

# Impact of Preliminary versus Final Economic Census Data on the Universe Extraction Process for Current Business Surveys<sup>1</sup>

Kari K. Clark and Carol S. King

U.S. Census Bureau, SSSD  
4600 Silver Hill Road, Washington DC 20233, USA  
Kari.K.Clark@census.gov

## Abstract

The U.S. Census Bureau selects a new sample for its current business surveys approximately once every five years. For use in constructing the sampling frame, the first step in the sample selection process is the extraction of establishment records from the Census Bureau's Business Register used in the creation of an establishment list. As part of this process, census data will also be extracted for establishments that were active during the time of the latest Economic Census. The data from the Census and the Business Register are used to determine each establishment's industry classification and the major kind of business for the sampling unit. Sampling units consist of aggregations of one or more establishments, based on the organization of the company. The industry classification is used to evaluate if the establishments could be considered in-scope to the current business surveys and is also used as a stratification variable in the sample design. Another stratification variable used in the sample design is a measure of size. This measure of size represents a full year of activity in terms of revenue at both the establishment and sampling unit level.

When the extraction process was done for the current sample, only preliminary data from the 2002 Economic Census was available. Using the final census data would result in the extraction process being done a year later. Research was conducted to evaluate how running the extraction process a year later would have affected the industry classification and the measure of size for both the establishment and the sampling unit. This paper will detail the extraction process and how the changes to the extraction process impacted the establishment list.

## 1. Introduction

Approximately once every five years, the U.S. Census Bureau reselects the samples for its annual, quarterly, and monthly economic surveys covering the retail, wholesale, and service sectors. This process is known as the Business Sample Revision and focuses on businesses classified in at least one of the three sectors as defined by the North American Industry Classification System (NAICS). The Annual Retail Trade Survey (ARTS), its monthly subsample (MRTS), and its advance monthly subsample (MARTS) cover the retail (NAICS 44-45) and the accommodation and food services (NAICS 72) sectors. The Annual Wholesale Trade Survey (AWTS) covers the wholesale (NAICS 42) sector and produces separate estimates based on the type of operation code (TOC). There are three different types of operation code: wholesale distributors that take the title of the goods they sell, agents or brokers that do not take the title of the goods they sell, and manufacturer sales branches and offices (MSBOs). The wholesale monthly subsample (MWTs) excludes agents or brokers and MSBOs. The Service Annual Survey (SAS) covers a wide range of service sectors, such as transportation and warehousing (NAICS 484, 492-493), information (NAICS 51), finance (NAICS 523), healthcare (NAICS 62), and other general service industries (NAICS 532, 54, 56, 71, and 81). The service quarterly subsample (QSS) covers NAICS 51, 54, 56, and part of 62.

Prior to selecting the samples, an initial list of business establishments is extracted for the purposes of creating a sampling frame. An *establishment* is defined as the smallest business unit at which transactions take place or services are provided and payroll and employment records are maintained. There are two types of establishments: singleunit (SU) and multiunit (MU). A SU establishment and a company can be used to define the same entity, whereas a MU establishment is a part of a larger company. The process of creating the list is known as the universe extraction, and combines information from the Census Bureau's Business Register (BR) and the most recent Economic Census. After the initial list is created and the establishments are grouped into SUs and MUs, the final establishment list is determined by evaluating all establishments as

---

<sup>1</sup> This report is released to inform interested parties of research and to encourage discussion. Any views expressed on methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau. We thank Katrina T. Washington and Anne Kearney for their helpful comments.

being in scope or out of scope to the surveys based on certain characteristics such as their NAICS codes, their business activity status, tax status, or TOC. For those identified as being in scope, each establishment is assigned a sampling recode based on its NAICS code, tax status, government status, and TOC. This recode is used as the primary stratification variable for sampling along with the measure of size (MOS). The MOS is calculated for each establishment to represent a full year of activity in terms of revenue data.

