Data-Driven Decision Making and the Design of Economic Census Data Collection Instruments

Eric B. Fink, Jennifer L. Beck, and Diane K. Willimack

U.S. Census Bureau[1], Washington, D.C. 20233

Proceedings of the 2015 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

As it is becoming increasingly common for surveys at the U.S. Census Bureau to shift to a single-mode electronic instrument, survey managers will have to think about both the opportunities and the risks associated with such a transition. At the extreme, the adoption of an electronic instrument allows a "form" to be uniquely customized to a respondent. This has the desired outcome of potentially reducing respondent burden to a substantial degree. In this paper, we explore this idea of a customized form by looking at how products may be pre-listed in the upcoming 2017 Economic Census.

Keywords: Instrument design and pretesting, multi-mode and web data collection, Economic Census, NAPCS

Introduction

The economic census, which the U.S. Census Bureau conducts every five years, is undergoing a major re-engineering effort. For the 2012 Economic Census, establishment-level data were collected using two-self-administered modes: mail-out/mail-back and electronic. For the 2017 Economic Census, only electronic reporting will be available. This presents unique challenges and opportunities for designing electronic instruments. For example, in past censuses, paper questionnaires contained long, detailed lists of products that filled multiple pages; respondents had to search for goods and services in this list and enter revenue data. A Web instrument offers automated features that may reduce the burden of sifting through long product lists, and improve data quality.

This paper focuses on the planned electronic collection of detailed product lines for implementing the North American Product Classification System (NAPCS), a comprehensive demand-oriented classification system that is being developed by the statistical agencies of the United States, Canada, and Mexico. As part of this effort, we analyzed product information that businesses reported in the 2012 Economic Census to discern reporting patterns associated with industries in the economic census. These reporting patterns may be leveraged in 2017 instrument design, providing a familiar starting point for respondents to begin supplying product-level receipts. We use the results from this analysis, in conjunction with usability testing, to aid development of an effective design for obtaining product-level detail.

We demonstrate how analyzing respondent reported data, paired with instrument design testing, may be used to inform instrument-design decisions for business surveys. As many Federal surveys increasingly rely on electronic-based data collection, electronic instruments can provide many benefits over traditional paper forms. The electronic instrument need not have the look and feel of its paper-based counterpart to be effective, but can be uniquely customizable to the respondent to improve the survey experience, thus reducing respondent burden and increasing data quality.

The paper will be organized as follows. First, background on both the economic census and NAPCS will be given, and then a discussion of the methods will follow. Finally results will be presented and discussed with ideas on next steps.

[1] Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Background

Economic Census

The data used for this paper come from the 2012 Economic Census (EC). This section covers basic 2012 EC methodology (for more information, please see http://www.census.gov/econ/census/), including basic sample design, data collection, and estimation.

The EC is conducted every five years, for reference years ending in 2 and 7. It covers the eight major trade areas of the U.S. economy including manufacturing, construction, mining, retail, services, wholesale, finance-insurance-real estate (FIRE), and utilities-transportation. Collected information consists of employment labor costs and output, assets, expenditures, inventory, and other industry-specific items. The information gathered by the EC serves as input to calculate gross domestic product, among other economic performance measures. The data are also used as a benchmark, serve to update the sampling frame for other economic survey programs, and is extensively analyzed by the business and academic communities.

Sample Design

There was a sampling component to the 2012 EC. Of the roughly 6.5 million U.S. business establishments in the eight major trade areas, approximately 3.9 million were selected to participate in the 2012 EC. Another 2.4 million were in scope, but not mailed a survey form. As estimation was at the sector level, the sample design differed among the trade areas. Table 1 summarizes the sample designs in the major trade areas[2] (for more information, please see https://www.census.gov/econ/census /help/sector/ sources _ of _ the _ data .html).

