

Problems and Prospects in the Use of ASI Data: A Study on ASI Panel Data

P C Mohanan and Anil Chopra¹, Computer Centre, New Delhi, India

Abstract

The Annual Survey of Industries (ASI) provides one of the longest continuous series of data on the manufacturing industries. Long term data like ASI is very helpful for time series studies to understand the changes in growth and structure of industries. It was expected that there should be a large sized panel data for ASI since the census units are being covered every year. This is not seen to be so from the data. Possible reasons for this are explored.

1. Introduction

1.1 The Annual Survey of Industries (ASI) provides one of the longest continuous series of data on the manufacturing industries anywhere in the world. Apart from the length of the data series, the most enduring aspect of the ASI is that it is specifically designed for economic analysis of the growth and structure of the manufacturing industries in great detail. As one of the basic objectives of the ASI is to meet the requirements of national income computation, the concepts and definitions used are in tune with the requirements of economic analysis.

1.2 However despite the wealth of details available in the ASI, the use of ASI data for research was rather limited till recently. The obvious reason for this was the difficulty in accessing firm level data. Now, with firm level data dissemination, the number of users is on the increase. Firm level data were accessed by 196 users from the computer centre till 2007. There have been 146 users since then. This has been despite the high cost of the ASI data.

1.3 There are several issues inhibiting wider and better use of ASI firm level data. Most of them have to do with the shortcomings in the data processing and metadata maintenance in the past. One such possibility is the availability of data for a panel of firms over the years. This paper discusses issues relating to preparation of panel data of firms from the ASI data.

2. Need for Panel Data

2.1 ASI data have been used for a variety of studies relating to employment growth, total factor productivity growth and input-output modeling in India. These studies have assumed importance in view of the significant changes in the industrial policies of the Government in the recent past. The existing data sets allow such studies at different geographical domains as also across industrial categories. This advantage of ASI is not shared by data from the corporate sector as the data from companies

¹ e-mail: anilc2k2@yahoo.co.in

cannot be disaggregated by geographical regions and industrial classifications based on the output of individual factories. At another level it is equally important for researchers to understand the changes taking place at the firm level in the changing economic environment. There have been demands from several users for providing a common firm level identifier in the unit level data so that a panel data for factories can be built up. Some of the users are also known to have developed algorithms for preparing a panel based on the opening and closing value of fixed assets in the ASI return to link the units in successive years.

3. Issues in Preparation of Firm Level Panel Data

3.1 The dissemination of firm level ASI data is required to comply with the provisions of confidentiality enjoined in the Collection of Statistics Act as also with the assurances given to the factories at the time of compiling the information from their records. In view of this, firm level identification is concealed while disseminating the unit level data. In view of this no efforts were made to disseminate data which provides any identification particulars of the unit. Therefore the first issue is to provide a factory identifier such that the identity of the factory is not revealed to the data users.

3.2 The second major issue relates to the data archived in the computer centre and has been anonymized to conceal their identity. In the absence of proper documentation and records, this makes it difficult to retrieve the original identification particulars.

4. Some Characteristics of Panel Data

4.1 In the ASI, the units are categorized as census units and sample units. As the name suggests the census units are surveyed every year and the sample units are surveyed depending on their selection during a particular year based on the sampling design adopted. The definition of the census unit is based on a threshold number of workers. The worker threshold for defining the census unit was changed during some years, from 100 or more workers to 200 or more workers. Table 1 gives the number of units appearing in the panel for different number of years.

4.2 The table-1 shows that there are only 4018 units figuring in all the 10 years considered. As can be seen from Table 2, most of these units come from the Census sector. As already noted, the definition of a census unit was changed for some of the Annual Surveys. Besides this, the number of workers noted in the ASI frame may also be updated with the number of workers reported in the ASI return each year.

4.3 In table-3 joint return has not been taken into account as there is no scope to include joint return because we are using unique PSL no. in panel data over the years.

5. Possible Reasons for the Small Number of Units in the Panel

5.1 As noted earlier, it was expected that there should be a large sized panel for ASI data since the census units are being covered every year. This is not seen to be so from the data. It is therefore necessary to explore the reasons for this.

