<div align="center">

Web Appendix for

Multi-product Firms and Product Turnover in the Developing World: Evidence from India

by Penny Goldberg, Amit Khandelwal, Nina Pavcnik, and Petia Topalova

</div>

A.1 Data Appendix

We compile a firm-level panel data set that spans the period from 1989 to 2003 based on the Prowess database, collected by the Centre for Monitoring the Indian Economy (CMIE).[1] The Prowess database contains information primarily from the income statements and balance sheets of about 9,500 publicly listed companies, almost 5,000 of which are in the manufacturing sector. The companies in the database together comprise 60 to 70 percent of the economic activity in the organized industrial sector and account for 75 percent of corporate taxes and 95 percent of excise duty collected by the Government of India (CMIE).

The Prowess database is the only Indian database, to our knowledge, that records detailed annual information on firms' product-mix.[2] Indian firms are required by the 1956 Companies Act to disclose product-level information on capacities, production and sales in their annual reports. The Prowess database compiles these detailed quantitative data and therefore enables us to track a firm's adding and dropping of products over time. Furthermore, for each product manufactured by the firm, the dataset provides the value of sales, quantity and units, allowing us to construct a time series of unit values at the firm-product level. Unlike the Annual Survey of Industries (ASI), the Prowess data is a panel of firms, rather than a repeated cross section. The Prowess is therefore particularly well suited

---

[1] The Prowess database has now been used in several studies including Bertrand et al. (2002), Khanna and Palepu (1999), Fisman and Khanna (2004), Topalova (2007), Dinc and Gupta (2007), and Chari and Gupta (2008).

[2] Product-level information is available for the 1997/98, 2000/01 and 2001/02 rounds of the Annual Survey of Industries (ASI), India's manufacturing survey of plants, but there is no information in years close to the economic reforms implemented in the early 1990s. Furthermore, the version of the ASI available to public does not include firm-level identifiers that are needed to create a panel.

for understanding how firms adjust their product lines over time and how their responses may be related to policy changes.[3]

As described in section A.2, CMIE uses an internal product classification that is based on the Harmonized System and National Industry Classification (NIC) schedules. There are a total of 1,886 *products* linked to 108 four-digit NIC *industries* across the 22 manufacturing *sectors* (two-digit NIC codes). As a comparison, the U.S. manufacturing data used by BRS (2006a), contain approximately 1,500 products, defined as five-digit Standard Industrial Classification (SIC) codes, across 455 four-digit SIC industries. Thus, our definition of a product is slightly more detailed than BRS (2006a).

Examples of products within the Basic Metals sector (NIC 27) of this hierarchical mapping are listed in Table A1. The table reports two industries within the sector: Manufacture of Basic Iron & Steel, which contains over 100 products, and Casting of Iron and Steel, which contains 7 products. As with all classifications, the degree of detail varies across industries and sectors. As documented in Table A2, the number of products ranges from a low of 6 products in the Tobacco industry (NIC 16) to 506 products in the Chemicals industry (NIC 24).

The product classification provides a concordance to the more familiar NIC industry codes used to classify economic activity in India. Each of the 1,886 product codes can therefore be mapped to a five-, four-, three-, two-, or one-digit NIC code. The concordance allows us to assess the relative degree of product disaggregation. Approximately 88 percent of the products map to the most detailed five-digit NIC and 10 percent of the products concord to four-digit NICs. Products mapping to four- or five-digit NIC codes account for 99 percent of total output. With the exception of Printing and Publishing (NIC 22), products within all sectors overwhelmingly map to four- or five-digit NIC codes.[4] This gives us confidence that the variation in product detail is a fundamental feature of sectors rather than a problem with data. Moreover, given that our industrial policy measures are specified at the

---

[3] The CMIE database is not well suited for understanding firm entry and exit because firms are under no legal obligation to report to the data collecting agency. However, since Prowess contains only the largest Indian firms, entry and exit is not necessarily an important margin for understanding these firms.
[4] These figures are available upon request.

four-digit NIC, the majority of our product information is specified *at least* at this level of aggregation.