For past Business Sample Revisions, the final data from the latest Economic Census was not available at the time of the extraction, so the preliminary data was used. Had we waited for the final census data, the extraction would have been conducted about a year later. This paper will discuss research conducted to determine the impact of running extraction a year later than the most recent universe extraction process. Section 2 discusses the changes due to the timing of the extraction, while sections 3 to 5 evaluate the changes in specific aspects of the universe extraction process: the determination of in scope versus out of scope, the assignment of sampling recodes, and the calculation of the MOS. Based on this evaluation, we will determine whether or not the changes have enough impact to consider delaying the extraction process until the final census data is available.

## **2. Changes due to Timing**

With the universe extraction process running a year later, the input data for the establishment list will change. Besides using the final census data as input, the Business Register will now have more current data and have been updated with information from the next Company Organization Survey (COS). The COS is used to update the Business Register each year by mailing questionnaires to a sample of MU companies asking for changes in their organization structure, such as if they had bought or sold an establishment. During an economic census year (years ending in 2 and 7), all MU companies are mailed.

Also, establishments that were active in 2004, but not in 2003, will now be included in the establishment list. These establishments are known as birth establishments, but will not be addressed in this paper, as we will only evaluate the changes that occur for establishments included in both extractions. There are also establishments that would be active in 2003 but are no longer active in 2004. These establishments are still included on the initial list but will later be identified as being out of scope on the final establishment list.

In December of 2004 and January of 2005, the most recent universe extraction process was conducted. To assess the impact of using the final census data, the extraction process was modified in the following ways:

- When calculating the MOS, 2004 data instead of 2003 data from the Business Register was used. This means a two-year lag from the time of the 2002 Economic Census. The deflation factors, used to bring MOS down to a 2002 basis, would have to be recalculated. Besides the deflation factors, there were other parameters used throughout the extraction process, such as the bounds used for editing the data and parameters used in the sample design. These would also have to be recalculated to account for the two-year lag. Due to time constraints for this research, we decided to make an assumption that these parameters would not have changed substantially when using a two-year lag instead of a one-year lag.
- When determining the NAICS code for each establishment, the BR NAICS code was used because the final census information had been applied to the BR. For the initial extraction, a hierarchy was used based on when the BR NAICS code was last updated. If the BR NAICS code had been updated more recently than the NAICS code from the 2002 Economic Census, then the BR NAICS code would have been used. Otherwise, the NAICS code from the Census would have been used.

## **3. Determination of Scope**

After determining each establishment's NAICS code, tax status, and/or TOC, we then identified if it was in scope or out of scope to our current surveys. An establishment is said to be in scope if it is active at the time of extraction, has paid employees, and has U.S operations in one of the industries covered by the current surveys. Because using the final census data would have resulted in the universe extraction being done a year later, how many establishments were still in scope to our current surveys? For establishments that were identified in both the initial extraction and the new extraction, tables 1 and 2 show the changes in the determination of scope and the reasons for the change. Under the Reason column, "Business Codes" combines different codes describing the establishment, such as its location or government status.

For establishments that were in scope in the initial extraction but became out of scope for the new extraction, the MOS from the initial extraction represents what these establishments would have contributed had they still been in scope. For establishments that were out of scope for the initial extraction but became in scope for the new extraction, the MOS from the new extraction represents what these establishments could have contributed had they been included in the initial extraction. For establishments that were in scope to both extractions, the MOS is from the initial extraction.

Table 1: Changes in Scope for MU Establishments				
Scope	Reason	# Estabs	% Estabs	MOS (in thousands)
In to In		1152481	47.29%	\$9,799,722,675
Out to Out		1196474	49.09%	\$0
In to Out	Activity Status	76181	3.13%	\$1,051,265,965
	NAICS code	1693	0.07%	\$27,150,725
	Payroll from Business Register	842	0.03%	\$3,135,070
	Payroll from Census	22	0.00%	\$0
	Receipts from Census	12	0.00%	\$0
	Business Codes	308	0.01%	\$3,528,999
	Total	79058	3.24%	\$1,085,080,759
Out to In	Activity Status	2299	0.09%	\$8,734,531
	NAICS code	1965	0.08%	\$45,275,349
	Payroll from Business Register	3800	0.16%	\$19,901,992
	Payroll from Census	0	0.00%	\$0
	Receipts from Census	0	0.00%	\$0
	Business Codes	1009	0.04%	\$21,905,932
	Total	9073	0.37%	\$95,817,804
Total		2437086	100.00%	\$10,980,621,238