5.2 One possible reason is that the Permanent Serial Number (PSL) of the factory recorded in the ASI return is not verified with the frame during the data processing. This is quite possible as the PSL number has no use in the processing especially for the census sector that does not require any extrapolation during estimation.

5.3 Mismatch of PSL numbers can also occur in the case of joint ASI returns where only one return is submitted by units having multiple under the same location. In such a situation the factory submits only one return noting all the PSL numbers covered under that specific return. In such cases the data records only one PSL number out of the many. If this number is different during different years, then it will be difficult to identify the panel from the data series.

5.4 Changes in the definition of census units and changes in the status of a census unit due to decrease/increase in the number of workers could also affect the length of the panel.

6. Panel Data and Data Quality

6.1 One of the major advantages of panel data from the point of view of the data producer is that it enables the validation of the data of one year with reference to the data submitted in the previous years. In table-4 the opening and closing value of fixed assets for a few randomly selected factories are provided. It may be checked if the closing value of one year is the same as the opening value of the next year.

7. Conclusion

7.1 The exercise in preparation of a panel dataset for ASI indicates some of the problems in making panel data due to the absence of a dependable linking identifier for each unit surveyed. Most importantly it shows that even for the units in the census sector which are supposed to be enumerated on an annual basis there are not many units that appear every year. Keeping this in mind, it is also necessary to consider if we need to maintain census status of a unit on a permanent basis as was the practice in the past.

Table 1: Length of Panel Data in Years (ASI: 1998-99 - 2007-08)

State/UT	length of panel in years									
	10	9	8	7	6	5	4	3	2	1
Number of units										
Jammu & Kashmir	63	112	149	185	231	275	379	496	654	876
Himachal Pradesh	70	115	162	210	275	358	513	767	1176	1741
Punjab	169	253	327	414	533	744	1468	3112	6382	12526
Chandigarh(U.T.)	71	95	101	113	139	188	232	325	403	520
Uttaranchal	25	54	174	216	255	309	423	670	1103	1857
Haryana	129	220	359	469	653	896	1434	2558	4615	7382
Delhi	52	104	177	264	385	601	1055	1932	3394	5181
Rajasthan	120	199	300	382	488	700	1082	2108	4241	7548
Uttar Pradesh	233	324	555	727	951	1283	1997	3842	7960	15096
Bihar	58	113	164	213	266	359	534	903	1462	2587
Sikkim										
Arunachal Pradesh										
Nagaland	44	100	104	112	117	126	147	173	189	194
Manipur	29	39	39	42	43	48	77	88	103	145
Mizoram										
Tripura	101	172	192	211	245	296	341	422	469	530
Meghalaya	20	25	31	34	42	48	60	67	80	109
Assam	126	209	320	403	477	581	816	1216	1823	2857
West Bengal	231	314	453	591	763	1022	1563	2785	4841	8329
Jharkhand	86	137	189	257	357	482	704	1080	1661	2576
Orissa	100	152	218	274	355	498	701	1084	1692	2663
Chattisgarh	45	91	182	251	324	423	551	872	1400	2216
Madhya Pradesh	152	233	341	441	568	775	1139	1860	2968	4784
Gujarat	254	388	630	809	1033	1345	2061	4153	9435	22093
Daman & Diu	34	69	113	153	211	292	450	771	1389	2421
Dadra & Nagar Haveli	30	56	88	149	207	300	469	737	1196	1886
Maharashtra	461	677	1001	1248	1576	2015	3099	5621	12089	25868
Andhra Pradesh	253	362	509	679	872	1167	1875	3982	9885	22552
Karnataka	167	281	497	699	965	1429	2221	3898	6980	12310
Goa	78	139	187	250	327	408	470	550	702	861
Lakshadweep										
Kerala	156	271	389	504	656	930	1381	2210	3891	6377
Tamil Nadu	548	809	1299	1719	2243	3092	4484	7253	14582	30665
Pondicherry	94	144	193	247	321	408	508	594	847	1126
Andaman & N. Island	19	19	20	21	21	21	24	26	29	29
Total	4018	6276	9463	12287	15899	21419	32258	56155	107641	205905