Several features of the product data give us additional confidence in its quality despite the self-reported and non-standardized nature of the dataset. First, as mentioned above, firms are required to report not just the names of the products, but also product-level details about installed capacity, production, sales quantity and value. Table A3 reports that product-level data are available for 85 percent of the firms; this accounts for more than 90 percent of output and exports of the manufacturing firms in Prowess. More importantly, the product-level information and overall output are in separate modules of the Prowess database which enables us to cross check the consistency of the data. The final two rows of Table A3 report that the total product-level sales account for 92 percent of the (independently) reported overall and 99 percent of the reported manufacturing output of the firm.[5] This implies that product-level sales account for virtually all of the firm's total output. Since our study predominantly analyzes firms' product mix, our final database includes the 4,216 manufacturing firms that report product-level information. The data span the period from 1989 to 2003.

We complement the data on firm product mix with various measures on trade policy at the industry level. Data on disaggregated tariffs for 1987-2001 have been compiled in Topalova (2007). Tariffs are reported at the six digit level of the Indian Trade Classification Harmonized System (HS) Code, which are then aggregated to the 116 NIC codes, using the concordance by Debroy and Santhanam (1993) to calculate average industry-level tariffs. To capture changes in the domestic industrial policy over this time period, we use Aghion et al. (2008) measures of industrial delicensing.

---

[5] There is some variation in the availability of product level information across sectors. However, with the exception of two of the smaller sectors (Publishing/Printing (NIC 22) and Office, accounting and computing machinery (NIC 30)), product details reporting is very high across sectors. Table A1 shows that in 14 of the 22 sectors, total product-level sales as a share of firm output exceed 85 percent.

A.2 Variable Construction

Product classification

The reporting of products by Indian firms is not governed by any particular product classification. Although CMIE has developed a classification of 5,800 codes based on the NIC and the HS schedule, the agency has not explicitly linked the product names reported by the firms to this classification. The names of products reported by the firm could differ in aggregation, or even in spelling (e.g., "Steel Rod" versus "Steel Rods"). We therefore standardized the approximately 8,500 product names to the 5,800 possible CMIE product codes. This mapping process was extremely time consuming and involved subjective calls. To minimize the scope for error and subjectivity in assigning the codes, the mapping was performed independently by two research assistants. The research assistants assigned product codes with identical NIC codes in 80% of the cases which represents 91% of output. A third research assistant resolved the differences between the mappings done by the first two research assistants by again manually checking the classifications.

Our final sample includes 1,886 product codes out of the universe of 5,800 product codes. The seemingly low coverage of products in the Prowess is not a source of concern. First, the distribution of unused codes is remarkably similar across sectors. Secondly, 25% of the unused codes are products in the agriculture and services sectors, which, of course, are not produced by manufacturing firms. The remaining unmatched codes appear to be a result of excess detail in the product codes. For many of these sectors, the number of potential CMIE products exceeds, by a large margin, the number of five-digit SIC products in BRS (2006a). The correlation between an industry's share in total unused codes and the number of possible codes in an industry is .99. Moreover, the chemicals sector alone account for 40% of the unused manufacturing product codes. Thus, if anything, the low coverage stems from the overly detailed CMIE product classification in certain sectors.

In all the analysis above, we refer to a *product* as the CMIE product code and not the reported product name.

Output tariffs. Data on disaggregated tariffs for 1987-2001 have been compiled in Topalova (2007). Tariffs are reported at the six digit level of the Indian Trade Classification Harmonized System (HS) Code, which are then aggregated to the 116 NIC codes, using the concordance by Debroy and Santhanam (1993), to calculate average industry-level tariffs.

Delicensed. Delicensed is an indicator equal to one if the industry is not subject to licensing requirements in that particular year and zero otherwise. The information is obtained from Aghion et al. (2008).

