RECOVERING THE ITEM-LEVEL EDIT AND IMPUTATION FLAGS IN THE 1977-1997 CENSUSES OF MANUFACTURES

by

T. Kirk White
U.S. Census Bureau

CES 14-37 September, 2014

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Republication in whole or part must be cleared with the authors.

To obtain information about the series, see www.census.gov/ces or contact Fariha Kamal, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies 2K132B, 4600 Silver Hill Road, Washington, DC 20233, CES.Papers.List@census.gov.

Abstract

As part of processing the Census of Manufactures, the Census Bureau edits some data items and imputes for missing data and some data that is deemed erroneous. Until recently it was diffcult for researchers using the plant-level microdata to determine which data items were changed or imputed during the editing and imputation process, because the edit/imputation processing flags were not available to researchers. This paper describes the process of reconstructing the edit/imputation flags for variables in the 1977, 1982, 1987, 1992, and 1997 Censuses of Manufactures using recently recovered Census Bureau files. The paper also reports summary statistics for the percentage of cases that are imputed for key variables. Excluding plants with fewer than 5 employees, imputation rates for several key variables range from 8% to 54% for the manufacturing sector as a whole, and from 1% to 72% at the 2-digit SIC industry level.

[*] Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been re-viewed to ensure that no con dential information is disclosed. I thank Randy Becker and Cheryl Grim for helping me obtain access to the raw data described in this paper, for helping me begin the process of translating the data into formats readable by modern econometrics software, and for providing helpful comments on the paper. Email: thomas.kirk.white@census.gov.

1 Introduction

Like nearly all economic censuses or surveys, plant-level records collected in the U.S. Census of Manufactures (CM) sometimes include errors. The Census Bureau mitigates the effects of errors in the data by checking to see if plant-level records are internally consistent (e.g., detail items sum to reported totals) and plausible (e.g., ratios of variables are within certain bounds)—these checks are called *edit rules*. Records that fail the Census Bureau's edit rules are sometimes corrected by a Census Bureau analyst and sometimes replaced with automated imputations that satisfy the edit rules. The Bureau also imputes for data that is missing due to unit or item non-response.[1]

The Census Bureau's main goal in collecting the CM is to publish industry aggregates. The Bureau's imputations are not designed to be suitable for multivariate regression analysis. Imputations that are not designed for multivariate regression analysis of the completed data can strongly impact the results of that analysis (Little and Rubin (2002)).

As part of its processing of the CM data, the Census Bureau creates item-level flags indicating whether an item was imputed or changed by analyst corrections. Starting with the 2002 CM, those item-level edit/impute flags are available to researchers via the Census Research Data Centers.[2] However, prior to the 2002 census, item-level edit/impute flags were not made available to researchers using the CM microdata. Researchers using the earlier years of the CM microdata sometimes used reverse-engineering methods to identify imputed items (see, e.g., Roberts and Supina (1996), Roberts and Supina (2000), Foster, Haltiwanger, and Syverson (2008)). However, for some industries in some years these reverse-engineering methods fail to identify a significant percentage of the imputed data (White, Reiter, and

[1] Unit non-response occurs when the intended respondent does not respond to the census or survey. Item non-response occurs when the respondent returns the census or survey form, but does not answer one or more items on the questionnaire.

[2] The edit/impute flags and the impact of imputations on data in the 2002 and 2007 CM and the 2003-2006 Annual Surveys of Manufactures have been analyzed in White, Reiter, and Petrin (2012).

Petrin (2012)).

Fortunately, the Census Bureau's Center for Economic Studies recently recovered some old files which do contain item-level edit/impute processing flags for the 1977 and 1982 Censuses of Manufactures. These are just two of many files recovered from a Census Bureau Unisys mainframe computer before it was decommissioned in 2010.[3] The original data was stored on tapes in a combination of formats not usable by modern computers. Section 4 describes how I converted the original data into a usable SAS dataset. My hope is that this section will be useful to other researchers who wish to convert some of the many other old data files recovered by the Center for Economic Studies' historical data recovery project.[4]

The Census Bureau's Center for Economic Studies also has data files with item-level edit/impute flags for the 1987, 1992, and 1997 CMs. In an internal Census Bureau memo, Dunne (1998) analyzed edit/impute flags for the 1987 and 1992 Censuses of Manufactures and showed that in many industries, significant percentages of the data are imputed.[5] In section 5, I analyze the edit/impute flags for the 1977, 1982, 1987, 1992, and 1997 CMs. Specifically, I report, by 2-digit SIC industry sector and year, the percentage of observations for key variables that are imputed as part of the editing process. White, Reiter, and Petrin (2012) show that imputations in the 2002 and 2007 CMs have a significant impact on conditional distributions of key variables as well as estimates of within-industry productivity dispersion in selected industries. My hope is that publishing these edit/imputation rates will be useful for other researchers who are considering submitting a proposal to gain access to these data.[6] Relatively high imputation rates for key variables in some industries may also suggest a need to reassess existing

[3]See Becker and Grim (2011) for an analysis of some of the other files that were recovered: plant-level micro data from the Annual Surveys of Manufactures from the 1950s and 1960s.

[4]See http://www.census.gov/ces/dataproducts/recovered/ for a list of the recovered datasets and the process for requesting access to them.

[5]Dunne (1998) is available to qualified researchers using these data on approved projects in the Research Data Centers.

[6]The SAS code from this project will be made available to other researchers working with recovered historical data in the Research Data Centers.

published research that used these data for these industries.

2 Related Research

In addition to Dunne (1998) and White, Reiter, and Petrin (2012), there have been at least a few other analyses of missing or imputed data in the CM or in the broader Economic Census. As part of the 1982 Economic Census Evaluation Report, Ramos, Waite, and Cole (1986) studied imputation for small establishments (plants with fewer than 10 employees) in the CM. The study found that imputation using administrative records data for these small plants resulted in upward biases (compared to the tabulations from reported values) of 8 to 15% for total value of shipments, 12 to 22% for the total cost of materials, and 57 to 63% for beginning of year inventories, and downward biases of 37 to 32% for total new capital expenditures. However, since these plants were small, their contribution to industry totals in most industries was also small. The current paper complements Ramos, Waite, and Cole (1986) in the sense that I focus on plants in the CM with more than 5 employees.

Greenberg and Petkunas (1986) study the effects of editing and imputation in the 1982 Economic Census, focusing on 6 detailed industries in the Wholesale, Services, and Retail sectors. Although not directly comparable to the CM, the imputation rates reported by Greenberg and Petkunas (1986) are similar in magnitude to the imputation rates I find in the 1982 CM. For example, for the Sales variable (which is roughly comparable to total value of shipments in the CM), for their 6 industries Greenberg and Petkunas (1986) report imputation rates (i.e., the percentage of observations that are imputed) of 47.7, 22.5, 7.8, 23.3, 35.6 and 22.8%. As I report below, in the 1982 CM, excluding very small plants that are not sent a survey form, the total value of shipments imputation rate for the entire manufacturing sector was 19%.

