The Center for Economic Studies Program to Assemble Economic Census Establishment Information

Alfred R. Nucci*

Center for Economic Studies
U.S. Bureau of the Census

The Bureau of the Census has been collecting information on U.S. business populations since about 1810 [Bohme, 1987]. Economic historians have used the published aggregates from the economic census as a primary source of information on the evolution of the U.S. economy. More recently, individual scholars have used the responses of individual businesses, located on economic census' schedules now in the public domain [Bertin, Bresnahan, and Raff, 1996].[1] However, the use of these data is hampered by lack of a central data file in machine-readable format, similar to the Integrated Public Use Microdata Series [IPUMS] project for decennial census data [Ruggles, Hacker, and Sobeck, 1995]. This paper reports on a Census Bureau project to assemble economic census information in such a format for the "recent past."

This paper has two purposes. The primary one is to acquaint the business history community with the business data base efforts at the Bureau of the Census Center for Economic Studies (hereafter, the Center). The second purpose is to begin/continue the discussion of the current state of availability of economic census response information from the past two centuries.

The outline of the paper is as follows. An introductory section is a brief discussion of the economic census program, with a review of how the individual economic census responses differ from published aggregates and how in recent years these business-level data on employment, wages, and sales over time have proved useful in describing changing economic situations. The next section, serving the major purpose of this paper, is a description of the business database activities at the Center. I then turn to a discussion of how the Center's efforts may be placed in the context of the complete economic census series. A concluding section presents some observations on where to go from here.

---

* The views expressed in this paper do not necessarily reflect those of the Bureau of the Census or the Center for Economic Studies. The author wishes to thank Timothy Dunne, John Haltiwanger, Robert McGuckin, John Ostenso, and Daniel Raff for their suggestions and encouragement.

[1] For a discussion on business data files collected by state Bureau of Labor Statistics, see Carter, Ransom, and Sutch [1991].

BUSINESS AND ECONOMIC HISTORY, Volume Twenty-seven, no. 1, Fall 1998.
 ISSN 0894-6825.

Economic Census Establishment Response Data

The Bureau of the Census surveys individual businesses for information on industry, location, employment, payroll, and sales, among other information, in its periodic economic census program. The frequency and industrial coverage of these business enumerations has been irregular over the past two centuries. From 1810 to 1890, these data were collected along with the decennial census enumeration [Bohme, 1987]. Beginning in 1900, the enumeration was shifted to a quinquennial basis for about two decades, became biannual for about a decade and half, was irregular during the depression years, was largely suspended during the war years, and has become quinquennial again since 1947.

For about the first 100 years, the business enumerations covered the agriculture and manufactures sectors, occasionally with minerals.[2] Beginning in the late 1920s the business enumeration was expanded to include nonmanufacturing sectors. Note that similar to the population program at the Census Bureau, these responses are protected under the confidentiality guidelines governed by U.S.C. Title 13 and are subject to similar restrictions on their use, although with unclear rules on duration of data retention.

There is no regular procedure at the Bureau of the Census or at the National Archives and Records Administration [NARA] whereby these individual responses are retained – either on micro-film or in manuscript form, analogous to the policies developed for population schedules. Hence, the availability of these data from past economic censuses, particularly those prior to 1958, is not consistent – that is, response data may be available for some years (e.g., the census of manufactures for 1929) and not others (e.g. the census of manufactures of 1921). Exacerbating this irregular retention policy is the lack of a centralized data file creation program in the relevant disciplines (e.g., economic history, social organization etc.), again particularly for economic census years prior to 1958. I return to this observation in my concluding remarks.

Why should we care about the underlying response data if we have the published aggregates? It is these business responses that allow us to understand the underlying business population dynamics (e.g., formation, dissolution) that characterize changes in the aggregate-level data across time. A number of scholars from a variety of disciplines have contributed to this work over the past quarter century [Hannan and Freeman, 1989; King and Wicker, 1993; Pakes and Ericson, 1989]. It is fairly accurate to say that we are witnessing a new approach to the study of businesses and organizations over this period, characterized by an increased interest in the contribution of the heterogeneity of individual firms (within larger populations of firms) to changes in the broad aggregates.

---

[2] Given the uncertainty over classification of businesses by industry during this period, this observation is at best an approximation.

The Census Bureau established the Center for Economic Studies in the early 1980s to maintain census of manufactures schedules from 1963 forward in machine-readable form in a centralized data file, given increased demand for use of these data by academics and other researchers. While this program is a welcome departure from previous, haphazard procedures for maintaining economic census establishment data, there remains no agreement between Census and NARA on retention of the underlying schedules for any given year.

