

Appendix 1

The Indian Manufacturing Data Preparation

The Annual Survey of Industries data collected by the Central Statistical Organization, Ministry of Statistics and Programme Implementation through annual surveys of registered plants in the Indian manufacturing sector was purchased for this project.

I purchased data for the nine years 1986-87, to 1994-95, and the data includes information on material inputs, labor, wages and benefits, output, capital, inventories, etc. The data preparations I undertake include the following steps:

- Merging in data on industry, state and other codes
- Reconciling the sampling procedure and standardizing the variable measurement across the years.
- Attempting to construct a panel from the census sector plants, using available data on location, industry, year of initial production, etc

1. Reading in the rawdata

This is done using dictionary files (dictionary_files.zip) in program **a_readin_all.do**.

2. Merging in the year-wise datasets

Data on state codes, industry codes, organization, ownership, management, power, etc were provided separately. I merge these data (correcting for variation across the years in the case of the state codes) in file “**b_merge_codes.do**”.

3. Sampling procedure and correcting for the multiplier

3.1. Sampling procedure

The key features of the sampling procedure followed for the ASI is outlined below:

Feature	Time period	Description
Unit of enumeration	All Nine Years (1986-87 to 1994-95)	<u>Factory</u> for manufacturing industries, <u>workshop</u> for repair services, and <u>undertaking</u> for electricity, gas, water supply or bidi/cigar industries.
Sampling Frame	All Nine Years (1986-87 to 1994-95)	List of registered factories/units maintained by the Chief Inspector of Factories in each state, under the section 2(m) (i) and 2(m) (ii) of the Factories Act 1948; list maintained by licensing authorities for bidi, cigar and electricity undertakings. “Factory” under the Factories Act is “any premises”: i. Wherein 10 or more workers are or were working on any day in preceding 12 months, and in any part of a manufacturing process is being carried on with the aid of power; <u>OR</u> ii. Wherein 20 or more workers are or were working on any day in the preceding twelve months, and in any part of a manufacturing process is being carried on without the aid of power. “Manufacturing process” is defined as:

		<ul style="list-style-type: none"> i. Making, altering, ornamenting, finishing packing, oiling, washing, cleaning, breaking up, demolishing or otherwise treating or adapting any article or substance with a view to its use, sale, transport, delivery or disposal; OR ii. Pumping oil, water or sewage; OR iii. Generating, transforming or transmitting power; OR iv. Composing, types for printing by letter press, lithography, photogravure or other similar process or book binding; OR v. Constructing, reconstructing, repairing, refitting, finishing or breaking up ships and vessels. <p>Important exceptions are:</p> <ul style="list-style-type: none"> i. Establishments under the control of the Defense Ministry ii. Oil storage and distribution units iii. Restaurants and cafes iv. Technical training institutions v. Services industries such as sanitary services, recreation services (like motion picture production), personal services (like laundries), job dyeing, etc vi. States of Arunachel Pradesh, Mizoram, Sikkim and Union territory of Lakshadweep.
Revisions		Every three years, but updated every year by the regional office of the NSSO. Updations involve only additions of new firms; revisions involve deletion of de-registered factories. Lot of firms do not exist, even though they have not been deregistered, and these continue in the sampling frame (blank cards are filled by the samplers).
Sampling sectors	1987-88 to 1994-95	<p>i. <u>Census sector (Complete Enumeration or 'CE' category):</u> All the units are completely enumerated for:</p> <ul style="list-style-type: none"> - _Factories employing 100 or more workers - _All electricity undertakings - _All factories located in relatively less industrialized states and Union territories in Annex 1 - Those 3 digit NIC industry-state strata where number of factories is 20 or less <p>ii. <u>Segment S1:</u> A fixed sample of 20 for, circular systematic with random start:</p> <ul style="list-style-type: none"> - Those 3 digit NIC industry-state strata where number of factories is between 21 or 60 <p>iii. <u>Segment S2:</u> A sampling of one in three, circular systematic with random start:</p> <ul style="list-style-type: none"> - Those 3 digit NIC industry-state strata where number of factories is between 21 or 60
	1986-87	
Reference period	For any year	Accounting year for the factory ending any day during the fiscal year of the survey. For eg, for 1989-90, data relate to their accounting year ended on any day between April 1, 1989 and March 31, 1990, with the survey conducted in the year 1990-91 (July to June).
Classification scheme	1986-87, 1987-88, 1988-89	NIC 1970, based on principal product manufactured.
	1989-90 to 1994-95	NIC 1987, based on principal product manufactured.

3.2. Correcting for the multiplier and some other codes.

The data for the multiplier was missing for the years 1987-88 and 1988-89. On enquiry with the CSO, I learned that for these years, the multiplier had already been “applied” and that the data was “estimated”. That is, the data had already been weighted by their multipliers.

This complicates the comparison across years. However, fortunately analyses of other years indicate that the field “scheme code” and “modified scheme code” generally match one-for-one with the multiplier. My analysis using these scheme codes suggest that the multiplier has indeed been applied to all the variables; the group means for these years appear to be systematically higher than the group means for the other years.

Accordingly, I unravel the effects of “estimation” by first deducing the multiplier from the scheme codes and the basis explained in section above, and then dividing the “estimated” data by the multiplier to arrive at the comparable data numbers. This is carried out and explained in the file “**c_mult_adj.do**”.

Also, in mult_adj.do, I correct the “inityr” record for the year 1991-92. I convert this from two digit entry to the 4 digit entry corresponding to the other years by assign 1900 to this record. This could potentially bias plants that entered earlier than 1900; but my examination of the other years indicate that this is very rare -- less than 500 plants (out of 45,000 to 50,000) report non-zero initial years less than 1900. In any case, it is useful to take extra care to correct for any bias from this source.