A Matrix Sample Design to Reduce Burden on Companies in the Manufacturing Sector of the Annual Integrated Economic Survey

Colt S. Viehdorfer

U.S. Census Bureau

2024 Federal Committee on Statistical Methodology Conference

Overview

- Brief Description of the Manufacturing Sector in the Annual Integrated Economic Survey (AIES)
- Matrix Sample Design
- Estimation and Variance Estimation Methods
- Simulation Study
- Production Sample
- Discussion

United States® Census Bureau

Manufacturing Sector in the AIES

- Employer businesses with manufacturing activity are classified in North American Industry Classification Systems (NAICS) sector 31-33

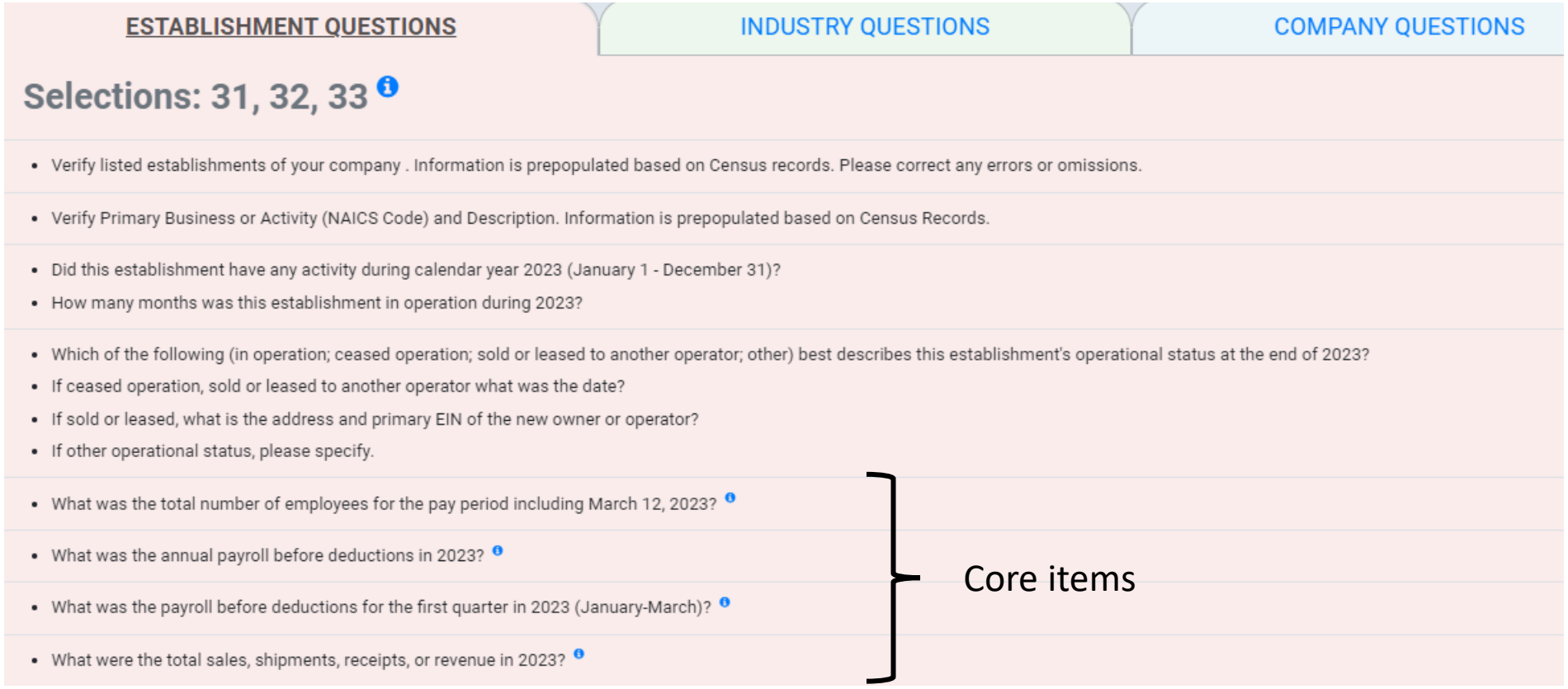
- Data items are collected for each establishment within a company

- Prior annual collections and estimates were done as part of the Annual Survey of Manufactures (ASM)
  - Excluding years ending in ‘2’ and ‘7’ when estimates were produced as part of the Economic Census

- Statistics are provided on employment, payroll, worker hours, payroll supplements, cost of materials, selected operating expenses, value added by manufacturing, capital expenditures, inventories, and energy consumption. Estimates of value of shipments for approximately 3,500 products manufactured are also included, as defined by the North America Product Classification System (NAPCS)

United States® Census Bureau

AIES Core Items



United States® Census Bureau

AIES Noncore Items

- What was the number of production workers (direct labor including first-line supervisors) for the pay period including March 12, 2023?
- What was the total amount paid for production worker annual payroll in 2023?
- For all production workers, what was the payroll before deductions for the first quarter (January-March 2023)?

