

Dosi, Giovanni; Grazi, Marco; Mathew, Nanditha

Working Paper

The cost-quantity relations and the diverse patterns of "learning by doing": Evidence from India

LEM Working Paper Series, No. 2016/26

Provided in Cooperation with:

Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies

Suggested Citation: Dosi, Giovanni; Grazi, Marco; Mathew, Nanditha (2016) : The cost-quantity relations and the diverse patterns of "learning by doing": Evidence from India, LEM Working Paper Series, No. 2016/26, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM), Pisa

This Version is available at:

<https://hdl.handle.net/10419/174535>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

The cost-quantity relations and the diverse patterns of “learning by doing”: Evidence from India

Giovanni Dosi [°]
Marco Grazzi ^{*}
Nanditha Mathew [¶]

[°] Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy

^{*} Department of Economics, University of Bologna, Italy

[¶] IBIMET-CNR, Florence, Italy, and University of Pisa, Italy

2016/26

June 2016

ISSN(ONLINE) 2284-0400

The cost-quantity relations and the diverse patterns of “learning by doing”: Evidence from India

Giovanni Dosi^{*1}, Marco Grazzi^{†2}, and Nanditha Mathew^{‡3,4}

¹Institute of Economics, Scuola Superiore Sant’ Anna, Pisa

²Department of Economics, University of Bologna

³IBIMET-National Research Council, Florence

⁴University of Pisa

June 16, 2016

Abstract

“Learning-by-doing” is usually identified as a process whereby performance increases with experience in production. The paper investigates different patterns of “learning by doing”, studying learning curves at product level. Cost-quantity relationships differ a lot across products belonging to sectors with different “technological intensities”. Moreover, such differential patterns are affected by firm spending on research and capital investments. Finally, our evidence suggests that “learning”, or performance improvement over time is not a by-product of the mere repetition of the same production activities, as sometimes reported in previous studies, but rather it seems to be shaped by deliberate firm learning efforts and by the interactions among firms themselves.

Keywords: Learning-by-doing, learning curves, power law, product innovation, process innovation

JEL codes: D22, D24, L6, O3

* *Corresponding author:* Piazza Martiri della Liberta 33, 56127 Pisa, Italy. email:gdosi@sssup.it

† Piazza Scaravilli 2, 40126 Bologna, Italy. email:marco.grazzi@unibo.it

‡ Via Giovanni Caproni, 8, 50145 Firenze, Italy. email:nandithamathew@gmail.com

Acknowledgements: We thank Martin Bell, Alex Coad, Gerald Silverberg, Tommaso Ciarli, Alessandro Nuvolari, Luciano Pietronero, Andrea Zaccaria, Emanuele Pugliese and several participants at the 11th ENEF meeting in Manchester (2014), the 9th EMAEE conference in Maastricht (2015), 13th Globelics conference in Havana (2015), the 2nd annual conference of the International Association of Applied Econometrics in Thessaloniki (2015), SPRU and LEM seminars for their insightful comments. We gratefully acknowledge the research support by the IBIMET-CNR (grant CrisisLab-ProCoPe). The usual disclaimers apply.

1 Introduction

Theoretical and empirical studies in economics consider “learning-by-doing” as a process in which an increase in experience in a particular type of production (‘doing’) yields an improvement in efficiency (‘learning’). Typically the postulated relation is a power law, linking some performance indicator (eg. unit costs, unit prices, productivity) with an experience indicator (eg. cumulated production). The evidence, which we shall review below is quite robust. However the simplified version of “learning by doing”, henceforth LBD, presents significant drawbacks. As discussed by Scott-Kemmis and Bell (2010), *first*, it is often considered that learning is a costless activity and an automatic by-product of continued production activities. *Second*, it is generally assumed that all organizations have the same capacity to learn and there are no differences in absorptive capacities that might lead to differences in the intensity of learning across different organizations. *Third*, the literature often assumes that, with time, the process linking experience to efficiency operates automatically without any purposeful activity by the firms (Scott-Kemmis and Bell, 2010).

In this work we investigate the existence, shape and slope of learning curves in a developing country, namely India, at the product level, conditioning on firms’ and sectoral characteristics. In particular, we shall analyze first how the slope of the learning curves are affected by R&D and fixed investment activities, and second, by the timing of entry of the firm in any one production activity, and hence, indirectly by the positioning of the firm along the life cycle of a product, the possible knowledge spillovers it gains from older incumbents. Needless to say, the understanding of the determinants of the very existence and slope of learning curves is not only important in its own right as a part of the microeconomics of innovation, but bear far reaching implications for the very analysis of the determinants of growth - in general and especially with respect to emerging economies. For example, would one find widespread and relatively uniform learning curves, that would give support to the view whereby knowledge is simply the “unintentional side effect of the production of a conventional good” (Romer, 1990) and, dynamically, the notion that the “the larger the rate of production, the greater the learning experience” (Rosen, 1972). Indeed a wide ensemble of growth models are build in such notion: among others, Rosen (1972); Lucas (1998); Romer (1986); Stokey (1988); Young (1993) and De Liso et al. (2001).

Conversely, were one to find a great inter product/inter-technological/inter-firm diversity in learning rates one would be forced to bring more “Schumpeterian” and “evolutionary” elements into the explanation, related to both the specificities of the different technologies and the characteristics and strategies of different firms and it would also carry different policy implications. More generally, such evidence should urge to a finer consideration of the complementarity between production-related learning on the one hand, along with other drivers of knowledge accumulation, on the other. So, as Romer (1990) in a self-critical mood puts it, considering the importance of determinants of knowledge accumulation other than sheer “learning-by-producing”, “if the fundamental policy problem is that we have too many lawyers and MBAs and not enough engineers, a subsidy to fixed capital accumulation is a weak, and possibly counterproductive, policy response” (p.S94).

The paper proceeds as follows. In the next section, we provide a critical review of previous studies on LBD. Section 3 describes the data and variables used in the paper. In section 4, we look at the cost-quantity relationships across products as revealed by learning curves and learning rates. Section 5 presents the observed heterogeneity in the cost-quantity relationships among different products, while section 6 attempts to relate it to different firm-specific characteristics. Section 7 deals with the effect of market experience in mode of entry of firms. Section 8 concludes.

2 Studies on LBD and its applications

The concept of “learning-by-doing” was first put forward in the 1930s and 1940s, thanks to studies of aircraft and ship production. The notion of learning formally entered economics with the empirical observation of the learning curve, a notion originally born in the engineering discipline when T. P. Wright, the director of engineering of the Curtiss-Wright Corporation, was apparently inspired by the mass production of airplanes as well as its declining unit cost when he began to plot out “the effect of quantity production on cost” (Wright, 1936). The resulting graphs reported a decreasing convex curve with cost or its components on the vertical axis and the sequence number of the airplanes on the horizontal axis, that is a log-linear relation between labour required per unit and the cumulative volume of production.

Wright suggested a continuing reduction of unit costs by 20% with each doubling of output volume. In most other studies that followed, the basic power law relation between costs and quantity appeared to fit the data quite well in a wide range of industries including, but not limited to, shipbuilding, machine tools, specialty chemicals, and semiconductors (Dutton and Thomas, 1984; McDonald and Schratzenholzer, 2001; Argote and Epple, 1990).

The mathematical relation has been broadly corroborated by studies carried out by the Stanford Research Institute and the Rand Corporation, including Arrow and Arrow (1950), Arrow (1951) and later Alchian (1963). Table 1 provides a summary of a few of the works on LBD. However, most of these studies provide limited information about the causes of improved performance. One interpretation focuses upon, we would say nowadays, some form of collective improvements in production activities, even holding the capital equipment unchanged. Lundberg (1961) called the “Horndal effect” the observation that at the Horndal steel works plant in Sweden with no new investment for a period of 15 years, still productivity (output per man hour) rose on the average close to 2 percent per annum. Therefore, he suggests, the increasing performance should be imputed to learning from experience. As known, that cumulative production-productivity relation has been a source of inspiration also for the seminal “learning by doing” theoretical contribution by Arrow (1962).