Wholesale	Manufacturing and Mining	Retail, Services, and Utilities-Transportation	Finance, Insurance, Real Estate	Construction
Take all units with certainty	Take all units with certainty. All multi-unit businesses receive a long form. Based on payroll cutoffs for single-unit establishments, unit receives a long form, short form, or no form. Impute units that receive no form with administrative data	Take those with the largest payroll with certainty Sort remaining cases on 8-digit NAICS code and payroll and take a systematic sample. Fill key items** of non-selected units with administrative data. Adjust other items using an expansion method	In select industries, take all units with certainty In all other industries: •Take those with largest payroll with certainty Sort remaining cases on 8-digit NAICS code and payroll and take a systematic sample. Fill key items** of non-selected units with administrative data. Adjust other items using an expansion method	For single-unit businesses with complete 6-digit NAICS, stratify by state and NAICS. Take small strata with certainty. Within remaining strata, take a probability-proportional-to-size (payroll) sample For single-unit businesses with incomplete 6-digit NAICS, take those with largest payroll with certainty. Take a simple random sample of remaining cases

Table 1: 2012 EC high level sampling schemes across the major trade areas.
**Key items include receipts, payroll, and number of employees

Data Collection

Although sample designs varied across trade areas, the 2012 EC data collection was one massive operation. In general, the EC reporting unit is the establishment. The main exception is multi-unit business operations that have a predetermined arrangement that one reporting unit will report for multiple establishments. For the 2012 EC, survey data were primarily collected using two self-administered modes: mail-out/mail-back paper and electronic. Electronic collection was through one of two custom-built Census Bureau software products, depending on the size of the business. For multi-establishment businesses and large single-establishment businesses, data were collected

[2] MU's are taken with certainty across all trade areas.

through the downloadable software Surveyor. For other single-establishment businesses, data were collected through the Web application Centurion. There were other collection modes, such as telephone, for a small portion of cases. For the 2017 EC, data collection will be electronic.

Each single-establishment business received one of the following three types of survey questionnaires: a classification questionnaire, a "long" questionnaire, or a "short" questionnaire. The classification form collects industry classification information. Establishments not receiving a long or short form, for which we have insufficient identifying information, receive a classification form. The long and short forms both collect information on key economic items such as receipts, payroll, and employment. The long form, however, collects this information at a more detailed level. For example, where the long form asks for annual payroll by job, the short form asks for total annual payroll. Single-establishment businesses that received the long form also receive a supplemental form, which collects information about the ownership and control of the establishment. Some long form recipients also receive a second supplemental form, which collects information on items such as foreign ownership, research and development activity, royalties, and manufacturing activities. Most reporting units received a long form. Only small reporting units for selected industries in the manufacturing and mining sectors may have received a short form.

Although response to the EC is required by law, historically, the Census Bureau has employed an intensive contact strategy, especially targeted to large, multi-establishment businesses, in order to maintain high return rates (roughly the number of forms returned over the number of forms mailed to establishments). The Census Bureau targets these large businesses, because we tend to estimate totals for variables of interest. It is for this reason, that totals are the estimate of choice in economic statistics that the analysis below focuses on totals as well.

For the 2007 and 2012 ECs, contact initially began in the autumn of the reference year, when large multi-establishment companies were first provided with EC forms and notified about electronic reporting procedures. The purpose of this early contact was to facilitate businesses' planning for EC response as the reference year ends. Mail-out to the rest of the units eligible for the EC occurs mid-December of the reference year, with a due date of February 12 the following year. There were as many as four follow-up mailings to non-respondents, the first of which began approximately one week after the due date. Intervals between subsequent mail-outs varied, but were roughly 30 days. The final mailing for the 2012 EC occurred in July 2013. Follow-up plans have not been solidified for the 2017 EC.

Estimation

In general, the 2012 EC produces estimates of totals at all North American Industry Classification System (NAICS) levels and on key subgroups. Key estimates across all trade areas include total revenue, payroll, and employment. Estimates are being released on a flow basis, with final estimates by industry and zip code. For the 2012 EC, no measures of sampling error are being published, except for the construction trade area (for more information, see https://www.census.gov/econ/census/help/methodology_disclosure/data_processing_and_treatment_of_nonresponse.html). For nonsampling error, the proportion of receipts that were imputed and derived from administrative data were published. For product estimates, we publish the proportion of an estimate that came from response and verified secondary source data.

North American Product Classification System (NAPCS)

The North American Product Classification System (NAPCS)

As part of the EC, businesses provide detailed information about the revenue generated through the goods and services they provide. In past censuses, this reporting has been geared toward the products and services within the businesses' main industry, determined by the NAICS. Questionnaires, both paper and Web modes, reflected this focus, containing detailed lists of industry-specific products, and providing only a small amount of questionnaire space to indicate anything else that might lie outside the main industry.

