

# INVESTIGATING POTENTIAL BIRTH COMPANIES FOR THE ANNUAL CAPITAL EXPENDITURES SURVEY

Marie C. Stetser, Ana Rodriguez, Amy M. Newman-Smith<sup>1</sup>, U.S. Bureau of the Census

Marie Stetser, U.S. Bureau of the Census, Washington D.C. 20233-8700

**Key Words:** Establishment survey; Frame; Births; Coverage; Capital Expenditures

## I. Introduction

Since 1993 the Census Bureau has conducted a full scale data collection effort of capital investment information from businesses in the United States by means of the Annual Capital Expenditures Survey (ACES). Several years ago, Census statisticians began looking at coverage of birth companies on the ACES frame and hypothesizing about these companies being a significant source of capital investment dollars. Users of these data have also been critical of the coverage of birth companies in the ACES frame. Discrepancies between ACES estimates and estimates from other surveys have given us reason to search for explanations and ways to improve the coverage of birth companies on the ACES frame. This paper will discuss two recent research efforts undertaken by Census Bureau staff to address this issue. Before we begin discussing the details of these efforts, we will explain a little more about the problem of undercoverage of birth companies for establishment surveys.

## II. Understanding the Problem

When designing frames for establishment surveys, we attempt to capture a population of businesses that were actively operating during a specific time frame, such as a month, a quarter, or a year. Creating an accurate frame has several challenges. First we must define the scope of our frame. How do we define when a business becomes active and when it becomes inactive? This definition may vary depending on what type of data we are trying to collect. Surveys that collect information on payroll and employment or sales and receipts commonly define a business as active when it first reports payroll information. This definition is reasonably clear cut. However, surveys

that collect information on capital investment may wish to define a business as active when it first makes capital expenditures. This definition makes sense, however, it is not easy to operationalize.

This leads to the next question, how do we create a frame that includes all businesses which meet our definition? For Census Bureau establishment surveys, we generally construct frames from administrative data that consist primarily of tax receipts. Statisticians use information from income tax returns and payroll tax returns to construct and maintain a large business register called the Standard Statistical Establishment List (SSEL). Each month the Census Bureau receives updates to its business register from the Internal Revenue Service's Business Master File (BMF) and from payroll tax returns. The BMF and other tax files are organized by Employer Identification Number (EIN), and in part, the SSEL is also organized by EIN. An EIN is a unique identifier for an administrative unit used for tax reporting. In most cases, a business will apply for an EIN just before it begins paying payroll. For most establishment surveys, births are identified and available for sampling when payroll or receipts information are reported for the first time. For surveys that collect sales or payroll information, the amount that these birth companies would contribute to the total estimates before being identified as active is negligible. However, our hypothesis is that capital expenditures data may be different. As part of the startup costs a birth company may make significant purchases for structures or equipment before any employees have been hired or sales have been transacted. Administrative data that define when a business first made capital expenditures are not available, and it is difficult to adequately cover birth companies through a frame constructed from tax records. However, the contribution of capital expenditures from birth companies may be significant.

Although a frame constructed primarily from tax records may not be ideal for ACES, we still do not have

---

<sup>1</sup> This paper reports the results of research and analysis undertaken by the Census Bureau staff. It has undergone a more limited review than Official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

any better alternative sources for our frame. However, we may be able to make better use of the available data to improve the coverage of birth companies. This leads us to discuss the first of the two research efforts.

### III. Small Companies with No Administrative Data

We examined the ACES frame and discovered that there was a group of 1.39 million records available on the SSEL that were not being included in the ACES frame selection. For each of these EIN's no administrative data (such as payroll or receipts) were available. We determined that at least some of these records might be birth companies that had applied for an EIN at the start of the business but had not yet reported payroll; however, there also seemed to be quite a few records for businesses that may be going out of business. We decided to sample this population and include the sampled cases as part of the regular ACES data collection, but first we wanted to narrow the group of records down to those that would most likely be in operation or have the potential to make capital expenditures. From the very small amount of administrative data available, we were able to classify the businesses into "active" and "inactive" categories. After doing this, were able to reduce the size of the frame by about one-half, giving a final frame size of 709,751. Table 1 shows the distribution for the 1.39 million records by activity status. We were still unsure what kinds of businesses were on the final frame since 78.8% of the cases in this frame had no industry classification and 30.9% had unknown legal form of organization classification.