- During 2023, what amount of the total sales, shipments, receipts, or revenue was from goods, services, or manufactured products that were ordered or whose movement was controlled or coordinated electronically?

- What was the amount paid for temporary staff and leased employees in 2023?
- What was the amount paid for software (prepackaged, custom-coded, or vendor-customized) in 2023?
- What was the amount paid for data processing and other purchased computer services in 2023?
- What was the amount paid for purchased communications services in 2023?
- What was the amount paid for expensed computer hardware and other equipment in 2023?
- What was the amount paid for purchased advertising and promotional services in 2023?
- What was the amount paid for purchased professional and technical services in 2023?
- What was the amount paid for purchased repairs and maintenance to machinery, vehicles,and equipment in 2023?
- What was the amount paid for purchased repairs and maintenance to buildings, structures, and offices in 2023?
- What was the amount paid for lease and rental payments for machinery, equipment, and other tangible items in 2023?
- What was the amount paid for lease and rental payments for land, buildings, and other structures in 2023?
- What was the amount paid for purchased electricity in 2023?
- What was the amount paid for purchased fuels (except motor fuels) in 2023?
- What was the amount paid for water, sewer, refuse removal, and other utility payments in 2023?
- What was the amount paid for governmental taxes and licensing fees in 2023?
- What was the amount paid for materials, parts, and supplies not purchased for resale in 2023?
- What was the amount paid for all other operating expenses in 2023?
- If all other operating expenses are greater than 50% of total operating expenses, specify their primary source.

United States® Census Bureau

AIES Noncore Items

- What was the total cost for all products bought and sold without further processing in 2023?
- What was the amount paid for production-related costs for work done for you by others on your materials (work contracted to others) in 2023?

- What were the total operating expenses in 2023?

- What was the quantity of purchased electricity in 2023?
- What was the quantity of electricity sold or transferred to other establishments in 2023?
- What was the quantity of generated electricity in 2023?

- What were the total capital expenditures for new buildings and other structures?
- What were the total capital expenditures for used buildings and other structures?
- What were the total capital expenditures for new machinery and equipment in 2023?
- What were the total capital expenditures for used machinery and equipment in 2023?

- What were the capital expenditures for new and used industrial robotic equipment, including software, installation, and other one-time costs?
- What was the number of industrial robots in operation at this establishment?
- If you are unable to provide the number of industrial robots in operation, please explain:
- What was the number of industrial robots purchased for this establishment?
- If you are unable to provide the number of industrial robots purchased, please explain:

- What was the value of inventories owned as of December 31, 2023, for finished goods?
- What was the value of inventories owned as of December 31, 2023, for work-in-process?
- What was the value of inventories owned as of December 31, 2023, for materials, supplies, fuels, etc.?
- What was the total value of inventories owned as of December 31, 2023?
- What was the value of inventories owned as of December 31, 2022, for finished goods?
- What was the value of inventories owned as of December 31, 2022, for work-in-process?
- What was the value of inventories owned as of December 31, 2022, for materials, supplies, fuels, etc.?
- What was the total value of inventories owned as of December 31, 2022?

United States® Census Bureau

AIES Noncore Items

- What was the amount of LIFO (Last-in, First-out) reserve (if any) as of December 31, 2023?

- How much of the total inventories as of December 31, 2023, were subject to the FIFO (First-in, First-out) cost valuation method?
- How much of the total inventories as of December 31, 2023, were subject to the average cost valuation method?
- How much of the total inventories as of December 31, 2023, were subject to the standard cost valuation method?
- How much of the total inventories as of December 31, 2023, were subject to other non-LIFO (Last-in, First-out) cost valuation methods?
- What was the total value of inventories owned as of December 31, 2023, using non-LIFO (Last-in, First-out) cost valuation methods?
- How much of the total inventories as of December 31, 2023, were subject to the LIFO (Last-in, First-out) cost valuation method?
- What was the total value of inventories owned as of December 31, 2023, using LIFO (Last-in, First-out) and non-LIFO cost valuation methods?
- How much of the total inventories as of December 31, 2022, were subject to the FIFO (First-in, First-out) cost valuation method?
- How much of the total inventories as of December 31, 2022, were subject to the average cost valuation method?
- How much of the total inventories as of December 31, 2022, were subject to the standard cost valuation method?
- How much of the total inventories as of December 31, 2022, were subject to other non-LIFO (Last-in, First-out) cost valuation methods?
- What was the total value of inventories owned as of December 31, 2022, using non-LIFO (Last-in, First-out) cost valuation methods?
- How much of the total inventories as of December 31, 2022, were subject to the LIFO (Last-in, First-out) cost valuation method?
- What was the total value of inventories owned as of December 31, 2022, using LIFO (Last-in, First-out) and non-LIFO cost valuation methods?
- What was the amount of LIFO (Last-in, First-out) reserve (if any) as of December 31, 2022?