Beginning in 2017, the Census Bureau will introduce a new way of collecting and disseminating the revenue information through the North American Product Classification System (NAPCS) framework. The goal of the framework is to provide an economy-wide picture of goods and services. The focus will be on where goods and services are sold, not just in the primary industry in which they are produced, capturing a snapshot of the demand-side economy.

Within the NAPCS collection structure, there are some changes to the way that goods and services are described, and therefore, differences in how respondents will identify and report goods and services when filling out the survey. One of the biggest changes to this process is how to address business activities and goods and services that may be atypical for a business. Under the NAPCS framework, businesses will be asked explicitly to report detailed information on all of the goods and services they offer, even if those goods and services are not typical for that business' industry. For example, a car dealership's primary source of revenue may be in car sales and car repairs. These services would make up most of the revenue the business generates. However, the dealership may also sell clothing in the form of t-shirts with the dealership's logo on it. T-shirts would be a primary good of a clothing retailer, but within the NAPCS framework, it is the goal to capture revenue from all t-shirt-sales, including sales that occur in car dealerships.

In previous censuses, the t-shirts sold at a car dealership would have been classified as "miscellaneous revenue," and reported in the limited write-in spaces allotted on the questionnaire. With the NAPCS structure, although these items might make up a small fraction of the business' revenue, respondents should still report these products, providing a description of them and their associated revenue.

The multi-mode approach to NAPCS collection

The 2017 Economic Census will usher in a new set of challenges for the Census Bureau. This new way of collecting product information necessitates a change in how we ask questions about detailed product information to ensure that we are able to capture information on all of the goods and services that businesses provide. The transition in the NAPCS collection structure prompted us to embark on a program of research to help identify an optimal way to collect this information that both maintained data quality and also eased respondent burden.

Analysis

Analysis Variables

In analyzing 2012 Economic Census data, we used data from the Census Bureau's microanalytical database (MADb). Product lines data are available for both the services sector (which includes wholesale and retail industries, among others) and the manufacturing sector (which also includes mining and construction). For this analysis, we focused solely on the services sector. The product lines come in two forms, as a broad line or a detail line. The detail line, as the name suggests, is a more detailed product description than the broad line. As an example, a broad line might be dog food, whereas a detail line might be organic soft dog food.

The analysis focuses on respondent data. We focus on respondent data because our goal is to create an instrument that will make the response task easier for the respondent. Using data that have been imputed, raked, etc. will not help us understand what the respondent actually reported. The data are weighted, and the weights are incorporated into the analysis. The weight assigned to an establishment indicates the number of similar units that establishment represents. The rules for how data are weighted vary by trade area. Furthermore, only lines that were reported and deemed usable were included in the analysis. Finally, we focus on receipts, as receipts, along with payroll and employment, is one of the key estimates produced from economic census data.

Analysis Questions

Much of the research we have conducted is exploratory in nature. As such, many of the initial research questions necessitate only descriptive statistics to answer. Our overarching research goal was to utilize 2012 Economic Census respondent data to inform the pre-listing of products in the 2017 Economic Census electronic instrument. We further refined this general goal into several more manageable parts about business-respondent reporting behavior of product lines. We develop the following, initial questions:

1. Do respondents have a preference as to how an electronic instrument is presented to them?
2. How many broad lines would need to be pre-listed in order to obtain 80% of reported receipts?
3. How many broad lines comprise the bottom 10% to 20% of reported receipts?
4. What would a pre-listing based on our research look like compared to how we have gone about this in the past?

Results

Instrument Testing With Respondents

As previously discussed, one of the overarching goals of the analysis of product reporting behavior is to encourage response and simplify the response process for respondents. Survey questionnaires represent a form of

communication with respondents (Haraldson, 2013). The questions not only communicate the information *we* wish to collect, it also communicates something to the respondents about the importance and value of the information they provide. Good questionnaire design has content that engages the respondent with relevant information, communicating that the respondent has important information to contribute. Our research with respondents on the presentation of NAPCS products to respondents further reinforced these principles of good design and illuminated the contribution of the product-reporting analysis.

As part of the NAPCS data-collection for 2017, we are redesigning our existing questionnaires and introducing a new design for collecting NAPCS information. In 2014, low-fidelity usability testing on potential designs for collecting NAPCS information from respondents compared two different question designs. One version of the questions contained a list of proposed NAPCS products that were listed on a single scrolling screen. For some respondents, this list could be quite long, containing more than 50 different products and services. Respondents also received a version of the question that presented the same list of NAPCS products and services across multiple screens. Not only were fewer products presented at one time, the products were also ordered to present the most relevant products on the first screens. (It should be noted that products tended to be grouped in order of most relevant when presented in a single list). The findings indicated that respondents preferred seeing fewer products at a time, as seeing the products listed on a single screen was overwhelming.