The Center for Economic Studies and Manufacturing Establishment Responses

Richard and Nancy Ruggles conducted a project at the Census Bureau in the late 1970s to identify linkages across census of manufactures establishment responses for 1972 and 1977 and for annual survey of manufactures respondents for intervening years [Govoni, 1982]. This project continued a Census Bureau tradition to engage members of the research community in its data collection activities.

However, access to these (and other) data had been limited, given their confidential nature, the requirement to conduct research at Census Bureau facilities in Suitland, Maryland, and the limited time available to Census Bureau staff to conduct disclosure of research output. The Center was established shortly after the conclusion of the Ruggles' project to "house" the resulting longitudinal data file, to provide facilities for outside researchers to use these data, to conduct studies on the longitudinal data file, and to continue adding information from later years' census of manufactures [McGuckin and Pascoe, 1988]. Note that the presence of these data at the Center does not reduce Bureau of the Census confidentiality restrictions, although staff are now available to guide researchers through the disclosure avoidance process.[3]

The Center has fulfilled these expectations, adding census of manufactures data from earlier years, 1963 and 1967, and from later years, 1987 and 1992, as the data became available from the operations area. What resulted was a longitudinal data file consisting of "operating" manufacturing establishments for the seven economic census years 1963 – 1992 and annual survey of manufactures data for the years from 1972 through 1995. Researchers could follow manufacturing establishments which changed ownership, although they occupied the same physical location, in this data file, in addition to a number of enhancements to the underlying data (e.g., harmonization of industry classification to the 1987 Standard Industrial Classification basis, standardization of variable names and values). In other words, the longitudinal research database (LRD) is a prototype for a centralized data file containing establishment information from a number of economic census years (along with other business surveys) similar in format and spirit, if not in scope, to the

[3] See Reznek, Jensen, and Cooper [1997] for more information. The Center site at the Census Bureau web site has information on how to use data housed at the Center.

IPUMS project database. Table 1 contains a brief outline of information available in the LRD.

Table 1: *Economic Census Variables*

Identification Items
- Establishment Identification Number
- Permanent Plant Number
- Tabulated Industry Code (SIC Based)
- Primary Product Class Code
- Legal Form of Organization Code

Location of Establishment
- FIPS State Code
- County Code
- Place Code

Employment
- Total Employment
- Total Salaries and Wages
- Production Workers
- Production Worker Wages
- Total Supplemental Labor Costs
- Legally Required Supplemental Labor Costs
- Voluntary Supplemental Labor Costs

Measures of Activity
- Total Value of Shipments
- Beginning and End of Year Inventories
- Beginning and End of Year Gross Assets
- Capital Expenditures
- Value Added

A wide variety of research[4] has been conducted with these manufacturing data. These include employment change [Dunne, Roberts, and Samuelson, 1988], gross job flows [Davis, Haltiwanger, and Schuh, 1996], the productivity effect of ownership transfer [McGuckin and Nguyen, 1995], and the spatial distribution of manufacturing activity [Nucci and Long, 1996].

The Center for Economic Studies – Expanding Beyond Manufacturing

The Center initiated an expansion of its manufacturing data base to include information on research and development, energy use, cost of pollution abatement, and "advanced" technology adoption in the manufacturing sector, available from Census Bureau surveys. The growing interest in these data caused an increase in the demand for economic census information outside the manufacturing area. This demand resulted in the initiation of a project to develop a longitudinal business database (LBD), using the Census Bureau's industrial directory [U.S. Bureau of the Census, 1978] as the central business register and the Bureau's quinquennial economic census program (and other

[4] See the Center for Economic Studies web page at the Census Bureau web site for a bibliography of Center working papers. Selected papers are also available at the location.

annual surveys) for detailed establishment information. In other words, we at the Center are now interested in the full range of activity covered in the more recent (i.e., for the years 1958 – 1992) economic census – mining, construction, manufactures, wholesale, retail, services.[5]

As noted above, the presence of these data at the Center does not reduce the Census Bureau disclosure avoidance requirement. Hence, access is restricted to "approved" projects and "cleared" individuals. How to obtain such access is outlined in a document on the Center web page at the Census Bureau web site and discussed in Reznek, Cooper, and Jensen [1997].

The Center's research is beginning to reflect the addition of information from manufacturing survey, economic census, and Census Bureau industrial directory sources. These include studies on construction [Belsky, Calabria, and Nucci, 1997], services [Foster, Haltiwanger, and Krizan, 1998], establishment dissolution by age in private, non-agriculture industries [Nucci, forthcoming], and wholesale [Nucci, 1997].