Table2: Changes in Scope for SU Establishments				
Scope	Reason	# Estabs	% Estabs	MOS (in thousands)
In to In		3027847	44.19%	\$3,890,076,381
Out to Out		3143376	45.87%	\$0
In to Out	Activity Status	248527	3.63%	\$166,247,574
	NAICS code	18531	0.27%	\$35,190,341
	Payroll from Business Register	96779	1.41%	\$28,461,366
	Business Codes	727	0.01%	\$922,921
	Total	364564	5.32%	\$230,822,202
Out to In	Activity Status	20415	0.30%	\$8,358,112
	NAICS code	86699	1.27%	\$78,003,123
	Payroll from Business Register	208974	3.05%	\$139,419,346
	Business Codes	659	0.01%	\$681,883
	Total	316747	4.62%	\$226,462,464
Total		6852534	100.00%	\$4,347,361,047

From looking at the tables, 96.38% of MU and 90.06% of SU establishments did not change scope between the two extractions. The following subsections examine the reasons behind the changes occurring in more than one percent of all establishments: activity status for both types of establishment (3.13% MU, 3.63% SU), the payroll from the Business Register for SU establishments (1.41%, 3.05%), and the NAICS code for SU establishments (1.27%). The last subsection examines changes when an establishment was a SU in one extraction but a MU in the other extraction.

### 3.1 Activity Status and Payroll

For both MU and SU establishments, those that were active in 2003 but not in 2004 make up the majority of the cases where they became out of scope based on activity status or payroll from the Business Register. For the SU establishments, activity status was determined based on whether or not the company still reports quarterly payroll to the Internal Revenue Service (IRS). For the MU establishments, activity status was extracted from the Business Register, which was updated based on responses to the Company Organization Survey (COS). With a different COS sample selected each year, it is possible that a MU company would not be mailed a COS form in one year but is mailed a form in the following year. Depending on when the MU company responds to the COS, it may be that the activity status codes for each of its establishments are not current for a given year. As an extraction is done at a certain point in time, there will be establishments that could be in scope but later be out of scope or vice versa.

### 3.2 NAICS codes

One of the changes made when running the new extraction was how the NAICS code was assigned to the establishments. In any given year, there could be a potential misclassification or not enough information available to determine the industry for the establishment. Table 3 shows the sources from which the NAICS codes were assigned for SU establishments originally out of scope but now in scope. The unclassified establishments were those for which we could not assign a NAICS code in the initial extraction, so the code was set equal to all zeroes.

Table 3: Source of NAICS Codes for SU Establishments			
NAICS - Initial	NAICS - Current	# Estabs	% Estabs
Census	Census	755	0.87%
	BR	2441	2.82%
BR	Census	5050	5.82%
	BR	51146	58.99%
Unclassified	BR	27307	31.50%
Total		86699	100.00%

Of the 5,825 establishments (5050 + 755) that had their NAICS code updated as a result of the final census data, only 1,016 of these occurred because the NAICS code from the preliminary census was different compared to the NAICS code from the final census. Waiting for the final census data to be finalized does not seem to affect the NAICS assignment, but with the updated Business Register, unclassified cases are now assigned a NAICS code.

### 3.3 Type of Establishment

The previous sections examined the cases for which the establishment was a SU in both extractions or a MU in both extractions. Depending on how a business is restructured or bought by another business, a MU establishment could become a SU establishment and vice versa. Table 4 shows the number of establishments that change their type between extractions as well as whether or not they were in scope. For those that remained in scope to both extractions regardless of their type, they will be included in sections 4 and 5 with the same type of establishment they were identified for the initial extraction when looking at the changes in the sampling recode and measure of size. These establishments could have an effect on the sampling units as explained in section 4.