**Table 1. Number of Records with No Administrative Data by Activity Status**

Activity Status	Number	Percent
Active	709751	50.8
Inactive	686148	49.2
Total	1395899	100.0

We selected a simple random sample of 2,409 potential birth companies to which we mailed the ACES short form (ACE-2) as part of the 1997 ACES data collection. As the data collection effort wound down, we began evaluating the response data and also evaluating the frame to determine if these companies were, in fact, active and if we should continue sampling from this frame in the future. First let's look at the basic results from the sample. For convention's sake in Table 2 we define the strata for the ACES ACE-2 frame using the names we commonly use for each stratum:

**Table 2. Summary of ACE-2 Strata**

Stratum	Definition
3B	Consists of companies with zero employees in the prior year, but prior year payroll > 0.
3C	Consists of companies with zero employees in the prior year and zero prior year payroll.
3D	Consists of nonemployer partnerships and corporations.
3E	Consists of nonemployer sole proprietorships.
3F	Consists of companies with EINs which were active but have no administrative data reported.

We compared the capital expenditures collected in each of the ACE-2 strata. As can be seen from Table 3, the stratum of interest, 3F, contributed \$11 billion in total capital expenditures or 11% of the total capital expenditures from the ACE-2 strata. The \$11 billion included \$2.8 billion in new structures and \$4.5 billion in new equipment. These results suggest that presence of payroll may not necessarily be the best predictor of capital expenditures, particularly for small companies.

**Table 3. Capital Expenditures By ACE-2 Stratum**

	Estimate (millions of dollars)	Relative Standard Error (%)	Percent ACE-2 Total Estimate
<i>Total Capital Expenditures</i>			
Stratum 3B	13112	41.0	13.7
Stratum 3C	34730	7.2	34.9
Stratum 3D	20874	13.8	21.0
Stratum 3E	19750	11.6	19.8
Stratum 3F	11119	12.8	11.1
Total	99586	7.2	100.0
<i>New Structures</i>			
Stratum 3B	6383	82.3	21.7
Stratum 3C	8058	14.7	27.4
Stratum 3D	10673	21.9	36.3
Stratum 3E	1508	45.5	5.1
Stratum 3F	2775	26.3	9.4
Total	29397	20.3	100.0
<i>New Equipment</i>			
Stratum 3B	5218	19.3	11.3
Stratum 3C	17335	7.4	37.6
Stratum 3D	6812	16.4	14.7
Stratum 3E	12209	11.5	26.4
Stratum 3F	4547	12.5	9.9
Total	46120	5.4	100.0

Table 4 lists the percent of respondent, nonrespondent, and out of scope cases for each ACE-2 stratum. Out of scope cases consist of companies that reported they were inactive during 1997 or were determined to be farms. The response rate for Stratum 3F is only 48%. The low response rate for the 3F cases is of particular concern since it affects the response rates for the survey as a whole and increases the

variability of the estimates. Only 6% of Stratum 3F cases were identified as out of scope for our survey, which is less than we previously expected and similar to the percent out of scope for the other ACE-2 strata. Unfortunately, we have no way of knowing what portion of the nonrespondent cases may be truly out of scope or if the in scope nonrespondent cases have significantly different capital expenditures than the respondent cases.

**Table 4. Percent Response for each ACE-2 Stratum**

Stratum	Respondent	Nonrespondent	Out of Scope
3B	76	19	5
3C	54	32	14
3D	71	24	5
3E	65	24	11
3F	48	46	6

Table 5 shows the percentage of respondent companies reporting varying amounts of capital expenditures by ACE-2 stratum. Of the Stratum 3F cases, 57% reported zero capital expenditures and 3% reported capital expenditures greater than \$100,000. These percentages are close to that of Strata 3B and 3C.

