Adaptive Survey Design with Multiple Criteria: the American Community Survey

Stephanie Coffey, Principal Statistician

Jonathan Eggleston, Senior Economist | Kendall Houghton, Economist | Carl Lieberman, Economist

Federal Committee on Statistical Methodology Conference

October 24th, 2023

United States Census Bureau

Motivation for Adaptive Survey Design in the ACS

- Information collected in the ACS is critical
  - Largest continuous household survey in the US
  - Collects variety of information on household- and population-based topics
- Large overall sample sizes (3.5 million housing units per year)
  - Sequential multimode survey design to control costs
  - About 20% of ACS sample still ends up CAPI mode
- In the past (e.g., in 2022) data collection budgets were exhausted
  - Data collection was stopped for all CAPI cases – no targeting
  - No chance to change the respondent set at time of work stoppage
  - Could exacerbate lead to nonresponse bias

- Goal: Create a Quality/Data-Driven Tool for Reallocating Effort

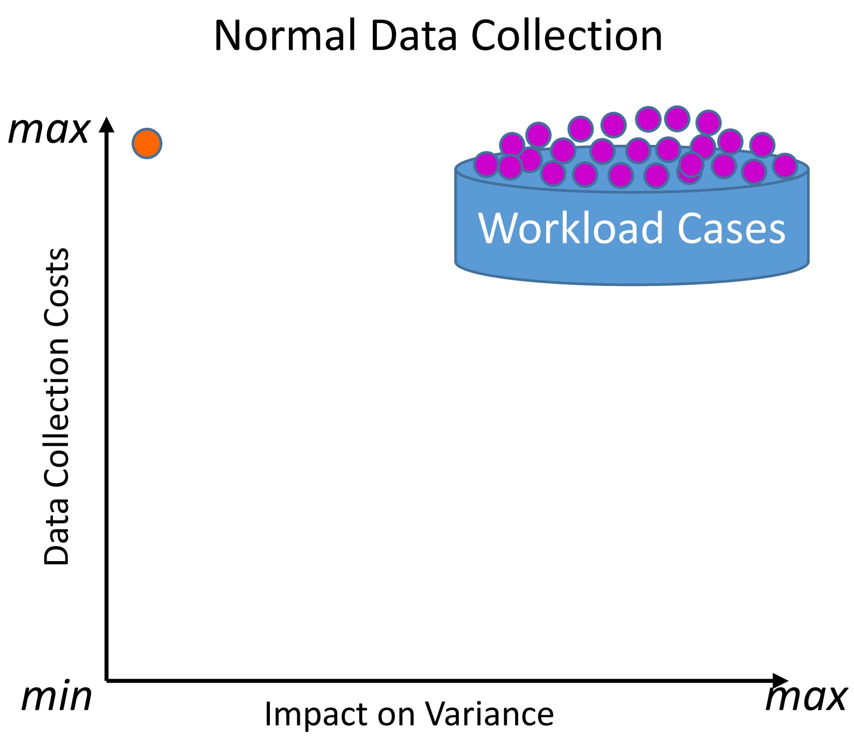
Adaptive Survey Design

- Framework for Data Collection
  - Leverage tradeoffs
  - Min(budget) for a fixed level of data quality
  - Max(data quality) for a fixed budget
  - Balance resources and quality
    If we can save resources, willing to give up "some" data quality

United States® Census Bureau

Adaptive Survey Design

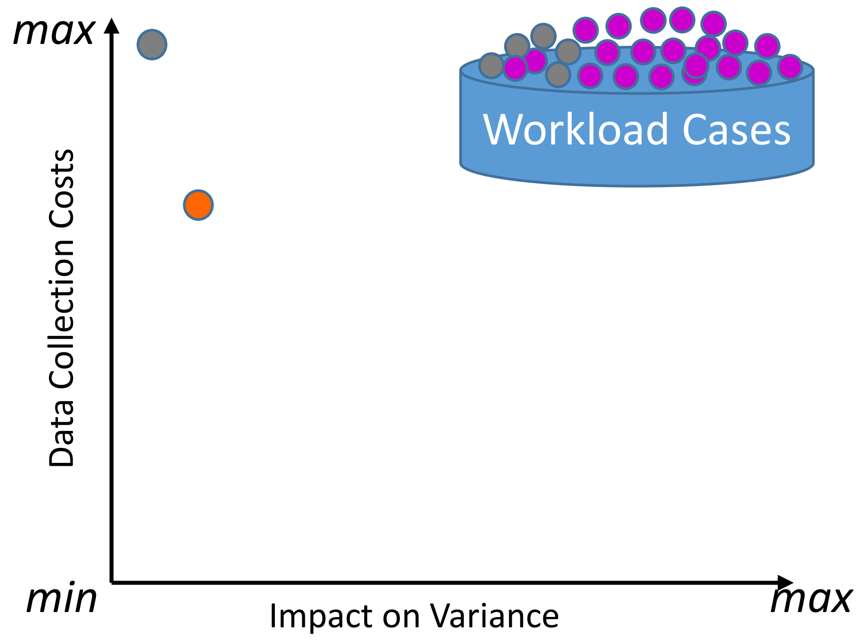
- Framework for Data Collection
  - Leverage tradeoffs
  - Min(budget) for a fixed level of data quality
  - Max(data quality) for a fixed budget
  - Balance resources and quality
    If we can save resources, willing to give up "some" data quality



Adaptive Survey Design

- Framework for Data Collection
  - Leverage tradeoffs
  - Min(budget) for a fixed level of data quality
  - Max(data quality) for a fixed budget
  - Balance resources and quality
    If we can save resources, willing to give up
    “some” data quality

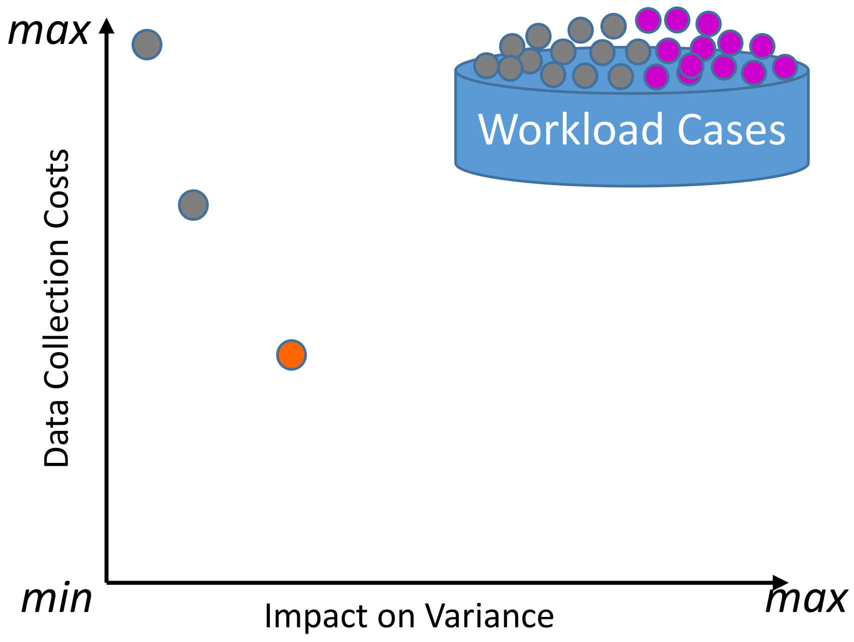
Stop “Some” of the Cases



Adaptive Survey Design

- Framework for Data Collection
  - Leverage tradeoffs
  - Min(budget) for a fixed level of data quality
  - Max(data quality) for a fixed budget
  - Balance resources and quality
    If we can save resources, willing to give up “some” data quality