- Did this establishment (or establishments in this industry) own inventories, regardless of where held, as of December 31, 2023?

- What was the value of products shipped and other receipts in 2023?

United States® Census Bureau

(Multiple)Matrix Sampling

- Goal: reduce response burden induced by a long questionnaire

- Procedure: divide the questionnaire into sections and administer sections to randomly selected subsamples of the original sample
  - Planned increase in sampling variance
  - Potential increase in quality of selected items by focusing efforts on subsamples

- Split questionnaire: AIES manufacturing
  - One randomly selected set of companies received the full suite of manufacturing questions for establishments (core + noncore)
  - The remaining companies only received the four core questions

United States® Census Bureau

Experimental Matrix Sampling

- Pilot is being conducted in year 1 of AIES on companies with business activity in the manufacturing sector

- 2nd Phase sample restricted to Manufacturing noncertainty (sample weight > 1) companies selected in AIES

- Estimation Strategies
  - Sampling weight adjustment (design-based) – reported here
  - Unit-level (multiple) imputation models – next talk

United States® Census Bureau

Sample Design

AIES

- Phase one
- Stratified by NAICS3 and geography
- Sequential probability proportional to expected size (payroll)
- Complexity, Allocation, and PPES certainties

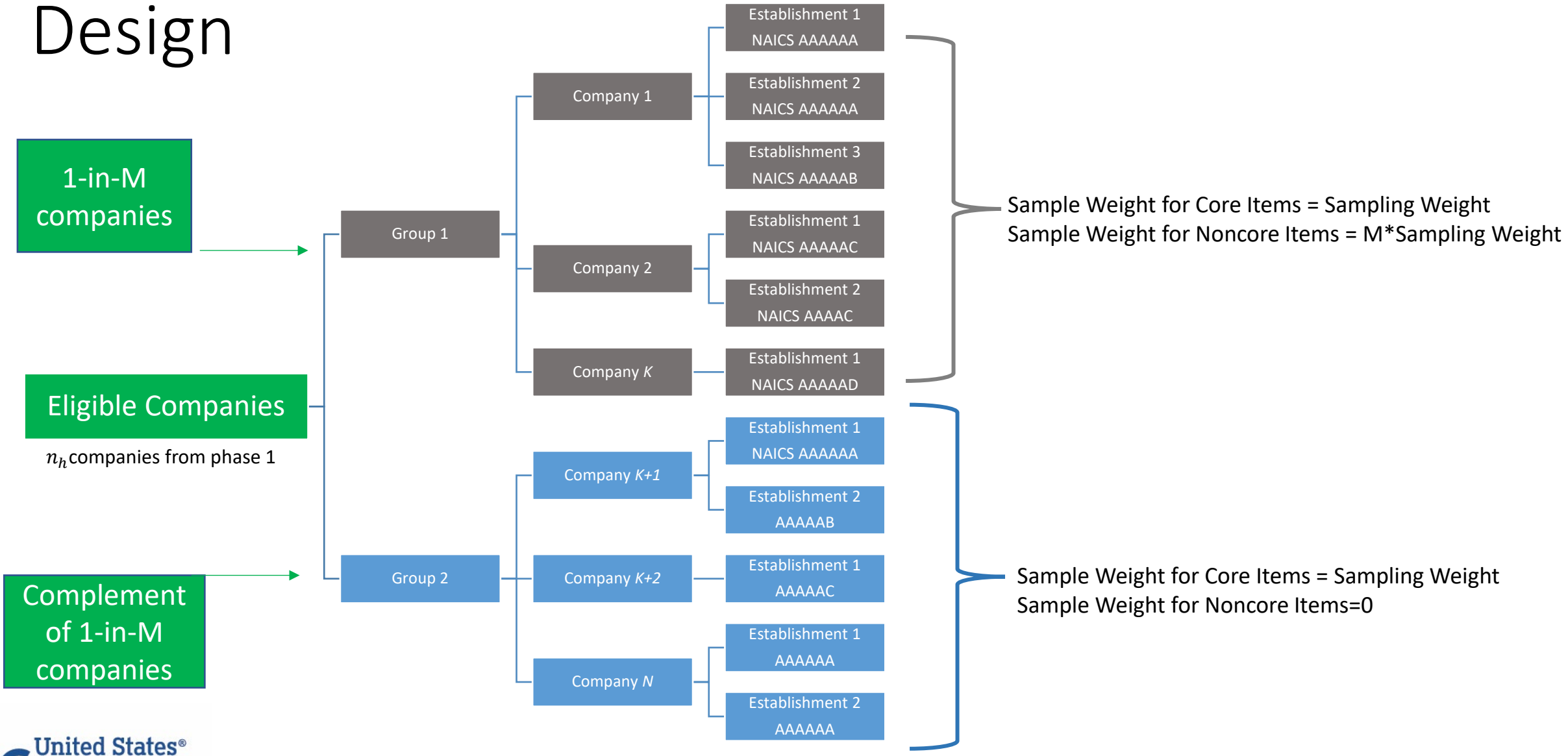
Subsample

Matrix Sample

- Phase two
- Manufacturing Sector
- Same stratum definitions
- Equal probability
- Phase one certainties were phase two certainties
- Strata with fewer than 15 companies selected in phase one were certainties in phase two

