From Research to Implementation of Product Estimation in the 2017 Economic Census: Hard, Harder, and Hardest

William C. Davie Jr[1] (U.S. Census Bureau), Scot Dahl (U.S. Census Bureau), Katherine Jenny Thompson (U.S. Census Bureau)

Economic Statistical Methods Division, U.S. Census Bureau
4600 Silver Hill Road, Washington DC 20233

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Conference

Abstract

The U.S. Census Bureau conducts an Economic Census every five years, producing key measures of American business and the economy. The Economic Census requests information on the revenue obtained from products sold from all large businesses and a sample of smaller businesses. Beginning in 2017, the Economic Census will use the North American Product Classification System (NAPCS) to produce economy-wide estimates of products sold. This marks a major departure from the prior collections, which explicitly linked product codes to industry, and required the development of a single imputation approach for all products that is statistically defensible and operationally feasible.

A research team was assembled to recommend a unified method for treating missing product data. The team's evaluative approach relied on simulation, using empirical data from a purposively selected, small subset of industries as the basis for the study. The research was complicated by the nature of product data, which are characterized by poor item response rates, few available predictors, additivity-within-establishment requirements, many rarely reported products in an industry, and sampling effects. To avoid confounding treatment effects with respondent size effects, the study restricted the analysis variables to a limited set of products within each studied industry. This greatly simplified the evaluations, but left potential implementation challenges uninvestigated. This paper describes the recommended missing data treatment methods and how these methods are being implemented into the 2017 Economic Census production system. Examples are provided to illustrate implementation issues and the modifications and enhancements needed to fully implement the research-based recommendations.

1. Introduction

Improvements to official statistics programs may require complicated changes to existing methods or procedures to address new – emerging – requirements or to accommodate new requests. Such changes are always constrained. Budget constraints could force an overall reduction in sample size; methodologists would need to revise the sampling design while maintaining predetermined reliability levels on key statistics. Project sponsors might commission the collection of additional data items, request preliminary tabulations (and publication) of survey estimates, or desire subdomain estimates not considered in the initial design. Again, these applications would be constrained by reliability and confidentiality mandates. In any case, the appropriate solution is rarely obvious; research is required. Of course, time and resource constraints can be prohibitive, so the research problem may be simplified to allow for a transparent and repeatable solution, thus delaying the unaddressed details or unforeseen nuances until the implementation stage, leaving little time or resources for additional research.

Efficiency frequently dictates the research and implementation processes. In the Economic Directorate of the U.S. Census Bureau, it is a common practice to establish a "dedicated" research team with a fixed duration comprising representatives from the relevant job series with differing experience levels, perhaps utilizing matrix management. Team responsibilities encompass defining and scoping the research problem, obtaining data, designing and conducting the research, writing and testing programs needed to carry out the research, documenting the findings, and presenting the "data-driven" recommendation. Assuming that the recommendation is endorsed, a subsequent

---

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

implementation team is established. This team's composition can differ greatly from the research team, as expert staff are required – and production programmers must be included in the discussions – although some overlap in membership between the research and implementation teams is desirable. As with the research team, the implementation team usually operates under a fixed deadline. Team responsibilities include writing specifications that implement the research recommendation while addressing the issues that were "ignored" in the research. Logistical issues such as coding, testing, and validation are likewise included. An education component is not unusual, as the implementation team members may be unfamiliar with the methods under consideration.

The 2017 Economic Census leadership team endorsed a number of innovative updates, each introducing a new set of methodological and production implementation challenges. Historically, the Economic Census was a paper (mail out/mail back) collection; in 2017, collection will be primarily via the Web. In prior censuses, the collection unit for detailed breakdowns of dollar-valued totals varied by sector (percentage of total, $1000s, or both). In 2017, all detailed breakdowns of dollar-valued totals are collected in $1000s, as are the associated totals. Standard unit response rates will be released for the first time with the 2017 Economic Census, as will imputation rates for key statistics. Variability estimates for selected sample-based statistics will be published for the first time as well.

And beginning in 2017, the Economic Census will use the North American Product Classification System (NAPCS) to produce economy-wide product tabulations. The Economic Census collects a core set of data items from each establishment called general statistics items: examples include total receipts or value of shipments, annual payroll, and number of employees in the first quarter. In addition, the Economic Census collects information on the revenue obtained from the sales of good and services (hereafter referred to as "products"). The introduction of NAPCS marks a major departure from the prior collections which explicitly linked product codes to industry, allowing for different missing-data treatments for products by sector. Implementing a NAPCS-based collection necessitated the development of a <u>single</u> imputation approach for all Economic Census products to allow production of cross-sector tabulations.