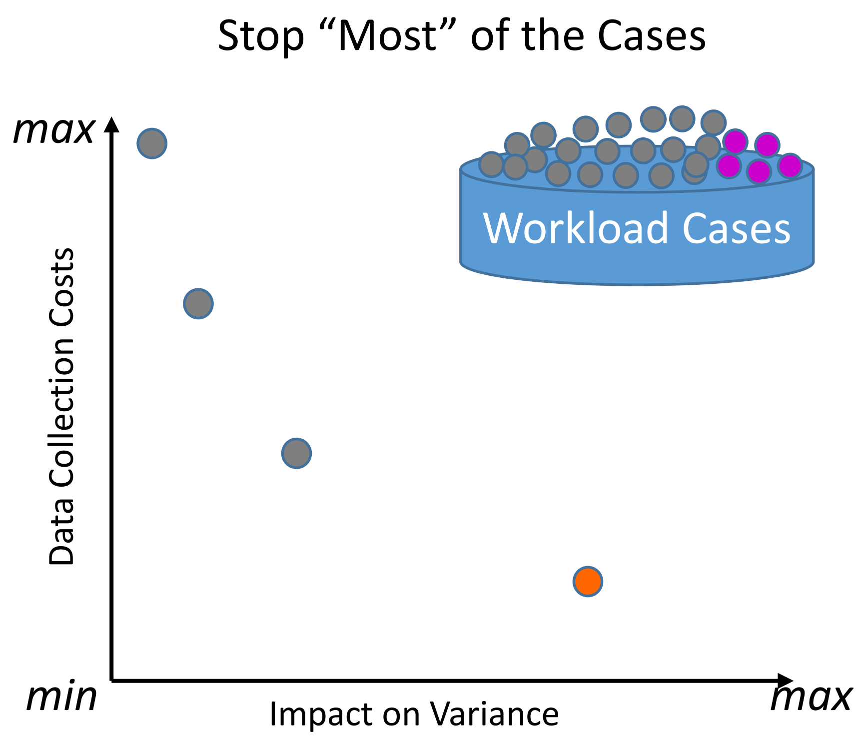
Stop “Half” of the Cases



United States® Census Bureau

Adaptive Survey Design

- Framework for Data Collection
  - Leverage tradeoffs
  - Min(budget) for a fixed level of data quality
  - Max(data quality) for a fixed budget
  - Balance resources and quality
    If we can save resources, willing to give up "some" data quality



Adaptive Survey Design

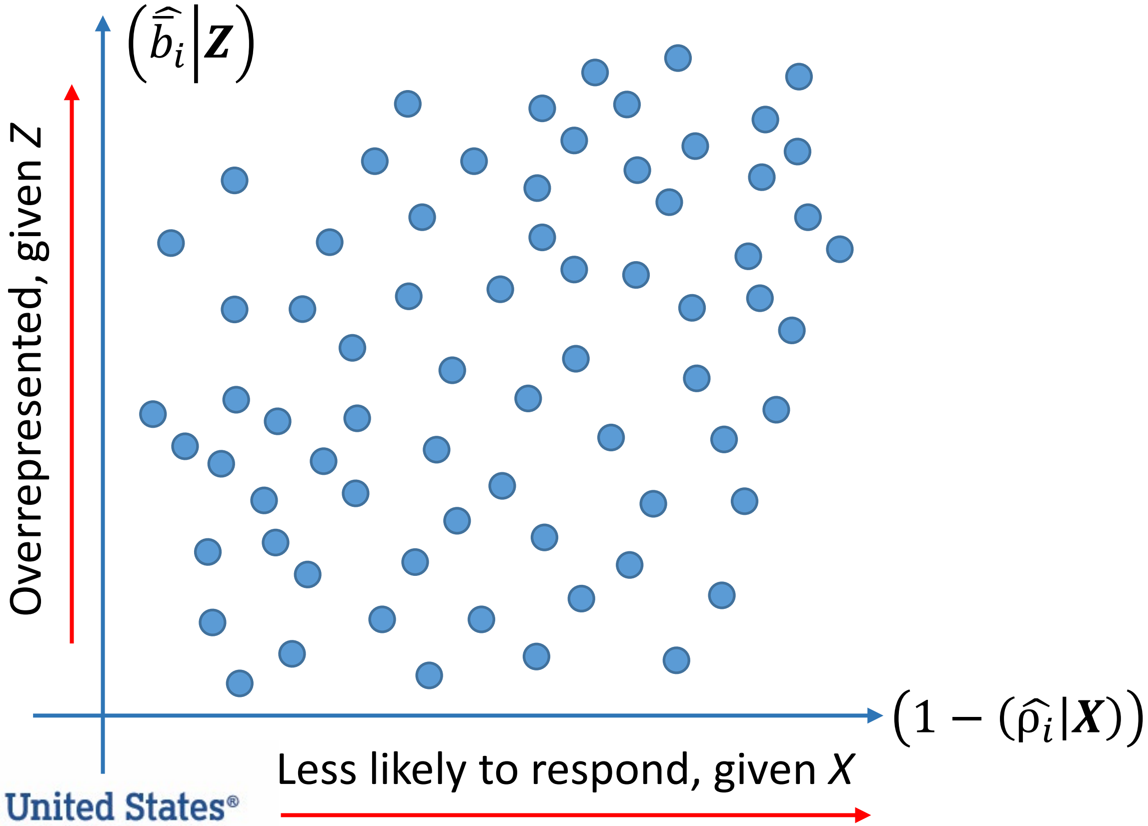
- Need predictive models to determine the impact of stopping cases
- Typically care about:
  - *Quality* – summary statistic (variance inflation, MSE(item), CVs, etc.)
  - *Cost* – cost-per-outcome (response, nonresponse, etc.) – 2 viewpoints
    - “Cut Costs” - other interviewer behavior stays the same, costs are reduced
    - “Reallocate Costs” – shifts resources from stopped cases to retained cases
  - *Response Behavior* – will a case actually respond?
    - Likely nonrespondents have impact on cost, but not on quality (vs baseline)
- Use model output to identify which cases to stop *and* their impact on quality

Adaptive Survey Design

- Need predictive models to determine the impact of stopping cases
- Typically care about:
  - *Quality* – summary statistic (**variance inflation**, MSE(item), CVs, etc.)
  - *Cost* – cost-per-outcome (response, nonresponse, etc.) – 2 viewpoints
    - "Cut Costs" - other interviewer behavior stays the same, costs are reduced
    - **"Reallocate Costs"** – shifts resources from stopped cases to retained cases
  - *Response Behavior* – will a case actually respond?
    - Likely nonrespondents have impact on cost, but not on quality (vs baseline)
- Use model output to identify which cases to stop *and* their impact on quality

Which cases to stop?