In Fall, 2015, we continued the research on the optimal design for presenting NAPCS product information, conducting high-fidelity usability testing on two different question designs for presenting products and services. Because there were some issues with the initial design that presented products across multiple screens, we sought to refine this design to decrease potential response problems. Keeping with the idea of simplifying the information that respondents have to process, we once again compared the single-screen list of products to a design that simplified the response process into multiple steps.

Presenting products and services on a single screen encompasses two separate response tasks. First, the respondent must read through the list of products and services and select the descriptions that match their business' activities. Once they have identified those activities, they then determine the revenue associated with these activities. When the list of products is long, containing products and services that are irrelevant and the descriptions of the activities are unfamiliar, the response task becomes more difficult. Respondents indicate having to review the list carefully before beginning to report revenue information, a burdensome process. Separating the selection task from the dollar-reporting task allows the respondents to select the relevant information and focus their attention on only the information that is most relevant to them.

To this end, we tested a design that employed two screens. On the first screen, respondents selected the products and services that captured their business activities. On the second screen, respondents saw and reported the revenue associated with *only* the products and services that they had selected on the initial screen. The findings from this testing once again showed support for the notion that "less is more" in questionnaire design. Respondents preferred the multi-screen design to the single screen design. Respondents expressed a desire to see more relevant information, implying that they would like to see only the goods and services that are relevant to their business activities. While the multi-screen approach does not tailor the instrument to this degree, it did provide the respondents with the opportunity to reduce the amount of information they had to process when reporting revenue, allowing them to tailor the content themselves. Not only is there less information on the reporting screen, the information that the respondent sees is highly relevant.

These findings, coupled with the analysis provide a way for us to reduce the amount of information that respondents see, and also to offer a way to tailor the content and increase the relevancy of the products and services that respondents see.

Respondent Data Analysis

In the 2012 Economic Census, there were 2,155,058 establishments reporting 7,306,347 usable broad lines across 530 different industries. We first summed the receipts values of the broad lines reported in each industry (six-digit NAICS), and ranked them from those contributing the most to an industry's estimated total, to the least. We found that over all industries (see Table 2), reporting just the top broad line (measured as the broad line accounting for the largest percentage of total receipts) accounted for a median value of 74.65% of total receipts. Reporting the top three broad lines accounted for a median value of 93.29% of total receipts.

	Mean	Median	Standard Deviation	Minimum	Maximum
Top Broad Line	69.75%	74.65%	23.45%	11.51%	100.00%
Top Two Broad Lines	82.27%	88.81%	17.38%	20.94%	100.00%
Top Three Broad Lines	88.14%	93.29%	13.47%	27.89%	100.00%

Table 2: Sum of the top three broad lines, and the percentage of total receipts they account for.

To get a better idea of the variability, as Figure 1 shows, we looked at the percentage of total receipts that the top three broad lines contributed to an industry, and summarized this at the sector level (two-digit NAICS). We notice substantial variability across sectors. The utilities sector (sector 22) is clearly the sector where reporting the top three broad lines contributes the most to an estimated total of receipts, with a median value of 98.96%. Looking at the range, we see there is at least one industry within that sector where reporting three broad lines accounts for 89.78% of total estimated receipts, and there is at least one industry where reporting the top three broad lines accounts for 100% of total estimated receipts. The sector with most variability is retail trade sector (sector 44-45), where reporting three broad lines accounts for a median value of 93.04% of total estimated receipts. There is at least one industry where reporting the top three broad lines accounts for only 27.89% of total estimated receipts, and another industry where reporting the top three broad lines accounts for 99.89% of total estimated receipts.