Roberts and Monahan (1986) report item imputation rates in the 1977 CM for several key variables. For the entire manufacturing sector, *including* very small plants which are not sent a survey form (sometimes called Ad-

ministrative Records cases), they report imputation rates of 53%, 54%, 36%, and 54% for, respectively, total employment, production worker wages, total value of shipments, and total cost of materials. Administrative Records (AR) cases are roughly a third of the plants in the CM, and almost all of the data (except payroll) for these plants is imputed. Taking into account the AR cases, the imputation rates reported by Roberts and Monahan (1986) are consistent with my results for the 1977 CM.

More recently, Census Bureau researchers have begun to develop new response measures for the Economic Census, including weighted item response rates (Lineback, Oliver, and Willimack (2014)). Although not currently used in Economic Census processing, weighted item response rates may be calculated and used as a data quality measure in processing future Economic Censuses (Lineback, Oliver, and Willimack (2014)). The results presented below for the 1977-1997 CMs would complement those data quality measures by providing a measure of the extent to which item response/nonresponse rates in the Economic Census have changed over time.

3 The Census of Manufactures

The quinquennial Censuses of Manufactures are available to researchers via the Census Research Data Center network. The Censuses include roughly 300,000 manufacturing plants in each year. Plants with fewer than five employees, which account for about a third of the plants in the census, are not sent a survey form. Hence, most data for these plants are imputed from administrative records (AR). Following most researchers who use the Census of Manufactures (CM hereafter), I drop these AR cases. The final sample sizes for the analysis below consist of approximately 200,000 plants in each year.

The CM microdata includes well over 200 variables, including plant-level and product-level variables collected on the survey form, data from administrative records, and variables constructed as part of processing the census. In this paper I focus on only a handful of key variables that are frequently used both to produce aggregate statistics and in research using

the microdata: total employment, total value of shipments, total cost of materials, production worker wages, and production worker hours. To avoid disclosure issues, I do not report imputation rates for total salaries and wages (payroll), because the impute flag for this variable is almost never set.[7] These are not the only the variables for which which item-level edit impute flags are available, but they are some of the most-frequently-used variables in the CM microdata files.

4 Converting Unisys Data Files to SAS Datasets

The data from the 1977 and 1982 CMs were stored on data tapes on the Unisys mainframe in two different formats, requiring two different conversion programs. Fortunately the record layouts for both years are available. I will address each year's data in turn.

4.1 The 1977 Census of Manufactures Data

Plant-level records from the 1977 CM were recovered from the Unisys data tapes in 4 files — one file each for the 2-digit SIC industry sectors 20-24, 25-29, 30-34 and 35-39. Descriptions of what is on the data tapes and (in some cases) record layouts for those data are stored in paper files associated with numbered "registers." In many cases there are multiple registers that contain the same data, because data may have been copied or backed up multiple times. For the 1977 CM records, I used the files described in register 81-244. Fortunately, this register also includes a record layout for the 1977 CM product file, which was crucial to converting the data to a usable format. These data include all of the plant-level variables published as part of the Census Bureau's General Summary statistics for the Economic Census. Importantly, the files also include variable-specific flags indicating whether or not the value recorded for a given variable was the value reported by the respondent or if it was imputed or changed by Census Bureau processing.

[7] The Census Bureau has administrative records data on payroll for almost every manufacturing plant which is used to replace any payroll data that is missing due to nonresponse.

The data were stored on the Unisys tapes in a combination of binary[8] and Excess-3 (XS3) formats. XS3 is a way of storing data that was readable by the Unisys mainframe but is different from the ASCII character set that is used by most modern computers. To speed recovery of the files, they were converted from the Unisys data tapes into ASCII text files using two alternative assumptions: (i) all of the data was stored in XS3; or (ii) all of the data was stored in another format called Fieldata. Separate files were created using each assumption.

Data stored in Fieldata format consists of 6-character "words", where each character is one of 64 characters in the Fieldata character set. Each character, in turn, corresponds to one of the 64 binary numbers representable by 6 bits. For example, the character # represents 000011 in binary or the number 3 in decimal notation. Table 1 presents the Fieldata character set and the corresponding decimal numbers and 6-bit binary numbers. One can think of the words and characters as representing a base-64 number system. The 6th (right-most) character represents the numbers 0 to 63; the 5th character represents 0 to 63 multiplied by 2^6; the 4th character represents 0 to 63 multiplied by 2^{12}; and so on. For example, the decimal number 1000 is represented by the Fieldata word @@@@J*. The @ symbol represents 0, J represents 15, and * represents 40:

$$1000 = (0 \times 2^{30}) + (0 \times 2^{24}) + (0 \times 2^{18}) + (0 \times 2^{12}) + (15 \times 2^{6}) + (40 \times 2^{0}) \quad (1)$$

If the first character in a 6-character word is one of the final 32 characters in the Fieldata character set, then the word represents a negative number, which is calculated by subtracting 63 from the number represented by each character in the word.

Since there is a straightforward mapping from Fieldata to binary, I used the converted Fieldata files to reconstruct variables originally stored as binary. To reconstruct the variables originally stored in XS3 format, I use the converted XS3 files. On the record layout for the 1977 CM file, portions of

[8] Some documentation of the data tapes state that the data was stored in a format called Binary Integer. For the discussion in this paper, there is no meaningful distinction between binary and Binary Integer, so I will use the more familiar term — binary.

the record stored as binary numbers are indicated with a "B" with arrows running in either direction. The rest of the record is stored in XS3 format. To reconstruct the variables, I began by reading in chunks of the record and converting them as appropriate. For example, the first 6 characters of each record contain the 4-digit SIC industry code and a 2-digit sub-industry code, which are already converted to ASCII numbers in the converted XS3 file. "Words" 7-90 and 97-180 were orginally stored in binary format, so for these variables I use the Fieldata files. For example, the record layout indicates that word 21 contains the plant-level values for production worker wages.

Among the variables stored in binary format, there are two types of special cases in the record layout that require additional processing. The first type of special case involves the flags variables in the 18th, 19th, and 20th words of the record. These words are subdivided into 1- or 2-bit variables, defined in detailed "word explanations" also included in register 81-244 documentation. For example, the final bit of word 18 indicates whether or not the plant is an Administrative Records imputed case. I used a 4-step process to convert these flags data items to a usable format. First, using the converted Fieldata file, I calculated a single decimal number from the 6-character Fieldata word. Next I converted the decimal number to a single binary number. Then I converted the binary number to a 36-digit character string of 0s and 1s. Finally, I reconstructed the flags by subdividing the character string to correspond to the variables in "word explanations" in the register.