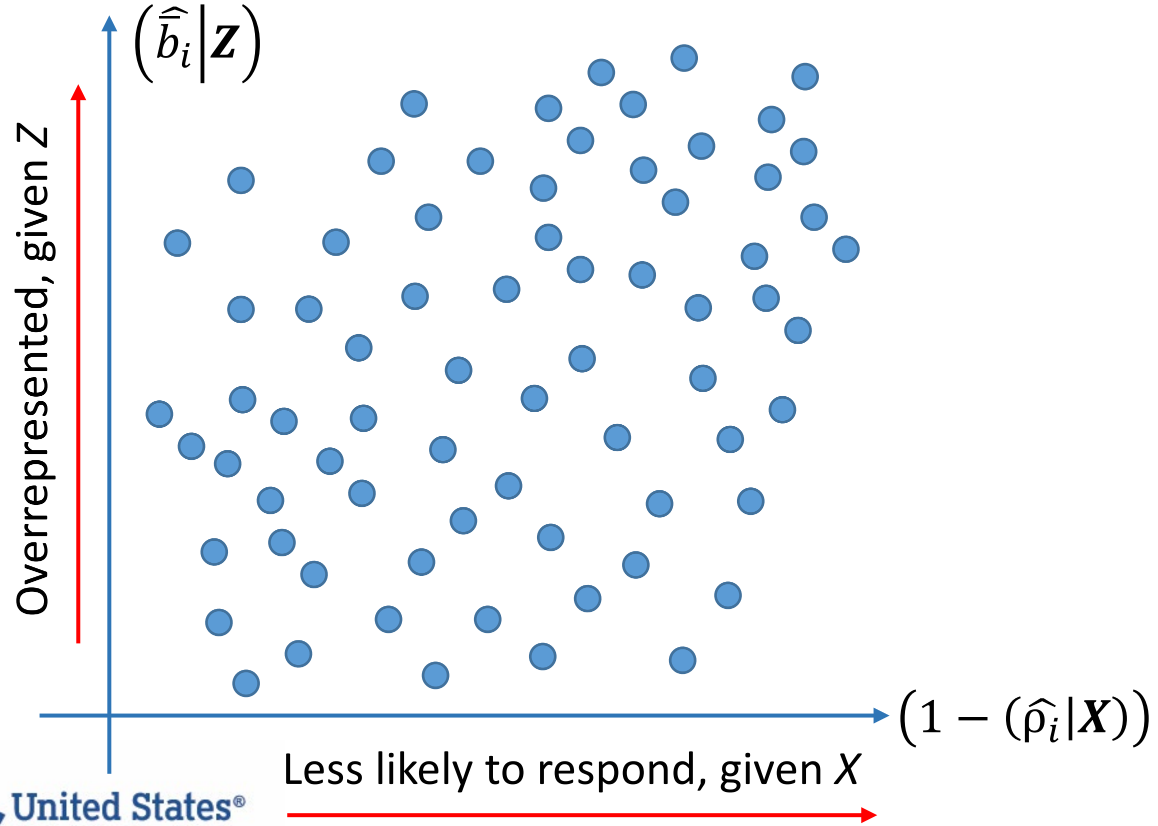
- Juxtapose "importance" of case with "likelihood to respond"

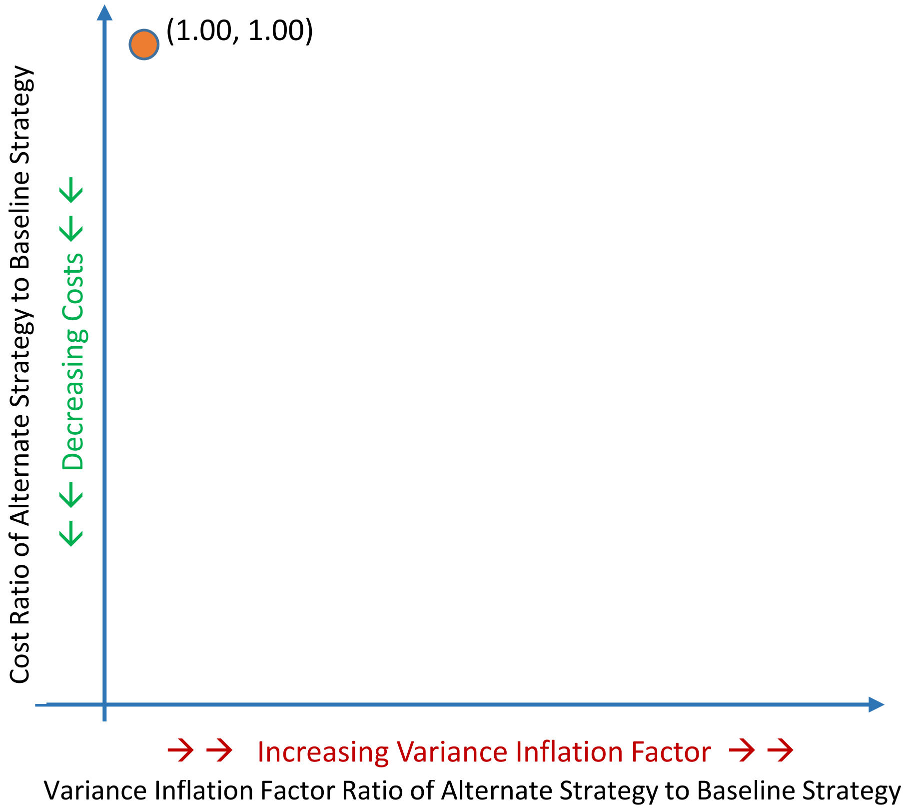


United States Census Bureau

Which cases to stop?

- Juxtapose "importance" of case with "likelihood to respond"