The Challenges in Creating an LBD

Assembling the information in order to create an LBD is a considerable task. Choosing the years 1958–1992 (and 1997 as it becomes available) not only allows for use of the more recent and "policy-relevant" information, but also takes advantage of the Census Bureau's "archiving" of these data in machine-readable format. (See Table 2 for an overview of economic census data available by sector and year.) This decision has implications for the contents of the business files. The files archived at the Census Bureau, for the most part, contain the "massaged" business information used in the tabulation phase of the Bureau's economic census program and hence, these data may be different from that on the individual schedules.

In principle, it ought to be a relatively straightforward exercise to obtain these data (i.e. a copy from archival tapes). This is not the case. The Census Bureau's legacy mainframe is a non-standard UNISYS (a descendant of the earlier UNIVAC computers). Further, the Census Bureau enhanced the capabilities of these computers with a proprietary file system. Hence the migration of data from earlier years not only has the more traditional problem of missing or incomplete file documentation and deteriorating tape files but requires the use of specialized software, often specially written for each file. This latter requirement has required often time-consuming review of the migrated data to establish the reliability of the conversion software.

Exacerbating this situation is the strain on Bureau resources, given that the Bureau is currently in the midst of an economic census cycle and anticipating the 2000 decennial census cycle. Further, the skills required for this

[5] The Census Bureau has separately prepared longitudinal data files from recent agriculture census. The 1992 economic census included an expansion of coverage to include finance, insurance, and real estate and to selected industries in the transportation and public utilities sectors.

task are no longer common at the Bureau and becoming increasingly rare as time progresses, in addition to the continuing diminution of Census Bureau institutional memory on the contents of earlier files and idiosyncracies on their creation. Hence, we at the Center now face bottlenecks based upon the inability to transfer data from machine-readable tape files located on the Bureau's legacy mainframe, in spite of the increase in computer technology over the past years.

Table 2: *Economic Censuses by Sector, 1958-1997*

	Agri-culture	Minerals	Construc-tion	Manufac-tures	Whole-sale	Retail	Selected Services
1958		W		W	W	W	W
1963		W		MR	W	W	W
1967	W	W	x	MR	W	W	W
1972	W	W	MR	MR	W	W	W
1977	W	W	MR	MR	W	W	W
1982	W	W	MR	MR	MR	MR	MR
1987	W	W	MR	MR	MR	MR	MR
1992	W	W	MR	MR	MR	MR	MR
1997		W	W	W	W	W	W

x = Economic Census Conducted, No Record of Establishment Responses
W = Economic Census Conducted, Machine-Readable File of Establishment Responses Awaiting Migration
MR = Economic Census Conducted, Machine-Readable File of Establishment Responses at Center
Source: Center for Economic Studies documents

How to resolve this problem? This is one of the reasons I am here today – to acquaint an audience interested in business and economic history with current developments and with the bottlenecks. I have come for assistance, either direct in material form (e.g., individuals knowledgeable in specific areas, human and financial resources) or indirect, and suggestions on how to proceed.

Back to the Future

Up to this point, I have discussed the assembling of a business database from economic census data currently archived at the Census Bureau in machine-readable form. From the perspective of this audience, these data are likely to be viewed as "current." Is it possible to "push back" beyond 1958 to 1954 or 1947? How about incorporating information from pre-World War II economic census years? And what about bringing to such a data series the efforts of economic historians who have created data files from selected portions of one or more earlier economic census? As many of you may know far better than I, such a task is far more daunting than the relatively "simple" creation of an establishment data file from post-World War II sources.

There are three observations here. One, such a project is no longer a Census Bureau initiative, insofar as these data are in the public domain. Two, hence, it is primarily a task for this organization and other organizations with

an interest in the study of businesses and organizations over time. This suggests that some type of steering committee may be in order. Short of this, is it possible to agree on standards (i.e., format or record layout) to be used in the creation of the individual years' data so that putting together a series over time is not cumbersome. Lastly, does the IPUMS project, and efforts that preceded it, serve as a potential model?

The remainder of this section is a sketchy outline of tasks that need to be performed. Two appear to me to take precedence. The most important of these is the assembly of information on the location of economic census responses – particularly for the years prior to 1958 – and an inventory of which data files have been created from which economic census. (Table 3 contains an outline of this information for the period 1900–1954; this same outline could be applied to data from the 1800s.) As I have mentioned above, this is important given the irregular frequency and coverage of the economic census program over the past 190 years. This is exacerbated by an idiosyncratic policy on retention of these data. The second task is the creation of a network of interested scholars and other researchers who have either begun this process and/or are interested in seeing it performed.