References

Aghion, P., R. Burgess, S. Redding and F. Zilibotti (2008). "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India" *American Economic Review,* 98(4): 1397–1412.

Bertrand, M, P. Mehta and S. Mullainathan (2002). "Ferreting Out Tunneling: An Application to Indian Business groups," *Quarterly Journal of Economics*, pp. 121-148.

Chari, A. and N. Gupta (2008). "Incumbents and Protectionism: The Political Economy of Foreign Entry Liberalization," *Journal of Financial Economics* 88, 633-656.

Debroy B. and A. T. Santhanam (1993). "Matching Trade Codes with Industrial Codes." *Foreign Trade Bulletin*, 24(1).

Dinc, I. S. and N. Gupta (2007). "The Decision to Privatize: Finance, Politics and Patronage," mimeo, MIT.

Fisman, R. and T. Khanna (2004). "Facilitating Development: The Role of Business Groups," *World Development*, 32(4).

Khanna, T. and K. Palepu (1999). "Policy Shocks, Market Intermediaries and Corporate Strategy: the Evolution of Business Groups in Chile and India." *Journal of Economics & Management Strategy*, 8(2), pp. 271-310.

Topalova, P. (2007). "Trade Liberalization and Firm Productivity: The Case of India." *IMF Working Paper, WP/04/28.*

Table A1: Examples of Industries, Sectors and Products

NIC		Description
27		Basic Metal Industries (Sector)
2710		Manufacture of Basic Iron & Steel (Industry)
	130101010000	Pig iron
	130101020000	Sponge iron
	130101030000	Ferro alloys
	130106040800	Welded steel tubular poles
	130106040900	Steel tubular structural poles
	130106050000	Tube & pipe fittings
	130106100000	Wires & ropes of iron & steel
	130106100300	Stranded wire
2731		Casting of iron and steel (Industry)
	130106030000	Castings & forgings
	130106030100	Castings
	130106030101	Steel castings
	130106030102	Cast iron castings
	130106030103	Maleable iron castings
	130106030104	S.G. iron castings
	130106030199	Castings, nec

(The word "Products" appears as a vertical label alongside the product rows for each industry.)

Notes: For NIC 2710, there are a total of 111 products, but only a subset are listed in the table. For NIC 2731, all products are listed in the table. Source: Prowess database and authors' matching of product names to product codes (see text).