However, other studies, as in the discussion in Thompson (2001), report that systematic observed cost-quantity relationships are often linked with other factors beyond learning by experience. He reviews the evidence on the sources of productivity growth in the Liberty ship program, discussed earlier by Rapping (1965), providing evidence that the capital stock was not unchanged during the production period when performance improvements were observed.

In fact, a *first* major issue still far from settled in the literature is the interaction between improvements in the production methods directly associated with production activities, on the one hand and other forms of “learning” such as explicit search, often ‘offline’ of new products and processes. As we shall see in the following, the latter might even imply apparent ‘de-learning’ in production efficiency, where in fact they yield products characterized by higher quality and performances. *Second*, but relatedly, crucial issues concerns the robustness of the linearity of the log-log curve itself (as pointed out long ago by Carr (1946)) and the inter-product differences in the slope of such curves (Middleton, 1945). In the following we shall address both issues.

Table 1: Major reviews, empirical and theoretical studies on learning curves and LBD

Wright (1936)	- Synthesizes earlier USAF and supplier company improvement curve studies.
Rigdon (1944)	- Labour productivity trends in the WW II US airframe industry.
Searle (1945)	- Labour and time requirements trends in WW II US shipbuilding industry.
Middleton (1945)	- Report of plant level observations of WW II airframe productivity performance.
Carr (1946)	- Critical review of airframe learning applications.
Mensforth (1947)	- Review of productivity trends and other developments, mainly in the airframe industry.
Stanley (1949)	- Empirical study of time to achieve peak output rates in the USA and UK WW II airframe industries.
Arrow and Arrow (1950)	- Productivity trends in WW II US airframe industry.
Arrow (1951)	- Labour productivity trends in WW II US airframe industry.
Asher (1956)	- Improvement trends in the WW II and post-war US airframe industry.
Zieke (1962)	- Review of progress curves in WW II and post-war US aircraft industry.
Alchian (1963)	- Labour productivity trends in WW II US airframe industry.
Brewer (1965)	- Labour productivity and cost trends in WW II and post-war US airframe industry.
Rapping (1965)	- Improvement trends in WW II US shipbuilding industry.
Young (1966)	- Critical review of airframe learning curve applications.
Colasuonno (1967)	- Review of progress curves largely in WW II and post-war US aircraft industry.
Brockman and Dickens (1967)	- Labour productivity trends in post-war US airframe production.
Hartley (1969)	- Discussion of the use of the learning curve in UK aircraft production.
Orsini (1970)	- Review of progress curves largely in the WW II and post-war US aircraft industry.
Dosi (1984)	- Models cost and pricing procedures under conditions of technical change.
Lieberman (1984)	- Documents variations in the slope of learning curve linked to differences in R&D & capital intensity.
Gruber (1992)	- Learning curve in semiconductor chips; heterogeneity of learning across products (chip types).
Irwin and Klenow (1994)	- Learning-by-doing spillovers within the semiconductor industry.
Jovanovic and Nyarko (1994)	- Develops one-agent Bayesian model of LBD and technology choice.
Argote (1996)	- Reviews organizational learning & forgetting and evidence about whether learning transfers across organizations.
Hatch and Mowery (1998)	- Analyzes the relationship between process innovation and learning by doing in the semiconductor industry.
McDonald and Schrattenholzer (2001)	- Estimates learning rates for energy conversion technologies.
Jovanovic and Rousseau (2002)	- Empirical learning curves for three general-purpose technologies: Computers, electricity, and the internal combustion engines.
Lapr�e and Tsirikrisis (2006)	- Explore whether customer dissatisfaction follows a learning-curve pattern looking at trends in customer complaints against 10 largest airlines.
Schoots et al. (2008)	- Learning curves using cost data for hydrogen production process; No cost reduction is found.
Grubler (2010)	- Cost trends in specific reactors in time; finds that reactor construction costs increase in time.

3 Data and Variables

The paper employs firm-level data from the Prowess database, provided by the CMIE (Centre For Monitoring Indian Economy Pvt. Ltd.). Prowess reports information from the financial accounts of Indian companies. Annual Reports of companies are the most important source of this database. The data are collected by the CMIE from company balance sheets and cover both publicly listed and unlisted firms from a wide cross-section of manufacturing, services, utilities, and financial industries. About one-third of the firms in Prowess are publicly listed firms.¹ The data used for the current study includes only firms in manufacturing sector and span from 1988 to 2012.

Total firm sales are broken down by the revenues generated by each product sold, in accordance with the requirements of the 1956 Companies Act according to which the firms in their annual reports have to disclose product-level information on production and sales. The product classification structure is detailed in Appendix A.1. The product-level information is available for 90 percent of the manufacturing firms, that collectively account for more than 90 percent of Prowess' manufacturing output and exports. Firms are required to report not just the names of the products, but also product-level details about production, sales quantity, sales value and unit prices. The coverage of product-level information - especially for sales - is extremely good: summing up sales at the product level yield more than 90 percent of total sales reported through firm balance sheet. Prowess is therefore particularly well suited to investigate how firms adjust their product lines over time. Table 2 reports some summary statistics covering different years to provide evidence of the representativeness of Prowess over time.

Table 2: Summary statistics for product-reporting firms

	1991	1996	2001	2006	2011
2 Firms	1875	3712	5281	6264	3492
3 Number of Products	1268	1758	1952	2114	1841
4 Product-Reporting Firms	1769	3560	4983	5640	3289
5 Share of single product firms	47.65	54.66	57.27	52.32	42.99
6 Share of multi-product firms	52.35	45.34	42.73	47.68	57.01
7 Share of sales of product reporting firms	0.89	0.91	0.90	0.92	0.93
8 Share of exports of product reporting firms	0.86	0.88	0.90	0.90	0.92

3.1 Variables

In the literature on learning curve, three variables can be employed to measure experience: 1) cumulative volume 2) calendar time and 3) maximum volume, however, the most used is cumulative production volume (Yelle, 1979; Argote et al., 2000). The idea behind using cumulative volume is that, the repetition in production allows organization to gain experience. Levin (2000) and Field and Sinha (2005) employ the calendar time elapsed since the start of operation and state that the time for reflection is more important for learning, so that the calendar time captures the notion of “learning by thinking”. Mishina (1999) proposed a third “experience” variable, i.e, the maximum output produced to date or maximum proven capacity to date. When a plant is scaling up production, the production system faces significant challenges and unprecedented difficulties. Factory personnel need to figure out how to solve such challenges

¹Appendix A provides additional information on the database.

during scale-up. In this respect, Mishina’s experience variable captures the notion of “learning by new experiences” or “learning by stretching” (Mishina, 1999; Lapré and Van Wassenhove, 2001).