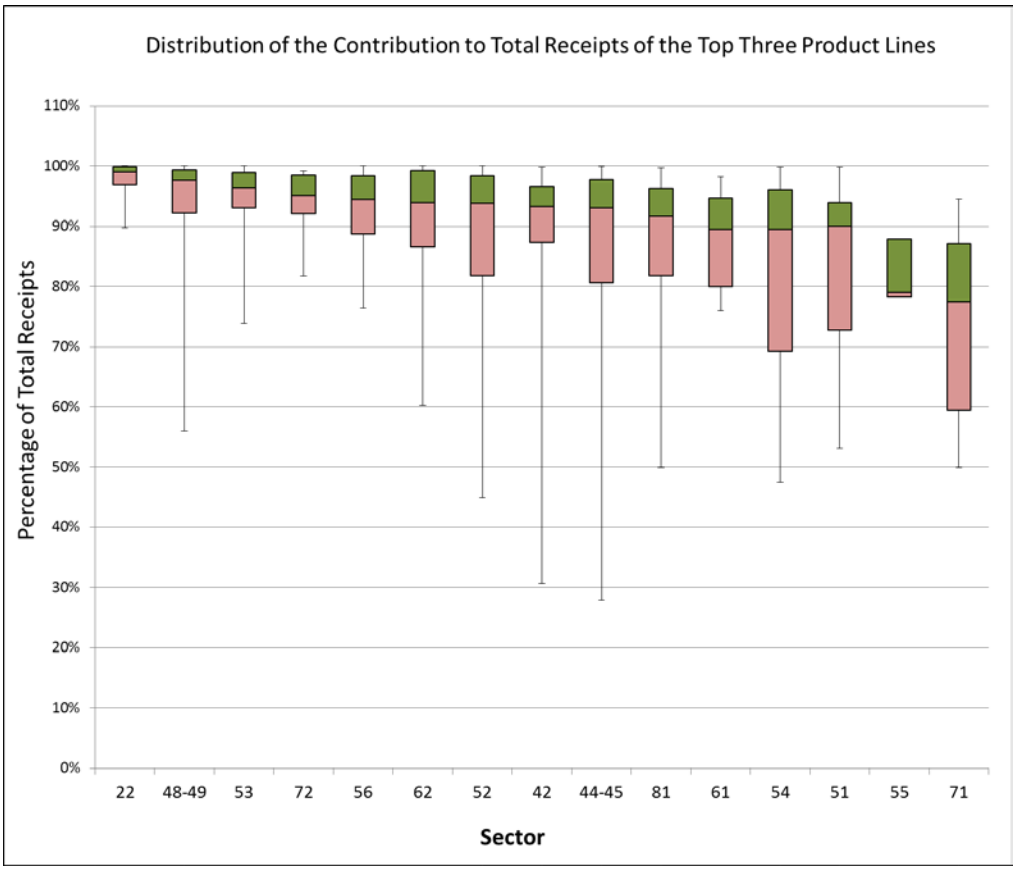

Figure 1: Top three broad lines, and the percentage of total receipts they account for, summarized at the sector level, arranged by decreasing median percentage value.

If the Census Bureau were to adopt a rule where we pre-list those broad lines accounting for 80% or 90% of total receipts, we want to know how many broad lines would typically be needed to capture the remaining receipts. As can be seen in the Table 3, we would still need to capture about 16 broad lines. There are some industries where we would need to only capture one more broad line, and others where we would need to capture between 117 and 131 broad lines, depending on the percentage of receipts we aim to obtain in the prelisting.

Broad Lines Comprising the Bottom Percent of Receipts	Mean	Median	Standard Deviation	Minimum	Maximum
10%	19.64	15	20.77	1	117
15%	20.37	16	21.05	1	124
20%	20.83	16	21.24	1	131

Table 3: Count of broad lines at the industry level comprising bottom 10% to 20% of total receipts.

Finally, we wanted to see how many establishments were reporting the top broad lines. Table 4 indicates that overall, 95.21% of establishments are reporting the top broad line, and we see that number drop to 23.31% for the second broad line, and 15.32% for the third broad line. Looked at another way, however, about one out of every four establishments are still reporting the second largest broad line.

	Mean	Median	Standard Deviation	Minimum	Maximum
Top Broad Line	82.20%	95.21%	23.97%	1.04%	100%
2nd Broad Line	31.50%	23.31%	25.38%	0%	100%
3rd Broad Line	22.63%	15.32%	21.73%	0%	100%

Table 4: The percentage of establishments reporting the top three broad lines as a percentage of the total number reporting in an industry.