Table 3: *Economic Censuses Years and Availability of Establishment Responses by Sector, 1900-1954*

	Agricul-ture	Minerals	Construc-tion	Manufac-tures	Whole-sale	Retail	Selected Services
1899	x			x			
1904				x			
1909	x			x			
1914				x			
1919	x			x			
1921				x			
1923				x			
1925				x			
1927				x			
1929	x		x	F	x	F	
1931				T			
1933			x	T			
1935			x	F	x	F	
1939	x	x		x			
1947				x	x	x	x
1949	x						
1954				x	x	x	x

X = Economic Census Conducted, No Record of Establishment Responses
F = Economic Census Conducted, Micro-film of Establishment Responses
T = Economic Census Conducted, Textual Record of Establishment Responses
Source: National Archives and Records Administration, Economic Censuses Procedural Guide (various years)

Other tasks are a miscellany of technical, linkage, and harmonization issues. Assessing the state of current scanning (or other) technologies is key here. (Hence, we need to know which economic census have either micro-film and/or manuscript copies of respondents.) Such technology can assist in the otherwise laborious (and expensive) transfer of information from micro-film and/or manuscript responses to data files. We need to establish linkages among these responses. And last but not least, we need to harmonize responses of individual establishments along a number of dimensions – industry, geography, measures of performance.

Conclusion

The increasing power of computers and the changes in the way researchers have studied business populations over the past few years have provided the framework for business historians to address in a more consistent and organized manner, the development of business databases drawn from [publicly available] archival sources. The economic census schedules provide an "anchor" to these efforts, to the extent that these data have been preserved and are indeed comprehensive. It is now time to determine the value of economic census business databases and to decide whether preparing centralized data files from these data can play an integral role in the study of business history. From the point of view of the Center, these data – particularly pre-1958 data – provide the necessary historical complement to our current efforts.

References

Belsky, E., M. Calabria, and A. Nucci., "Survivorship in the Construction Industries: Evidence from the Census of Construction," Paper Presented at the 44th North American Meetings, Regional Science Association International, Buffalo, New York, November 7-9, 1997.

Bertin, A.L., T.F. Bresnahan, and D.M.G. Raff, "Localized Competition and the Aggregation of Plant-Level Increasing Returns: Blast Furnace Operations 1929-1935," *Journal of Political Economy*, 104 (April 1996), 241-266.

Bohme, F.G., "U.S. Economic Censuses, 1810 to the Present," *Government Information Quarterly*, 4 (1987), 221-243.

Carter, S.B., R.L. Ransom and R. Sutch, "The Historical Labor Statistics Project at the University of California," *Historical Methods*, 24 (1991), 52-65.

Davis, S., J. Haltiwanger, and S. Schuh, *Job Creation and Destruction* (Cambridge, MA, 1996).

Dunne, T., M.J. Roberts, and L. Samuelson, "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *Rand Journal of Economics*, 19 (1988), 495-511.

Foster, L., J. Haltiwanger, and C.J. Krizan, "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," Paper Presented at the Conference on Research in Income and Wealth, National Bureau of Economic Research, March 20-21, 1998, Silver Spring, MD.

Govoni, J. "Methodological Problems Related to Time Series File Developed from Census of Manufactures/Annual Survey of Manufactures," *Proceedings of the Workshop on the Development and Use of Longitudinal Establishment Data* (U.S. Government Printing Office, 1982).

Hannan, M.T. & Freeman, J.H., *Organizational Ecology* (Cambridge, MA, 1989).

King, J.C. and A.W. Wicker, "Demography for Organizational Populations: Methodological Review and Application," *Advances in Entrepreneurship, Firm Emergence, and Growth*, 1 (1993), 83-143

McGuckin, R. and G. Pascoe, "The Longitudinal Research Database: Status and Research Possibilities," *Survey of Current Business*, 68 (1988), 30-37.

McGuckin, R. and S.V. Nguyen, "On Productivity and Plant Ownership Change: New Evidence from the Longitudinal Research Database," *Rand Journal of Economics*, 26 (1995), 257-276.

Nucci, A. and L. Long, "Formation, Dissolution, and Survival of Metropolitan and Non-Metropolitan Manufacturing Establishments," Paper Presented at 43rd North American Meetings, Regional Science Association International, Washington, 1996.

Nucci, A., "The Demography of Business Closings," *Small Business Economics* (Forthcoming).

Pakes, A. and R. Ericson, "Empirical Implications of Alternative Models of Firm Dynamics," National Bureau of Economic Research, Working Paper 2983 (1989).

Reznek, A., J.M.R. Cooper, and J.B. Jensen, "Increasing Access to Longitudinal Survey Microdata: The Census Bureau's Research Data Center Program," Paper Presented at the Joint Statistical Meetings, Anaheim, California, August, 1997.

Ruggles, S., J.D. Hacker, and M. Sobeck, "Overview of the Integrated Public Use Microdata Series," *Historical Methods*, 28 (1995), 33-39.

United States Bureau of the Census, *The Standard Statistical Establishment List Program* (Washington, D.C., 1979), Technical Paper 44.