Table 4: Changes in Type of Establishments		
Scope	MU to SU	SU to MU
	# Estabs	# Estabs
In to In	669	822
Out to Out	7	69
In to Out	75	64
Out to In	17	522
Total	768	1477

#### 4. Assignment of Sampling Recode

After determining which establishments were in scope to our current surveys, we then assigned a sampling recode. For establishments in the wholesale sector, the sampling recode was assigned based on their NAICS code and TOC. For establishments in the retail sector, the sampling recode is assigned based on their NAICS code. For establishments in the service sector, the sampling recode is assigned based on their NAICS code, tax status, and government status. Table 5 shows the changes in the sampling recode for both MUs and SUs, and the reasons for the change. Around 98% of establishments do not have any changes in their sampling recode. For establishments where the sampling recode did change, the majority appeared to be because of the NAICS code.

Table 5: Changes in Sample Recode				
Reason for Change	MU		SU	
	# Estabs	% Estabs	# Estabs	% Estabs
No Change	1142206	98.97%	2959017	97.50%
NAICS Code	9591	0.83%	71393	2.35%
Type of Operation Code	881	0.08%	4	0.00%
Tax Status Code	1328	0.12%	4515	0.15%
Government Code	69	0.01%	43	0.00%
Total	1154075	100.00%	3034972	100.00%

For singleunits, there is no difference between the sampling unit and the establishment. However, for multiunits, the sampling units can either be the entire company or the level at which payroll for the establishments are reported to the IRS, also known as an EIN or Employer Identification Number. Summing up the MOS for each sampling recode within the sampling unit by trade area and finding the maximum MOS determines the major kind of business for the sampling unit. To approximate the possible changes at the sampling unit level, the major kind of business was calculated for each company and EIN in both extractions. Table 6 shows that almost 99% of the sampling units do not have a change in their major kind of business.

Table 6: Breakout of MU Sampling Units			
	Sample Recode	#	%
Companies	No Change	146108	98.76%
	Change	1833	1.24%
	Total	147941	100.00%
EINs	No Change	200104	98.53%
	Change	2985	1.47%
	Total	203089	100.00%

## 5. Calculation of MOS

Once a sampling recode was assigned to an establishment, we then calculated its measure of size to represent a full year of revenue in terms of the most recent data available. During the initial extraction, the MOS was calculated for the 2003 year while the MOS was calculated for the 2004 year during the new extraction. Depending on how the MOS was calculated, the data used could be payroll or receipts from either the Business Register or the Economic Census. After the MOS was calculated, a deflation factor was used to adjust the MOS to a 2002 level. Table 7 shows the number and percentage of establishments within a certain range of percent changes and absolute differences between the initial MOS and the new MOS. 34.60% of MU establishments and 27.47% of SU establishments had an absolute difference greater than or equal to 100,000 and a percent difference greater than 10%.

Table 7: Changes in Measure of Size					
Percent Change	Absolute Difference (in thousands)	MU		SU	
		# Estabs	% Estabs	# Estabs	% Estabs
No change		58342	5.06%	90110	2.97%
Change from Zero	< 100	3221	0.28%	5	0.00%
	>= 100	16205	1.40%	18	0.00%
<= 5%	< 100	258077	22.36%	600470	19.79%
	>= 100	66113	5.73%	36676	1.21%
5% - 10%	< 100	128540	11.14%	461873	15.22%
	>= 100	109532	9.49%	101490	3.34%
10% - 25%	< 100	89736	7.78%	589406	19.42%
	>= 100	212588	18.42%	322679	10.63%
> 25%	< 100	24967	2.16%	321098	10.58%
	>= 100	186754	16.18%	511147	16.84%
Total		1154075	100.00%	3034972	100.00%

For establishments that had their MOS changed, we looked at the reasons for the change to see if they were based on the method used to calculate the MOS or the actual data used in the calculation as shown in table 8. However, during the initial extraction, subject matter analysts could make corrections to the data, which would change how the MOS was calculated. Thus, these are put into a separate category. With the timing difference, it is not surprising that a majority of the establishments had their MOS changed because the data also changed. These results do not factor in the possibility that the establishment could have had its sampling recode changed, which would mean that the deflation factor was different. Thus, the change in timing is the main factor for the changes in MOS.