**Table 5. Percent of Respondents for each ACE-2 Stratum by Amount of Capital Expenditures Reported<sup>2</sup>**

Stratum	\$11,000			
	\$0	\$1,000 to \$10,000	to \$100,000	More than \$100,000
3B	54	27	17	2
3C	57	20	19	4
3D	73	15	9	3
3E	75	20	5	0
3F	57	22	18	3

In addition to evaluating the 3F sample, we attempted to evaluate the frame with data available from the current SSEL. The 1.39 million records in the 3F frame were extracted around December 1997. In November 1998 we matched those records back to the current SSEL and tried to evaluate what happened to these companies after about one year. We looked at the current activity status and the presence of administrative data reported for the prior year. We found that 98% of the inactive cases that we excluded from sampling in 1997 were still inactive in 1998, and 73% of the active cases that we included in the 3F sampling frame were still active in 1998.

Based on the administrative data that were reported on the SSEL since December 1997, companies on the

3F frame were grouped into five mutually exclusive categories in the following order: 1) if prior year payroll had been reported; 2) if current year payroll had been reported; 3) if prior year sales, receipts, rent or assets had been reported; 4) if current year sales, receipts, rent or assets had been reported; 5) if no administrative data had been reported.

Table 6 shows the percent reporting administrative data by activity status. As expected, 92.9% of the inactive cases in 1997 had no administrative data reported in 1998; only 0.5% had reported any payroll.

What we did not expect was that 68.3% of the active cases still had no administrative data; but of those that did have administrative data 16.2% had reported current year payroll.

**Table 6. Percent Reporting Administrative Data by 1997 Activity Status.**

Dec., 1997 Activity Status	Payroll > 0		Sales Receipts > 0		No Administrative Data
	Prior Year	Current Year	Prior Year	Current Year	
Active	0.3	16.2	0.7	14.5	68.3
Inactive	0.1	0.4	1.0	5.5	92.9
All	0.2	8.5	0.9	10.1	80.4

Next we examined the current SSEL data with the sample data collected for 3F companies. Surprisingly 38.0% came from companies that had no administrative data. We also looked at Stratum 3F capital expenditures by industry category and learned that a large portion of capital expenditures came from companies that were still unclassified by industry on the SSEL.

Table 7 shows the total capital expenditures by the groups described above. The largest portion of the 3F capital expenditures (44.9%) came from companies that had current year payroll.

**Table 7. Stratum 3F Capital Expenditures by Reporting of Administrative Data**

Payroll > 0	Capital Expenditures (millions of dollars)	Percent of Total	
			Prior Year
Current Year	5492.9	44.9	
Sales/Receipts > 0	Prior Year	21.6	0.2
Current Year	2068.3	16.9	
No Administrative Data	4650.0	38.0	
Total	12241.5	100.0	

<sup>2</sup> On the ACE-2 questionnaire, companies report capital expenditures in thousands of dollars.

Table 8 shows the response rates by the category of reporting administrative data. Companies with current year administrative data responded at a higher rate than those that had prior year or no administrative data.

		Respon- dent	Non- respon- dent	Out of Scope
Payroll > 0	Prior Year	60.0	40.0	0.0
	Current Year	82.1	17.4	0.5
Sales/ Receipts > 0	Prior Year	41.2	58.8	0.0
	Current Year	77.4	21.0	1.7
No Administrative Data		32.6	59.5	7.9

Table 9 shows the percentage reporting administrative data by response category. Companies with no administrative data had the highest percentage of nonrespondents (86.1%) and out of scope cases (94.2%). Ironically, companies with no administrative data also had the highest percentage of the respondents (46,6%).

	Payroll >0		Sales/Receipts > 0		No Admin- istrative Data
	Prior Year	Cur- rent Year	Prior Year	Cur- rent Year	
	Respondent	0.5	28.1	0.6	
Non- respondent	0.4	6.0	0.9	6.7	86.1
Out of Scope	0.0	0.5	0.0	1.7	94.2

In summary we concluded the following about companies that exist on the SSEL with no administrative data:

- Capital expenditures collected from these companies are large enough to continue sampling these companies for ACES in the future.
- There is a high rate of nonresponse for these companies.
- Companies that are inactive are likely to remain inactive. About a quarter of the active cases later became inactive and are unlikely to have significant capital expenditures. Removing

inactive cases is a good first step in weeding out the undesirable cases from the frame, however, we would like to have more administrative data that could help us sort out more of the inactive companies.

