Presentation on BU's Indian Firm Data

Amrit Amirapu and Michael Gechter

April 3, 2014
(updated Aug 7, 2015)

Outline

The SCC Cluster

- For basic information about the Cluster and how to access it:
    - http://sites.bu.edu/mrysman
        - download: "High Performance Computing for BU Economists" slides
- To get access to the cluster:
    - students don't have their own accounts - must use the Econ dept account
    - e-mail Marc with: "your BU ID, your login name, your country of citizenship, and the e-mail address that you want to use"
    - will need to download some software
        - telnet-type software, eg: XQuartz
        - FTP type software, eg: FileZilla

The Data (Part 1)

- Annual Survey of Industries (1998/9 to 2009/10)
  - Dougherty, Frisancho and Krishna, 2013, "State-level Labor Reform and Firm-level Productivity in India", India Policy Forum

- Economic Census (2005)
  - Novosad and Asher, 2013, Working Paper

- NSS Unorganized Manufacturing Surveys (2000/1 and 2005/6)
  - Chemin, 2012, JLEO

The Data (Part 2) - Datasets Procured Since April 2014

- Annual Survey of Industries (2010/11 and 20011/12)
    - not yet read into Stata/formatted
- Economic Census (1998)
    - has been read into Stata
- NSS Unorganized Manufacturing Surveys (2010/11) - 67th round

ASI - Basics

- "[T]he principal source of industrial statistics in India." - MOSPI
- About the data:
    - FY1999 to FY2010,
    - country-wide, state-level
    - "panel" data
    - "factory level"
    - "scheme_code":
        - Census Sector: factories with 100+ workers & all factories in 5/6 less developed Statues/UTs
        - Sample Sector: $\approx 20\%$ sample of registered factories with <100 workers
        - "inflation_multiplier" - sampling variable

ASI

- 10 Blocks:
  - Block A: Identification particulars
  - Block B: Particulars of the factory
  - Block C: Fixed Assets
  - Block D: Working Capital and Loans
  - Block E: Employment and Labour Cost
  - *Block F : Other Expenses - (not yet added)*
  - *Block G: Other Output/Receipts -(not yet added)*
  - Block H: Indigenous input items consumed
  - Block I: Imported input items consumed - directly only
  - Block J: Products and by-products manufactured by the unit
- Note: Only have A-E for FY2010

Eg: Blocks I & J

DSL No ☐☐☐☐☐ PSL No ☐☐☐☐☐

Block I: Imported input items consumed - directly only (if needed, additional sheets may be used for recording input items with serial nos. starting from 8)

Sl. No.	Item description (Major five imported items)	Item code (ASICC)	Unit of quantity	Quantity consumed	Purchase value (in Rs.)	Rate per unit (in Rs.)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.						
2.						
3.						
4.						
5.						
6.	**Other imported items**	**99221**				
7.	**Total imports (consumed) (items 1 to 6)**	**99940**				

DSL No ☐☐☐☐☐ PSL No ☐☐☐☐☐

Block J: Products and by-products manufactured by the unit (if needed, additional sheets may be used for recording output items with serial nos. starting from 14)

Sl. No.	Products/By-products description (First ten major items as per value - no brand name)	Item code (ASICC)	Unit of quantity	Quantity manu-factured	Quantity sold	Gross sale value (Rs) (including subsidy received)	Distributive expenses (Rs.)				Per unit net sale value (Rs. 0.00) (col. 7-col.11) ÷ col. 6	Ex-factory value of quantity manufactured including subsidy received (Rs.)
							Excise duty	Sales tax/ VAT	Others	Total		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1.												
2.												
3.												
4.												
5.												
6.												
7.												
8.												
9.												
10.												
11.	Other products/ by-products*	99211										
12.	**Total (items 1 to 11)**	**99950**										
13.	Share (%) of products/by-products directly exported											

* Full description of items not in ASICC:

Accessing the ASI

- ▶ Once on the cluster...
- ▶ cd */projectnb/econdept/asi –> 3 sub folders:*
  - ▶ Unit Level Data from MOSPI [read only]
  - ▶ Master Construction and Cleaning Do-Files [read only]
  - ▶ User-Added Do-Files
- ▶ Within "Unit Level Data from MOSPI" –>
  - ▶ raw text data in folders by year
  - ▶ folder: "Panel_data_supporting_Documents"
  - ▶ folder "Stata Datasets"
    - ▶ "ASI_1999_2010_clean.dta"