Comparing Two Pre-Listing Approaches

Finally, as we wanted to compare the results suggested by our methodology to a pre-listing that subject-matter experts wanted to present respondents, we developed what we term the 80-80 approach. As a general rule, electronic instruments will be developed at the industry level. Our approach suggests that a broad line qualifies for pre-listing if it contributes towards 80% of total receipts for an industry, or 80% of establishments reported a particular broad line within that industry. The 80% threshold is a somewhat arbitrary number, but a high enough value that survey programs should be happy to obtain. For instance, Census Bureau (statistical standard D3) and Office of Management and Budget (Guidelines 1.3.4 and 1.3.5) Standards specify that unit response rates falling below 80%, or item-level response rates falling below 70% should result in a survey program conducting a nonresponse bias analysis. However, it must be stressed that our 80-80 approach is just that, an approach. We recommend more research be conducted to decide on appropriate approaches for each industry. However, the approach does offer a business decision rule that can be easily implemented and automated.

The figures below show the number of broad lines to be pre-listed on the x-axis. Those items colored in red make up 80% of total receipts and potentially have 80% of establishments reporting that particular broad line. All other items listed are those that have been suggested by subject-matter analysts. As can be seen from the figures, the additional lines above and beyond our 80-80 rule only marginally add to the captured revenue. At the extreme, only one product in Figure 2 is needed to satisfy our 80-80 approach, highlighted in red. All other products listed do not satisfy the 80-80 criteria.

Looking at Figures 3 and 4, both of which are in NAICS sector 72 (Accommodation and Food Services), one industry needs eight products to be listed in order to satisfy the 80-80 approach, whereas the other only needs two products to be pre-listed. This is just one example of many where we see that within sectors, there is a lot of variability between industries that can be exploited and leveraged in an electronic instrument.

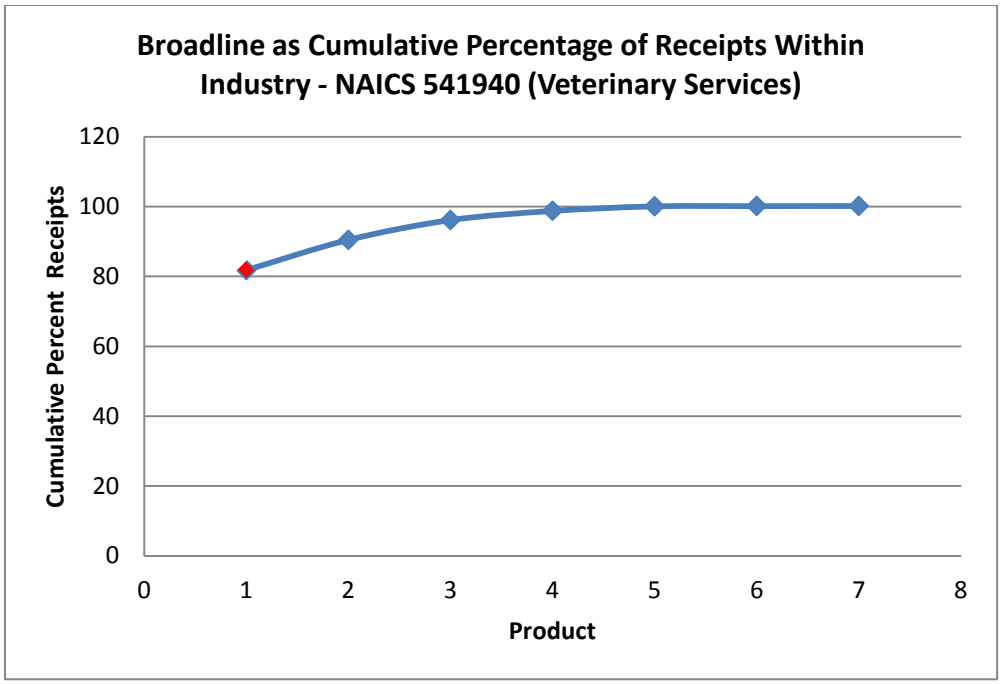

Figure 2: A proposed listing of products for NAICS 541940 (Veterinary Services). The lone product in red satisfies the proposed 80-80 approach.