The second type of special cases are the item-level edit/impute flag variables mentioned above. For example, the record layout indicates that words 101-110 contain "Current Year Edit Flags". Page 14 of register 81-244 includes detailed "word explanations" for these words. The word explanations show that each character in each of these words corresponds to a 6-bit variable referred to as a CANBIRS flag variable, which I describe in detail below. For example, the first character of word 101 contains the CANBIRS flags for production worker wages. For each of the CANBIRS flag variables, instead of converting each 6-character word to a decimal number, I convert each

Fieldata character to a 6-bit binary string — a string of 6 zeros and ones.

Table 2 presents the flags and their meanings. The first three letters of the acronym — C, A, and N — correspond to the first two bits of the 6-bit variable. The letter C, representing 11 (first two bits both set to 1), means that the value of the variable in question was corrected from correspondence with the respondent, while the letter A (first bit set to 1, second bit set to 0), means that the value was corrected by a Census Bureau analyst. The character N (representing the first bit set to 0 and the second bit set to 1) means "analyst accepted computer action." Note that the C, A, and N values are mutually exclusive by definition. The letter B in the CANBIRS acronym represents the *third* bit being set to 0, and means that the value was not reported (blank). Note that if the respondent *did* report a value for the variable in question, the third bit would be set to 1, and the character version of the flag variable would be blank. The fourth bit in the CANBIRS flag was set to 1 if the corresponding variable was imputed. Note that it is possible for a variable to be both "reported" and "imputed" (I), because both the third and fourth bits could be set to 1. For example, if the reported values for a particular plant fail the Census Bureau's edit rules, then the Bureau may replace one or more reported values with imputed values. However, in most cases, when the "imputed" flag is set to 1, the "reported" flag is 0, indicating that most imputations in the 1977 CM were imputations for non-response. The fifth and sixth bits of the CANBIRS flags were set to 1 if the variable in question was, respectively, rounded (R) or if the sum of detail items was substituted for a reported total (S). Although all of the flags may be of interest to researchers, this paper focuses on the impute flags. If the fourth bit of the CANBIRS flag for a given variable is set to 1, then I classify that observation as imputed.

4.2 The 1982 Census of Manufactures Data

Like the 1977 CM files, plant-level records with item-level edit/impute flags from the 1982 CM were also recovered from the Unisys data tapes in 4 files — one file each for the 2-digit SIC industry sectors 20-24, 25-29, 30-34 and 35-39. These files are described in data storage register 86-077.

Unlike the recovered 1977 CM files, the 1982 files were stored on the Unisys data tapes in a combination of binary and ASCII formats. Instead of the 6-bit characters used in the Fieldata character set, the ASCII character set uses 9-bit characters. This poses a problem, because the files were copied from the Unisys data tapes and converted to ASCII under the assumption that the original files were in Fieldata format. Consequently, variables that were one ASCII character in length were converted to 1.5 Fieldata characters — that is, one Fieldata character and the first three bits of a second Fieldata character.

For variables that were stored on the Unisys machine in ASCII format, I used a five-step process to convert the Fieldata characters back to the original ASCII characters or numbers. First, in cases where the original variable was an odd number of ASCII characters in length, I combined contiguous variables in the record so that the total length of the combined variables was an integer length in Fieldata characters. For example, the eighth, ninth, and tenth ASCII words in the file correspond to three variables – the 1-character employment size class variable, the 10-character Permanent Plant Number (PPN), and the 1-character record type (rectype). I read these three variables into SAS as a single character string of 15 Fieldata characters. Second, for each character in each variable, I converted the Fieldata character to its corresponding 6-digit binary number and stored it as a character string. For example, I converted the Fieldata character "[" to the binary string "000001" (see table 1). Third, I concatenated the binary strings to form variables or combinations of variables. Using the employment size class-PPN-rectype example again, I concatenated fifteen 6-digit binary strings to form one binary string of 90 digits (stored as a character string). Fourth, I subdivided the binary strings into 9-bit chunks and converted each 9-bit binary string to the corresponding ASCII character. For example, the 9-digit binary string "000110001" converts to the ASCII character "1". Finally, for variables that were more than one ASCII character in length, I concatenated the ASCII character to form the original variable. For example, for the variable PPN, I concatenated 10 ASCII characters to reconstruct the original variable.

For numeric variables that were originally stored on the Unisys machine

in binary, I used the same process I described for the 1977 CM file to convert the Fieldata characters to decimal numbers. As with the 1977 files, item-level edit flags were stored as 6-bit CANBIRS flag variables, and some other "words" consisted of layout-defined flags in which each bit represents a different processing flag. As I did for the 1977 file, I reconstructed these flags by converting the 6-character Fieldata words to 36-bit strings of 0s and 1s, and then subdividing the bit strings according to the "word explanations" in the record layouts for the 1982 CM files.

4.3 Interpreting the 1987 and 1992 Censuses of Manufactures Item-level Edit/Impute Flags

Recovering the edit/impute flags for the 1987, 1992, and 1997 CM product files was easier than for the 1977 and 1982 files, because the Census Bureau's Center for Economic Studies had archived copies of these files in ASCII text format. Converting the ASCII text files into SAS datasets was straightforward. However, interpreting the edit/impute flags requires some knowledge of editing and imputation process for the Census of Manufactures as well as information gleaned from the record layouts scanned from the data storage registers. In this subsection I describe how I interpreted the edit/impute flags for the 1987 and 1992 CM product files. The flag definitions for the 1997 CM are somewhat different from 1987 and 1992, so I address them in the next subsection.

The 1987 and 1992 CMs continued to use the CANBIRS flag variables for the item-level edit/impute flags. As mentioned above, each of these CANBIRS variables was *originally* stored as a 6-bit binary number on the Unisys computer. However, in the ASCII text files that were archived by CES, these flag variables take one of the character values that represent the original combinations of bits (see column 3 of table 2): C, A, N, B, I, R, or S.

As noted above in subsection 4.1, the C, A, and N values are mutually exclusive by definition. However, in the original binary data it would be possible for a variable to be both "not reported" (B) and "imputed" (I), because the third bit could be set to 0 and the fourth bit could be set to 1.

In the ASCII data for 1987 and 1992 the edit/impute flags can be set to B (not reported) or I (imputed), but not both. Thus some information may have been lost when the data was converted from binary to the ASCII text files that are archived at CES. We still know whether the recorded value for an item was imputed or not; we just don't know if that imputation replaced missing data or reported data that was judged to be erroneous. For the imputation rates reported below, I define an item in the 1987 and 1992 CMs as imputed if the corresponding CANBIRS flag value is "I".

4.4 Interpreting the 1997 Census of Manufactures Item-level Edit/Impute Flags

Beginning with the 1997 Economic Census (which includes the Census of Manufactures), the Census Bureau switched to a new system of codes for the edit/impute flags. In the new system, each edit/impute flag consists of two or three characters. The first character is either a blank, indicating that the item was not reported on the survey form, or an 'R', indicating that it was reported. The second and (if applicable) third characters take one of 22 values. Table 3 list the 22 codes (including blank) and the names of each code. Tables 4 and 5 briefly describe when each code is set. This new system of flag codes is used for all the main plant-level variables that are edited and imputed in the CM. In the 1997 CM file, I define an item as imputed if the second character in its edit/impute flag falls into one of the "Imputed" categories in table 3, with one exception: for the total employment variable, if the second character of the flag is the letter A (indicating imputation from administrative records data), I categorize the total employment value as non-imputed. For payroll and total employment, the administrative records data tends to be reasonably accurate (when compared to reported data for the same plant).

