

Developing a Sampling Frame for the Annual Integrated Economic Survey

Justin Z. Smith, Adam Smeltz

Economic Statistical Methods Division

James Burton, Economic Management Division

U.S. Census Bureau

Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (CBDRB-FY22-ESMD010-020).

Outline

- Annual Integrated Economic Survey (AIES) Background
- Definitions
- Initial Frame
- Sampling Frame
- Challenges and Solutions

AIES Background

- Historically, the U.S. Census Bureau's annual economic statistics programs developed independently
- National Academies of Sciences panel recommendations helped lead to the AIES
 - See National Academies of Sciences, Engineering, and Medicine (2018). Reengineering the Census Bureau's Annual Economic Surveys. Washington, DC: The National Academies Press.

AIES Background

- The AIES program aims to create a single frame that can be used by the integrated survey covering the target populations of the following precursor annual surveys:
 - Annual Capital Expenditures Survey (ACES)
 - Annual Retail Trade Survey (ARTS)
 - Annual Survey of Manufacturers (ASM)
 - Annual Wholesale Trade Survey (AWTS)
 - Manufacturers' Unfilled Orders (M3UFO) Survey
 - Service Annual Survey (SAS)

Definitions

- Establishment: an economic unit, generally at a single physical location, where business is conducted or where services or industrial operations are performed
- Enterprise: an economic unit comprising one or more establishments under common ownership or control, the top-level U.S. parent company and all subsidiary U.S. companies of which the parent owns or controls a majority (more than 50 percent) interest plus all EINs and establishments affiliated with the parent and its subsidiaries—i.e., a domestic enterprise group (For this presentation we will use the term company to mean enterprise.)
- Employer Identification Number (EIN): an administrative unit that the IRS has assigned a unique identifier for use in tax reporting

Definitions

- Single unit (SU): If an enterprise, an EIN, and an establishment represent one and the same thing
- Multiunit (MU): If an enterprise has
 - One or more EIN(s) and
 - Two or more establishments

Frame Creation

- General flow of the AIES frame creation process:
 - create the full frame
 - apply the in-scope criteria to create the initial AIES frame
 - create out-of-scope table
 - create the broad line table for AIES establishments
 - create AIES classification variables for sampling purposes
 - create the sample frame

Approximate Company and Establishment Counts

Source	Companies	Establishments
Full Frame	18,800,000	20,000,000
In-scope frame	7,450,000	9,200,000
Sample	400,000	2,000,000

Full Frame

- Extract a comprehensive list of employer establishments from the Business Register (BR)
- The BR is a central, multipurpose, business list, a primary source of survey frames, and contains administrative data, such as tax records from the Internal Revenue Service (IRS)
- An employer establishment is an establishment that has paid employees

Initial Frame – Apply In-scope Criteria

- Type
 - Include active establishments classified as SU or MU entities
- Business Activity
 - Include establishments that have either
 - payroll in the current year, or
 - payroll or receipts from the most recent Economic Census
- Geography
 - Include establishments operating in the 50 states plus DC
- Classification
 - Include employee establishments in the manufacturing, mining, utilities, construction, wholesale, retail, agricultural services, and services sectors

Initial Frame - Exclusions

- To exclude:
 - Out-of- scope industries such as Crop production, Animal Production and Aquaculture, Post Office, Private households (home businesses with no paid employees), and Offices of notaries
 - Establishments classified in the public sector (government)
 - Establishments with incomplete/inadequate NAICS classification (partial NAICS)
- Establishments that do not meet in-scope criteria will be output to an out-of-scope dataset with flags indicating which criteria was not met

Challenges – Partial NAICS Codes

- Matching 6-digit codes in AIES to an official list of all valid 6-digit codes showed about 300 in-scope establishments have a partial code
- If the code was partial, then the national tabulation industry was set to blank
 - Unless the partial code is an ACES-only level that is needed at less than 6-digits
- This will set the company initial probability of selection at the national industry level to blank, and the initial probability of selection will be that associated with the geography level

Initial Frame – Assign Variables

- Assign state
 - Best Federal Information Processing System (FIPS) code set for SUs based on state where establishment is physically located (if known), otherwise set to mailing state
 - MU establishments should always have a physical FIPS
- Assign wholesale type of operation code (TOC)
 - TOC is necessary for assigning wholesale establishments to AIES sampling recodes
 - Identifies between Distributors, Manufacturers' Sales Branches and Offices, and Agents, Brokers, and Electronic Markets
- Legal form of organization is used to assign tax status for selected service industries