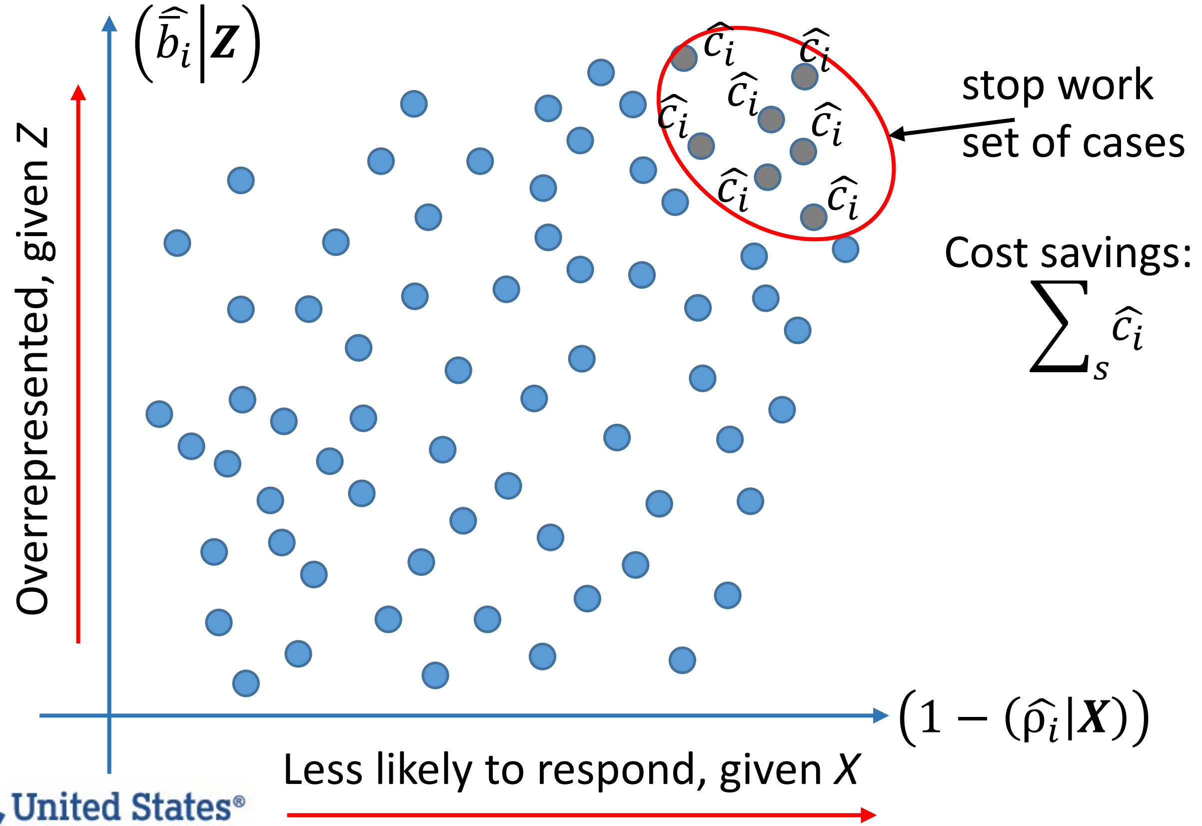


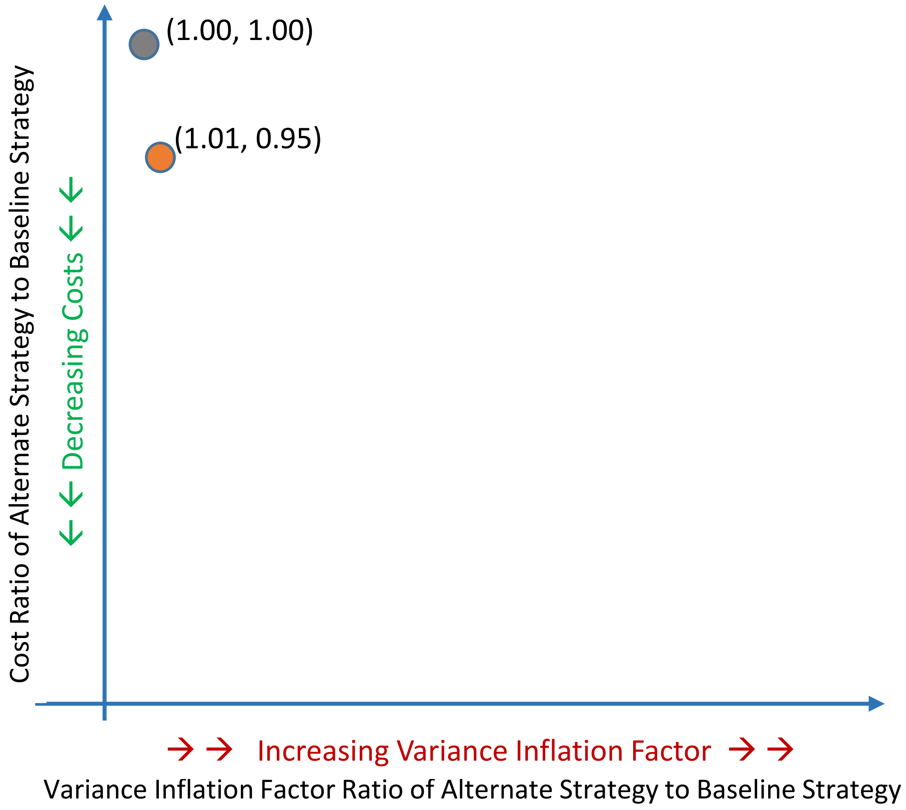


United States® Census Bureau

Which cases to stop?

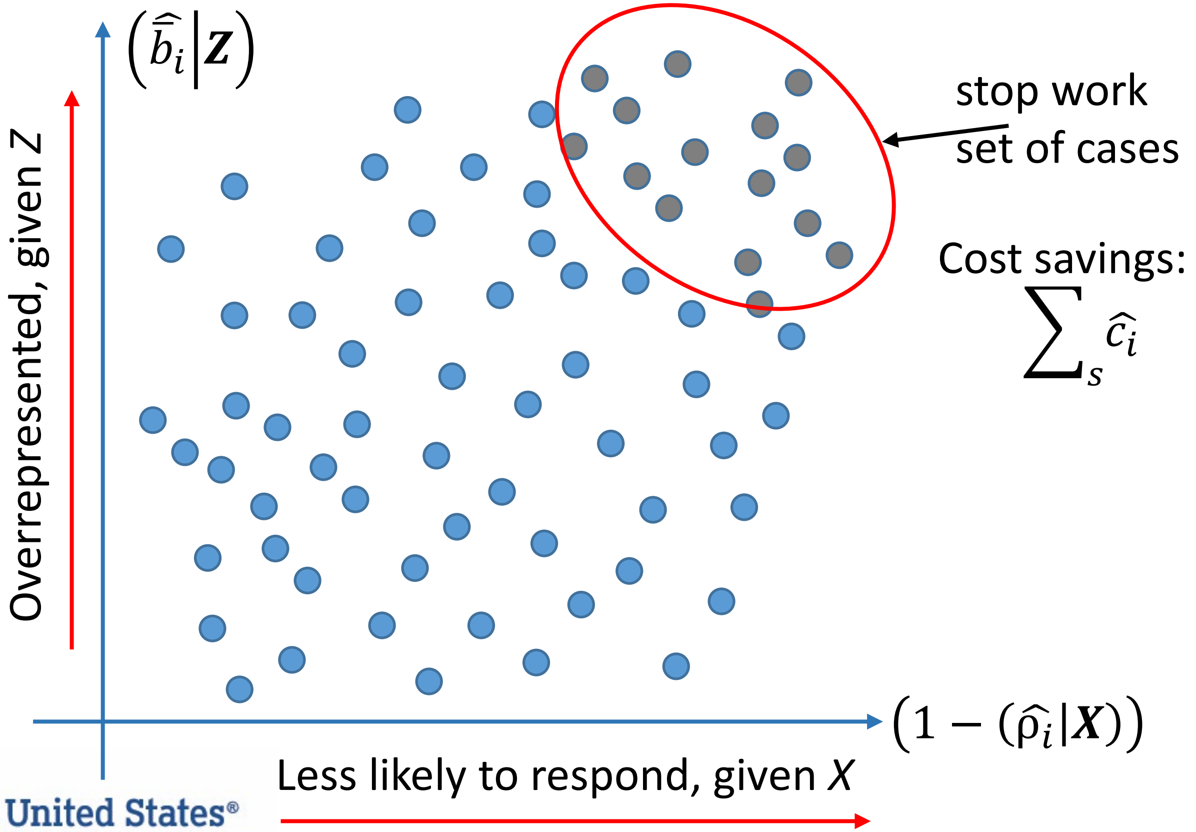
- Juxtapose "importance" of case with "likelihood to respond"

