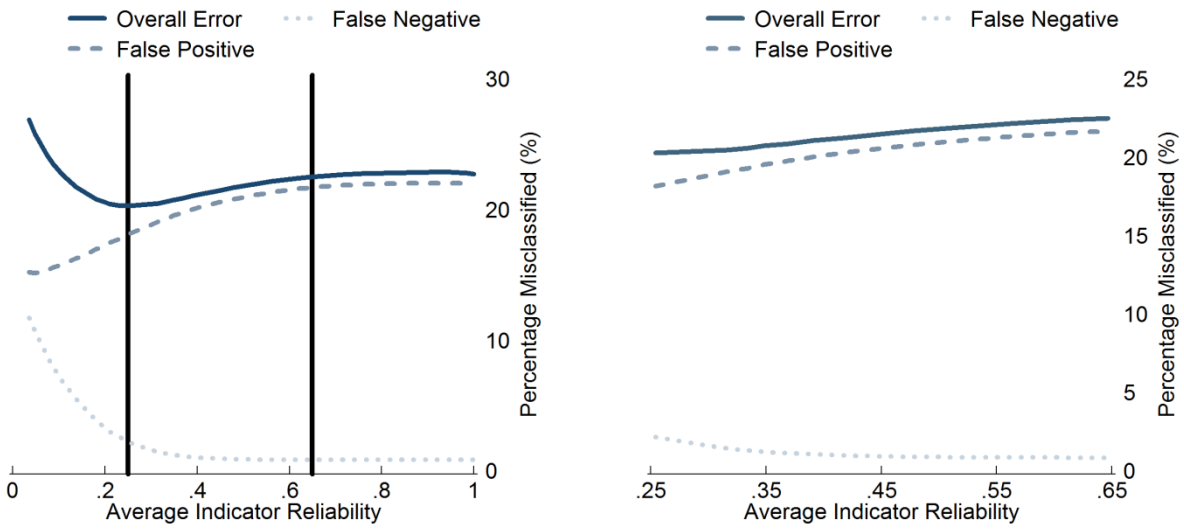
(b) Likely Reliability Range in 2010

Finally, Figure 3 plots misclassification rates associated with a significance-based rule that requires statistically significant evidence, at the 90% level,* that block groups are not eligible for TACPA before labeling them as such. As discussed in the *Briefly Stated*, this approach minimizes the risk of incorrectly classifying distressed block groups as ineligible. To illustrate the consequences of such as approach, Figure 3 breaks the overall misclassification rate into two types. The first, labeled *False Negative*,

* Statisticians use confidence levels to describe the amount of uncertainty associated with a sample estimate of a population parameter. Each estimate is uncertain and contains a margin of error, the size of which is determined by our choice of a confidence level. For example, the Census Bureau reports 90% margins of error with all ACS data. Researchers balance the likelihood of committing Type I (false positive) and Type II (false negative) error when choosing a confidence level. Type I error refers to situations where we incorrectly reject a null hypothesis when it is, in fact, true. Type II error, on the other hand, occurs when we fail to reject a null hypothesis when an alternative is true. The two are inversely related: as the rate of Type I error decreases, the rate of Type II increases. Type I error in this situation occurs whenever a block group that truly meets a given criterion, is judged to be outside the upper quartile on that indicator. Type II error occurs when a block group that does not actually meet a given criterion is judged to be in the upper quartile. By setting a confidence level, we decide how willing we are to commit Type I error. A 90% confidence level in this application means that approximately 10% of block groups that actually meet a given criterion will be incorrectly determined to be outside of the upper quartile on that item. The analysis presented here relies on a 90% confidence level, a choice that is relatively arbitrary but consistent with common practice. The confidence level could be altered depending on the goals of the program. In general, a high confidence level will limit the number of block groups that should be eligible but are judged ineligible (i.e. false negatives). However, a high confidence level will also increase the total number of "distressed" block groups and the number of block groups that are classified as distressed even though they are not (false positives).

represents block groups that are distressed but incorrectly classified as ineligible. The second, labeled *False Positive*, occurs when nondistressed block groups are classified as eligible. Figure 3(a) plots the rate of these types of error for a wide range of reliabilities. It is only when the average reliability of the TACPA indicators dips below .2 that more than 1-2% of distressed block groups are mistakenly classified as ineligible. And as Figure 3(b) demonstrates, false negatives never exceed 2% across the range of likely reliability observed in the 2010 ACS data. The negative consequence of the significance-based approach is that a large number of nondistressed block groups incorrectly receive TACPA eligibility.

Figure 3: Estimated Misclassification Rates Associated with Significance-Based Rule, Based on Simulation Analyses



(a) Full Reliability Range

(b) Likely Reliability Range in 2010

Works Cited

1. U.S. Census Bureau. 2002. *Census 2000 Summary File 3: Technical Documentation.* U.S. Census Bureau, Washington, DC.

2. U.S. Census Bureau. 2009. *A Compass for Understanding and Using American Community Survey Data: What State and Local Governments Need to Know*. U.S. Government Printing Office, Washington, DC.

3. U.S. Census Bureau. 2011. *American Community Survey: Multiyear Accuracy of the Data*. American Community Survey Office, Washington, DC.

4. U.S. Census Bureau. 2011. *The 2006-2010 ACS 5-Year Summary File: Technical Documentation*. American Community Survey Office, Washington, DC.