5 Imputation Rates in the 1977-1997 Censuses of Manufactures

Table 6 shows imputation rates by 2-digit SIC sector[9] for key variables in the 1977 CM: total employment, total value of shipments, total cost of materials, production workers wages, and production worker hours. The imputation rate for a given variable is the percentage of observations of that variable that are imputed. For these calculations (and for all the imputation rates reported in the tables described below), I excluded Administrative Records (AR) cases (which make up about a third of the plants in the CM). For AR cases, all variables except payroll and employment are imputed. Therefore the percentages in table 6 are lower bounds for the overall imputation rates for all of these variables except employment. There is significant variation in the imputation rates both across variables and across industry sectors for a given variable. For example only 8% of non-AR plants have imputed total value of shipments (TVS) in textiles, but 29% of non-AR plants had imputed TVS in apparel manufacturing. Except in the textiles sector, the production worker hours variable has the highest imputation rates of these variables. In fact, in 6 of the 19 sectors, more than 50% of non-AR plants have imputed values for production worker hours: apparel manufacturing; wood products; printing and publishing; stone, clay, and glass manufacturing; industrial machinery manufacturing; and miscellaneous manufacturing. Across all manufacturing sectors, nearly half of all non-AR plants (48%) had imputed values for production worker hours in 1977.

Since the Census Bureau mostly focuses on publishing aggregate statistics from the CM, the Bureau devotes more resources to obtaining complete responses from larger establishments. As a result, item-level response rates are higher (and thus imputation rates are lower) for larger manufacturing plants. Table 7 shows the imputation rates in the 1977 CM by 2-digit SIC sector, where each plant is weighted by its share of total value of shipments.[10] Except for the total cost of materials in the textiles manufacturing

[9] To avoid disclosure issues, I combined SIC sector 20 and 21 into one sector.

[10] As in table 6, Administrative Records cases (plants with fewer than 5 employees,

sector, and production worker wages in the primary metals manufacturing sector, the shipments-weighted imputation rates are lower than the unweighted imputation rates in every sector and for every variable in tables 6 and 7. However, even on a shipments-weighted basis, in several sectors a significant percentage of the plants have imputed data for these key variables. For example, in apparel manufacturing, the shipments-weighted imputation rates for total employment, total value of shipments, total cost of materials, are, respectively, 29%, 22%, 28%, and 33%. In the printing and publishing sector, the shipments-weighted imputation rates for total employment, production worker wages, and production worker hours, are, respectively, 21%, 27%, and 35%.

The final row of table 7 shows shipments-weighted imputation rates for the entire manufacturing sector (excluding Administrative Records plants). Except for the total cost of materials variable, the shipments-weighted imputation rates for the manufacturing sector as a whole are significantly lower than the imputation rates for most or all of the 2-digit sectors. This shows that sectors with larger values of shipments also tend to have higher response rates (and lower imputation rates).

Table 8 shows the unweighted imputation rates by 2-digit SIC sector in the 1982 CM, again excluding Administrative Records plants. With the exception of total empoyment, in most industries the imputation rates for key variables in the 1982 census are similar to for 1977. For total employment, the imputation rates are 10 to 20 percentage points lower in every sector except textiles manufacturing, for which the 1982 rate is 4 percentage points higher than in 1977. Textiles manufacturing is outlier for other variables as well — this sector's imputation rates for total value of shipments, total cost of materials and production worker hours are, respectively, 14, 34, and 41 percentage points higher in 1982 than in 1977.

Table 9 shows the shipments-weighted imputation rates for the 1982 CM. Surprisingly, the shipments-weighted imputation rates for total employment are within a few percentage points of the unweighted imputation rates in every 2-digit sector, implying that larger plants (as measure by total value of

which are not sent a survey form) are excluded from these calculations.

shipments) are no less likely to have their employment imputed than smaller plants in 1982. With the exception of production worker wages in the paper products sector, the shipments-weighted imputation rates are lower than unweighte imputation rates for every variable and every sector. The differences between the weighted and unweighted rates are especially large for production worker hours. However, even the shipments-weighted imputation rates for production worker hours range from 9% in the paper products manufacturing sector to 40% in apparel manufacturing sector. Compared to 1977, the shipments-weighted imputation rates for total employment are between 4 and 15 percentage points lower in all but four sectors (textiles, paper products, petroleum & coal products, and primary metals manufacturing). For the other key variables, the imputation rates in 1982 are quite similar to those in 1977, with the exception of production workers wages in primary metals manufacturing (23 percentage points lower in 1982).

Table 10 presents the unweighted imputation rates by 2-digit SIC sector in the 1987 CM. Compared to the unweighted imputation rates in 1982, the most noticeable changes are that the imputation rates for total employment are several percentage points lower in every sector, and the imputation rates for production worker wages are on average 19 percentage points *higher* in 1987. Table 11 shows the shipments-weighted imputation rates for 1987. With the exception of total employment (for which the unweighted imputation rates in are quite low), we again see that the shipments-weighted imputation rates tend to be significantly lower than the unweighted rates. In a few sectors (e.g., apparel manufacturing, wood products, and printing & publishing) the shipments-weighted imputation rates for total cost of materials, production worker wages and production worker hours continue to be in the 20 to 30% range.

The imputation rates for 1992 are presented in tables 12 and 13. Compared to 1987, the unweighted imputation rates in 1992 (table 12) are about the same for total employment and on average 13 percentage points lower for total value of shipments. With a few exceptions at the 2-digit industry level, the unweighted rates for total cost of materials, production worker wages and production worker hours are a few percentage points points lower than

in 1987. One notable exception is the leather products industry, in which the imputation rates for production worker wages and production worker hours both increased 17 percentage points in 1992 compared to 1987. Shipments-weighted imputation rates (table 13) continued to be significantly lower than the unweighted rates for every variable except employment.

Finally, tables 14 and 15 present, respectively, the unweighted and shipments-weighted imputation rates for 1997. Compared to 1992, the unweighted imputation rates in 1997 increased for every variable and in almost every sector. For the manufacturing sector as a whole, the imputation rates for total employment, total value of shipments, and total cost of materials increased, respectively, 7, 19 and 14 percentage points. The unweighted imputation rates for production worker wages and production worker hours also increased by, respectively, 6 and 8 percentages points. It is important to note that for 1992 and earlier years I classified items as imputed or not using the CANBIRS flag variables presented in table 2. In 1997, much of the editing and imputation processing for the Census of Manufactures was done by a new set of computer programs which used the Plain Vanilla editing and imputation flags presented in table 3. It is possible that some or all of the change in the item-level imputation rates from 1992 to 1997 is due to this change in the definition of when an item is imputed or not. Although it is not clear how much of the increase in imputation rates was due to changes in the definition of imputation, the shipments-weighted imputation rates also increased 6 to 10 percentage points for each variable in table 15.