Challenges – Wholesale Type of Operation Codes

- The type of operation code (TOC) identifies a wholesale establishment as either a Distributor (1X), Manufacturers' Sales Branches and Offices (2X), or Agents and Brokers (4X)
- Merchant Wholesalers (NAICS 423 and 424) can be either 1X or 2X while Electronic Markets and Agents and Brokers (NAICS 425) can be only 4X
- On the initial frame, there were many instances of establishments in NAICS 423 or 424 with TOC=4X and establishments in NAICS 425 with TOC=1X or 2X
- The TOC codes were thoroughly reviewed and rules were established for assigning them during the frame extraction

Challenges – Tax Status for Services

- While reviewing the initial tests of sampling, we noticed some certainty strata were created because they contained a single company on the frame
- Some of these strata had a NAICS and tax status combination that we believed should not exist. The tax status on the frame is correct because a company in any industry can meet requirements to become tax-exempt
- We reviewed which industries for which the Service Annual Survey (SAS) publishes either only taxable or only tax-exempt estimates. If an establishment was in a NAICS that SAS only publishes taxable estimates (e.g., general automotive repair), but it was listed as tax-exempt on the frame, its tax status was reassigned to taxable for sampling

Initial Frame – Payroll As Measure of Size

- On the frame we have payroll, employment, and receipts
- Payroll was selected to use for measure of size (MOS) because:
 - Employment has many zeroes, because it is employment as of March 12th
 - No receipts at establishment level, available at EIN level but not at establishment level, outside of the Economic Census
 - Due to timing of 2024 mailout, data review for 2022 Economic Census will probably not be complete

Initial Frame – Measure of Size Creation

- Create a Measure of Size (MOS) variable for each establishment based on administrative annual payroll from the current reference period (2022 for the production frame)
- The MOS at the establishment level is input to key sampling parameters
 - Strata allocations
 - Company inclusion probability

Initial Frame – Measure of Size Creation

Establishment MOS hierarchy

Step	Description	Notes
1	Sum current annual payroll and annual payroll for leased employees	SU and MU
2	Annualize payroll using quarterly administrative payroll	SU only For example, if only Q1-Q3 are present, then $ANNUALIZED_PAYROLL_2020 = (4/3) * \text{sum}(Q1, Q2, Q3)$
3	Adjust Economic Census payroll (2017) to current year payroll	SU and MU For example, calculate adjustment factors $\text{median}(ANNUAL_PAYROLL_2020/ANNUAL_PAYROLL_2017)$ at the NAICS4 level using those cases that have positive $ANNUAL_PAYROLL_2020$ and $ANNUAL_PAYROLL_2017$
4	Sum EIN's Quarterly Payrolls for the year / # of EIN's establishments	MU only For example, for an MU as part of an EIN with 6 establishments: $\text{sum}(Q1, Q2, Q3, Q4)/6$

Challenges – Auxiliaries

- Some MU companies have auxiliary establishments (e.g., a warehouse that only services the company) which are considered out-of-scope to AIES
- Including or excluding auxiliary payroll could change the probability of inclusion
- Research is underway to determine if payroll from these auxiliaries should be included in the company MOS, and if so, how

Sampling Frame

- From the SU and MU in-scope establishment frames, the following are created:
 - Combined establishment frame
 - Certainty company frame
 - Certainty national industry roster
 - Certainty industry by geography roster
 - Noncertainty company sampling frames by sector
 - Noncertainty national industry roster by sector
 - Noncertainty industry by geography roster by sector

Rosters

Company (Sampling) Unit

Company	Sampling Stratum	Company MOS _c
1	2 Industries	181
2	> 1 State	64
3	AAAA_MO	54
4	AAAB_MO	100

Roster 1: National Industry Tabulation Units

Unit	Company	"Sampling Stratum"	MOS
1_AAAA	1	AAAA	104
1_AAAB	1	AAAB	77
2_AAAA	2	AAAA	64
3_AAAA	3	AAAA	54
4_AAAB	4	AAAB	100