United States® Census Bureau

Design



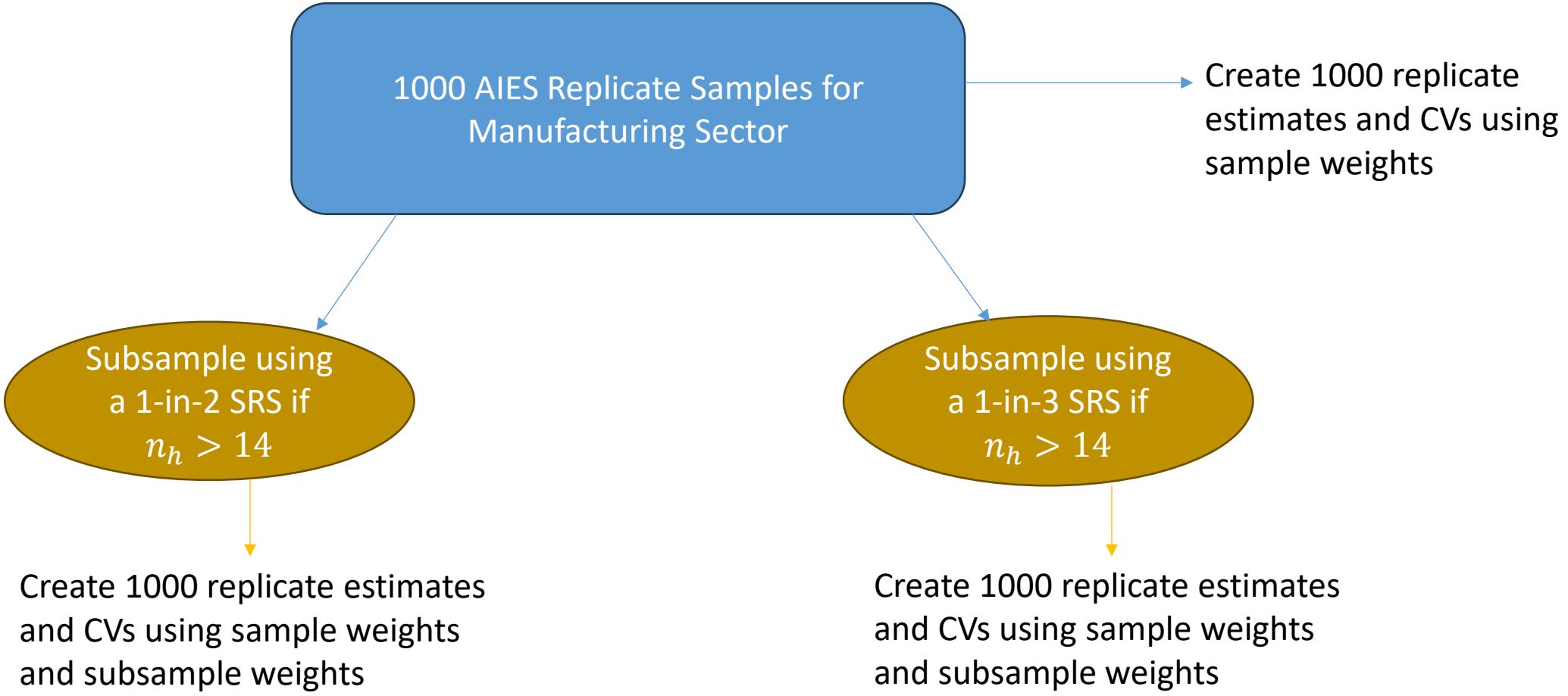
United States® Census Bureau

Simulation Study Design

- Selected 1000 AIES replicate samples from the research frame using the AIES design
- The same certainty units were included in all 1000 replicate samples
- Sample allocation n_h was fixed for each noncertainty stratum across the replicates
- Subsampled within the noncertainty strata if $n_h > 14$

United States® Census Bureau

Simulation Study Design



United States® Census Bureau

Simulation Study Analysis

- Included a subset of the Manufacturing Sector
  - NAICS 312 (Beverage and Tobacco Product Manufacturing)
  - NAICS 321 (Wood Product Manufacturing)
  - NAICS 325 (Chemical Manufacturing)
  - NAICS 336 (Transportation Equipment Manufacturing)
- Used Receipts (Sales), Employment, Cost of Materials, and Total Inventories for analysis
  - Receipts and Employment collected from all companies in production
  - Industry average imputation was used when frame data not available
- Compared estimates and CVs from each AIES replicate sample to the estimates and CVs produced from the corresponding replicate sample with subsampling