5.1 Imputation Rates by Type in the 1997 Censuses of Manufactures

The Census Bureau uses a variety of different types of imputation methods and models to impute for missing or faulty data in the Economic Census. The CANBIRS flags used in the 1977-1992 Census only tell us whether or not an item was corrected by an analyst, reported, imputed, rounded, or raked. However, as noted above, the Plain Vanilla flags used in the 1997 Census of Manufactures provide more information about what *type* of imputation (if any) was done for each item. The Plain Vanilla flags can take on any of the

values listed in table 3. However, some of these values are set infrequently. To avoid disclosure of confidential information, I put the flags into one of nine categories.

Reported data that pass all the edit rules for the Census of Manufactures are the first category. For these cases, the first character of the flag is set to "R" and the second character is blank. I define the remaining eight categories of flags as follows. Administrative Records Data (second character of flag is "A"), Analyst Corrected ("C"), Direct Substitution ("S"), Historic Values ("H"), Industry Average ("V"), and Logical ("L") each get their own category. I group the flag values "D" (Donor Model Record), "E" (Endpoints of Upper/Lower Limits), "J" (Subject Matter Rule), "M" (Midpoints of Limits), "P" (Prior Year Administrative Records Data), "T" (Trim and Adjust Algorithm), and "W" (Warm Deck Statistical) into one category, "Other Imputation Methods." I group the remaining non-imputation flag values — "G" (Goldplated), "K" (Raked), "N" (Rounded), "O" (Override Edit with Reported Data), "U" (Unable to Impute), "X" (Unusable), and "Z" (Acceptable Zero) into the final category, which I call "Other Non-imputes". These flags are defined in tables 4 and 5. Note that the flag value "B" (Beta Cold Deck Statistical) is defined by the Plain Vanilla system, but was not used in the 1997 Census of Manufactures. Table 16 presents the percentage of items (for the entire manufacturing sector, excluding plants that are not sent a questionnaire) in each of the above nine categories for each of five key variables. Reported data that passes all the edit rules accounts for 86% of total employment cases, 56% of total value of shipments observations, and only 42% of production worker hours observations. The categories Direct Substitution, Historic Values, Industry Average, Logical, and Other Imputation Methods are different types of imputation. For each of the key variables, the majority of imputations in the 1997 CM were constructed using the Industry Average method.[11] For example, 64% of the imputations for total value of shipments in 1997 were created using the Industry Average method. For total cost of materials, production worker

[11] This method imputes for a variable X for plant i by multiplying another variable Y for plant i by an industry average of the ratio of X over Y.

wages, and production worker hours, respectively, 76%, 86% and 85% of the imputations were constructed using the industry average ratio method.

6 Conclusions and Future Research

This paper has described how I converted recently recovered historical data from the 1977-1997 Censuses of Manufactures (CM), including item-level edit/impute flags. I also report average imputation rates for key variables in the 1977-1997 CMs. The (unweighted) imputation rates (which exclude very small plants that are not sent a survey form) range from 8% to 54% for the entire manufacturing sector, depending on the year and the variable. At the 2-digit SIC industry level, the rates range from 1% to 72%. Shipments-weighted imputation rates are considerably lower than unweighted rates, but still range from 20 to 30% for cost of materials, production worker wages, and production worker hours in some 2-digit industry sectors.

In this paper I have focused on the edit/impute flags for several key variables in the CM. The recovered historical files contain edit/impute flags and other types of processing flags for many other variables from the CM. Future research could fruitfully focus on these other flags and other variables.

References

Becker, R. A., and C. Grim (2011): "Newly Recovered Microdata on U.S. Manufacturing Plants from the 1950s and 1960s: Some Early Glimpses," Working Papers 11-29, Center for Economic Studies, U.S. Census Bureau.

Dunne, T. (1998): "CES Data Issues Memorandum 98-1," CES data issues memorandum, Census Bureau Center for Economic Studies.

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

Greenberg, B., and T. Petkunas (1986): "An Evaluation of Edit and Imputation Procedures Used in the 1982 Economic Censuses in Business Division," Statistical Research Division Report Series Census/SRD/RR-86/04, Statistical Research Division, U.S. Bureau of the Census.

Lineback, J. F., B. Oliver, and D. K. Willimack (2014): "Developing Response Metrics for the Economic Census," Federal Committee on Statistical Methodology Research Conference Paper, U.S. Bureau of the Census.

Little, R., and D. Rubin (2002): *Statistical Analysis with Missing Data, Second Edition*. John Wiley, New York.

Ramos, M., P. J. Waite, and S. J. Cole (1986): "Evaluation Study of the Imputation of Small Manufacturing Companies in the 1982 Census of Manufactures," 1982 Economic Census Evaluation Task Force Report, U.S. Department of Commerce, Bureau of the Census.

Roberts, M., and J. Monahan (1986): "The Effects of Nonsampling Errors on the Development and Use of the Longitudinal Establishment Data (LED) File," in *Second Annual Research Conference Proceedings*, Washington, D.C. Bureau of the Census, U.S. Department of Commerice.

Roberts, M. J., and D. Supina (1996): "Output Price, Markups, and Producer Size," *European Economic Review*, 40(3), 909–921.

——— (2000): "Output Price and Markup Dispersion in Micro Data: The Roles of Producer Heterogeneity and Noise," in *Advances in Applied Microeconomics, Vol. 9, Industrial Organization*, ed. by M. R. Baye, chap. 4. JAI Press.

White, T. K., J. P. Reiter, and A. Petrin (2012): "Plant-level Productivity and Imputation of Missing Data in U.S. Census Manufacturing Data," NBER Working Paper No. 17816.