Roster 2: Geography Industry Tabulation Units

Unit	Company	Ind	State	MOS
1_AAA_MO	1	AAA	MO	118
1_AAA_TN	1	AAA	TN	63
2_AAA_MO	2	AAA	MO	37
2_AAA_TN	2	AAA	TN	27
3_AAA_MO	3	AAA	MO	54
4_AAA_MO	4	AAA	MO	100

Combined Establishment Frame

- National tabulation industry and geographic tabulation industry get set
 - national tabulation industry is the 6-digit NAICS plus a sector-specific detail
 - geographic tabulation industry is the 3-digit NAICS plus a sector-specific detail
- Each establishment is assigned to one of 27 geographies based on its state FIPS code. The 27 geographies include 23 direct-use states and 4 balances of region
- Indicators are set on some establishments to select them with certainty (probability = 1)

Geographic Groups for AIES Sampling Design

Region	Total States	Direct-Use States	Balance of Region States
Northeast	9	Massachusetts, New York, New Jersey, Pennsylvania	5 (Connecticut, Maine, New Hampshire, Rhode Island, Vermont)
Midwest	12	Illinois, Indiana, Michigan, Minnesota, Missouri, Ohio, Wisconsin	5 (Iowa, Kansas, Nebraska, North Dakota, South Dakota)
South	16	Florida, Georgia, Maryland, North Carolina, Tennessee, Texas, Virginia	9 + DC (Alabama, Arkansas, Delaware, District of Columbia, Kentucky, Louisiana, Mississippi, Oklahoma, South Carolina, West Virginia)
West	13	Arizona, California, Colorado, Oregon, Washington	8 (Alaska, Hawaii, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming)

Challenges – Invalid State Codes

- There are approximately 350 establishments with invalid state codes
- Because we have strata defined as 3-digit NAICS by Geography, they do not currently fit into these noncertainty strata
- A decision is still pending on if they should be excluded or how they could be included

Company Frame

- Aggregated values of quantitative establishment data
- Categorical variables such as NAICS and geography levels contain
 - A single value (e.g., state = MO) if all establishments associated with the company have the same value
 - “Mult” otherwise
- A variable that contains the company’s most detailed NAICS

AIES Designation of Initial Certainty Stratum by Complexity

Stratum	Contains Establishment Certainty	Number of Sectors	Number of 4-digit NAICS Industries	Number of Geographies
Certainty	Y	Not considered	Not considered	Not considered
	N	≥ 2	≥ 2	Not considered
	N	1	≥ 3	Not considered
	N	1	2	≥ 2
	N	1	2*	1

*Excluding retail (sector 44-45) and healthcare (sector 62)

Company Frame Stratification

- Remaining companies are each assigned to a noncertainty stratum within their sector
- Because we are selecting a PPES sample and some companies may have MOS = 0, we create separate stratum to select some of these companies
- Noncertainty companies in sectors which are in-scope for only national level capital expenditures (sectors 11, 21, 23, and 55) are assigned to their single sector stratum

Company Frame Stratification Continued

- Noncertainty companies in retail (sector 44-45) and healthcare (sector 62) that operate in more than one 4-digit NAICS are assigned to their sector's multi-NAICS stratum
- Noncertainty companies that operate in more than one state are assigned to the multi-geography by 3-digit NAICS stratum
- The rest of the noncertainty companies are assigned to one of the 27 geography by 3-digit NAICS strata

Noncertainty Strata Definitions

Stratum	Company MOS	Capital Expenditures Sector	Number of 4-digit NAICS Industries	Number of Geographies
Sector MOS=0 stratum	0	Not considered	1 or 2**	≥ 1
Sector stratum	> 0	Y	1	≥ 1
Sector multi-NAICS stratum	> 0	N	2**	1
3-digit NAICS multi-geo stratum	> 0	N	1	≥ 2
3-digit NAICS by geo stratum	> 0	N	1	1

**Only sectors 44-45, 62

Challenges – Sampling NAICS level differences between ACES and the other annual surveys

- The ACES survey overlaps NAICS codes with the other five surveys but samples at a less detailed level
- Create final strata as 3-digit NAICS by Geography and evaluate the national level CVs at the 6-digit level, except for the NAICS that are only included in ACES

Thank you!

justin.z.smith@census.gov