- Companies that reported current year payroll provided the largest portion of 3F expenditures; however, companies that have not reported administrative data still account for a considerable portion of the 3F expenditures. This suggests that absence of payroll does not necessarily indicate absence of capital expenditures.

#### **IV. Matching the Frames from Two Consecutive Years**

For our second research effort to identify potential birth companies, we matched the 1997 ACES employer frame against the 1996 ACES frame. It should be noted that the ACES employer frame for a given year consists of companies that had payroll in the prior year or indicated that they would have payroll in the future. We created a list of 987,764 potential births from single establishment employer companies that were in the 1997 ACES frame but could not be matched to any company in the 1996 frame. To simplify our data collection efforts, we matched the potential births to our 1997 ACES sample, and obtained 3,134 matched companies for our sampling frame. We selected a random sample of 500 companies from the frame and conducted a brief telephone survey to collect information about start-up behavior. Since this was an initial look at these companies, we kept the sample size small so that the data collection efforts could be done in-house. The results from this survey would determine whether we would do a more extensive survey in the future.

Table 10 shows the distribution of the respondents and the nonrespondents by category. The overall response rate was around 56 percent. The response rate was actually quite good considering that participation was voluntary, we conducted the survey in mid-December, and we instructed the interviewers (ACES data analysts) to make only three attempts to contact each company. Surprisingly, 6.6 percent (33 cases) of the sample was determined in the interview to be out-of-scope. The majority of the out of scope cases were farms. Some of the farms have the word 'farm' or 'ranch' included in their business name while others have no obvious way to identify the establishment as a farm. The majority of the nonrespondents were 'no contact' cases; that is, after three attempts, the analysts were unable to contact the business by phone. Table 10

also shows that 35 (12.6%) of the 278 respondents were no longer in business. These results did not provide much information about birth companies but did identify problems with the ACES frame.

**Table 10. Distribution of the Respondents and Nonrespondents**

	Number	Percent
Total Sample	500	100%
Respondents	278	55.6
Responded	243	48.6
Out of Business	35	7.0
Nonrespondents	189	37.8
No Contact	140	28.0
No Listing	21	4.2
Refusal	28	5.6
Out of Scope	33	6.6

The following are the definitions of the response and nonresponse categories:

- Responded - the respondent answered at least one of the main items in the questionnaire.
- Out of business - the establishment was already out of business at the time the interview was conducted
- No contact - the phone number was available but after three attempts by the analysts, they still were not able to contact the business
- No listing - there was no phone number on file, and hence, no way to contact the business
- Refusal - the respondent refused to give out information over the phone or declined participation in the survey upon learning that it was voluntary
- Out-of-scope - the company was a farm or ranch or its usual place of business was outside the United States

Our assumption for the birth companies was that they started operation in 1996 and that they had payroll in 1996 or would have payroll at some point. Table 11 shows that 29.9 percent of the respondents started operation before 1996 and 16.9 percent paid wages before 1996. These results suggest that there was an undercoverage problem with the 1996 frame. Why weren't these cases on the 1996 ACES frame? This is something we would like to investigate further. Table 11 also shows that over 26.3 percent of the respondents first had capital expenditures in 1996 or before.

**Table 11. Select Variables by Year**

Variables	Prior to	In	After	Missing
	1996	1996	1996	
	(%)	(%)	(%)	(%)
Started Operations	29.9	14.7	44.2	11.2
First Paid Wages	16.9	7.6	38.8	36.7
First Had Capital Expenditures	18.7	7.6	28.4	45.3

We can see from Table 12a the relationship between the year wages were first paid and the year the business was started. There were 29.9 percent that started operations before 1996, and about half (50.6%) of these also first paid wages prior to 1996. Once again, our original assumption is questionable. Our study did not target only birth companies.