Table 1: Mapping from Fieldata to Decimal and Binary

Fieldata	Decimal	Binary	Fieldata	Decimal	Binary
@	0	000000	)	32	100000
[	1	000001	−	33	100001
]	2	000010	+	34	100010
#	3	000011	<	35	100011
ˆ	4	000100	=	36	100100
	5	000101	>	37	100101
A	6	000110	&	38	100110
B	7	000111	$	39	100111
C	8	001000	∗	40	101000
D	9	001001	(	41	101001
E	10	001010	%	42	101010
F	11	001011	:	43	101011
G	12	001100	?	44	101100
H	13	001101	!	45	101101
I	14	001110	,	46	101110
J	15	001111	\	47	101111
K	16	010000	0	48	110000
L	17	010001	1	49	110001
M	18	010010	2	50	110010
N	19	010011	3	51	110011
O	20	010100	4	52	110100
P	21	010101	5	53	110101
Q	22	010110	6	54	110110
R	23	010111	7	55	110111
S	24	011000	8	56	111000
T	25	011001	9	57	111001
U	26	011010	’	58	111010
V	27	011011	;	59	111011
W	28	011100	/	60	111100
X	29	011101	.	61	111101
Y	30	011110	”	62	111110
Z	31	011111	_	63	111111

Table 2: Interpreting the CANBIRS Flag Variables

Bit #'s	Bit Values	Meaning (and associated letter value)
1-2	11	Correction from correspondence with respondent (C)
	10	Analyst correction (A)
	01	Analyst accepted computer action (N)
3	1	Reported
	0	Not reported – blank (B)
4	1	Imputed (I)
	0	Not imputed
5	1	Rounded (R)
	0	Not rounded
6	1	Detail raked to total or sum of detail substituted for total (S)
	0	Not raked or substituted

The table shows the meaning of the CANBIRS edit/impute flags in the Census of Manufactures.

Table 3: Plain Vanilla Edit/Impute Flags in the 1997 Census of Manufactures

Code	Name	Category
(blank)	Flag Not Set	Non-imputed
A	Administrative Records Data	Imputed
B	Beta (Cold Deck Statistical)	Imputed
C	Analyst Corrected	Non-imputed
D	Donor Model Record	Imputed
E	Endpoints of Limits (Upper/Lower)	Imputed
G	Goldplated	Non-imputed
H	Historic Values	Imputed
J	Subject Matter Rule	Imputed
K	Raked	Non-imputed
L	Logical	Imputed
M	Midpoints of Limits	Imputed
N	Rounded	Non-imputed
O	Override Edit with Reported Data	Non-imputed
P	Prior Year Administrative Records Data	Imputed
S	Direct Substitution	Imputed
T	Trim and Adjust Algorithm	Imputed
U	Unable to Impute	Non-imputed
V	Industry Average	Imputed
W	Warm Deck Statistical	Imputed
X	Unusable	Non-imputed
Z	Acceptable Zero	Non-imputed

Table 4: Definitions of Plain Vanilla Edit/Impute Flags

Edit/Impute Action	Occurs when...
Administrative (A)	the item is imputed by direct substitution of corresponding administrative data (for the same establishment/record).
Cold Deck Statistical (B)	the item is imputed from a statistical (regression/beta) model based on historic data.
Analyst Corrected (C)	the reported value fails an edit, and an analyst directly corrects the (reported or imputed) value.
Model (Donor) Record (D)	the item is imputed using hot deck methods.
High/Low (E)	the item is imputed by direct substitution of value near (high or low) endpoints of imputation range.
Goldplated (G)	the reported value for the item is "protected" from any changes by the edit. The value of a goldplated item is not changed by the editing system, even if the item fails one or more edits. In general, the goldplate flag is set by an analyst.
Historic (H)	the item is imputed by ratio imputation using historic data for the same establishment (for example, prior year data imputation in Manufacturing)
Subject Matter Rule (J)	the item is imputed using a subject matter defined rule (e.g. y=1/2x).
Raked (K)	the sum of a set of detail items do not balance to the total. The details are then changed proportionally to correct the imbalance. This preserves the basic distribution of the details.
Logical (L)	the item's imputation value is defined by an additive mathematical relationship (e.g., obtaining a missing detail item by subtraction).
Midpoint (M)	the item is imputed by direct substitution of midpoint of imputation range.

Table 5: Definitions of Plain Vanilla Edit/Impute Flags (continued)

Edit/Impute Action	Occurs when...
Rounded (N)	the reported value is replaced by its original value divided by 1000.
Restore Reported Data (O)	the reported value fails an edit. Either an analyst interactively restores the originally reported value of an edit (set by the interactive update system) or the ratio module later "imputes" originally reported data for an item which was imputed in the previous edit pass.
Prior Year Administrative (P)	the item is imputed by ratio imputation using corresponding administrative data from prior year (for same establishment).
Direct Substitution (S)	the item is imputed by direct substitution of another item's value (from within the same questionnaire.)
Trim-and-Adjusted (T)	the item was imputed using the Trim-and Adjust balancing algorithm (balance module default).
Unable to Impute (U)	the reported item is blank or fails an edit, and the system cannot successfully substitute a statistically reasonable value for the original data.
Industry Average (V)	the item is imputed by ratio imputation using an industry average.
Warm Deck Statistical (W)	the item is imputed from a statistical (regression/beta) model based on current data.
Unusable (X)	the sum of a set of detail items cannot be balanced to the total because none of the scripted solutions achieved a balance.
Acceptable Zero (Z)	the reported value for an item is zero, and the item has passed a presence (zero/blank) test. This often occurs with part time reporters (e.g., births, deaths, idles). The zero value will not be changed, even if it fails one or more edits.

Table 6: Imputation Rates for Key Variables, by 2-digit SIC sector, 1977 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	17,037	22%	19%	27%	19%	33%
Textiles	5,138	5%	8%	3%	19%	1%
Apparel	17,694	34%	29%	48%	23%	53%
Wood Products	17,778	28%	24%	33%	35%	61%
Furniture	5,865	26%	21%	29%	20%	47%
Paper Products	5,021	17%	15%	19%	7%	25%
Printing & Publishing	25,110	29%	20%	32%	43%	64%
Chemicals	8,208	22%	18%	24%	23%	32%
Petroleum & Coal	1,882	23%	22%	25%	20%	35%
Rubber & Plastics	7,850	21%	18%	25%	20%	44%
Leather Products	2,073	26%	23%	32%	16%	43%
Stone, Clay & Glass	11,434	24%	21%	27%	30%	54%
Primary Metals	5,185	18%	17%	20%	15%	31%
Fabricated Metal	21,840	21%	17%	25%	27%	47%
Industrial Machinery	26,163	22%	19%	30%	39%	56%
Electronic Equipment	8,766	26%	22%	33%	17%	37%
Transportation	5,661	26%	23%	34%	19%	41%
Instruments	4,140	25%	21%	32%	27%	43%
Miscellaneous	9,198	27%	24%	33%	36%	62%
All	206,043	24%	20%	30%	28%	48%

The table shows unweighted imputation rates by 2-digit SIC industry sector for several key variables. The imputation rate for a given variable is the percentage of "observations" for which the impute flag for that variable is set (indicating that the value was imputed by the Census Bureau Administrative Records cases (plants with fewer than five employees) are excluded from these calculations. I have combined industry sectors 20 and 21 to avoid disclosure of confidential information. The final row of the table shows the imputation rates for the entire manufacturing sector.