In the present study, we measure experience as the cumulated (physical) quantity of a given product manufactured by the firm. The performance variable that we will mostly employ is the unit price of the product. It is to be noted that production costs per product for multi product firms are not available and probably unknown with precision to the firms themselves. In fact, several previous empirical studies have used price data to construct experience curves (Boston Consulting Group, 1970; Abell and Hammond, 1979; Ayres and Martinas, 1992; Neij, 1997; Gruber, 1992, 1998; Irwin and Klenow, 1994; Chung, 2001; Wene, 2007).² And, indeed there is a good matching between unit cost dynamics and unit prices whenever the latter are fixed according to mark-up pricing procedure (Dosi, 1984). Consider the following equation where price is a multiplicative markup over marginal (or better, average) cost (for a similar pricing structure, see, among others, Amiti et al. 2014).

$$P_{ijt} = C_{ijt}(1 + MU_{ijt}), \quad (1)$$

where the three terms are, respectively, the unit price (P_{ijt}), the unit production cost (C_{ijt}) and the markup (MU_{ijt}) of firm i , for product j at time t . In this work we are not interested in the estimation of the markup per se, however, to the extent that equation (1) offers an accurate approximation of the pricing behavior of firms, that provides support to our choice of price as a performance measure. We can test the validity of our conjecture for the case of single-product firms.³ In the setting of the equation (1), a large share of variation of the change in price explained by changes in cost, that is, the R^2 of the log transformed linear model, would provide evidence of the validity of the assumption entailed by equation (1), and in turn, of our choice for the performance measure. Table 11 in appendix B shows the results from a fixed effects regression of equation (1). The R squared value of 86.9% supports our hypothesis, hence in the rest of the paper we will be using data from all firms, including multiple products ones, and price as a proxy for cost of production. Finally, monetary variables are deflated with 3 digit industry output deflators.

4 Learning curves and learning rates: Product level analysis

Let us start by investigating the cost-quantity⁴ relationships at the product level by plotting “learning curves”, i.e, the relation between cumulative quantities and prices (or cost) of products. Usually, the learning curve is expressed in the form of a power law⁵:

$$P = a * Q^{-\beta} \quad (2)$$

²There are many other related fields, such as energy economics, where it is common to use price as a proxy for cost in the learning-by-doing literature, see for example, Berry (2009); Coulomb and Neuhoff (2006); Junginger et al. (2005); Kobos et al. (2006).

³For multiple-products firms, on the output side, one finds a sale price for each product sold, but on the input side, only the cost for the purchase of inputs. It is only for single product firms that it is possible to directly relate the cost of production to the price of the output sold.

⁴Throughout the paper we use the expression cost-quantity instead of price-quantity, since we use price as a proxy to measure cost. Comments and interpretations in the paper rest on the assumption that cost-price margins of products remain roughly constant over time. The hypothesis is tested in section 3.

⁵Among others, Dutton and Thomas (1984); McDonald and Schrattenholzer (2001) and Argote and Epple (1990) use this formulation.

where P is the price of the product, a the normalization constant (which can be understood as the initial costs), β is a the scaling factor and Q the quantity produced. Taking the logs, one of course gets a straight line. Here we focus on estimation of the learning parameters using power and also other 3 functional forms, generally suggested by previous studies, which include, linear, exponential and logarithmic functions.⁶ The equations are reported in Table 3.

Table 3: Equations representing Linear, Exponential, Logarithmic and Power functions

Function Type	Equation
Linear	$P = a - \beta Q$
Logarithmic	$P = a - \beta \log Q$
Exponential	$P = (a)e^{-\beta Q}$
Power	$P = (a)Q^{-\beta}$

We start by investigating the “aggregate” cost-quantity relationship, that is, we pool together observations from all firms producing a given product. We exploit the panel structure of the data, and we look at the price of a given product and its cumulative output. We proceed to perform a panel fixed effects regression with the four different functional forms and we check which functional form provides the best representation of the cost-cumulated quantity relationship. The equations in Table 3 are rewritten below (except for linear, other equations are log-transformed), including controls for size and time effects.

Linear Form:

$$p_{ijt} = a_{ij} - \beta_j q_{ijt} + \gamma_j s_{it} + d_t + \epsilon_{ijt} \quad (3)$$

Logarithmic form:

$$p_{ijt} = a_{ij} - \beta \log(q_{ijt}) + \gamma_j s_{it} + d_t + \epsilon_{ijt} \quad (4)$$

Exponential Form:

$$\log(p_{ijt}) = \log(a_{ij}) - \beta q_{ijt} + \gamma_j s_{it} + d_t + \epsilon_{ijt} \quad (5)$$

Power Form:

$$\log(p_{ijt}) = \log(a_{ij}) - \beta \log(q_{ijt}) + \gamma_j s_{it} + d_t + \epsilon_{ijt} \quad (6)$$

where p_{ijt} is the price of product j produced by firm i at time t , q_{ijt} is the cumulated quantity of product j produced by firm i at time t , s_{it} is gross fixed assets of the firm, as proxy for the size of the firm, d_t represents time dummies, a_{ij} the intercept and ϵ_{ijt} is the error term.

To investigate which of the functional forms described before fits the learning pattern best, as standard in the literature (see, among others, Heathcote et al. 2000), we compare the goodness-of-fit using *R squared* as a fit criterion. In line with previous literature, we find that the goodness-of-fit of the power and exponential functions is higher than the linear functions. Similar findings have been reported by Anderson and Schooler (1991) and Wixted and Ebbesen (1997). The average value of *R squared* is around 0.9 for power and exponential functions, while for linear and log functions, it is around 0.7. In what follows, we will be using the parameters of the power law estimation, but also using the exponential parameter our general results do not change.

⁶Koh and Magee (2006, 2008) claim that an exponential function of time predicts the performance of several different technologies. According to Goddard (1982) costs follow a power law in production rate rather than cumulative production. Multivariate forms involving combinations of production rate, cumulative production, or time have been examined by Sinclair et al. (2000) and Nordhaus (2014).

Table 4 reports the learning rates estimated using the power law function. *First*, note that the learning coefficient vary a lot among products. *Second*, we observe that for quite a few products, the learning coefficient is positive, and thus, hints at an upward sloping cost-quantity curve. Figure 1 plots

Table 4: Learning coefficients using Power function

Product name	Power Coeff.	Std error
Mineral Water	-0.551***	(0.050)
Detergent Soaps	-0.217***	(0.055)
Writing & Printing Paper	-0.235***	(0.043)
Passenger Cars	-0.062***	(0.016)
Portable Electronic Typewriters	-0.249***	(0.067)
Automotive Diesel Engines	-0.169***	(0.062)
Pollution Control Equipment	-0.071**	(0.032)
Cranes	-0.264***	(0.059)
Road Construction & Maintenance Machines	-0.251***	(0.057)
Washing Machines	-0.170***	(0.025)
Refrigerators	-0.150***	(0.020)
Air Coolers	-0.044**	(0.021)
Wiring Accessories	-0.205***	(0.043)
Electrical Porcelains And Insulators	0.714***	(0.139)
Tyre Pressure Gauges	0.257***	(0.056)
Can Making Machinery/Industrial machinery	0.619***	(0.123)
Coke Oven	0.403***	(0.084)
Sprocket Wheels	0.219**	(0.044)
Food Processing Machines	0.549***	(0.200)
Metal Cutting Incl. Grinding Machines	0.738**	(0.264)
Oil Purifiers	0.218***	(0.083)
Perfumery Compounds Aromatic Spices & Herbal Etc.	0.262***	(0.081)
Stainless Steel Forging,Flanges & Allied Pipe Fitt	0.011***	(0.004)
Radio Sets,Tape Rec.,Combination Sets & Rec.Player	0.165***	(0.016)
Hand watches& manufactured components-watches	0.799***	(0.196)

the canonical cost-quantity curves obtained by pooling together the observations of firms producing the same product, for some products displaying a downward sloping cost-quantity curve. Dots with different symbols represent different products. Each dot with the same symbol represents a pair of (log) cumulated quantity and (log of) price for a given firm in a given year. This evidence is well in line with several other studies that detect a cost-quantity relationship that is accounted for by a power-law and that applies to a wide variety of technologies (Dutton and Thomas, 1984; McDonald and Schrattenholzer, 2001; Argote and Epple, 1990). Figure 2 provides a graphical account of the “positive” learning rates, that is, a positive relationship between unit price (or cost) and experience.