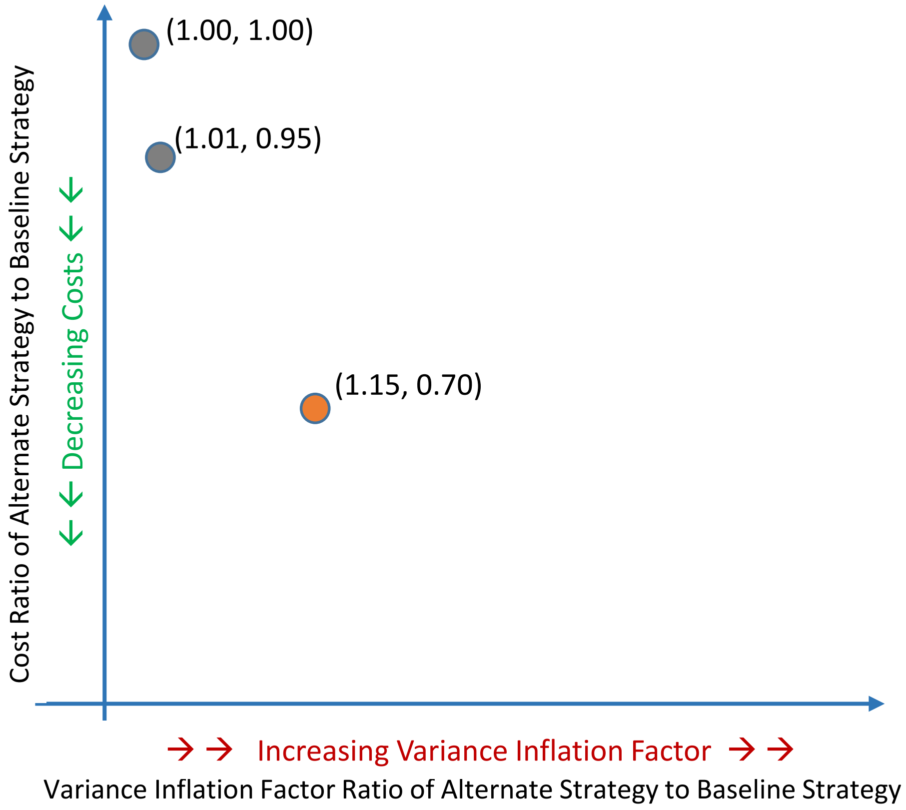




Which cases to stop?

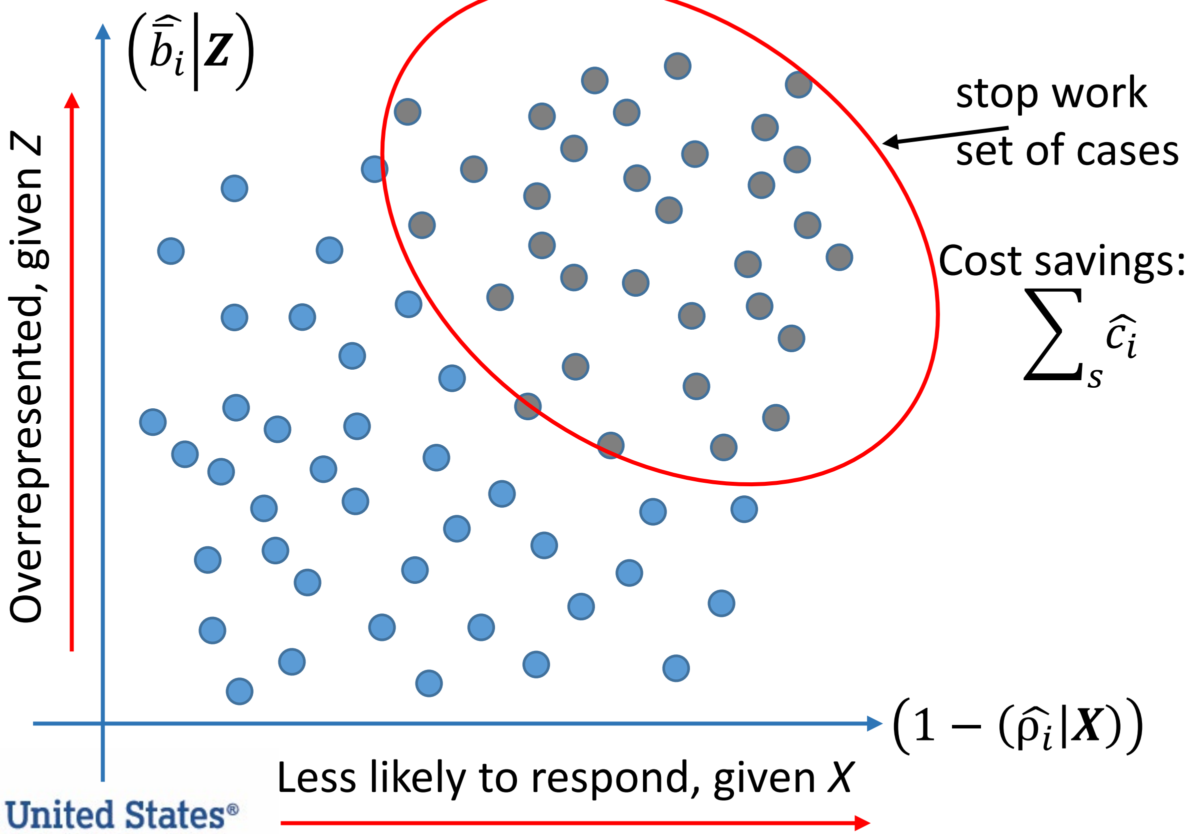
- Juxtapose “importance” of case with “likelihood to respond”

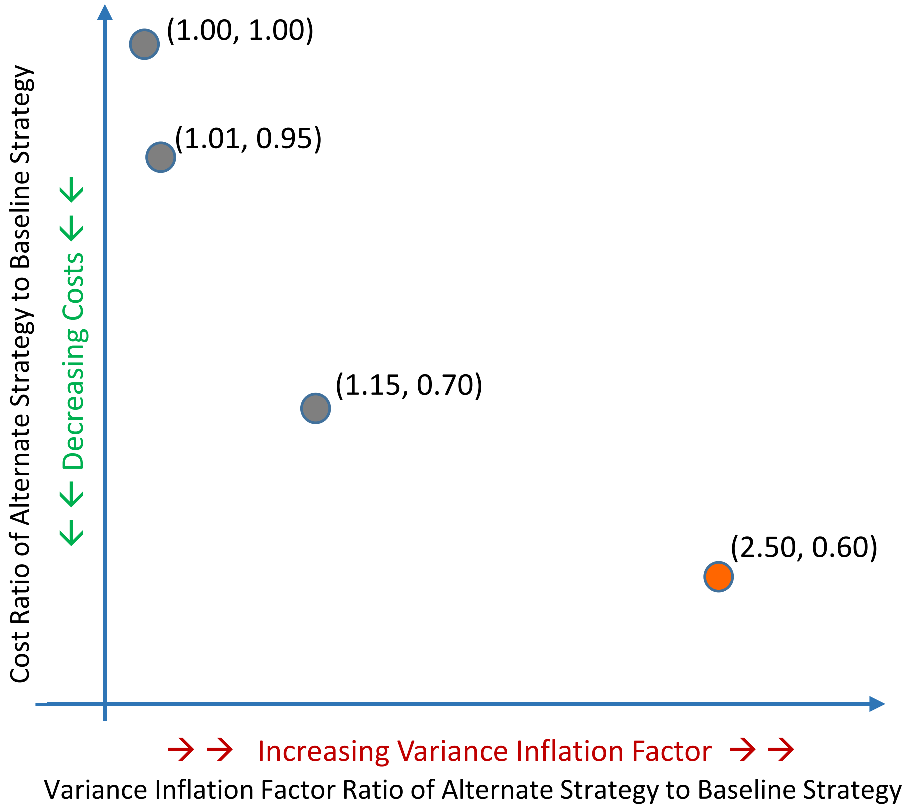




Which cases to stop?