United States® Census Bureau

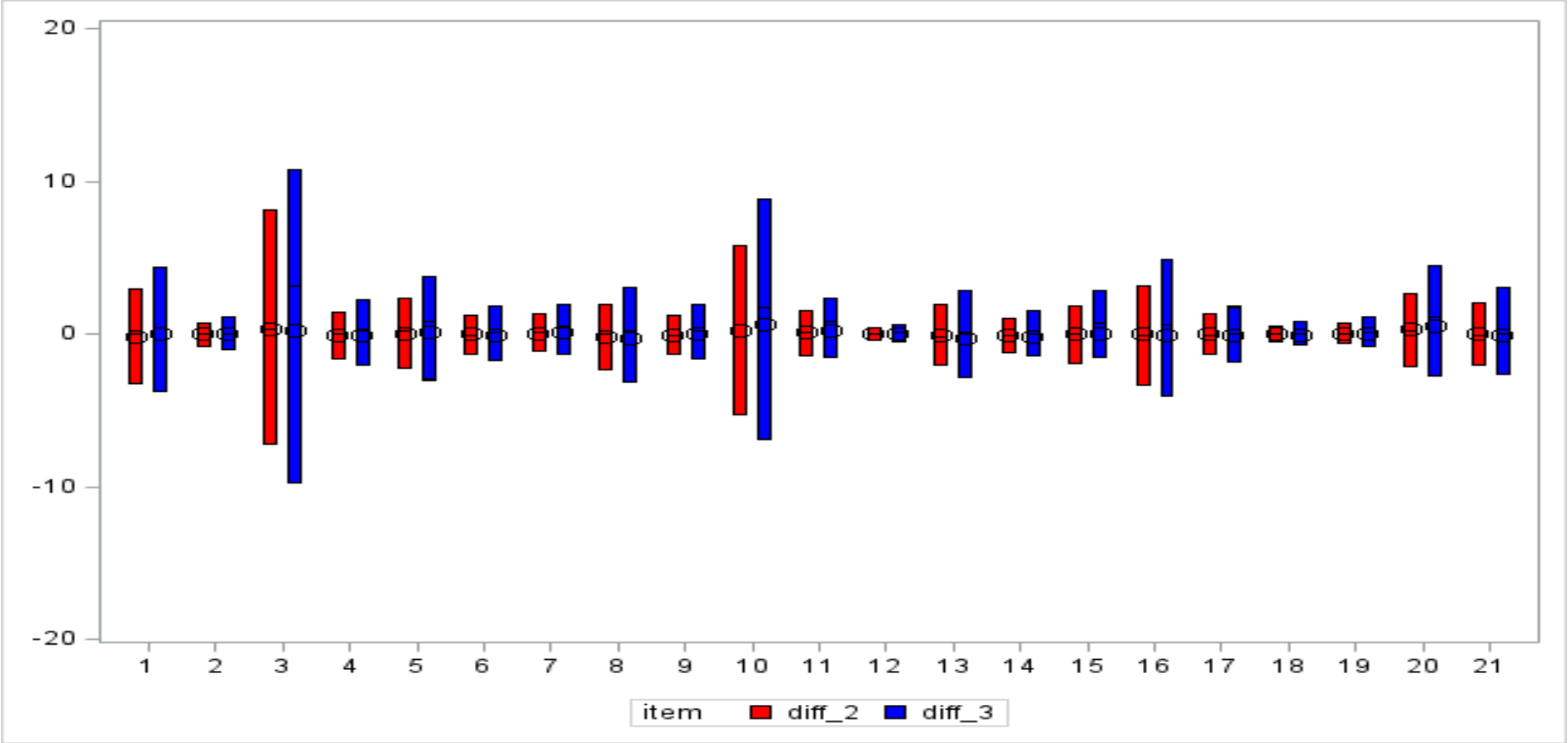
Estimation

- Horvitz-Thompson estimates with design weights accounting for both AIES and matrix sample designs
  - Ratio estimation will be used for production tabulations

- Estimates in the Manufacturing Sector are produced by industry (6-digit NAICS at the national level) and geography (NAICS by State)
  - Smaller domain estimates are more variable (not controlled for in sampling)
  - Primary focus for the simulation study was on evaluating estimates and CVs for NAICS3 by State, although national level industry estimates were produced and examined