The analyses so far have shown that, in most of the cases, the cost-quantity relationship displays a non-linear nature and that such relationship could differ across products. Some products might require larger time than others for cost reduction, even within the same firms. In the following section we look at the heterogeneity in learning patterns across products and firms, classifying them on the basis of the different technological intensities, as captured by the Pavitt taxonomy. (Pavitt, 1984).

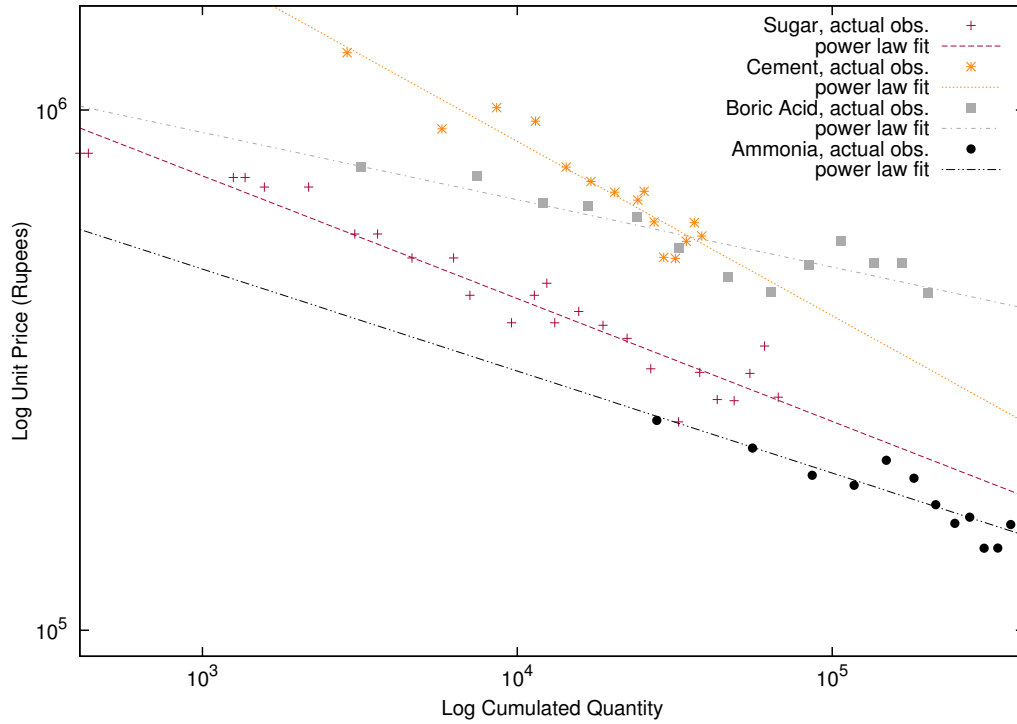


Figure 1: The relation between cost and quantity (on log scales) together with power law fit for selected products; the ‘canonical’ downward sloping learning curve

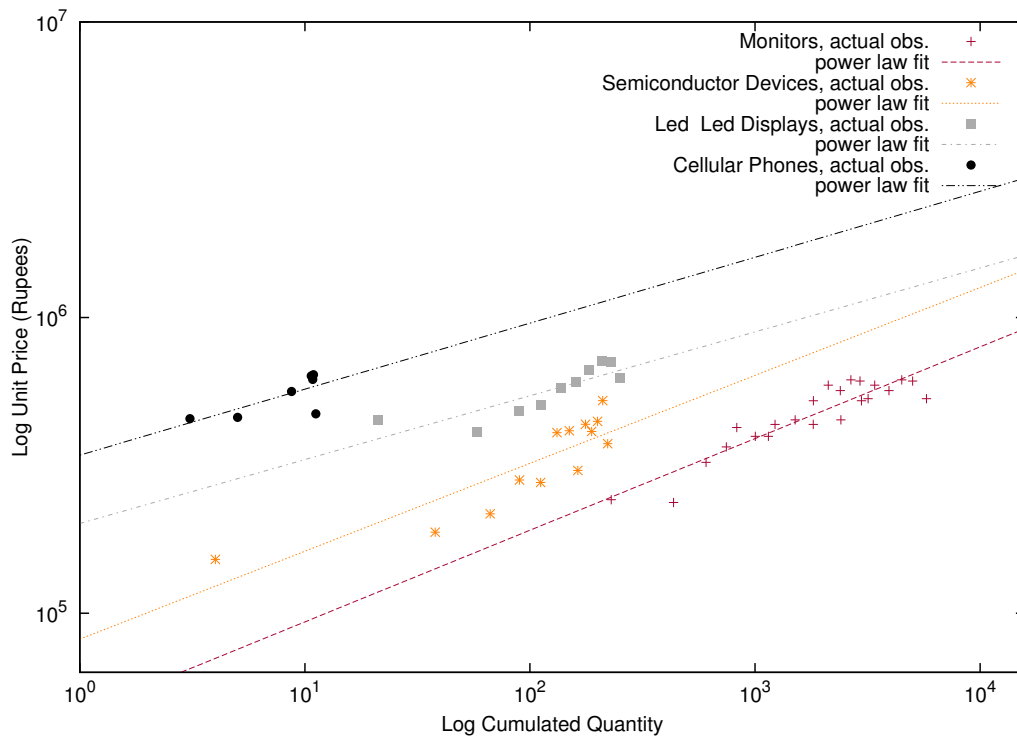


Figure 2: The relation between cost and quantity (on log scales) together with power law fit for selected products; the upward sloping learning curve

5 Heterogeneity in cost-quantity relationships

In this section, let us try to investigate the differences in cost-quantity relationships, or, to put it in other words, the heterogeneity in the learning rates, and we further try to associate such differential learning

rates to different sectors, based on which sectors the products belong to. Moreover we shall address the presence of both downward and upward sloping cost-quantity curves. At a first look, an increasing relation in cost-quantity curves for some products would indicate the absence of learning. However, our suggested interpretation is that increasing cost-quantity relationships could also be pointing to learning, although of a different kind than the one usually identified in the ‘learning-by-doing’ literature.

There could be various product-specific characteristics which might lead to such differential learning patterns. Here, we are interested to investigate the technological characteristics of the product. Our hypothesis is that the products which show a positive relationship between cost and experience undergo systematic quality upgrading over time, hence the observed positive cost-quantity relationship. Resorting to the Pavitt taxonomy, let us investigate if the products displaying a positive cost-quantity relation belong to more research intense sectors.

Pavitt’s Taxonomy distinguishes between sectors and technologies according to sources of technological knowledge, requirements of the users, and appropriability regimes (Pavitt, 1984). It identifies four categories:

(1) Supplier-Dominated sectors includes most traditional activities such as textiles, clothing and agriculture which rely on sources of innovation external to the firm.

(2) Scale-Intensive sectors, characterized by scale-biased technical advances and covering both basic materials and consumer durables, e.g. automobiles. Sources of innovation are both internal and external to the firm.

(3) Specialized Suppliers are smaller, more specialized firms, e.g. industrial machinery instruments, used in most other industrial sectors.

(4) Science-based sectors, rely on both in-house R&D and on university research; they include industries such as pharmaceuticals and electronics.

Figure 3 show the violin plots of the distributions of learning rates across different Pavitt categories. The violin plot is a combination of box plot and kernel density distributions. The median of the product level learning rate for each sector is marked by the central bar and the box indicates the interquartile range as in standard box plots. Notice that the distribution of learning rates in the science based category and specialized suppliers (S-B and S-S in the figure) is shifted upwards, implying that there are more cases of positively shaped cost-quantity curve.