- Juxtapose "importance" of case with "likelihood to respond"

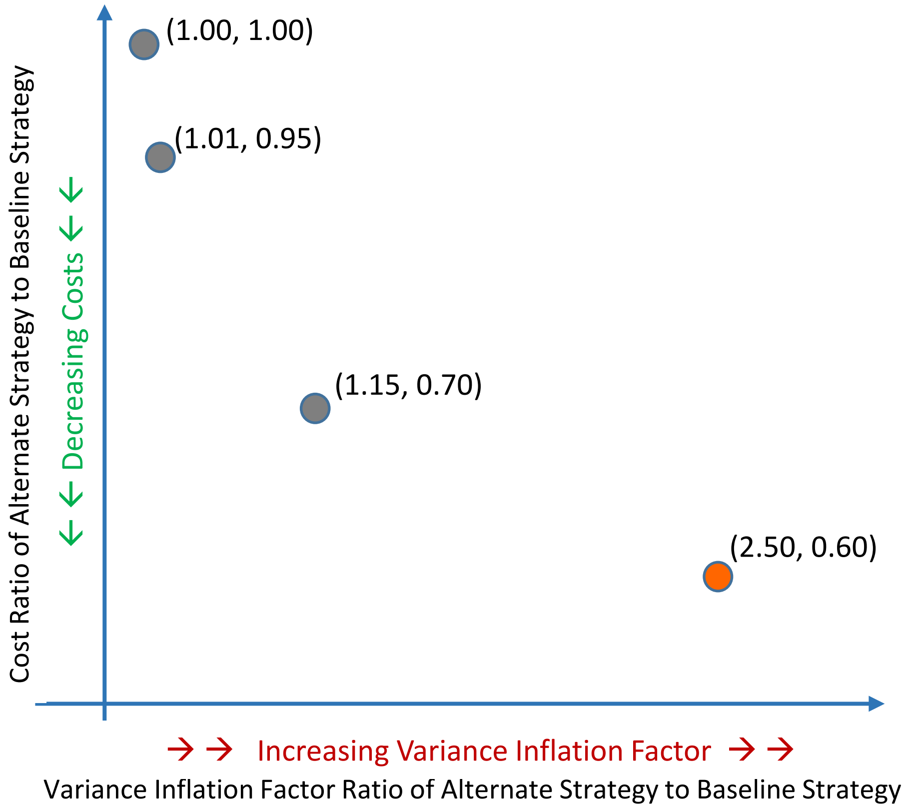




United States Census Bureau

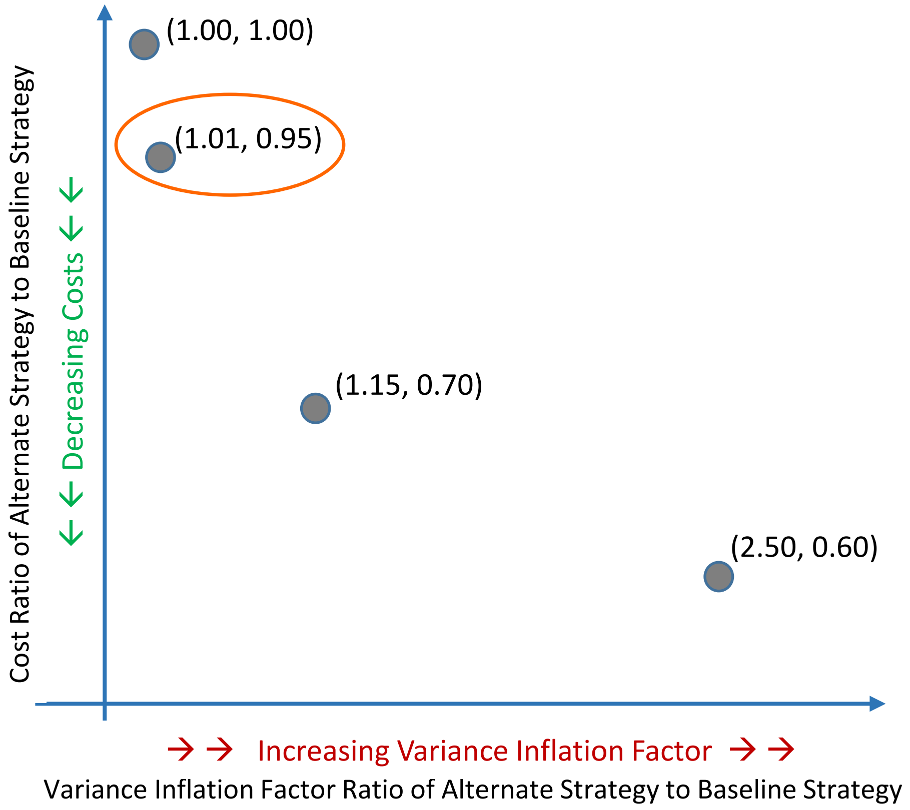
When to stop stopping cases?

- Optimization tool is flexible
- Can be adjusted given goals



When to stop stopping cases?

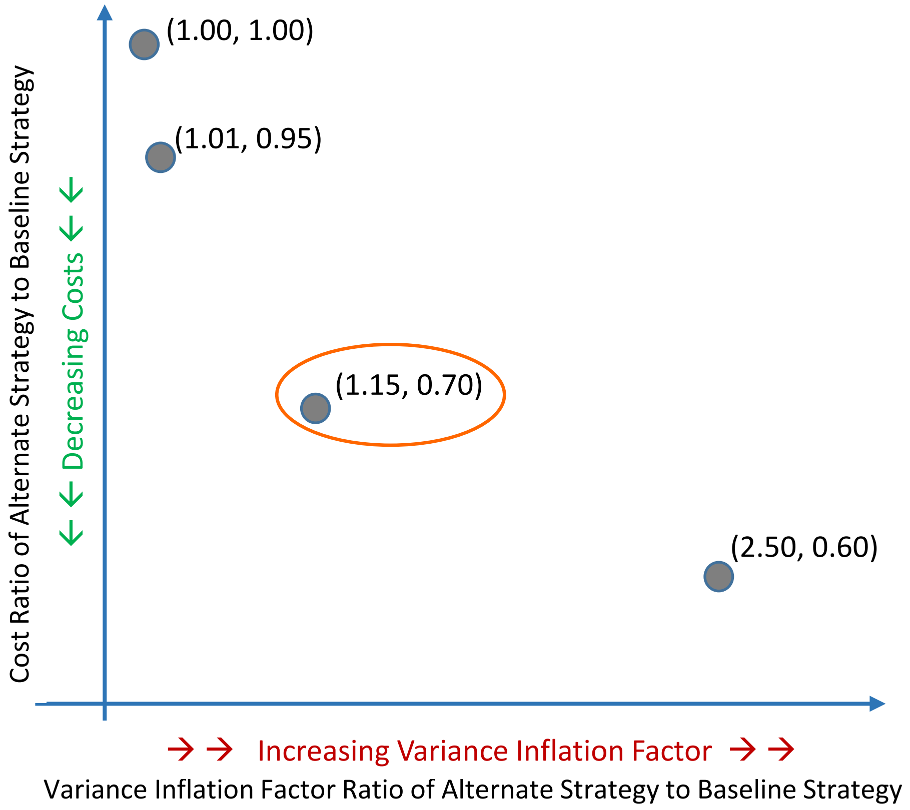
- Optimization tool is flexible
- Can be adjusted given goals
- Goal 1: Need to reallocate 5% of budget
  Need to accept a 1% variance inflation