United States® Census Bureau

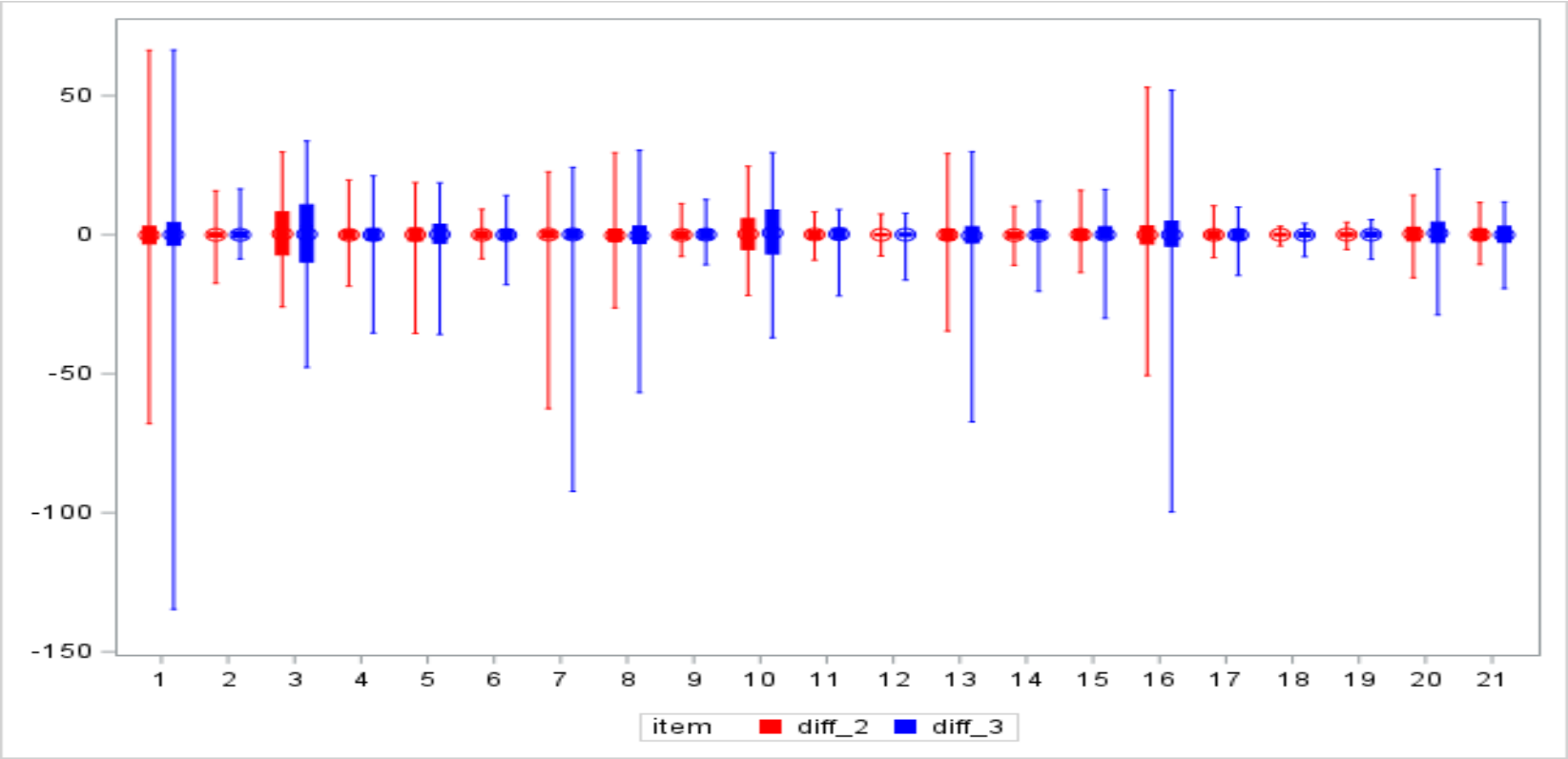
Percentage Change in Replicate Domain Estimates of Receipts



Source: U.S. Census Bureau, Project No. 7529180/Approval CBDRB-FY25-ESMD001-002.
diff_2 = the percentage change between the phase 1 replicate estimate and the phase 2 replicate estimate in domain *k* with 1-in-2 subsampling
diff_3 = the percentage change between the phase 1 replicate estimate and the phase 2 replicate estimate in domain *k* with 1-in-3 subsampling

United States® Census Bureau

Percentage Change in Replicate Domain Estimates of Receipts



Source: U.S. Census Bureau, Project No. 7529180/Approval CBDRB-FY25-ESMD001-002.
diff_2 = the percentage change between the phase 1 replicate estimate and the phase 2 replicate estimate in domain *k* with 1-in-2 subsampling
diff_3 = the percentage change between the phase 1 replicate estimate and the phase 2 replicate estimate in domain *k* with 1-in-3 subsampling