Table 7: Shipments-weighted Imputation Rates for Key Variables, by 2-digit SIC sector, 1977 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	17,037	11%	10%	11%	7%	11%
Textiles	5,138	0%	4%	20%	1%	0%
Apparel	17,694	29%	22%	28%	16%	33%
Wood Products	17,778	14%	13%	16%	12%	22%
Furniture	5,865	15%	13%	15%	6%	19%
Paper Products	5,021	6%	5%	7%	4%	8%
Printing & Publishing	25,110	21%	12%	16%	27%	35%
Chemicals	8,208	10%	8%	9%	8%	9%
Petroleum & Coal	1,882	3%	9%	6%	4%	3%
Rubber & Plastics	7,850	11%	8%	10%	4%	16%
Leather Products	2,073	15%	13%	17%	4%	19%
Stone, Clay & Glass	11,434	12%	9%	11%	10%	23%
Primary Metals	5,185	7%	7%	8%	27%	8%
Fabricated Metal	21,840	10%	9%	11%	13%	14%
Industrial Machinery	26,163	9%	8%	9%	11%	11%
Electronic Equipment	8,766	12%	11%	9%	13%	9%
Transportation	5,661	9%	19%	16%	13%	6%
Instruments	4,140	11%	9%	9%	9%	11%
Miscellaneous	9,198	16%	14%	17%	11%	24%
All	206,043	1%	4%	20%	1%	1%

The table shows shipments-weighted imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 8: Imputation Rates for Key Variables, by 2-digit SIC sector, 1982 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	15,151	9%	22%	31%	16%	37%
Textiles	5,345	9%	22%	37%	7%	42%
Apparel	18,344	14%	23%	48%	21%	64%
Wood Products	17,602	9%	21%	39%	36%	68%
Furniture	6,469	6%	17%	36%	22%	56%
Paper Products	5,248	6%	18%	28%	12%	29%
Printing & Publishing	26,918	10%	16%	39%	40%	66%
Chemicals	8,650	9%	24%	30%	21%	36%
Petroleum & Coal	2,205	11%	32%	37%	20%	38%
Rubber & Plastics	8,708	5%	16%	26%	17%	42%
Leather Products	2,093	7%	22%	36%	10%	40%
Stone, Clay & Glass	11,144	9%	19%	27%	24%	48%
Primary Metals	5,705	6%	18%	27%	11%	36%
Fabricated Metal	24,549	6%	15%	29%	23%	49%
Industrial Machinery	30,712	6%	17%	33%	31%	54%
Electronic Equipment	10,915	8%	19%	34%	15%	43%
Transportation	6,110	7%	22%	39%	16%	48%
Instruments	5,039	7%	22%	37%	17%	45%
Miscellaneous	9,167	10%	20%	44%	28%	66%
All	220,074	8%	19%	35%	25%	52%

The table shows (unweighted) imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 9: Shipments-weighted Imputation Rates for Key Variables, by 2-digit SIC sector, 1982 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	15,151	6%	12%	14%	9%	14%
Textiles	5,345	7%	10%	13%	3%	15%
Apparel	18,344	16%	19%	30%	21%	40%
Wood Products	17,602	3%	14%	20%	13%	28%
Furniture	6,469	3%	12%	22%	6%	26%
Paper Products	5,248	5%	10%	10%	14%	9%
Printing & Publishing	26,918	6%	13%	20%	29%	33%
Chemicals	8,650	6%	16%	12%	9%	11%
Petroleum & Coal	2,205	4%	13%	10%	6%	12%
Rubber & Plastics	8,708	2%	11%	15%	5%	17%
Leather Products	2,093	7%	13%	17%	5%	18%
Stone, Clay & Glass	11,144	6%	10%	14%	10%	18%
Primary Metals	5,705	5%	9%	8%	4%	10%
Fabricated Metal	24,549	4%	12%	17%	7%	20%
Industrial Machinery	30,712	4%	13%	16%	10%	17%
Electronic Equipment	10,915	5%	9%	12%	12%	13%
Transportation	6,110	2%	7%	12%	5%	10%
Instruments	5,039	2%	12%	12%	6%	15%
Miscellaneous	9,167	7%	17%	27%	11%	28%
All	220,074	5%	12%	14%	9%	16%

The table shows shipments-weighted imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 10: Imputation Rates for Key Variables, by 2-digit SIC sector, 1987 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	12,259	2%	19%	23%	27%	29%
Textiles	3,850	3%	21%	27%	31%	32%
Apparel	13,961	5%	32%	32%	42%	55%
Wood Products	16,952	5%	35%	37%	53%	65%
Furniture	6,132	4%	28%	35%	48%	53%
Paper Products	4,526	2%	14%	19%	22%	25%
Printing & Publishing	29,861	4%	32%	37%	56%	63%
Chemicals	7,681	3%	20%	23%	29%	29%
Petroleum & Coal	1,894	5%	26%	27%	35%	34%
Rubber & Plastics	9,201	2%	23%	27%	35%	37%
Leather Products	1,070	1%	11%	15%	18%	19%
Stone, Clay & Glass	10,156	4%	24%	29%	43%	46%
Primary Metals	4,780	3%	24%	27%	35%	37%
Fabricated Metal	22,290	2%	25%	29%	46%	50%
Industrial Machinery	27,594	3%	26%	31%	53%	58%
Electronic Equipment	8,972	3%	24%	32%	37%	39%
Transportation	5,638	4%	28%	34%	41%	44%
Instruments	5,075	3%	27%	34%	37%	38%
Miscellaneous	7,373	4%	32%	38%	55%	61%
All	199,265	3%	27%	31%	44%	50%

The table shows (unweighted) imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 11: Shipments-weighted Imputation Rates for Key Variables, by 2-digit SIC sector, 1987 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	12,259	1%	6%	9%	10%	11%
Textiles	3,850	1%	5%	9%	10%	10%
Apparel	13,961	2%	11%	23%	30%	29%
Wood Products	16,952	2%	19%	21%	24%	28%
Furniture	6,132	1%	11%	16%	18%	20%
Paper Products	4,526	1%	3%	7%	8%	9%
Printing & Publishing	29,861	1%	18%	22%	32%	29%
Chemicals	7,681	1%	4%	7%	8%	8%
Petroleum & Coal	1,894	0%	4%	4%	5%	4%
Rubber & Plastics	9,201	1%	11%	15%	16%	16%
Leather Products	1,070	0%	2%	4%	5%	6%
Stone, Clay & Glass	10,156	1%	11%	17%	20%	21%
Primary Metals	4,780	0%	5%	8%	9%	9%
Fabricated Metal	22,290	2%	10%	15%	18%	19%
Industrial Machinery	27,594	1%	7%	13%	14%	16%
Electronic Equipment	8,972	0%	7%	13%	13%	14%
Transportation	5,638	0%	2%	5%	4%	4%
Instruments	5,075	1%	5%	12%	11%	12%
Miscellaneous	7,373	1%	13%	19%	22%	24%
All	199,265	1%	7%	11%	12%	13%