United States® Census Bureau

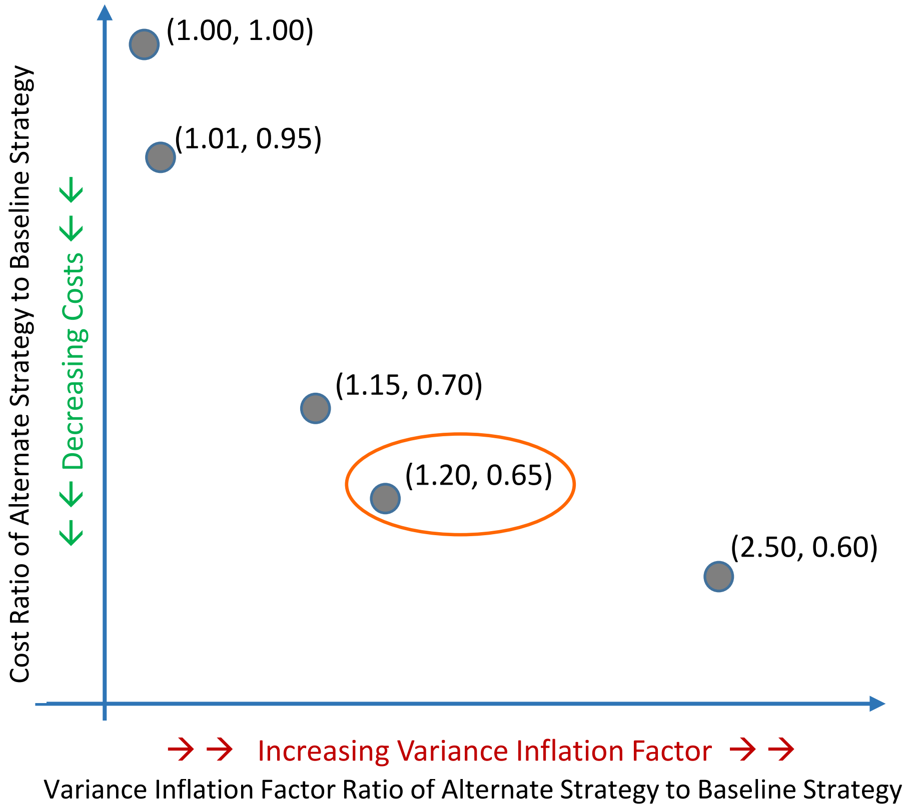
When to stop stopping cases?

- Optimization tool is flexible
- Can be adjusted given goals
- Goal 1: Need to reallocate 5% of budget
  Need to accept a 1% variance inflation
- Goal 2: Maximize resource shift for a 15% increase (or less) in variance inflation
  You can obtain a 30% resource shift



When to stop stopping cases?

- Optimization tool is flexible
- Can be adjusted given goals
- Goal 1: Need to reallocate 5% of budget
  Need to accept a 1% variance inflation
- Goal 2: Maximize resource shift for a 15% increase (or less) in variance inflation
  You can obtain a 30% resource shift
- Goal 3: Balance cost and variance inflation
  Find best tradeoff – minimum of product



Mathematical Optimization

- Formalize Idea
  - $\varphi(\mathrm{A}) = \min_{A(i \epsilon s)} \left(\left(\frac{\hat{C}^A | s^A, R^A}{\hat{C}^0 | s^0, R^0} \right) \left(\frac{\hat{V}^A | s^A, R^A}{\hat{V}^0 | s^0, R^0} \right) \right)$ where *0* is the baseline strategy (normal data collection), and *A* is the alternate strategy (some set, *s*, of cases stopped)
  - Response propensity model $\mathrm{r_{it}} = p(R = 1 | \boldsymbol{X})$
  - Balancing propensity model (for ranking) $\mathrm{b_{it}} = p(R = 1 | \boldsymbol{Z})$
  - Cost model $\hat{c}_i^o + \hat{c}_f^o$
  - Variance inflation formula $\hat{V}^A / \hat{V}^0$

- Examples in the literature
  - National Survey of College Graduates – minimize RMSE of key statistic (salary)
  - National Survey of Family Growth – minimize the MSE of several key statistics
  - Dutch Labor Force Survey – minimize mode effects in multimode survey

United States Census Bureau

What is Unique about the ACS?

- ACS is cross-sectional, with no past response data for modeling
  - Need covariates for the balancing model
  - Characteristics related to outcomes of interest
  - Broad set of administrative data
    [MAF, Commercial Housing Data, IRS, SSA, Demographic Data from 2010/2020 Census, etc.]
  - Assign characteristics to sample units for the balancing model
    [sex, age, race/ethnicity, marital status, income, program participation flags, housing structure vars, etc.]

United States® Census Bureau

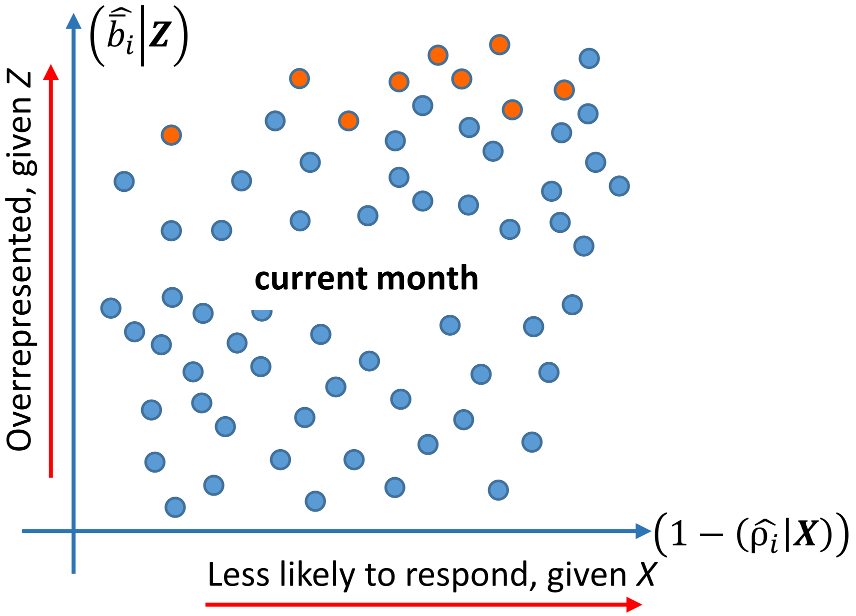
What is Unique about the ACS?

- ACS is Pseudo-longitudinal
  - Data collected for one panel is not released independently
  - 12 months of data combined into annual estimates (or 60 months for 5-yr)
  - Interventions we make could impact estimates for a long time
  - Use two balancing propensity models to account for time: $\mathrm{b}_{1\mathrm{it}}, \mathrm{b}_{3\mathrm{it}}$

United States® Census Bureau

What is Unique about the ACS?