Table 8: Reasons for the Change in Measure of Size				
Reason	MU		SU	
	# Estabs	% Estabs	# Estabs	% Estabs
Analyst Correction	348	0.03%	71	0.00%
Change in Data	1071165	97.76%	2588620	87.90%
Change in Calculation	24220	2.21%	356171	12.09%
Total	1095733	100.00%	2944862	100.00%

We also examined the total MOS for the retail, services, and wholesale sectors to see the overall impact of the changes in MOS. Tables 9 and 10 show the number of establishments and total MOS, and the wholesale sector is divided into the three different TOC types. The percentages are not different by more than 10%. For the singleunit establishments, there are no MSBOs because they are defined to be multiunit establishments working for a parent company. The 5 establishments in the new extraction are those where they originally were singleunit establishments in the initial extraction, and were multiunit establishments in the new extraction.

Table 9: MOS by Trade Area for Multiunit Establishments						
Trade Area	Estabs - Initial	MOS - Initial	% MOS	Estabs - Current	MOS - Current	% MOS
Agents & Brokers	2983	\$215,444,397	2.19%	3205	\$225,778,860	2.36%
MSBOs	16473	\$1,441,111,866	14.67%	16116	\$1,525,562,825	15.93%
Retail	537495	\$2,178,778,538	22.18%	536985	\$2,238,264,921	23.38%
Services	506984	\$2,826,227,201	28.78%	507370	\$3,574,141,561	37.33%
Wholesale Distributors	90140	\$3,159,914,245	32.17%	90399	\$2,010,328,299	21.00%
Total	1154075	\$9,821,476,247	100.00%	1154075	\$9,574,076,466	100.00%

Table 10: MOS by Trade Area for Singleunit Establishments						
Trade Area	Estabs - Initial	MOS - Initial	% MOS	Estabs - Current	MOS - Current	% MOS
Agents & Brokers	22491	\$238,989,123	6.10%	26490	\$256,825,876	5.95%
MSBOs	0	\$0	0.00%	5	\$2,017,670	0.05%
Retail	857768	\$1,187,809,426	30.34%	857994	\$1,254,692,789	29.05%
Services	1926429	\$1,582,253,202	40.41%	1926360	\$1,801,351,435	41.70%
Wholesale Distributors	228284	\$906,543,185	23.15%	224123	\$1,004,442,354	23.25%
Total	3034972	\$3,915,594,936	100.00%	3034972	\$4,319,330,124	100.00%

## 6. Conclusion

The goal of this paper was to assess the impact of the final versus preliminary census data on the universe extraction process. However, the timing of the new extraction also included an updated Business Register. Thus, any changes on the final establishment list due to the determination of scope, the assignment of sampling recodes, and the calculation of MOS were based on either the final census data or data from the updated Business Register. With the changes in the MOS, it is recommended that the extraction process be run a year later. Also, the combined changes in MOS and sampling recodes could have an effect on the samples selected for the current surveys, but more research would need to be conducted to verify this.

## References

- King, Carol S. (2007), "An Evaluation of Changes to Universe Extraction Editing for Current Business Surveys at the U.S. Census Bureau," paper presented at the third International Conference on Establishment Surveys, Montreal, Canada.
- Kinyon, D., D. Glassbrenner, J. Black, and R. Detlefsen (2000), "Designing Business Samples Used for Surveys Conducted by the United States Bureau of the Census," paper presented at the second International Conference on Establishment Surveys, Buffalo, NY.
- Walker, Ed (1997), "The Census Bureau's Business Register: Basic Features and Quality Issues," paper presented at the 1997 Joint Statistical Meetings.