United States® Census Bureau

Horvitz-Thompson Variance Estimation

$$\hat{\boldsymbol{v}}\left(\hat{\boldsymbol{Y}}^{\boldsymbol{k}}\right) = \hat{\boldsymbol{v}}_{\boldsymbol{1}}(\boldsymbol{E}_{\boldsymbol{2}}) + \boldsymbol{E}_{\boldsymbol{1}}(\hat{\boldsymbol{v}}_{\boldsymbol{2}})$$

$$\hat{\boldsymbol{v}}_{\boldsymbol{1}}(\boldsymbol{E}_{\boldsymbol{2}}) = \sum_h \sum_{i>l}^{m_h} \left(\frac{\pi_{hi}\pi_{hl} - \pi_{h,il}}{\pi_{h,il}}\right)\left(\frac{y_{hi}^k}{\pi_{hi}\pi_{2|hi}} - \frac{y_{hl}^k}{\pi_{hl}\pi_{2|hl}}\right)^2$$

$$= \sum_h \sum_{i>l}^{m_h} \left(\frac{\pi_{hi}\pi_{hl} - \pi_{h,il}}{\pi_{h,il}}\right)\left(w_{1i}w_{2i}y_{hi}^k - w_{1l}w_{2l}y_{hl}^k\right)^2$$

$$\boldsymbol{E}_{\boldsymbol{1}}(\hat{\boldsymbol{v}}_{\boldsymbol{2}}) = \sum_h {n_h}^2 \left(1 - \frac{m_h}{n_h}\right)\left(\frac{s_y^2}{m_h}\right)$$

Where $s_y^2 = \dfrac{\sum_{m_h}\left(\frac{y_{hi}^k}{\pi_{hi}} - \breve{y}_h^k\right)^2}{m_h - 1}$ And $\breve{y}_h^k = \dfrac{\sum_{m_h}\left(\frac{y_{hi}^k}{\pi_{hi}}\right)}{m_h}$

y_{hi}^k= value of item y in domain k for company i in stratum h

$n_h =$ the sampled units in stratum h in the first phase

$m_h =$ the subsampled units in stratum h in the second phase

United States® Census Bureau

Horvitz-Thompson Variance Estimation

$$\hat{v}\left(\hat{Y}^k\right) = \hat{v}_1(E_2) + E_1(\hat{v}_2)$$

$$\hat{v}_1(E_2) = \sum_h \sum_{i>l}^{m_h} \left(\frac{\pi_{hi}\pi_{hl} - \pi_{h,il}}{\pi_{h,il}}\right)\left(\frac{y_{hi}^k}{\pi_{hi}\pi_{2|hi}} - \frac{y_{hl}^k}{\pi_{hl}\pi_{2|hl}}\right)^2$$

$$= \sum_h \sum_{i>l}^{m_h} \left(\frac{\pi_{hi}\pi_{hl} - \pi_{h,il}}{\pi_{h,il}}\right)\left(w_{1i}w_{2i}y_{hi}^k - w_{1l}w_{2l}y_{hl}^k\right)^2$$

Phase two weights

y_{hi}^k= value of item y in domain k for company i in stratum h

$n_h =$ the sampled units in stratum h in the first phase

$m_h =$ the subsampled units in stratum h in the second phase

United States® Census Bureau

Horvitz-Thompson Variance Estimation

$$\hat{v}\left(\hat{Y}^k\right) = \hat{v}_1(E_2) + E_1(\hat{v}_2)$$

$$E_1(\hat{v}_2) = \sum_h {n_h}^2 \left(1 - \frac{m_h}{n_h}\right)\left(\frac{s_y^2}{m_h}\right)$$