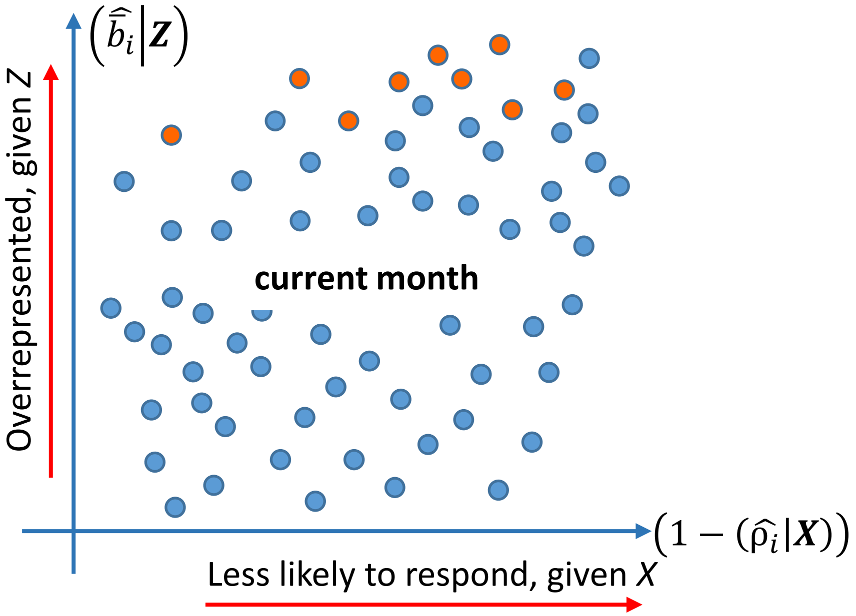
- ACS is Pseudo-longitudinal
  - Data collected for one panel is not released independently
  - 12 months of data combined into annual estimates (or 60 months for 5-yr)
  - Interventions we make could impact estimates for a long time
  - Use two balancing propensity models to account for time: $\mathrm{b}_{1it}, \mathrm{b}_{3it}$

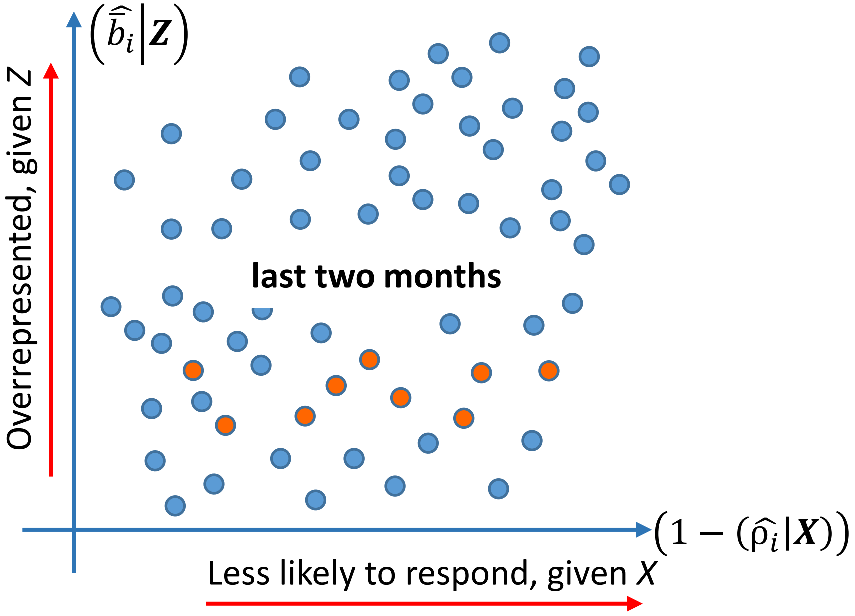


United States® Census Bureau

What is Unique about the ACS?

- ACS is Pseudo-longitudinal
  - Data collected for one panel is not released independently
  - 12 months of data combined into annual estimates (or 60 months for 5-yr)
  - Interventions we make could impact estimates for a long time
  - Use two balancing propensity models to account for time: $\mathrm{b}_{1\mathrm{it}}, \mathrm{b}_{3\mathrm{it}}$





United States® Census Bureau

What is Unique about the ACS?

- ACS estimation geographies are Census tracts (average pop ~4,000)
  - Sample is spread across 12 months – very small sample sizes
  - Can’t run balancing propensity models at the tract level
    - Unstable, too noisy, sometimes only 1 or 2 cases!
  - Run balancing propensities at the state level
    - Could lead to stopping work on all cases in a tract

What is Unique about the ACS?

- ACS estimation geographies are Census tracts (average pop ~4,000)
  - Sample is spread across 12 months – very small sample sizes
  - Can’t run balancing propensity models at the tract level
    - Unstable, too noisy, sometimes only 1 or 2 cases!
  - Run balancing propensities at the state level
    - Could lead to stopping work on all cases in a tract
- Tract eligibility rules. At time of stop work, tracts must have:
  - At least 3 sample cases
  - At least one occupied interview
  - Retain enough sample to have two complete interviews

What is Unique about the ACS?

- ACS estimation geographies are Census tracts (average pop ~4,000)
  - Sample is spread across 12 months – very small sample sizes
  - Can't run balancing propensity models at the tract level
    - Unstable, too noisy, sometimes only 1 or 2 cases!
  - Run balancing propensities at the state level
    - Could lead to stopping work on all cases in a tract
- Tract eligibility rules. At time of stop work, tracts must have:
  - At least 3 sample cases
  - At least one occupied interview
  - Retain enough sample to have two complete interviews
- This has led to a conservative stop work algorithm
  - Allowed us to develop models, code, evaluation metrics
  - Release version 1 of process – for future improvements

United States® Census Bureau

What have we done so far?

- Development lasted from December 2022 – June 2023
  - Data access and linkage
  - Model development
  - June 2023 – end-to-end test – stopped one case per state
- Implementation began July 2023
  - Stop Work Interventions Delivered for July, August, September, October
  - Midway through CAPI, carried out optimization procedures
  - Stopped 50% of the optimal number of cases for stop work (after restrictions)

What are we monitoring?

- Initial Monitoring
    - Unweighted CAPI completion rate
    - Mean balancing propensities
    - CV(mean) balancing propensities
    - R-Indicators (overall- and state-level)
- Future Monitoring
    - Attempts / hours on cases retained after stop work
- Longer-Term Monitoring
    - ACS 1-year estimates

United States® Census Bureau

Next Steps

- Consider different geographies for the balancing (quality) model
  - Urban-Rural? MSA? Something else?
  - How does changing the geography change stop work patterns?
- Improving cost/resource model
  - Add in additional covariates, geography, interviewer/workload characteristics
  - Incorporating mileage into the resource model
- Investigating impacts of resource reallocation
  - More hours on cases retained? More attempts?
  - Impact on response propensities for cases that are retained?
- Continuing to monitor outcomes
  - Completion rates, hours spent on cases, R-indicators, etc.

United States® Census Bureau

Contact & Acknowledgements

Stephanie Coffey
stephanie.coffey@census.gov

Thanks to the entire ACS program…

survey operations, survey design, field operations, household nonresponse teams

…for their interest and support of our work!!!

United States® Census Bureau