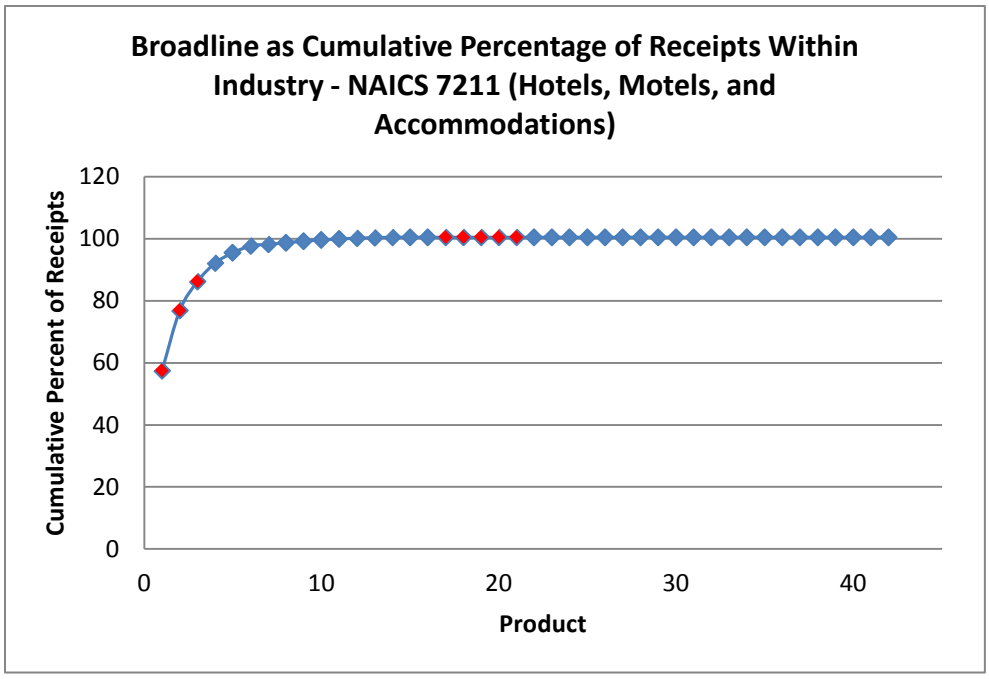

Figure 3: A proposed listing of products for NAICS 7211 (Hotels, Motels, and Accommodations). The products in red satisfy the proposed 80-80 approach. The reason for so many products between the first set that satisfy the 80-80 approach and the second, is the second set are more detailed lines of products given in the 2012 Economic Census that will now be broad lines in the 2017 Economic Census.

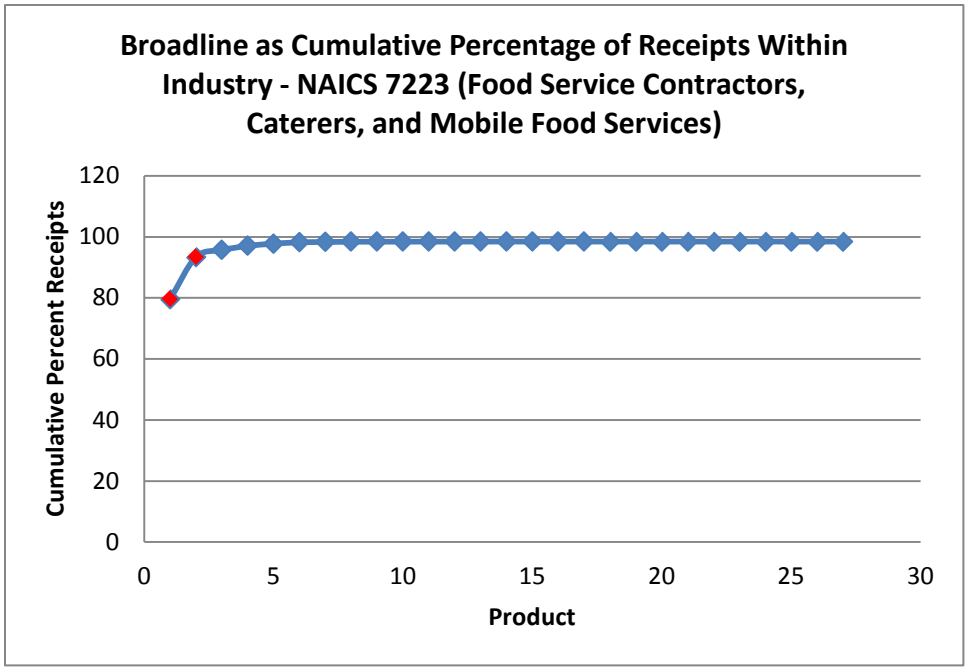

Figure 4: A proposed listing of products for NAICS 7223 (Food Service Contractors, Caterers, and Mobile Food Services). The two products in red satisfy the proposed 80-80 approach.