Making changes to the do files

- The folder "Master Construction and Cleaning Do-Files" [read only] contains:
  - basic do file for constructing the data from the raw text files
  - basic do file that does minor cleaning
- The folder "User-Added Do-Files" contains nothing
  - if you create do files that make the data construction or cleaning process better, please save those do files here

Economic Census of India, 2005

- Complete enumeration of all non-agricultural **enterprises** (plants): $\approx$ 42 million observations
- Administered by state statistical offices
    - Ad-hoc enumerators did the data collection
    - The background of ad-hoc enumerators varied by state
        - School teachers in Bihar
        - Unemployed graduates in Maharashtra
- Very few variables
    - Number total workers
    - Number of non-hired workers
    - 4-digit NIC code
    - Geographical information
    - Power usage
    - Some owner information

Economic Census of India, 2005

The data files

- ▶ Text files of unit-level data for each state, questionnaire and the layout of the text files are in
  - ▶ /projectnb/econdept/ec/ec05/Unit Level Data From MOSPI/EC05ENTP

Economic Census of India, 2005

What's available

- Assembled data in Stata format
  - /projectnb/econdept/ec/ec05/Unit Level Data From MOSPI/Stata Datasets/all_india.dta
- Cleaned data in Stata format
  - projectnb/econdept/ec/ec05/Unit Level Data From MOSPI/Stata Datasets/ec_05_all_india_cleaned.dta
- Code to construct data from the raw text files is coming soon

NSSO Unorganised Manufacturing

- We have two waves
  1. Round 56 (2000-2001)
  2. Round 62 (2005-2006)
- Years refer to years when the survey was conducted
- The recall period was 12 months prior to the date of the interview
- Covers all manufacturing units not in the ASI
- Professional enumerators
- Detailed information
  - Inputs, outputs
  - Financials
  - Industry classification

NSSO Unorganised Manufacturing Round 62 (2005-2006)

- Two sampling frames:
  - "List" frame of 8000 large units identified in a previous "census" (MSME Census 2002-2003)
  - "Area" frame for the remaining units
- Very complicated sampling procedure
  - /projectnb/econdept/nsso/0506/Unit Level Data From MOSPI/NSSO Round 62/Nss62_2.2/Supporting Documents/Estimation Procedure 62.doc
- Multipliers are provided to make the data representative at the national and state levels
  - An adjustment should be made to produce district-level estimates
  - Depend on the 1998 Economic Census, Census MSME 2002-2003 and 2001 population census in a complicated way
    - Understanding this constitutes work in progress
- Non-response does not appear to be a huge issue

NSSO Unorganised Manufacturing Round 62 (2005-2006)

The data files

- The dataset is split up into **levels**, each of which contains one or more **blocks** from the questionnaire
- /projectnb/econdept/nsso/0506/Unit Level Data From MOSPI/NSSO Round 62/Nss62_2.2/Supporting Documents/Layout_62_2.2.XLS describes the layout:
  - Level 1 contains Blocks 1 and 10
  - Level 2 contains Block 2
  - ...

NSSO Unorganised Manufacturing Round 62 (2005-2006)

- /projectnb/econdept/nsso/0506/Unit Level Data From MOSPI\all_levels.dta
    - Levels 2, 5, 7, 8
    - Partially cleaned
- "/projectnb/econdept/nsso/0506/Master Construction and Cleaning Code" contains do files for creating Stata files for all levels
- Many variables are currently read in as strings to avoid problems with coding errors
    - These should be modified

Round 56 (2000-2001)

- Only one sampling frame, based on the 1998 Economic Census
- Non-response appears to be a bigger issue

- The dataset is split up into **workfiles**
- Layout:
  - /projectnb/econdept/nsso/0001/Unit Level Data From MOSPI/NSSO Round 56/Nss56_2.2/Supporting Documents/Layout_56_2.2.doc
- Questionnaire:
  - /projectnb/econdept/nsso/0001/Unit Level Data From MOSPI/NSSO Round 56/Nss56_2.2/Supporting Documents/Schedule_56_2.2.doc
- Sampling procedure:
  - /projectnb/econdept/nsso/0001/Unit Level Data From MOSPI/NSSO Round 56/Nss56_2.2/Supporting Documents/Instrn. to Field Staff/Appendix-3.doc

What's available

- /projectnb/econdept/nsso/0506/Unit Level Data From MOSPI/all_workfiles.dta
  - Workfile 2
  - Partially cleaned
- "/projectnb/econdept/nsso/0001/Master Construction and Cleaning Code" contains do files for creating Stata files for all workfiles
- Again many variables are currently read in as strings to avoid problems with coding errors