This paper describes the research process used to determine the product imputation method and the process for implementing the research recommendations into the 2017 Economic Census production systems. Research and implementation were accomplished by two different teams, with a small fraction of membership overlap. Together, both teams developed and implemented the methodology that will be used in the 2017 Economic Census production systems.

Section 2 provides general background on the Economic Census along with more detailed background on product collection and estimation. Section 3 summarizes the research approach and resultant recommendations. In Section 4, we discuss the implementation of the recommended methods into a production system, specifically focusing on some of the unaddressed or unforeseen – but important – details that were excluded from the research study. We conclude in Section 5 with a few general observations about implementing research-based results in a production system.

2. Background on Economic Census and Product Data

The Economic Census is the U.S. Government's official five-year measure of American business and the economy. The term "Economic Census" is a bit of a misnomer as a sample of small single-unit establishments is surveyed in addition to all multi-unit establishments[2]. As mentioned in Section 1, the Economic Census collects a core set of data items from each mailed establishment called general statistics items: examples include total receipts or value of shipments ("receipts"), annual payroll, and number of employees in the first quarter. In addition, the Economic Census collects data on the revenue obtained from product sales. All sectors construct a *complete universe of general statistics values* by using administrative data (or imputation) in place of respondent data for non-mailed units. In contrast, sample weights are used to account for the non-mailed single-unit establishments when producing product sales estimates. Finally, for each industry, the weighted sample estimates of product sales are further

[2] A single-unit (SU) establishment owns or operates a business at a single physical location; whereas, multi-unit (MU) companies comprise two or more establishments that are owned or operated by the same company.

calibrated to ensure the sum of the product sales equals the total receipts (based on the complete universe) for the industry.

The 2017 Economic Census requests data on over 8,000 different products based on the North American Product Classification System (NAPCS); see https://www.census.gov/eos/www/napcs/more.html. However, evidence from prior economic censuses indicates that many products are rarely reported. Legitimate zero values are expected for the many products in an industry, at both the individual establishment and total industry levels. Respondents can report data from a long, pre-specified list of potential products in a given industry – some lists contain more than 50 potential products – and can write in descriptions of other products that were not pre-specified. Product lists can differ by industry within a sector. Furthermore, some product descriptions are quite detailed, and some products are mutually exclusive. Consequently, some establishments choose not to report any product data (complete product nonresponse). Those that do report often provide the same products typically reported by other responding establishments within a given industry (Fink, Beck, and Willimack 2015).

Several industries collect both "broad" and "detailed" products. Figure 1 presents a fictional illustration of broad and detailed products.