Table 2: Length of Panel in Years for Census Units (ASI 1998-99 - 2007-08)

State/UT	length of panel data in years									
	10	9	8	7	6	5	4	3	2	1
	Number of units									
Jammu & Kashmir	12	13	26	35	47	60	157	284	356	493
Himachal Pradesh	18	30	45	49	58	66	205	455	713	1105
Punjab	131	175	224	258	299	341	745	1426	2075	2689
Chandigarh(U.T.)	5	5	10	10	11	13	91	163	177	259
Uttaranchal	19	22	29	36	41	60	187	414	633	965
Haryana	90	137	228	275	329	413	643	1090	1676	2349
Delhi	10	18	44	63	99	150	391	648	933	1448
Rajasthan	69	101	162	190	226	266	397	633	903	1380
Uttar Pradesh	207	243	447	499	628	773	1082	1543	2210	3679
Bihar	23	24	43	49	55	70	185	306	365	649
Sikkim										
Arunachal Pradesh										
Nagaland	44	100	104	112	117	126	147	173	189	194
Manipur	29	39	39	42	43	48	77	88	103	145
Mizoram										
Tripura	101	172	192	211	245	296	341	422	469	530
Meghalaya	20	25	31	34	42	48	60	67	80	109
Assam	89	125	209	240	269	302	410	581	711	1069
West Bengal	200	240	346	403	469	545	691	982	1245	1819
Jharkhand	47	55	82	108	123	144	318	504	659	911
Orissa	37	48	70	81	102	132	264	393	493	654
Chhattisgarh	25	30	50	61	74	109	212	358	501	718
Madhya Pradesh	100	124	179	230	278	322	467	704	916	1402
Gujarat	234	311	517	613	779	918	1245	1754	2433	4122
Daman & Diu	5	11	24	39	57	76	163	283	336	540
Dadra & Nagar Haveli	3	10	29	34	51	76	154	300	372	558
Maharashtra	432	594	876	1011	1219	1457	2092	2735	3491	5338
Andhra Pradesh	206	250	398	447	569	696	945	1274	1766	2587
Karnataka	140	204	385	484	639	901	1267	1869	2611	3770
Goa	6	13	29	44	61	81	204	265	346	536
Lakshadweep										
Kerala	88	130	204	255	336	482	660	931	1180	1659
Tamil Nadu	504	680	1082	1319	1667	2158	3229	3989	5131	7396
Pondicherry	18	26	53	74	86	116	230	310	443	681
Andaman & N. Island	19	19	20	21	21	21	24	26	29	29
Total count	2931	3974	6177	7327	9040	11266	17283	24970	33545	49783

Note: In Table 1 and Table 2, States are uniform for all the 10 years from 1998-99 to 2007-08 as three states Jharkhand, Chhattisgarh and Uttaranchal data is also available from 1998-99. Districts are not included in the panel data to maintain the confidentiality.

Table 3: Length of Panel by NIC (ASI: 1998-99 - 2007-08)