Non-parametric analysis: Fligner Policello statistics

The difference between the learning parameters across Pavitt sectors is further tested using Fligner-Policello location test (Fligner and Policello, 1981). The null hypothesis for the test is $H_0 : \theta_X = \theta_Y$, where θ_X and θ_Y are the population medians of the two Pavitt sectors in each pair.⁷ The test assumes that the distribution in each class is symmetric around the class median, but it does not require that the two class distributions have the same form or that the class variances be equal.⁸ Table 5 reports the pairwise Fligner-Policello statistics of the distributions of learning rates across different Pavitt categories. A positive and significant F-P statistics suggests that the distribution of learning rates of the second sector in each pair (for instance, scale intensive sector in the first row) statistically dominates the other. In all the cases we observe that the “learning” rates in specialized suppliers and science based sectors statistically dominate the supplier dominated and scale intensive sectors: That is, there is a higher

⁷The first column of table 5 gives different pairs of Pavitt sectors.

⁸Refer Hollander et al. (1999) and Juneau (2007) for details.

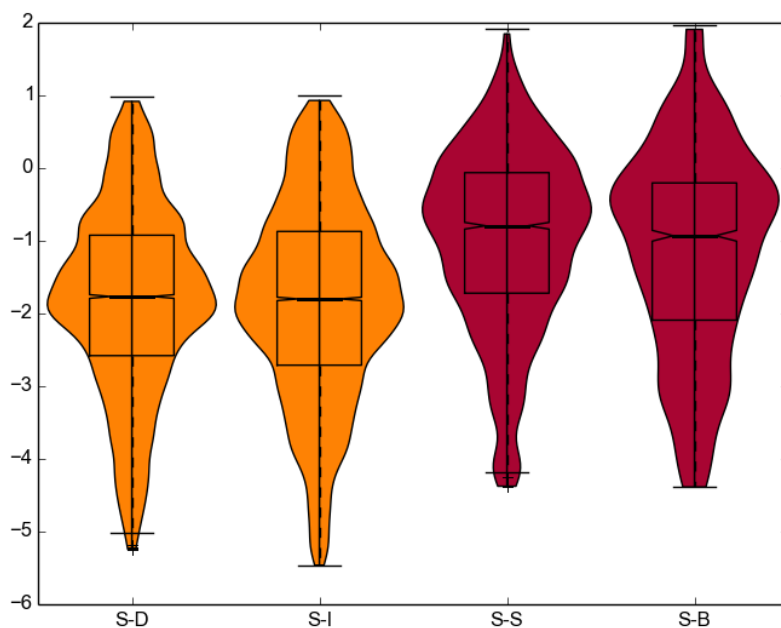


Figure 3: Distribution of learning rates across different Pavitt categories at firm level

probability to observe a positively shaped learning curve, i.e, an apparent “anti-learning”.

Table 5: Heterogeneity of learning rates across different Pavitt sectors: Two Sample Fligner-Policello Robust Rank Order Test

Sector	Obs.	Average placement	Index of variability	F-P statistic	Two-tailed p-value
<i>Supplier-dominated</i>	20885	2.0e+04	2.4e+12	10.163	0.000
<i>Scale-Intensive</i>	37123	9.9e+03	1.4e+12		
<i>Supplier-dominated</i>	20885	5.5e+03	1.2e+11	28.413	0.000
<i>Specialized suppliers</i>	9101	8.3e+03	3.7e+11		
<i>Supplier-dominated</i>	20885	4.2e+03	6.2e+10	37.234	0.000
<i>Science-based</i>	6497	7.3e+03	2.4e+11		
<i>Scale Intensive</i>	37123	5.3e+03	2.0e+11	24.168	0.000
<i>Specialized Suppliers</i>	9101	1.5e+04	1.2e+12		
<i>Scale Intensive</i>	37123	4.2e+03	1.0e+11	37.289	0.000
<i>Science-based</i>	6497	1.3e+04	7.8e+11		
<i>Specialized suppliers</i>	9101	3.7e+03	3.2e+10	14.673	0.000
<i>Science-based</i>	6497	3.9e+03	4.4e+10		

Interestingly, we also find that for multi-product firms, in 90 percent of the cases, all products produced by the firm display similar cost-quantity trend. This is circumstantial evidence of the influence of the nature of underlying technologies which the firms masters and a sign of its ‘learning’ coefficient. Given the

existence of such inter-sectoral differences, we next investigate the presence of firm-specific characteristics associated with such learning patterns.

6 Cost-quantity relationships and firm characteristics: Product vs Process innovation

The main sector of activity of the firms deeply influence the propensity to undertake *R&D* and the balance between products and process innovation. But, still inter firm variability remains high. Thus, here we investigate the relation between firms' *R&D* and investment spending to the observed learning patterns at the firm-product level.

We perform an ordinary least squares regression to investigate the relationship between the observed learning parameters and innovative activities, separately for each Pavitt sectors. The proxies we consider for innovative activities are cumulated *R&D* intensity (cumulated *R&D* over cumulated sales) for product innovation and cumulated investment intensity (cumulated investment over cumulated sales) for process innovation.

The estimated model is the following cross-sectional regression:

$$LP_{ij} = \alpha + \beta_1 R\&D_i + \beta_2 Inv_i + controls_i \quad (7)$$

where LP_{ij} is the estimated learning parameter for product j produced by firm i , Inv_i is investment intensity of firm i and $R\&D_i$ is the *R&D* intensity of firm i . In equation 7, RHS variables are at firm level, since we observe the *R&D* and investment spending at firm level and not disaggregated at the product level. The LHS variable, i.e, the learning parameter is computed at the firm-product level, as in equation 6. Combining these in the same equation is not a major concern because, as recalled before, products produced by the same firm tend to display similar cost-quantity relation.

Table 6 shows the regression results for all four Pavitt sectors, namely, Supplier Dominated (S-D), Scale Intensive (S-I), Specialized Suppliers (S-S) and Science Based (S-B). The first column (I) shows the results from the regression of equation 7 without any controls. The second column (II) shows the results from the regression while controlling for size, third column (III) also includes year dummies. Standard errors are in brackets. We calculate cluster-robust standard error that permit general heteroskedasticity and within-cluster error correlation.

The first row shows the coefficient values for *R&D*. In all the sectors, the coefficients are positive and significant suggesting that, even within the same Pavitt sector, higher spending on *R&D* is associated with higher values of the learning parameter, that is, higher probabilities to observe a positively shaped "learning curve". The higher spending on *R&D* in turn suggests an underlying quality upgrade of the product in time, i.e, product innovation.

Concerning the relation between learning parameter and investment intensity, as seen in the second row, in three out of four Pavitt sectors, the coefficient is significant and negative, as one would expect on the grounds of capital embodied process innovation. This suggests that higher spending on investment is associated to a higher probability to observe the traditional learning curve. These results corroborates, in fact the suggestion that process learning, or conversely that apparent 'anti-learning' plausibly associated with product upgrades are not a costless and automatic process, or a mere joint output of production activity, but they involves conscious firm intervention.