Table A2: Products and Multi-product Firms by Sector

NIC Sector	Products	Industries	Products per Industry	Products per Firm	NIC4 per Firm	MP Firm Share	Multiple-Industry Firm Share	Multiple-Sector Firm Share	MP Share of Output	Multiple-Industry Share of Ouput	Multiple-Sector Share of Ouput	Products per MP Firm	Industries per MP Firm	Sectors per MP Firm
15 Food products and beverages	135	17	7.9	2.12	1.55	0.53	0.35	0.21	0.67	0.54	0.35	3.11	2.03	1.51
16 Tobacco products	6	1	6.0	2.04	1.73	0.58	0.49	0.45	0.85	0.84	0.84	2.78	2.25	2.17
17 Textiles	83	7	11.9	1.76	1.42	0.45	0.32	0.28	0.63	0.46	0.42	2.67	1.92	1.76
18 Wearing apparel	14	1	14.0	1.24	1.18	0.18	0.17	0.17	0.14	0.14	0.14	2.38	2.03	2.03
19 Tanning and dressing of leather	21	3	7.0	2.01	1.51	0.50	0.34	0.13	0.82	0.73	0.10	3.03	2.02	1.32
20 Wood and products of wood	13	2	6.5	2.20	1.77	0.61	0.45	0.37	0.73	0.39	0.38	2.94	2.21	1.82
21 Paper and paper products	32	3	10.7	1.40	1.21	0.24	0.19	0.15	0.50	0.47	0.44	2.68	1.88	1.72
22 Publishing/printing	13	3	4.3	1.61	1.29	0.36	0.64	0.29	0.22	0.84	0.19	2.71	1.48	1.88
23 Coke, refined petroleum products	24	2	12.0	2.77	1.77	0.60	0.44	0.44	0.98	0.81	0.81	3.97	2.29	2.12
24 Chemicals	506	9	56.2	2.26	1.43	0.53	0.32	0.17	0.79	0.61	0.34	3.36	1.79	1.38
25 Rubber and Plastic	85	3	28.3	1.68	1.34	0.40	0.26	0.21	0.67	0.33	0.31	2.72	1.82	1.71
26 Non-metallic mineral products	63	8	7.9	1.62	1.41	0.37	0.26	0.20	0.59	0.48	0.28	2.69	2.12	1.76
27 Basic Metal	103	3	34.3	1.85	1.34	0.46	0.26	0.20	0.85	0.44	0.40	2.84	1.71	1.54
28 Fabricated metal products	50	6	8.3	1.70	1.52	0.38	0.38	0.30	0.61	0.61	0.53	2.84	2.29	2.13
29 Machinery/equipment n.e.c.	195	14	13.9	2.20	1.81	0.55	0.48	0.32	0.78	0.74	0.62	3.20	2.46	1.87
30 Office, accounting and computing machines	19	1	19.0	1.56	1.36	0.29	0.20	0.20	0.37	0.10	0.10	2.96	2.26	2.28
31 Electrical machinery and apparatus	105	6	17.5	2.20	1.76	0.49	0.39	0.37	0.71	0.63	0.62	3.44	2.54	2.11
32 Radio, TV and communication	91	3	30.3	1.93	1.52	0.40	0.31	0.28	0.68	0.58	0.56	3.29	2.29	2.07
33 Medical, precision and optical instruments	71	5	14.2	1.63	1.38	0.30	0.22	0.18	0.48	0.45	0.42	3.10	2.25	1.95
34 Motor vehicles, trailers	96	2	48.0	2.03	1.52	0.51	0.39	0.35	0.63	0.59	0.57	3.02	1.97	1.97
35 Other transport	22	4	5.5	2.12	1.86	0.60	0.57	0.50	0.63	0.52	0.47	2.88	2.36	2.32
36 Furniture	22	5	4.4	1.53	1.29	0.24	0.09	0.09	0.34	0.07	0.07	3.21	2.20	2.01
Total	1,886*	108	17.5	1.97	1.48	0.47	0.33	0.24	0.80	0.62	0.54	3.06	2.00	1.68

Notes: Table reports summary statistics, by sector, for product-reporting firms. Column 1 our reports the total product codes by industry. Note that this column sums to 1,769 products, but there are 117 products in non-manufacturing industries (but produced by manufacturing firms) which we include in the analysis; a total of 1,886 product codes comprise the multiple-product Prowess sample. Column 2 reports the number of industries within each sector. Column 3 is the first column divided by the second column. Columns 4 and 5 report average products per firm and industries per firm, respectively. Column 6-8 report the fraction of firms that produce multiple products, industries and sectors, respectively. Columns 9-11 report the multiple-product (or industry or sector) firms' share of total output. Columns 12-14 report mean products, industries and sectors per MP firm. All figures are unweighted averages over 1989-2003. Industry refers to a 4-digit NIC code. Sector refers to a 2-digit NIC code. Source: Authors' calculations from the Prowess database.

Table A3: Summary Statistics of Product-Reporting Firms

Description	Statistic
Years	1989-2003
Firms	4,971
Product-Reporting Firms	4,216
Share of output of reporting firms	0.92
Share of exports of reporting firms	0.91
Median product share of total output	0.92
Median product share of total manufacturing output	0.99

Notes: Table summarizes aggregate statistics of product reporting manufacturing firms. Row 4 and 5 report the share of total manufacturing output and exports in Prowess by product-reporting firms, respectively. Rows 6 and 7 report the median share of product-reporting firms' total output and total manufacturing output, respectively, accounted by the listed products. Source: Authors' calculations from the Prowess database.