

Building a Sampling System for the Annual Integrated Economic Survey

Discussion of papers at
2022 Federal Conference on Survey Methodology

Washington, DC
October 27, 2022

BR

Dr. David Marker
Marker Consulting, LLC
MaryandDavidMarker@Gmail.com

Creating the Sampling System for the Annual Integrated Economic Survey (AIES)

- ▶ Congratulations to all the speakers and co-authors
- ▶ Really difficult undertaking to create the AIES
- ▶ Summarize key recommendations of NAS Panel
- ▶ Identify those being addressed
- ▶ Raise questions about those not addressed

Recommendation 8-1 (page 159)

- ▶ To improve the utility of annual business data,... the U.S. Census Bureau should develop a detailed concept and implementation plan for an Annual Business Survey System
 - ▶ Use a rotating panel sample drawn from a redesigned comprehensive Business Register
 - ▶ Use administrative records to the maximum extent possible for greater efficiency and accuracy of data and reduced respondent burden
 - ▶ Incorporate small-area models to produce subnational estimates

Business Register

- ▶ Using a common single register
 - ▶ Bringing in lots of administrative data, e.g., tax records
 - ▶ No mention of updating register from information collected from data collection efforts
- ▶ Recommendation 3-1d (page 49)
 - ▶ “Adopting processes for routinely incorporating in the register updated information about business structure”
 - ▶ These presentations focus on the first round of data collection, but need to incorporate plans for
 - ▶ Updating sampling frame and sample
 - ▶ Rotating non-certainty

Common Sampling Frame

National - 6-digit NAICS

State - 3-digit NAICS

Easy to create for single-establishment companies

How to handle for complex firms?

- Multiple locations

- Multiple states

- Multiple industries

Questions about what is Missing

- ▶ Companies with partial NAICS codes, still have $\pi_i > 0$?
- ▶ Over 10,000,000 companies are out of scope (Smith p8)
- ▶ Self-employed without employees?
 - ▶ BLS reporting problem, not Census
 - ▶ Missing payroll MOS?

Complexity Certainties

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)

Complexity Certainties

- ▶ Absolutely have to include such companies in sample
- ▶ Why the entire company, however?
- ▶ Why every Home Depot? 19 in Central Maryland
- ▶ Only if easier for respondent
 - ▶ OK for big companies; but for mid-sized?
- ▶ Recommendation 5-2 (page 99)
 - ▶ “Review whether the establishment-based ASM and MOPS could sample in two stages - enterprises followed by establishments within enterprises”

Overall Sample

- ▶ Previously the 6 surveys had a sample size ~250,000 companies
 - ▶ Actually 200,000 companies & 50,000 ASM establishments
- ▶ AIES has a sample size of ~400,000
 - ▶ To address state-level estimates
- ▶ Certainties reported on their industry, now all industries
 - ▶ Why should Costco and Barnes & Noble report on food sales?
 - ▶ If they can separate, just collect on certainty industry
- ▶ What proportion of sample are certainties?

Complexity Certainties

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)

Is Balance of Region Useful?

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)

Account Managers - Recommendation 3-3 (page 51)

- ▶ “The Census Bureau should establish a centralized and coordinated Account Manager Program that serves as a single point of contact for the largest enterprises ... [obtain] up-to-date information about these companies but also the coordination and facilitation of company responses.”

Account Managers

- ▶ No mention in any of these talks
- ▶ You are asking a great deal from larger companies
 - ▶ How do they organize the data you need?
 - ▶ Is it easier to report on all locations in a state or only a few?
 - ▶ For all business lines, or just where they are large?
- ▶ How do those answers change by industry?

Probability of Selection and Survey Weights

- ▶ Square root of MOS is a reasonable compromise
- ▶ What proportion of MOS from certs? Otherwise $n=10$ is very small

		2+	MD	MO	MT	OH
AAA	Companies (N_h)	21	245	244	245	245
	Allocation (n_h)	6	10	10	11	10
	Sampling fraction (f_h)	0.29	0.04	0.04	0.04	0.04
AAB	Companies (N_h)	23	3,330	11,654	1,665	3,328
	Allocation (n_h)	23	202	381	146	202
	Sampling fraction (f_h)	1.00	0.06	0.03	0.09	0.06

Winsorization

- ▶ I agree with the idea of adjusting probabilities of selection rather than trimming weights
- ▶ I was surprised with how much of the sample is being adjusted, 40-50%
- ▶ Should dramatically improve precision
- ▶ Better to see this as consistent with Recommendation 5-3

Recommendation 5-3 (page 101)

- ▶ “The Census Bureau should study the statistical efficiency of probability proportional to size sample designs ... and convert them to stratified, equal probability designs.”
- ▶ Have you considered doing this for mid-sized strata as well?

Key Issues to Address

- ▶ Is large increase in sample size really necessary?
- ▶ How much of the sample are certainties? Can this be reduced
- ▶ Why ask certainties for one industry to report on all others?
- ▶ Plans for rotating non-certainties out of sample
 - ▶ Hopefully not in for more than 2 or 3 years
- ▶ Account Managers can be very helpful in getting large companies to participate in timely fashion