Discussion

The above results are a wonderful example of an effective interplay between qualitative and quantitative analysis. The qualitative results showed support for literature results hypothesizing that respondents preferred to see less information presented to them in a survey instrument, and wanted a survey instrument they found relevant to their type of business. The quantitative results demonstrated diminishing marginal returns to the amount of receipts captured after offering just a few broad lines. It also demonstrated that, on average, at least one out of every five establishments report the top three broad lines giving establishments a pre-listing relevant to their business.

There is, as we saw, substantial variability both between and within industries with respect to how many broad lines it would take to satisfy the 80-80 approach. This is where having an electronic instrument becomes a very powerful tool, and one that should be utilized to its fullest potential. We now have the power to customize an instrument at any level of granularity we want, from the sector, to the industry, all the way down to the individual establishment.

The potential drawbacks of prelisting a small number of questions are twofold: (1) Respondents may stop reporting products that are not prelisted, and (2) Respondents may write in products that are not pre-listed and the Census Bureau has to convert the write-in to an actual product code. The first drawback can potentially be mitigated with smart questionnaire design. An electronic instrument can be designed to encourage the reporting of products not prelisted. Additionally, one could make the argument that prelisting too many products could encourage respondents to stick whatever they have in the existing plethora of products because they assume that the list is exhaustive.

The second drawback can potentially be mitigated by using an autocoder. The American Community Survey (ACS) used an autocoder for industry and occupation codes (Day, 2014). The Canadian Census of Population in 1991 used autocoding for the write-in portion of its instrument (Ciok, 1993). Both experiences using an autocoder saved the agencies a great deal of money and human resources.

The approach we are proposing need not be restricted to the two dimensions of receipts and frequency of reporting, and these criteria can be programmed for production purposes. Indeed, we have already incorporated a third dimension whereby we look at the contribution of the industry to the total receipts for a particular product. The analysis for this third dimension was conducted too late to be incorporated into this paper. The take-home point,

however, is that whatever criteria one can think of for a pre-listing of items in an electronic survey, a thorough analysis can be conducted that be automated for production.

Increasingly, statistical organizations no longer have to be bound by a paper world, and as such, we should adjust our thinking to take advantage of the full potential of an electronic instrument.

Future Research

The above quantitative and qualitative results show promise for designing electronic instruments that reduce respondent burden and encourage response. Moving forward, there are several avenues we are exploring.

We had the benefit of conducting this analysis post-production. As not all establishments are correctly classified in an industry when they are selected in sample, a slightly broader set of products may be necessary so that establishments can be correctly classified. Therefore, in addition to the 80-80 approach, either discriminant or classification analysis may be necessary to figure out if additional products are necessary to properly assign establishments to an industry.

We may try to use embed this pre-listing approach as an experiment in the upcoming 2017 Economic Census and/or one of our economic surveys, such as the Annual Survey of Manufactures. Another potential avenue is to see if we can do a simulation of 2012 Economic Census data to examine the impact of pre-listing using our approach. Finally, we want to examine the literature to see how other countries and organizations tackle these issues.

Acknowledgements

The authors wish to thank Cha-Chi Fan, Suzanne Conard, and Scot Dahl for their helpful comments in reviewing this paper. We also wish to thank the analysts for a lot of thought-provoking discussion, and for putting together a pre-listing that could be used for comparison purposes. Finally, we thank John Kern for helping to guide us through, and answering numerous questions about the data set.

References:

Ciok, R. (1993). The Results of Automated Coding in the 1991 Canadian Census of Population. *Proceedings of the 1993 Annual Research Conference at the Bureau of the Census*, March 21-24, 1993.

Day, J. (2014). Using an Autocoder to Code Industry and Occupation in the American Community Survey. *Federal Economic Statistics Advisory Committee Meeting*, June 13, 2014.

Haraldsen, G. (2013). Business survey communication. In G. Snijkers, G. Haraldsen, J. Jones, & D. Willimack (Eds). *Designing and Conducting Business Surveys*, (pp. 303-357). Hoboken, NJ: Wiley.

Office of Management and Budget. 2015. Standards and Guidelines for Statistical Surveys. Located at: https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

United States Census Bureau. 2013. U.S. Census Bureau Statistical Quality Standards. Located at: http://www.census.gov/content/dam/Census/about/about-the-bureau/policies_and_notices/ quality/statistical-quality-standards/Quality_Standards.pdf