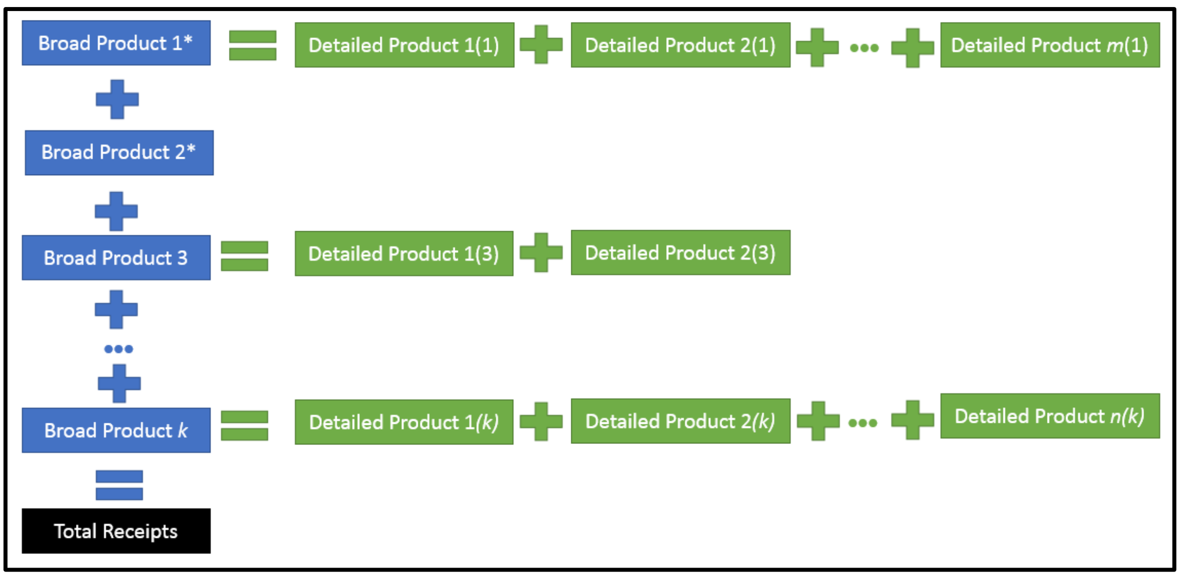

Figure 1: Illustration of nested balanced complexes defined by Broad and Detailed products in an industry. Broad products 1 and 2 are "must products" that have to be reported by an establishment to be classified into a particular industry; the "must product" status is indicated by an asterisk. Broad products 1 through k sum to the establishment's total receipts. Not all broad products request detailed product breakdowns on the questionnaire, and the number of requested details differs by item (broad product). For example, there are no detailed product breakdowns associated with broad product 2. Finally, detailed product values are expected to sum to the associated broad product value. The establishment will still be considered to have reported product values as long as Broad products 1 and 2 are reported; other broad products are considered optional.

Broad and detailed products comprise nested one-dimensional balance complexes. The broad product values within a given establishment are expected to sum to the total receipts value reported earlier in the questionnaire. Under NAPCS, the same broad products can be collected in different industries. Detailed product values are expected to sum to their associated broad product value. Additionally, a particular detailed product is associated with only one broad product. Missingness tends to be higher with detailed products than broad products. Lastly, many industries have required "must products" for establishment classification. Thus, although the same product can potentially be produced in different industries, product reporting is intertwined with industry classification. Certainly, the product distributions will differ between industries, even when the same products are reported. And of course, the detailed products will differ by industry. As a result, imputation cells cannot be collapsed beyond the most detailed North American Industry Classification System (NAICS) category provided by the sector.

It is not easy to develop viable imputation models for products. Auxiliary product data are not readily available. Moreover, other predictors such as total receipts are often weakly related. In most industries, the frequently reported products are highly correlated with total receipts and generally make up the majority of the total value of receipts, whereas the remaining products are not. Thus, the best predictors of an establishment's products are the industry assigned to the establishment from the sampling frame, which may change after collection, and the total receipts value (Ellis and Thompson 2015). Product distributions within the same industry tend to differ by unit type (single- or multi-unit), probably because multi-unit establishments must complete a separate questionnaire for each of their establishments. Product reporting propensities likewise differ within industry by unit type.

Any missing data treatment for products must ensure that the product totals sum to the industry receipt totals. This is not guaranteed automatically, as the industry receipts totals are computed from the complete population frame described above; whereas, products are only collected from sampled units. After processing, the product totals are calibrated to the industry totals via a simple ratio adjustment. In previous Economic Censuses, this final calibration was performed in tandem with the missing-data treatment for products for all sectors except manufacturing, mining, and construction.

3. Missing Data Treatment Research for Broad Products

Thompson and Liu (2015) give an overview of the large scale research project conducted to determine a single, unified imputation method for Economic Census broad products under NAPCS: more details are provided in Ellis and Thompson (2015), Garcia, Morris, and Knutson (2015), Tolliver and Bechtel (2015), Bechtel, Morris, and Thompson (2015), and Knutson and Martin (2015). The research was undertaken by a commissioned team whose members included methodologists, subject matter experts, and classification experts. The latter two groups developed the test data used for all analyses and provided expertise on the 2012 Economic Census procedures; the 2017 collection procedures and NAPCS-based collection structure were under development during the time of the research. The methodologists' familiarity with the subject matter and expertise on the current procedures ranged from completely novice (the majority) to extremely knowledgeable about a selected subset of industry-specific procedures. Both team leads were methodologists who were familiar with Economic Census processing procedures and methods in general but had little or no experience with the specific procedures used in product processing.

The research team was initially given six months to complete its work to allow time for implementation into the production system. The six months was eventually extended to nine months. During this time, the team had to learn about product data and distributions, develop the research methodology, write and test programs, conduct data processing, analyze the results, and prepare a report containing a final recommendation. Finally, the research project was conducted during a peak production processing time for the 2012 Economic Census, which meant the team's computing resources were shared with other production processes.

The team divided the project into the three separate components listed in Table 1, each lasting approximately two to three months. The project started slowly, with the acquisition of a processing environment and historical data sets. Subject matter experts extracted the Economic Census test data from industries provided by the classification experts. They also provided classification rules for donor records (whose values can be used for imputation) and recipient records (need an imputed value), thus ensuring that industry-specific "must-product" rules would be enforced by any imputation method. The classification experts provided industries whose product distributions were expected to remain largely the same under NAPCS. Even so, the historical product data were not expected to be perfect predictors of the 2017 product data because (1) some of the 2012 Economic Census product data were reported as percentages of total receipts; whereas, in 2017, all product data values will be collected in \$1000s; and (2) in 2012 businesses could respond by paper and electronically, but in 2017 only electronic reporting will be used.