The table shows shipments-weighted imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 12: Imputation Rates for Key Variables, by 2-digit SIC sector, 1992 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	12,589	3%	14%	22%	29%	30%
Textiles	3,959	4%	14%	22%	28%	29%
Apparel	15,194	9%	21%	38%	49%	56%
Wood Products	19,221	6%	16%	30%	48%	61%
Furniture	6,962	5%	15%	29%	44%	47%
Paper Products	4,987	1%	10%	16%	20%	22%
Printing & Publishing	34,706	4%	13%	30%	52%	60%
Chemicals	8,401	3%	13%	17%	30%	29%
Petroleum & Coal	1,912	5%	17%	16%	30%	31%
Rubber & Plastics	10,928	2%	14%	21%	30%	32%
Leather Products	1,243	5%	12%	23%	35%	35%
Stone, Clay & Glass	11,142	3%	15%	24%	39%	43%
Primary Metals	4,672	2%	11%	18%	26%	28%
Fabricated Metal	24,580	2%	12%	21%	40%	44%
Industrial Machinery	31,580	3%	11%	24%	46%	50%
Electronic Equipment	10,175	3%	14%	26%	33%	34%
Transportation	6,358	3%	15%	29%	38%	40%
Instruments	6,312	3%	15%	28%	35%	36%
Miscellaneous	9,091	5%	15%	34%	56%	61%
All	224,012	4%	14%	26%	42%	46%

The table shows (unweighted) imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 13: Shipments-weighted Imputation Rates for Key Variables, by 2-digit SIC sector, 1992 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	12,589	0%	6%	8%	12%	12%
Textiles	3,959	0%	3%	5%	10%	10%
Apparel	15,194	2%	12%	20%	28%	30%
Wood Products	19,221	1%	8%	15%	20%	24%
Furniture	6,962	1%	8%	13%	17%	18%
Paper Products	4,987	0%	3%	5%	6%	7%
Printing & Publishing	34,706	1%	11%	17%	29%	29%
Chemicals	8,401	0%	3%	5%	8%	9%
Petroleum & Coal	1,912	0%	3%	2%	4%	6%
Rubber & Plastics	10,928	0%	8%	12%	14%	14%
Leather Products	1,243	1%	4%	14%	26%	26%
Stone, Clay & Glass	11,142	1%	8%	13%	18%	19%
Primary Metals	4,672	0%	3%	5%	7%	8%
Fabricated Metal	24,580	1%	7%	11%	16%	17%
Industrial Machinery	31,580	1%	5%	10%	13%	14%
Electronic Equipment	10,175	0%	5%	10%	11%	14%
Transportation	6,358	0%	1%	3%	4%	4%
Instruments	6,312	1%	7%	13%	15%	15%
Miscellaneous	9,091	1%	9%	17%	22%	26%
All	224,012	0%	5%	8%	12%	13%

The table shows shipments-weighted imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 14: Imputation Rates for Key Variables, by 2-digit SIC sector, 1997 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	13,001	14%	31%	38%	36%	42%
Textiles	3,974	14%	37%	35%	36%	43%
Apparel	15,981	18%	51%	59%	66%	72%
Wood Products	20,260	8%	37%	43%	47%	59%
Furniture	7,267	10%	33%	43%	47%	55%
Paper Products	5,102	7%	20%	23%	19%	24%
Printing & Publishing	35,651	9%	38%	43%	64%	69%
Chemicals	8,546	13%	28%	37%	33%	36%
Petroleum & Coal	1,770	13%	29%	34%	34%	37%
Rubber & Plastics	11,983	9%	28%	32%	36%	42%
Leather Products	1,009	15%	32%	45%	43%	50%
Stone, Clay & Glass	11,438	11%	28%	36%	41%	47%
Primary Metals	4,749	9%	19%	23%	28%	33%
Fabricated Metal	26,672	9%	25%	33%	46%	51%
Industrial Machinery	34,469	9%	31%	40%	54%	59%
Electronic Equipment	10,439	12%	31%	40%	37%	41%
Transportation	6,598	11%	34%	43%	42%	45%
Instruments	6,545	13%	32%	41%	39%	43%
Miscellaneous	9,116	13%	36%	43%	59%	65%
All	234,570	11%	33%	40%	48%	54%

The table shows (unweighted) imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 15: Shipments-weighted Imputation Rates for Key Variables, by 2-digit SIC sector, 1997 Censuses of Manufactures, non-Administrative Records cases only

Sector Name	Sample Size	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Food & Tobacco	13,001	10%	14%	19%	18%	21%
Textiles	3,974	10%	17%	15%	16%	20%
Apparel	15,981	14%	24%	27%	33%	40%
Wood Products	20,260	5%	18%	22%	21%	25%
Furniture	7,267	9%	13%	22%	23%	23%
Paper Products	5,102	4%	9%	10%	8%	9%
Printing & Publishing	35,651	11%	25%	28%	27%	29%
Chemicals	8,546	7%	9%	15%	14%	15%
Petroleum & Coal	1,770	17%	13%	26%	20%	23%
Rubber & Plastics	11,983	7%	16%	19%	19%	22%
Leather Products	1,009	9%	14%	22%	20%	27%
Stone, Clay & Glass	11,438	9%	13%	23%	22%	25%
Primary Metals	4,749	9%	5%	11%	14%	15%
Fabricated Metal	26,672	6%	9%	18%	20%	23%
Industrial Machinery	34,469	7%	8%	19%	23%	29%
Electronic Equipment	10,439	9%	8%	19%	22%	23%
Transportation	6,598	5%	7%	12%	14%	8%
Instruments	6,545	10%	10%	18%	22%	21%
Miscellaneous	9,116	9%	19%	25%	28%	33%
All	234,570	8%	11%	18%	19%	20%

The table shows shipments-weighted imputation rates by 2-digit SIC industry sector for several key variables. See notes for table 6.

Table 16: Unweighted Percentages, by Edit/Impute Flag Type, for Key Variables, 1997 Censuses of Manufactures, non-Administrative Records cases only

Type of edit/impute flag	Total Employment	Total Value of Shipments	Total Cost of Materials	Production Worker Wages	Production Worker Hours
Reported Data (Passes All Edits)	85.6	55.6	51.8	46.5	41.5
Administrative Records Data	1.3	0.0	0.0	0.0	0.0
Analyst Corrected	2.3	10.9	7.7	4.2	3.9
Direct Substitution	0.0	6.3	0.0	0.0	0.0
Historic Values	2.6	3.9	4.9	6.1	6.0
Industry Average	6.4	21.3	30.3	41.9	46.3
Logical	1.2	0.9	4.8	0.0	0.1
Other Imputation Methods	0.6	0.8	0.0	0.4	1.9
Other Non-imputes	0.1	0.4	0.4	0.8	0.4
All imputation types	10.7	33.1	40.1	48.5	54.2
All flag types	100.0	100.0	100.0	100.0	100.0

The table shows percentages of plants in the 1997 Censuses of Manufactures, by type of edit/impute flag, excluding Administrative Records cases (plants that are not sent a questionnaire). See text for descriptions of the categories of edit/impute flags.