NIC-2004 code	Number of units for different length in years									
	10	9	8	7	6	5	4	3	2	1
Number of units										
01	6	13	48	73	106	176	346	838	2040	4401
14			3	11	17	22	39	69	109	209
15	899	1346	1980	2410	2933	3664	5244	8974	17668	33623
16	72	91	127	217	293	433	590	988	2097	4979
17	602	854	1217	1502	1852	2441	3537	5735	10708	22122
18	78	153	290	424	613	892	1419	2308	4085	8040
19	53	88	149	241	338	448	649	1053	2032	3986
20	69	133	173	219	290	378	624	1220	2461	4828
21	72	126	193	264	341	459	707	1332	2733	5651
22	65	132	200	276	363	486	736	1333	2546	4802
23	43	71	100	129	178	251	369	625	978	1561
24	365	579	929	1196	1553	2005	2872	4625	8601	16211
25	62	129	226	302	393	561	945	2048	4869	10929
26	274	465	650	805	1009	1400	2471	4682	9810	20098
27	150	253	404	547	724	1051	1576	2927	5756	11580
28	81	162	290	404	556	777	1244	2468	5781	13496
29	191	333	501	671	933	1345	2036	3628	7103	14524
30		10	28	39	57	89	124	179	282	493
31	107	198	303	399	538	783	1178	2080	3757	7015
32	52	93	144	196	270	362	510	781	1294	2318
33	48	110	169	242	315	422	570	784	1121	1755
34	90	157	246	335	425	568	788	1321	2455	4711
35	101	158	206	256	352	464	650	1050	1742	3358
36	60	125	211	296	391	536	839	1410	2487	4849
37			1	3	5	16	30	60	91	265
40		4	5	30	47	89	146	222	362	587
41				3	7	16	26	38	69	121
45									2	8
50		32	57	92	141	242	403	814	1978	4312
52		4	11	18	24	37	60	92	141	273
63			6	10	20	39	104	344	860	1695
72					1	6	9	9	12	22
73						1	1	1	1	2
74			2	5	8	17	29	37	54	96
90		3	3	6	16	22	26	33	35	40
92		12	19	20	24	25	29	46	61	88
93		7	10	20	35	55	78	113	163	223
YY										573
ZZ										176
Total Count	3540	5841	8901	11661	15168	20578	31004	54267	106344	214020

Note: The differences in the total number of units in Table 1 and table 3 can arise due to the NIC of the unit changing over the years.

Table 4: Opening and Closing Values of Fixed Assets (in Rs.) for Different Years for Selected Factories

State	Factory No	year	Opening Value	Closing Value
Andhra Pradesh	1	1998-99	130571116	135357344
	1	1999-2000	136354668	133237727
	1	2000-01	132740401	129477605
	1	2001-02	129477609	146051504
	1	2002-03	146051506	153685994
	1	2003-04	153685994	154182220
	1	2004-05	154182218	153078294
	1	2005-06	153078294	200232094
	1	2006-07	200232094	199223232
	1	2007-08	199206847	559476590
	2	1998-99	49053658	46257176
	2	1999-2000	46257175	43818073
Maharashtra	2	2000-01	43818073	41987303
	2	2001-02	41577646	37559848
	2	2002-03	37559848	34025273
	2	2003-04	34025273	31266317
	2	2004-05	31266317	31984509
	2	2006-07	107956045	183927582
	2	2007-08	144594141	235882866
	3	1998-99	1146638825	987087462
	3	1999-2000	987086803	827463977
	3	2000-01	873375369	1267305839
Gujarat	3	2001-02	1267305939	1151710478
	3	2002-03	1156415182	1046679775
	3	2003-04	1046679775	936919276
	3	2004-05	936919276	875904458
	3	2005-06	875904458	782433023
	3	2006-07	782433023	701620102
	3	2007-08	701620102	672170569
	4	1998-99	186269133	192769308

Table 4 (cntd.): Opening and Closing Values of Fixed Assets (in Rs.) for Different Years for Selected Factories

State	Factory No	year	Opening Value	Closing Value
Maharashtra	4	1999-2000	192769307	198169844
	4	2000-01	198169845	200308619
	4	2001-02	200308617	191133414
	4	2002-03	191133414	189449864
	4	2003-04	189449864	200718603
	4	2004-05	200718604	211685063
	4	2005-06	211685065	245426369
	4	2006-07	245426365	260529298
	4	2007-08	260529298	273337023
	5	1998-99	52217	48495
Tamil Nadu	5	1999-2000	48495	45145
	5	2000-01	45145	45755
	5	2001-02	72885	67096
	5	2002-03	42679	39911
	5	2003-04	39911	37419
	5	2004-05	37419	35177
	5	2005-06	35177	33160
	5	2006-07	33160	31343
	5	2007-08	31343	29708
	6	1998-99	23679602	23111230
Uttar Pradesh	6	1999-2000	23111229	22585668
	6	2000-01	22585848	21725941
	6	2001-02	21722147	19860515
	6	2002-03	19854572	18403424
	6	2003-04	18403424	16869087
	6	2004-05	16869087	18159195
	6	2005-06	18159195	20002693
	6	2006-07	20002693	22158287
	6	2007-08	22158287	25635281