Table 6: Learning rates of products and innovative characteristics of firms in different Pavitt sectors

<i>Supplier dominated</i>				<i>Scale intensive</i>			
	I	II	III		I	II	III
R&D	0.003*** (0.000)	0.002* (0.000)	0.002* (0.000)	<i>R&D</i>	0.002*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
Investment	-0.002*** (0.000)	-0.004*** (0.000)	-0.005*** (0.000)	Investment	-0.003* (0.002)	-0.002 (0.002)	-0.002 (0.002)
Size	No	0.006*** (0.001)	0.006*** (0.001)	Size	No	-0.003***	-0.003*** (0.000)
Year Dummy	No	No	Yes	Year Dummy	No	No	Yes
Observations	3383	3383	3383	Observations	4676	4676	4676
R^2	0.008	0.015	0.021	R^2	0.006	0.008	0.008

<i>Specialized suppliers</i>				<i>Science based</i>			
	I	II	III		I	II	III
R&D	0.003** (0.001)	0.004*** (0.001)	0.005*** (0.001)	<i>R&D</i>	0.003** (0.001)	0.004*** (0.001)	0.003* (0.001)
Investment	-0.003* (0.000)	-0.002 (0.002)	-0.001 (0.001)	Investment	0.007*** (0.001)	0.008*** (0.002)	0.008*** (0.002)
Size	No	-0.003* (0.001)	-0.002* (0.001)	Size	No	-0.002 (0.002)	-0.002 (0.002)
Year Dummy	No	No	Yes	Year Dummy	No	No	Yes
Observations	2289	2289	2289	Observations	2613	2613	2613
R^2	0.003	0.005	0.005	R^2	0.032	0.032	0.036

7 Firm entry and learning in the market

Recall that the foregoing analysis regard product-level learning curves, generally involving unbalanced panels of diverse firms. The results therefore summarize information about entry, the initial prices of which firm enter in any one product category - possibly with a distinct product quality, and the learning process thereafter. In terms of industrial dynamics, entry of course involves a challenge to the market predictions of the incumbents, while thereafter learning as such represents a barrier to entry as it establish a cost differential between incumbents and would be entrants over the same product quality.

In Dosi (1984), one sketch out a model of industrial evolution inspired by semiconductors, but as such applicable to a wide range of industries, both on the frontier and in the catching up phase - characterized by the co-existence of product innovation and product-specific learning-by-doing. The latter continuously induces advantages to incumbents, but that can be always overcome by introduction of new/improved products in the same family, but with improved functionalities. The prediction of the model is a persistent process of entry, jointly with subsequent learning by doing, and possibly with the new/improved product.

Of course, on the grounds of our data we have no way of accessing the techno-economic features of each product, but a story of persistent late entry cum higher entry prices is consistent with such conjecture. This is what we find indeed in the Indian case. At a finer level of resolution, one would expect on the

grounds of a capability-based theory of the firm (Dosi et al., 2000; Teece and Pisano, 1994) that late entrants which however have learned in related products know also how to produce “better products” at lower costs from the start.

Table 7 shows summary statistics on firm entry price with respect to the average price of the incumbents, or in other words, the market price. Market price is defined as the weighted average price of the product at the market level (incumbents) where the weights are the physical quantities of sales of the product.⁹ Column 1 shows the different product tenure, i.e, the number of years the incumbents are producing one product.¹⁰ Column 2 shows the percentages of firm-product combinations, where the firm-product entry occurs at a higher price than the market average. Column 3 shows the number of firm-product combinations, where the firm-product entry price is lower than the market price. Column 4 shows the total observations in each product tenure class. Column 5 shows the log difference of price of the entrant and market price. We observe in table 7 that with increasing product tenure, i.e, the higher the number of years the incumbents are producing a product, the higher the difference between entrant and market price.

Table 7: Descriptive evidence of firm entry in a given product

Product Tenure	Higher Price (%)	Lower Price (%)	Observations	Log-difference of price (Avg)
0-5	53.92	46.08	102	0.211
5-10	45.95	54.05	407	0.604
10-15	44.44	55.56	504	0.833
15-20	41.87	58.13	492	0.880

Note: Values in brackets are percentages

Let us estimate the model:

$$(P_{ij}/\bar{P}_j) = \alpha_j + \beta_1 Tenure_j + \beta_2 Product\ sales\ share_{ij} + \beta_3 Firm\ Size_i + year_{ij} + sector_i + e_{ij} \quad (8)$$

where P_{ij} is price of product j of firm i , \bar{P}_j is the weighted average price of the product at the market level as defined before, $Tenure_j$ is the (log) number of years the product has been produced by incumbent firms and is our proxy for market experience or knowledge stock at the market level. We control for product sales share, i.e, the share of sales of the product in total sales of firm in the year of entry. We also control for size of the firm and we include year and sector¹¹ effects. Since we are looking at firm entry, a one-time event, we pool the entry events in different years together and hence there is no time dimension in our analysis, in this respect the year dummies allow to control for temporal effect. Therefore, we perform a cross sectional analysis and employ an ordinary least squares regression with product fixed effects. We calculate cluster-robust standard error that permit general heteroskedasticity and within-cluster error correlation.

The results are reported in Table 8 and overall, they suggest that the higher the “age” of the product family, the higher is the ratio of entrant to market price. Conversely, the size of the “product entrant” has a negative effect on the price-ratio.

⁹The average price is computed excluding the new entrant.

¹⁰In order to precisely measure the tenure of the product, we drop the products that were present in the first year of our sample. For those products indeed, we have no information about the actual year of introduction. Note however that in a robustness check in which we include all observations, results remain the same.

¹¹Here, the sector is defined as the main sector of economic activity of the firm, at 2 digit National Industry Classification.

Table 8: Relation between “market experience” and firm entry pricing

	I	II	III	IV
Tenure	0.001** (0.000)	0.001* (0.000)	0.002*** (0.000)	0.001** (0.000)
Product Sales Share	No	0.001 (0.001)	No	0.001 (0.001)
Firm Size	No	No	-0.003** (0.001)	-0.003** (0.001)
N	1867	1561	1794	1501
R^2	0.392	0.484	0.395	0.488
Adjusted R^2	0.072	0.163	0.062	0.154

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Conclusions

Persistent technological and organizational learning is most likely the fundamental driver of economic dynamics since the Industrial Revolution and underlines all episodes of catching-up ever since (for a recent discussion, see Cimoli et al. 2009). Learning takes various forms which the economics of innovation has investigated in detail (a critical survey is in Dosi and Nelson 2010).

One of such ways, statistically identified even earlier than the return of economists to study “inside the blackbox” of technological change, is learning-by-doing, that is some relation between experience by making, usually proxied by cumulative production and increasing production efficiencies/falling cost and prices. The evidence is quite robust, mostly collected so far on industries in developed economies. But does it properly apply also to developing ones? And what explains the inter-sectoral/inter-product differences in the apparent learning patterns, if any?

In this work, on Indian manufacturing, we do corroborate in a good deal of cases the power law relation cost/prices vs cumulated produced quantities. At the same time, we find, *first*, a wide variation in learning coefficients which still demands satisfactory explanations. *Second*, relatedly, some relations appear to be positive, that is, an apparent “anti-learning”. Such patterns, however are consistent with some circumstantial evidence according to which learning tends to relate more to product than process innovation. *Third*, product innovation also explains why late entry in the same product family occurs notwithstanding learning curves, which as such represent entry barriers in favour of incumbents.

