

Technical Appendix

Automated and Manual Plant Matching Procedures for The Case of the Disappearing Large Employer Manufacturing Plants: Not Much of a Mystery After All

The automated matching algorithm

The goal of the matching algorithm is to match the 1,503 large-employer plants from 1997 to their corresponding plants in 2007, for those cases where the plant is still open in 2007 and a match exists. Analogously, the 1,014 large employers from 2007 are linked to 1997. Large employers are defined as plants with 1,000 or more employees. The data used are the public tabulations in the Subject Series: Location of Manufacturing Plants. For 1997, this is file E9731e2 from the 1997 Census of Manufactures (U.S. Bureau of the Census 2001). For 2007, this is file EC0731SA11 from the Census FTP site. (See the link in the references.)

The procedure—done automatically via computer—uses an algorithm to match plants by location, industry and plant size. The geographic information in the file is the Census five-digit place code and county. The industry information is the six-digit NAICS code. For each location and industry, the files report numbers of plants in the following employment size categories: “2,500 or more,” “1,000-2,499,” “500-999,” “250-499” and smaller categories that I do not use. The following table shows the number of plants in each of these categories for the two years.

Table A1
Plant Counts in Location of Manufacturing Plants Files

Employment Size Category	1997	2007
250 to 499 employees	7,854	6,154
500 to 999 employees	3,279	2,410
1,000 to 2,499 employees	1,187	822
2,500 employees or more	316	192

Part 1 of the automated matching algorithm takes the 192 plants from 2007 in the “2,500 or more” category and looks backward, chronologically, for matches in 1997. The following sequential procedure is used.

Step 1 looks for matches in the 1997 file in the “2,500 or more” category, looking first for an exact match on place, county and six-digit NAICS and then second for a match on just county and six-digit NAICS code.

For plants not matched at this stage, step 2 looks for matches in the “1,000-2,499” category in the same sequence as in step 1.

Step 3 repeats steps 1 and 2, only it loosens the NAICS match criterion to the four-digit level rather than the six-digit level.

Step 4 is to match remaining plants to the “500-999” size category at the four-digit level, first by place and county and then by county.

Part 2 of the algorithm is analogous to Part 1, but going in the other direction: taking plants from 1997 and linking them to their matches in 2007. It starts with the remaining unmatched plants in the “2,500 or more” category from 1997 and matches them to “2,500 or more,” “1,000-2,499” and “500-999” in 2007 using the same steps as in Part 1.

Part 3 takes the remaining unmatched plants from 2007 in the “1,000-2,499” category and matches them to 1997, using the same steps as in Part 1.

Part 4 takes the unmatched plants from 1997 in the “1,000-2,499” category and matches them to 2007, using the same steps as in Part 2.

Part 5 repeats all of the steps above in the same order, only now it looks for matches in the “250-499” category instead of the categories with 500 or more employees.

Manual matching

After running Part 1 of the procedure just described, I found that of the 192 plants in the “2,500 or more” category for 2007, 21 plants remained unmatched. I examined this list manually to see if the automated procedure described above was missing matches.

To determine identities of plants, I looked up plants in Dun and Bradstreet’s Million Dollar Directory, which provides information about address and employment. I also used information in the Toxic Release Inventory (TRI) from the Environmental Protection Agency, which is a database of pollution emissions of plants.

There are seven defense industry plants among the 21 unmatched plants from the “2,500 or more” category for 2007. With a very high degree of confidence, I was able to manually match six of these to existing large-employer plants at the same location in the 1997 file. The algorithm missed these matches because of a significant change in industry classification for these six defense plants. For example, two plants were in NAICS 334220, “Radio & TV broadcasting & wireless communication,” and they moved to NAICS 334511, “Search, detection, and navigation instruments.” While these industries are the same at the *three*-digit level, the industries differ at the *four*-digit level, and so were not matched in the automated procedure.

It is possible that in Parts 2 through 5 above, the automated algorithm misses additional defense industry matches for similar reasons. Time constraints prevented manually investigating each of these cases. Because defense plants are not a large portion of the total, I do not regard this issue as an important limitation of the analysis.

The Bensley, Va., plant in Table 2 in the “Non-Woven Fabric Mills” industry is another plant that existed in 1997 missed by the algorithm. I left this plant unmatched, and in footnote 6 of the paper, I refer to this case as an example of an imperfection in the matching algorithm. The DuPont Corp. at its website explains that the plant, called the

“Spruance Plant,” currently has 2,700 employees at the location.¹ The plant was built in 1929, and it corresponds to a plant in 1997 at the same location in the “1,000-2,499” category. The algorithm missed the match because the industry classification switched from “Noncellulosic Organic Fiber” to “Nonwoven Fabric Mills.” The records in the TRI file mentioned above show that the plant indeed produces in both of these industries.

References

U.S. Bureau of the Census. 2001. 1997 Economic Census. CD-ROM. Washington, D.C.: U.S. Dept. Commerce. C1-E97-NAFI-17-US1.

U.S. Bureau of the Census. 2007. Sector 31: EC0731SA1: Manufacturing: Subject Series: Location of Manufacturing Plants: Employment Size for Subsectors and Industries by U.S., State, County and Place: 2007. (Downloaded at U.S. Bureau of the Census, FTP <http://www2.census.gov/econ2007/EC/sector31/>.)

U.S. Environmental Protection Agency. Toxic Release Inventory. <http://www.epa.gov/tri/tridata/indexdata.htm>

¹ See http://www2.dupont.com/Spruance/en_US/.