From the beginning, the team agreed to study only broad products, as the rate of missing data for detailed products was quite high. Moreover, the team decided to limit the study to national-level industry estimates of products, even though industry-by-state level estimates are available in selected industries. A slow start was inevitable, as the majority of methodologists were unfamiliar with the Economic Census. Indeed, by the completion of the exploratory data analysis component, the team members realized the enormity of the task and the restrictiveness of the deadline.

Table 1: Research Components

Component	Purpose	Leaders
Test Data Preparation and Knowledge Sharing	• Find test data with comparable products under 2012 Economic Census and NAPCS • Define donors/recipients • Bring staff "up to speed" on data collections	Subject Matter and Classification Experts
Exploratory Data Analysis (Empirical Data)	• Understand the "nature" of reported data to assess potential imputation methods • Understand the "nature" of missing data to assess potential imputation cells and to develop response propensity models	Methodologists
Evaluation Study	• Evaluate the performance of considered imputation methods over repeated samples	Methodologists

The dearth of predictors was frightening. The volume of the problem and resultant analysis appeared to be overwhelming. At this point, the team had identified and programmed the candidate imputation methods:

- Ratio imputation (simple linear no-intercept weighted regression of each product on total receipts, a procedure also referred to as "expansion" used by several sectors for the 2012 and prior economic censuses)
- Sequential Regression Multivariate Imputation (SRMI) as described in Raghunathan et al (2001) and
- Two variations of hot deck imputation (random and nearest neighbor), both which imputed the multivariate distribution of products from donor establishments.

The team decided on a simulation approach to create industry "populations" from historical sample data in 39 industries by applying each candidate imputation method to replace the missing data as suggested by Dr. Trivellore Raghunathan (University of Michigan). Product nonresponse was induced in 50 independent replicates in each completed population, and all four candidate methods were used to "complete" the datasets, using multiple imputation to obtain the imputed estimates, standard errors, and evaluation statistics (imputation error and fraction of missing information). See Figure 2 for an illustration of the simulation study procedure.