References

- Abell, D. F. and J. S. Hammond (1979). Cost dynamics: Scale and experience effects. In D. F. Abell and J. S. Hammond (Eds.), *Strategic Planning: Problems and Analytical Approaches*, pp. 103–133. Prentice-Hall Englewood Cliffs, New Jersey.
- Alchian, A. (1963). Reliability of progress curves in airframe production. *Econometrica* 31(4), 679–693.
- Amiti, M., O. Itskhoki, and J. Konings (2014, July). Importers, Exporters, and Exchange Rate Disconnect. *American Economic Review* 104(7), 1942–78.
- Anderson, J. R. and L. J. Schooler (1991). Reflections of the environment in memory. *Psychological Science* 2(6), 396–408.
- Argote, L. (1996). Organizational learning curves: persistence, transfer and turnover. *International Journal of Technology Management* 11(7-8), 759–769.
- Argote, L. and D. Epple (1990). Learning curves in manufacturing. *Science* 247(4945), 920–924.
- Argote, L., P. Ingram, J. M. Levine, and R. L. Moreland (2000). Knowledge transfer in organizations: Learning from the experience of others. *Organizational Behavior and Human Decision Processes* 82(1), 1–8.
- Arrow, K. and S. S. Arrow (1950). Methodology problems in airframe cost-performance studies. Technical Report RM-456-PR (DDC No. AD 116556).
- Arrow, K. J. (1951). *Cost Quality Relations in Bomber Airframes*. Santa Monica, CA: Rand Corporation.
- Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 155–173.
- Asher, H. (1956). *Cost-quantity relationships in the airframe industry*. Number R-291. CA: Rand Corporation.
- Ayres, R. U. and K. Martinus (1992). Experience and the life cycle: some analytic implications. *Technovation* 12(7), 465–486.
- Balakrishnan, P., K. Pushpangadan, and M. S. Babu (2000). Trade liberalisation and productivity growth in manufacturing: Evidence from firm-level panel data. *Economic and Political weekly*, 3679–3682.
- Bernard, A. B., S. J. Redding, and P. K. Schott (2010). Multiple-product firms and product switching. *American Economic Review* 100(1), 70–97.
- Berry, D. (2009). Innovation and the price of wind energy in the US. *Energy Policy* 37(11), 4493–4499.
- Boston Consulting Group (1970). *Perspectives on experience*. Boston, MA: Boston Consulting Group.
- Brewer, G. M. (1965). The learning curve in the airframe industry. Technical report, Ohio: School of Systems and Logistics, Air Force Institute of Technology, Wright Patterson Air Force base.
- Brockman, W. F. and F. D. Dickens (1967). Investigation of learning curve and cost estimate methods for cargo aircraft; air force institute of technology (ad665464). Technical report, Ohio: Wright Patterson Air Force Base.
- Carr, G. W. (1946). Peacetime cost estimating requires new learning curves. *Aviation* 45(4), 220–228.
- Chung, S. (2001). The learning curve and the yield factor: the case of korea’s semiconductor industry. *Applied Economics* 33(4), 473–483.
- Cimoli, M., G. Dosi, and J. Stiglitz (2009). *Industrial Policy and Development: The Political Economy of Capabilities Accumulation*. Oxford University Press.

- Colasuonno, V. (1967). An analysis of progress curve conceptual advances and progress curve uses, since 1956. Technical report, Unpublished Thesis, Air Force Institute of Technology (AD665464), Wright Patterson Air Force base, OH.
- Coulomb, L. and K. Neuhoﬀ (2006). Learning curves and changing product attributes: the case of wind turbines. Technical report, Cambridge Working Papers in Economics, University of Cambridge.
- De Liso, N., G. Filatrella, and N. Weaver (2001). On endogenous growth and increasing returns: modeling learning-by-doing and the division of labor. *Journal of Economic Behavior & Organization* 46(1), 39–55.
- Dosi, G. (1984). *Technical Change and Industrial Transformation – The Theory and an Application to the Semiconductor Industry*. London, Mcmillan.
- Dosi, G., R. Nelson, and S. Winter (2000). *The nature and dynamics of organizational capabilities*. Oxford, Oxford University Press.
- Dosi, G. and R. R. Nelson (2010). Technological change and industrial dynamics as evolutionary processes. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Chapter 4. Elsevier: Amsterdam.
- Dutton, J. M. and A. Thomas (1984). Treating progress functions as a managerial opportunity. *Academy of Management Review* 9(2), 235–247.
- Field, J. M. and K. K. Sinha (2005). Applying process knowledge for yield variation reduction: A longitudinal field study. *Decision Sciences* 36(1), 159–186.
- Fligner, M. A. and G. E. Policello (1981). Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association* 76(373), 141–206.
- Goddard, C. (1982). Debunking the learning curve. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* 5(4), 328–335.
- Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010a). Imported intermediate inputs and domestic product growth: Evidence from India. *The Quarterly Journal of Economics* 125(4), 1727–1767.
- Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010b). Multiproduct firms and product turnover in the developing world: Evidence from India. *The Review of Economics and Statistics* 92(4), 1042–1049.
- Gruber, H. (1992). The learning curve in the production of semiconductor memory chips. *Applied Economics* 24(8), 885–894.
- Gruber, H. (1998). Learning by Doing and Spillovers: Further Evidence for the Semiconductor Industry. *Review of Industrial Organization* 13(6), 697–711.
- Grubler, A. (2010). The costs of the french nuclear scale-up: A case of negative learning by doing. *Energy Policy* 38(9), 5174–5188.
- Hartley, K. (1969). Estimating military aircraft production outlays: The British experience. *The Economic Journal* 79(316), 861–881.
- Hatch, N. W. and D. C. Mowery (1998). Process innovation and learning by doing in semiconductor manufacturing. *Management Science* 44(11-part-1), 1461–1477.
- Heathcote, A., S. Brown, and D. Mewhort (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review* 7(2), 185–207.

- Hollander, M., D. A. Wolfe, and E. Chicken (1999). *Nonparametric statistical methods*. Wiley Series in Probability and Statistics. Second edition.
- Irwin, D. A. and P. J. Klenow (1994). Learning-by-doing spillovers in the semiconductor industry. *Journal of Political Economy*, 1200–1227.
- Jovanovic, B. and Y. Nyarko (1994). Learning by doing and the choice of technology. Technical report, National Bureau of Economic Research.
- Jovanovic, B. and P. L. Rousseau (2002). Moore’s law and learning by doing. *Review of Economic Dynamics* 5(2), 346–375.
- Juneau, P. (2007). Nonparametric methods in pharmaceutical statistics. In A. Dmitrienko, C. Chuang-Stein, and R. D’Agostino (Eds.), *Pharmaceutical Statistics Using SAS: A Practical Guide*, pp. 117–150. SAS Press, Cary, NC.
- Junginger, M., A. Faaij, and W. C. Turkenburg (2005). Global experience curves for wind farms. *Energy Policy* 33(2), 133–150.
- Kobos, P. H., J. D. Erickson, and T. E. Drennen (2006). Technological learning and renewable energy costs: implications for US renewable energy policy. *Energy Policy* 34(13), 1645–1658.
- Koh, H. and C. L. Magee (2006). A functional approach for studying technological progress: Application to information technology. *Technological Forecasting and Social Change* 73(9), 1061–1083.
- Koh, H. and C. L. Magee (2008). A functional approach for studying technological progress: Extension to energy technology. *Technological Forecasting and Social Change* 75(6), 735–758.
- Krishna, P. and D. Mitra (1998). Trade liberalization, market discipline and productivity growth: new evidence from India. *Journal of Development Economics* 56(2), 447–462.
- Lapré, M. A. and N. Tsikriktsis (2006). Organizational learning curves for customer dissatisfaction: Heterogeneity across airlines. *Management science* 52(3), 352–366.
- Lapré, M. A. and L. N. Van Wassenhove (2001). Creating and transferring knowledge for productivity improvement in factories. *Management Science* 47(10), 1311–1325.
- Levin, D. Z. (2000). Organizational learning and the transfer of knowledge: An investigation of quality improvement. *Organization Science* 11(6), 630–647.
- Lieberman, M. B. (1984). The learning curve and pricing in the chemical processing industries. *The RAND Journal of Economics* 15(2), 213–228.
- Lucas, R. (1998). On the mechanics of economic development. *Econometric Society Monographs* 29, 61–70.
- Lundberg, E. (1961). *Produktivitet och rdntabilitet*. Stockholm: P. A. Norstedt and Soner.
- McDonald, A. and L. Schrattenholzer (2001). Learning rates for energy technologies. *Energy Policy* 29(4), 255–261.
- Mensforth, E. (1947). Airframe production. *Proceedings of the Institution of Mechanical Engineers* 156(1), 24–38.
- Middleton, K. A. (1945). Wartime productivity changes in the airframe industry. *Monthly Labor Review* 61(2), 215–225.
- Mishina, K. (1999). *Learning by New Experiences: Revisiting the Flying Fortress Learning Curve*. University of Chicago Press.