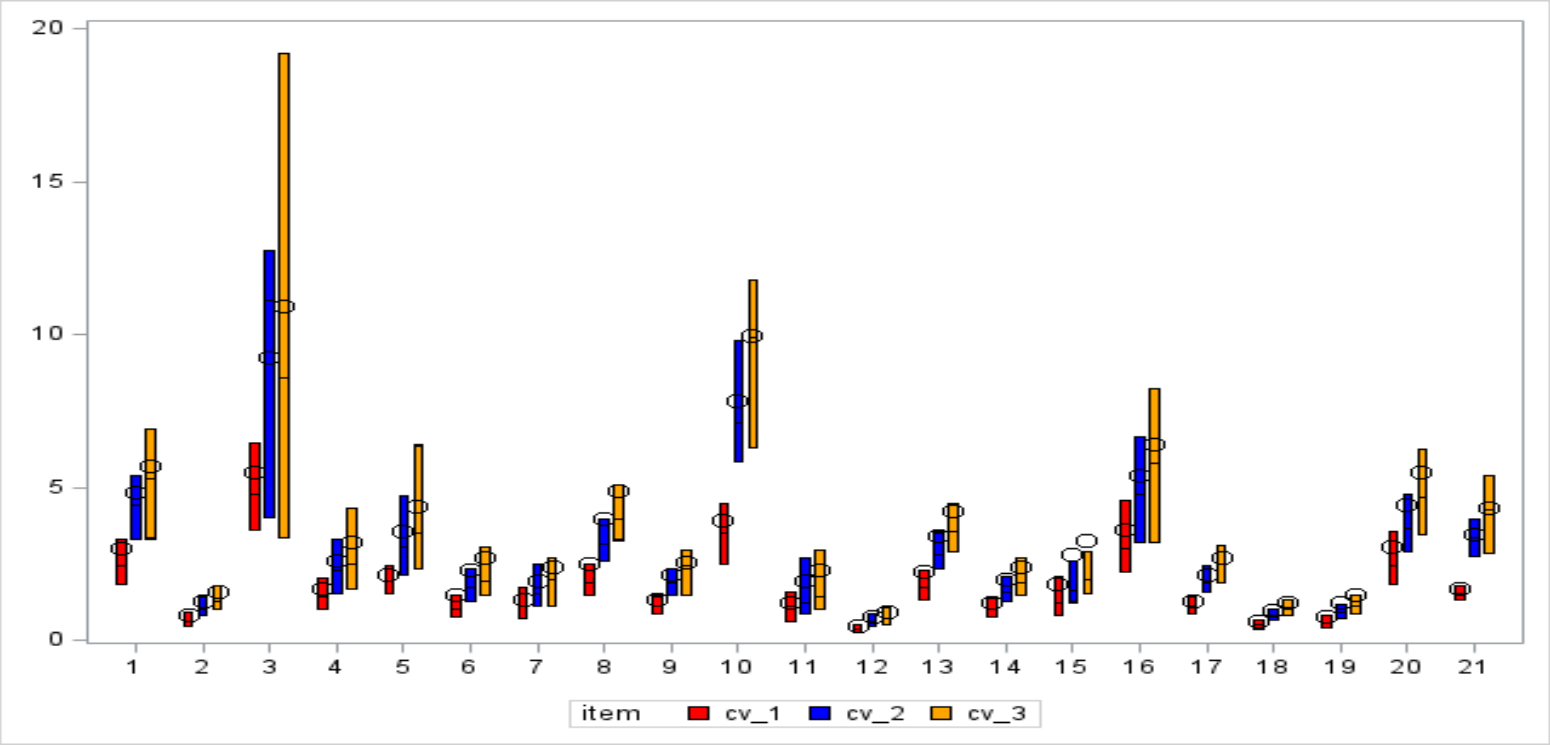
Where $s_y^2 = \frac{\sum m_h \left(\frac{y_{hi}^k}{\pi_{hi}} - \breve{y}_h^k\right)^2}{m_h - 1}$ And $\breve{y}_h^k = \frac{\sum m_h \left(\frac{y_{hi}^k}{\pi_{hi}}\right)}{m_h}$

United States® Census Bureau

y_{hi}^k= value of item y in domain k for company i in stratum h

$n_h =$ the sampled units in stratum h in the first phase
$m_h =$ the subsampled units in stratum h in the second phase

Replicate CVs ((SE/EST)*100)



Source: U.S. Census Bureau, Project No. 7529180/Approval CBDRB-FY25-ESMD001-002.
cv_1=CVs for receipts in domain *k* from replicate samples
cv_2=CVs for receipts in domain *k* from replicate samples with 1-in-2 SRS subsampling
cv_3=CVs for receipts in domain *k* from replicate sample with 1-in-3 SRS subsampling

United States® Census Bureau

2023 AIES Production Sample for the Manufacturing Sector

- The decision was made to use a 1-in-2 subsample for noncertainty strata with more than 14 phase one companies

- Estimates and CVs for the 1-in-2 subsample design had lower variability and fewer outliers than the 1-in-3 subsample design
  - CVs with the 1-in-3 subsample were still below the 15% target for most geographic domain estimates, indicating that the design could be adjusted in future AIES collections
  - CVs that were extreme from the phase one replicate samples became more pronounced with subsampling

- Modeling for nonresponse was a consideration even though the amount and quality of response was unknown during the design phase

United States® Census Bureau

2023 AIES Production Sample for the Manufacturing Sector

	Certainty Companies	Noncertainty Companies – Full Set of Items*	Noncertainty Companies – Four Core Only	Total
Used a 1-in-2 SRS Subsampling for Noncertainty Strata	22,000	11,500	**10,000**	43,500

*Includes second-phase certainties (companies in strata with fewer than 15 companies selected in the first phase)
* Subsampling with a 1-in-3 SRS moved an additional 3,500 noncertainty companies to the collection for the four core only
Note: Over 90% of the noncertainty companies receiving only the four core items were single-location companies (one establishment)

United States® Census Bureau

Future Considerations

- In Year 2 of AIES, switch the set of questions for the noncertainty panels
- Analyze Response Rates and Quality Measures
  - Hypothesis is noncertainty companies reporting data for only the four core items should have more and higher quality response data
- Compare results with imputation-based approach
  - Weighting causes oddities in micro data review
    - Ex. receipts and inventories have different weights within the same unit
  - Increased variability of estimates with weighting approach
- Research expansion of matrix sampling in AIES to reduce burden

United States® Census Bureau

Thank You!

- Mark Jost
- Stephen Kaputa
- Robert F. Miller
- Katherine J. Thompson
- Yeng Xiong

Contact: Colt.S.Viehdorfer@census.gov

United States® Census Bureau