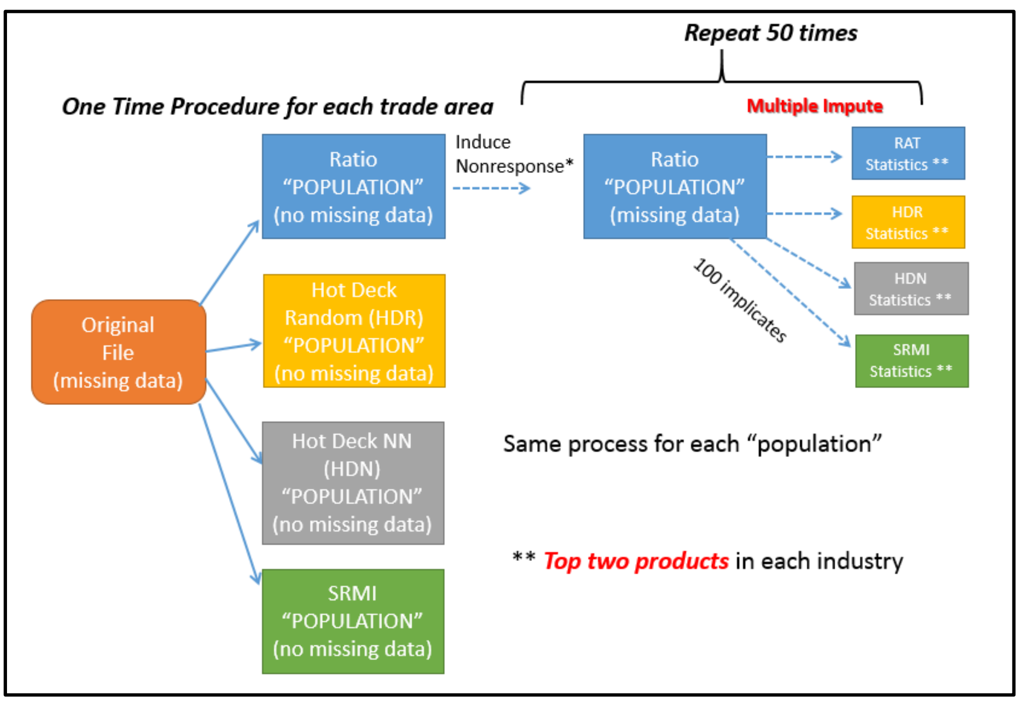

***Figure 2**:* Illustration of the Simulation Procedure for the Imputation Evaluation Research

By design and necessity, the simulation study made some simplifying assumptions beyond those already mentioned. Small sample size effects were controlled by choice of estimate level (national) and the selection of study industries. These choices sidestepped issues that would arise from small respondent sample sizes in imputation cells. The evaluation was restricted to the two best-reported broad products in each studied industry in terms of number of establishments that reported the product. Rescaling the size of the problem reduced computation time and increased available time for analysis, although it did impact the study's "representativeness." Lastly, the evaluation used rank-based tests within industry to compare the procedures, so that substantive improvements or deficiencies in specific situations were largely ignored. Instead, the evaluation procedures found common patterns among the methods on each evaluation criterion on the best-reported products.

The team recommended using hot deck imputation for broad products in all industries, allowing different hot deck variations by industries. This recommendation was endorsed by the project stakeholders. That said, the recommendation was incomplete. No guidance was provided in terms of optimal imputation cells, minimum cell size (or collapsing rules), backup imputation methods (in the event of no donors), or dollar value and additivity requirements for final imputed data (no imputed values of less than $500 were allowed, but all rounded values were required to add to the associated value of total receipts). Imputation of detailed products was not addressed in the research study, nor was calibration to industry totals.

4. Implementation Team

A new team was established for implementation. Leadership was provided by project management experts with extensive familiarity with the subject matter and with the planned Economic Census processes. A large representation of (industry) subject matter experts were included, as were production programmers. Four methodologists from the research team were retained as consultants, with a production methodologist recruited to develop the final specifications. [Note: This methodologist directly supervised a research team member and was not unfamiliar with the research project, even without working directly on the team.] The team lead and one subject matter expert had participated on the research team, but the remaining team members were recruited from other Economic Census projects and were not familiar with the earlier research or methods.

As with the research team, the project began slowly with an educational component. Team members each had their own set of implementation issues that needed to be addressed. The production programmers were concerned about the computing resource demands of fully implementing hot deck imputation; there were also staffing concerns as team members had to juggle working on this project with other high priority projects. Fortunately, the methodologists were able to provide concrete examples of efficient hot deck systems, which partially alleviated these concerns.

Presentation of the hot deck methods was generally met with acceptance. Some team members were hesitant about using random hot deck, as many of the subject matter experts intuited that product distribution was related to unit size. [Note: The Research Team was unable to find any evidence of this.] However, the subject matter experts balked at cell-collapsing procedures for two reasons. First, they were not convinced that imputation cells with sufficient observations should be combined with imputation cells with insufficient observations. Instead, they insisted on using the original imputation cells in the former case, reserving the collapsed imputation cells only for the latter case. Second, they did not have resources to research alternative cell collapsing procedures. Eventually, the implementation team compromised under strong protest from several methodologists, agreeing to the unconventional preferred cell-collapsing procedure but using a minimum cell size of one establishment. Although the specifications called for parameterized imputation cell definitions with three levels of collapsing, the imputation cell definitions that were coded into the production program varied little by industry.

The subject matter experts were emphatic on one point: they wanted to maximize the use of unit-level reported data in the imputation procedures whenever possible. However, businesses are much more likely to report broad product categories than detailed categories (there are some notable exceptions in some retail trade industries). Restricting donor records to establishments that provided usable[3] values for broad and detailed products was too restrictive for

[3] usable = all products classified and "in scope" for the industry and all additivity constraints satisfied

many industries and would likely lead to inefficient estimates. An inspection of 2012 Economic Census counts of reported broad and detailed products within the most restrictive imputation cell definitions (industry by state by unit type) confirmed this suspicion. The majority of imputation cells contained at least five establishments that reported usable broad products; this was not the case with the detailed products.

Moreover, the research team had not studied imputation methods for detailed products. The lack of available reported data – and the differences in types of detailed products between 6-digit industries – was a prohibitive barrier. Methodologists on the team recommended using the ratio imputation model for each detailed product x_{ij}, where i indexes the associated broad product (x_i) and j indexes the detailed product within the broad product:

$$x_{ij} = \beta_{ij} x_i + \varepsilon_i, \varepsilon_i \sim (0, x_i \sigma_i^2)$$

This model is commonly used by business surveys (for examples, see Beaumont and Bocci 2009 and Shao and Thompson 2009). Under this model, the B.LU.E of $\widehat{\beta_{ij}} = \sum x_{ij} / \sum x_i$ (Lohr 2010, Ch.4.6), known in-house as "category average" imputation. Although this method did not prove optimal with the referenced broad product research, the model is supported by the literature. Furthermore, the ratio-imputed broad line products were almost always unbiased estimates in the studied industries (Garcia, Morris, and Diamond 2015).