**Table 12a. Year Wages Were First Paid by Year Business Started**

Year Wages Were First Made	Year Business Started				
	Prior to 1996 (%)	In 1996 (%)	After 1996 (%)	Missing (%)	All (%)
Before 1996	50.6	0	3.2	3.2	16.9
In 1996	4.8	39.0	1.0	0	7.6
After 1996	12.0	22.0	70.7	6.4	38.8
Missing	32.5	39.0	25.2	90.3	36.7
All	29.9	14.7	44.2	11.2	100

Table 12b shows the year companies first made capital expenditures by the year the business started. None of the businesses that started operations in 1996 made capital expenditures before that year. Similarly, only 3.7 percent of the 123 businesses that started operating after 1996 spent money on capital investments in 1996 or earlier.

**Table 12b. Year Capital Expenditures Were First Made by Year Business Started**

Year Capital Expenditures First Made	Year Business Started				
	Prior to 1996 (%)	In 1996 (%)	After 1996 (%)	Missing (%)	All (%)
Before 1996	59.0	0	1.3	3.2	18.7
In 1996	4.8	34.1	2.4	0	7.6
After 1996	4.8	24.4	52.0	3.2	28.4
Missing	31.3	41.5	43.9	93.5	45.3
All	29.9	14.7	44.2	11.2	100

Table 13 shows the year businesses started operations and the source year. Source year is the processing year when payroll was reported to the IRS and added to the Census Bureau's Business Register. We could assume that the source year would be the

same as or later than the year the business first started operations. For those businesses that started operations in 1996 or later, only 1 percent had a source year that is earlier than the year the business started. However, altogether there are 14 companies (or 5 percent) that have a source year earlier than 1996. Why were they not included in the 1996 ACES frame? Did these companies exist on the SSEL with an out of scope industry classification or ZIP code?

**Table 13. Source Year by Year Business Started**

Source Year	Year Business Started				
	Prior to 1996 (#)	In 1996 (#)	After 1996 (#)	Missing (#)	All (#)
Before 1996	8	0	0	6	14
In 1996	5	5	2	0	12
After 1996	51	27	114	18	210
Zero	5	0	0	1	6
Missing	14	9	7	6	36

We can see from Table 14 the 1996 capital expenditures of the respondent companies by type and amount. A total of 35 companies (13 percent) spent money on structures. Twice that (70 companies, 25 percent) bought equipment. It is also interesting to note that 4 of them spent from \$100,000 to \$1 million on capital expenditures in 1996. It is hard to ascertain exactly how much each company's total capital expenditures were because the ranges we asked in the survey were too wide and the exact dollar amount was not provided. As a result we see that we were underestimating 1996 capital expenditures. However, we will need to do further analysis to quantify the contribution of the birth companies.

**Table 14. Number of Companies by Amount and Type of 1996 Capital Expenditures**

Amount	Structures	Equipment
Less than \$100,000	30	58
\$100,000 to \$1,000,000	3	1
Did not specify amount	2	11
No expenditures	211	170
Missing	32	38

In conclusion, we need to investigate some of the problems with the frame creation before we undertake a

larger scale survey of this type. Here are some things we may look into:

- A better method for determining out-of-scope cases before the frame is created. If we can eliminate some of the farms before sample selection we may have a better frame of potential births and get a more reliable estimate of missing birth activity.
- A method for identifying companies which are no longer in business.
- A procedure for adjusting for undercoverage since our results indicated that some companies that paid wages before 1996 were not included in the 1996 ACES frame. The inclusion of the 3F cases may eliminate some of the problems but not all.
- A better method for identifying birth companies. This should lead to a larger scale survey of this type, but with a better understanding of what constitutes a birth company. This would result in a more reliable estimate of capital expenditures for birth companies.

## Bibliography

Champion, Elinor J. and Berry, John H. (1991), "Annual Capital Expenditures Survey," paper presented to the Census Advisory Committees of the American Economic Association and the American Marketing Association, at the Joint Advisory Committee Meeting, April 25-26, 1991.

Stetser, Marie C., Goodloe, John Michael, Sands, Mark S. (1998), "The Evolution of Survey Methodology for a Company Survey of Capital Expenditures," *1998 Proceedings of the American Statistical Association*.

U.S. Bureau of the Census (1999), *Annual Capital Expenditures: 1997*, ACE/97, U.S. Government Printing Office, Washington, DC.

Walker, Edward (1997), "The Census Bureau's Business Register: Basic Features and Quality Issues," *1997 Proceedings of the American Statistical Association*.