Table 2 presents a categorization of donor and recipient establishments based on the presence of usable broad and detailed products. Using 2012 Economic Census data, we estimated the percentage of donors that would fall in each of these categories as follows: complete (86.6%), partial (7.4%), and minimal (6.0%). Obviously these percentages may be different in 2017, given the changes to the questionnaire.

Table 2: Establishment Classification for Imputation

Donors	Broad products usable
Complete	All broad and detailed products usable (contribute to category average)
Partial	All broad products usable and some detailed products usable (contribute to category average)
Minimal	All broad products usable; detailed products missing and required (receive detailed products from category average)
Recipients	Missing products
Full	Need broad and detailed products (receive all products from hot deck)
Partial	Need some (designated) detailed products (receive detailed products from category average)
Ineligible	All products usable, but not "typical"; excluded from donor pool

To simplify the operational procedure, the implementation team decided to create "complete" donor records (i.e. fill in the missing detailed products for partial and minimal donors) prior to hot deck imputation. To accomplish this, category averages were computed for each detailed product within a broad product for each potential imputation cell (generally industry-by-state-by-unit type, industry-by-state, industry) with a required minimum of one establishment in the cell reporting the detailed products. Designated missing detailed products were imputed from their associated broad product total by using the appropriate category averages. Once this process was finished and all donors were made "Complete," the hot deck process was performed to impute products for all "Full" recipients. This approach of completing the partial and minimal donors maximized the use of reported data in the hot deck imputation procedures, but later complicated the variance estimation of detailed products due to the partial donor/recipient establishments (Thompson and Thompson to appear).

After hot deck imputation, all imputed values were rounded to multiples of $1000 using an algorithm that ensured additivity at the establishment level. The basic approach of the method is to use a process of controlled rounding to balance the broad products of an establishment to its (rounded) total receipts value and then balance the detail products to the (now rounded) broad product values while also satisfying the requirement that some product values (the "must-products") cannot be rounded to zero. The example below shows how the six broad product values imputed for an establishment were rounded to multiples of $1,000. The two "must" products are placed first, followed by all other products.

NAPCS Product Code	Imputed (Hot-deck) Distribution	Total Estab Receipts ($000)	Required ("Must") Product? (Y/N)	NAPCS Product Value ($000)				Note
				Exact *(Original)*	Adjusted	Rounded	Cumulative Rounding Error	
4000225000	0.5%	6	Y Asc	0.030		**1.000**	-0.970	(1)(2)
4000115000	12.3%	6	Y	0.738	-0.232	1.000	-1.232	(3)(4)
4000335000	57.1%	6	N	3.426	2.194	**2.000**	0.194	(5)
4004045000	19.0%	6	N Desc	1.140	1.334	1.000	0.334	
4008550000	10.1%	6	N	0.606	0.940	1.000	-0.060	
4006005000	1.0%	6	N	0.060	0.000	0.000	0.000	(6)
Totals	100.0%			6.000		6.000		

***Figure 3*:** Illustration of Controlled Rounding

Notes:
(1) Even though the Adjusted value < 0.5, it is Rounded to 1.0 because this is a "must" product.
(2) Cumulative rounding error = 0.030 – 1.000 = -0.970
(3) Adjusted = 0.738 + (-0.970) = -0.232. However, the Rounded value is set to 1.0 because this is also a "must" product.
(4) Cumulative rounding error = -0.232 – 1.000 = -1.232
(5) Adjusted value = 3.426 + (-1.232) = 2.194 which is Rounded to 2.0
(6) Since the rounded product value is zero, this product will be removed from this establishment.

The implementation team met regularly over a two-year period. During this collaborative period, methodologists met separately each week (along with the team leader) to develop the missing data procedures and treatments that were not addressed by the research team. Specifications were reviewed first by this subgroup, then by the entire team.

Testing was a larger problem. Using small, single industry test decks, we were able to verify that the category average and hot-deck processes were working correctly. However, one of the concerns about hot-deck imputation was the time it would take to run the process to impute missing products for the entire Economic Census. To determine estimated run-times, as well as test more scenarios, we created a full size test deck with roughly 2.4 million donors (with over 20 million products) and 1.1 million full recipients covering all NAICS sectors in-scope to the Economic Census. Using a concordance that mapped 2012 product codes to 2017 NAPCS codes, we converted the 2012 Economic Census product data to a 2017 NAPCS basis. There were several challenges involved with this:

- The concordance did not cover all mapping situations encountered in the data. Missing detailed products were a particular problem. Sometimes this is acceptable, other times not[4].
- While many products mapped one-to-one, in other cases the mapping was very complex. See Figure 4 below for an example.
- Certain data flags could take on new values for 2017 so we had to predict the conditions under which these new values would be assigned.
- Ensuring that (broad) products sum to total receipts for every establishment and that detailed products sum to broad products.
- Ensuring that certain specific scenarios were included in the test data.

The performance testing using this test deck took approximately 80 minutes. (Recent improvements in processing have reduced this time to about 45 minutes.) This was a reassuring result, although it might not directly translate to run times using actual 2017 Economic Census production data and systems.

[4] For example, a barbershop need only report their total automobile sales (a broad product) without having to report the detailed product (used, new, sold, leased, cars, trucks, etc.). An automobile dealer is expected to report the details.

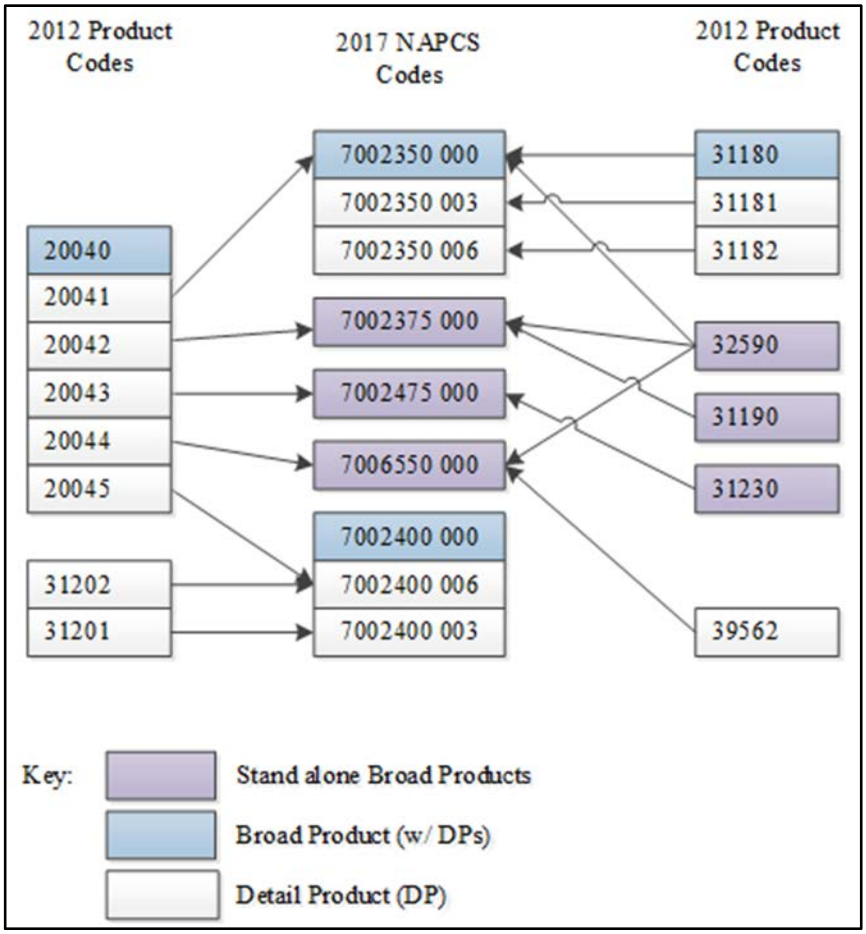

Figure 4: Illustration of Mapping of 2012 Product Codes to 2017 NAPCS-Based Collection Codes

5. Conclusion (Hard, Harder, Hardest)

When developing a research plan that applies to an ongoing survey, finding balance is hard. On one hand, making the scenario as simple as possible reduces the probability of treatment effects (solutions) being confounded by factors such as sample size or random noise. On the other hand, oversimplification can lead to very impractical solutions. Of course, it is crucial to limit the scope so that the research can be timely enough to be relevant when completed. However, it should be acknowledged that compressing the scope can lead to hasty decisions later in the implementation process, when there is no time left for careful further investigation.

Bluntly put, the hard problems require research. Even harder is figuring out how to conduct this research and who should be included. Should the research team be restricted entirely to methodologists with limited consultations with subject matter experts or should subject matter experts be integrated into the project? If the former approach is taken, what safeguards will be put in place to ensure that key requirements are met? And the hardest tasks tend to come from implementation, where every "cut corner" in the research needs filling in, and not every situation is ideal (e.g. small sample sizes, fewer donors than recipients, limited predictors).

There are real advantages in establishing (almost) separate research and implementation teams as discussed in this paper. Having two teams approach the same problem from different perspectives leads to innovative applications. Often, these teams provide practical opportunities for methodologists to learn about data and data collection and for subject matter experts to learn about alternative methodologies. From an administrative perspective, these teams can help with succession management planning, especially when junior staff are included. Lastly, they provide justification for the production procedures under the umbrella of data-driven decision making.

Of course, there are equally real disadvantages. The limited scope in research can lead to missed requirements, which can be revealed as unexpected results in implementation testing or in production. Delaying decisions until implementation can preclude having sufficient time for careful investigation, and quick decisions are made for convenience based on anecdotal justification, with no alternatives tested. Having two separate teams increases management challenges as well, as appropriate leaders need to be recruited and team members struggle with competing duties (and on occasion, motivation and morale challenges).

When the end-product is a theoretically solid and operationally viable system, this approach is a success. It certainly was in the case study presented in this paper. The two-phase team approach has been used for other 2017 Economic Census applications such as determining and implementing a variance estimation method for product estimates (Thompson and Thompson to appear; Knutsen, Thompson, and Thompson 2017; Thompson and Thompson 2016; Thompson, Thompson, and Kurec 2016) and for developing standard response rates (Lineback, Oliver, and Willimack 2012). Certainly in these examples, the advantages outweighed the disadvantages, with workable solutions and buy-in as well as shared understanding of implemented methods. And of course, the imperfect solutions provide plenty of exciting research ideas and opportunities for the next Economic Census.

Acknowledgements

The authors thank Andrea Chamberlain, Michael Kornbau, and John Ward for their useful comments on earlier versions of this manuscript. We also acknowledge the efforts of both of the Economic Census teams discussed in this paper, with special thanks to Joshua Buchheit for his extensive testing efforts.

References

Beaumont, J.-F. and Bocci, C. (2009), Variance Estimation When Donor Imputation Is Used To Fill In Missing Values. Can Journal of Statistics 37: 400–416. doi:10.1002/cjs.10019

Bechtel, L., Morris, D.S., and Thompson, K.J. (2015). Using Classification Trees to Recommend Hot Deck Imputation Methods: A Case Study. *Proceedings of the FCSM Research Conference.*

Ellis, Y. and Thompson, K.J. (2015). Exploratory Data Analysis of Economic Census Products: Methods and Results. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Fink, E.B., Beck, J.L. and Willimack, D.K. (2015). Data-Driven Decision Making and the Design of Economic Census Data Collection Instruments. *Proceedings of the FCSM Research Conference.*

Garcia, M., Morris, D.S., and Diamond, L.K. (2015). Implementation of Ratio Imputation and Sequential Regression Multivariate Imputation on Economic Census Products. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Knutson, J. and Martin, J. (2015). Evaluation of Alternative Imputation Methods for Economic Census Products: The Cook-Off. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Knutson, J., Thompson, M., and Thompson, K.J. (2017). Developing Variance Estimates for Products in the Economic Census. *Proceedings of the Government Statistics Section*, American Statistical Association.

Lineback, F., Oliver, B., and Willimack, D.K. (2012). Developing Response Metrics for the Economic Census. *Proceedings of the FCSM Research Conference.*

Lohr, S. (2010). Sampling: Design and Analysis (2nd Edition). Boston, MA: Brooks/Cole.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., and Solenberger, P. (2001). A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology*: 27(1), pp. 85–95.

Shao, J. and Thompsoon, K.J. (2009). Variance estimation in the presence of nonrespondents and certainty strata. Survey Methodology 35(2): 215-225.

Thompson, K.J. and Liu, X. (2015). On Recommending a Single Imputation Method for Economic Census Products. *Proceedings of the Government Statistics Section*, American Statistical Association.

Thompson, M. and Thompson, K.J. (To appear). Variance Estimation for Product Sales in the 2017 Economic Census: Utilizing Multiple Imputation to Account for Sampling and Imputation Variance. *Proceedings of the Government Statistics Section,* American Statistical Association.

Thompson, K.J. and Thompson, M. (2016). Estimating the Variance Due to Hot Deck imputation for Product Value Estimates in the 2017 Economic Census. *Proceedings of the Government Statistics Section,* American Statistical Association.

Thompson, M., Thompson, K.J., and Kurec, R. (2016). Variance Estimation for Product Value Estimates in the 2017 Economic Census Under the Assumption of Complete Response. *Proceedings of the Government Statistics Section*, American Statistical Association.

Tolliver, K. and Bechtel, L. (2015). Implementation of Hot Deck Imputation on Economic Census Products. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.