- Neij, L. (1997). Use of experience curves to analyse the prospects for diffusion and adoption of renewable energy technology. *Energy Policy* 25(13), 1099–1107.
- Nordhaus, W. D. (2014). The perils of the learning model for modeling endogenous technological change. *The Energy Journal* 35(1), 1–13.
- Orsini, J. A. (1970). An analysis of theoretical and empirical advances in learning curve concepts since 1966. Technical report, Masters Thesis, Air Force Institute of Technology (AD875892), Wright Patterson Air Force Base, OH.
- Pavitt, K. (1984). Sectoral pattern of technical change: Towards a taxonomy and a theory. *Research Policy* 13, 343–373.
- Rapping, L. (1965). Learning and World War II production functions. *The Review of Economics and Statistics* 47(1), 81–86.
- Rigdon, C. (1944). Analysis of progress trends in aircraft production. *Aero Digest* 132, 138.
- Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy* 94(5), 1002–1037.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* 98(5 pt 2), S71–S102.
- Rosen, S. (1972). Learning by experience as joint production. *The Quarterly Journal of Economics* 86(3), 366–382.
- Schoots, K., F. Ferioli, G. Kramer, and B. Van der Zwaan (2008). Learning curves for hydrogen production technology: an assessment of observed cost reductions. *International Journal of Hydrogen Energy* 33(11), 2630–2645.
- Scott-Kemmis, D. and M. Bell (2010). The mythology of learning-by-doing in World War II airframe and ship production. *International Journal of Technological Learning, Innovation and Development* 3(1), 1–35.
- Searle, A. D. (1945). Productivity changes in selected wartime shipbuilding programs. *Monthly Labor Review* 61, 1132–1147.
- Sinclair, G., S. Klepper, and W. Cohen (2000). What’s experience got to do with it? Sources of cost reduction in a large specialty chemicals producer. *Management Science* 46(1), 28–45.
- Stanley, P. (1949). The time to achieve peak output with special reference to aircraft production. Technical report, College of Aeronautics, Cranfield.
- Stokey, N. L. (1988). Learning by doing and the introduction of new goods. *The Journal of Political Economy* 96(4), 701–717.
- Teece, D. and G. Pisano (1994). The Dynamic Capabilities of Firms: an Introduction. *Industrial and corporate change* 3(3), 537–556.
- Thompson, P. (2001). How much did the liberty shipbuilders learn? New evidence for an old case study. *Journal of Political Economy* 109(1), 103–137.
- Topalova, P. and A. Khandelwal (2011). Trade liberalization and firm productivity: The case of India. *The Review of Economics and Statistics* 93(3), 995–1009.
- Wene, C.-O. (2007). Technology learning systems as non-trivial machines. *Kybernetes* 36(3/4), 348–363.
- Wixted, J. T. and E. B. Ebbesen (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & cognition* 25(5), 731–739.

- Wright, T. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences* 3, 122–128.
- Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision Sciences* 10(2), 302–328.
- Young, A. (1993). Invention and bounded learning by doing. *Journal of Political Economy* 101(3), 443–472.
- Young, S. L. (1966). Misapplications of learning curve concept. *Journal of Industrial Engineering* 17(8), 410.
- Zieke, R. P. (1962). *Progress Curve Analysis in the Aerospace Industry*. Ph. D. thesis, Department of Industrial Engineering, Stanford University.

A Few details on the data used

As we mentioned in Section 3, the data is taken from Prowess database, provided by the CMIE (Centre For Monitoring Indian Economy Pvt. Ltd.). Prowess is a database of active business entities for which information related to their financial performance is available. By “active” business entities, CMIE means those business entities that are not mere registrations without any activity. By “business entities” CMIE implies that it is not restricted to only registered companies. Prowess do not cover the universe of active business entities, even though it is the largest and most comprehensive database on the financial performance of Indian business entities. The companies covered account for around 70 percent of industrial output, 75 percent of corporate taxes, and more than 95 percent of excise taxes collected by the Government of India. Earlier studies, such as Topalova and Khandelwal (2011); Balakrishnan et al. (2000) and Krishna and Mitra (1998) have used the same database, at firm level, to study the impact of reforms on productivity growth of firms. Few studies which used the same firm-product level data include Goldberg et al. (2010a) to study the relationship between declines in trade costs, the imports of intermediate inputs and domestic firm product scope; and Goldberg et al. (2010b) focusing in the characteristics of multi product firms and the link between product rationalization and trade reforms in India.

Below, we present some description on the product data used for the study. Table 2 gives summary statistics of product-reporting firms. In the following parts, we describe on the product classification and provide some examples of product hierarchy.

A.1 On the product classification

All companies in the Prowess database are mapped to a product or a service in CMIE’s standardized products and services classification. This mapping reflects the company’s main economic activity during a year¹². Out of these, 90 percent of firms report products they produce. Among the firms that report products produced, in 99 percent of the cases, the sum of sales of products account for more than 95 percent of firm sales.

The product and services classification developed by CMIE is based on the Indian Trade Classification (ITC) which, in turn is based on the Harmonized Commodity Description and Coding System, commonly known as the HS. The ITC system covers only commodities and no services or utilities. CMIE has added these for its classification system. CMIE’s standardised products and services classification is a tree-like organisation of all products and services. The structure can be picturised as a set of groups of products/services at the broadest level. For example, chemicals or base metals are broad groups. Each such group consists of sub-groups of products/services. A sub-group can again consist of sub-sub-groups and, so on. Finally, all groups, sub-groups, sub-sub-sub groups, etc. consist of individual products or services. The groups and sub-groups are a way of organizing products/services into logical collections.

There are a total of 2411 products linked to 293 five-digit NIC industries across the 22 manufacturing sectors (two-digit NIC codes). As a comparison, the U.S. manufacturing data used by Bernard et al. (2010), 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 Bernard et al. (2010).

Examples of products with its hierarchical structure within 2 sectors, namely, manufacture of bever-

¹²A company is classified under a particular industry if more than half of its sales originates from the particular industry or industry group. The industry group could be any product or a product group in the CMIE products and services classification structure.

ages and manufacture of leather and leather products are listed in table 9. The table 9 reports one four digit industry within the beverages sector: Manufacture of malt liquors, beer and other alcohol. The products include malt spirit distilled, soda/carbonated water, beer, sparkling wine, potable alcohol etc. Similar to other classifications, the degree of detail varies across industries and sectors. As we see in table 9, within the manufacture of leather and leather products, there are two different four digit sectors listed, along with the products within each sector.

Table 9: Examples of Industries, Sectors and Products

NIC	Product Code	Description
11		<i>Manufacture of beverages</i>
1103		Manufacture of malt liquors, beer and other alcohol
	050703010000	Malt spirit distilled
	051401040000	Soda/Carbonated water
	051403000000	Beer
	051404010000	Sparkling wine
	051406000000	Potable alcohol
	051406010000	Indian made foreign liquor
	051406010200	Heritage liquor
	051406010300	Scotch & whiskey
15		<i>Manufacture of leather and leather products</i>
1512		Manufacture of consumer goods of leather and substitutes of leather
	070202040000	Shopping bags/ carry bags
	070202060000	Leather hand bag
	070202070000	Wallets and leather purses
	070203000000	Leather garments and accessories
	070203010000	Leather jackets
	070203020000	Leather gloves
	070203040000	Leather belts
	070203050000	Industrial leather hand gloves/apron
1520		Manufacture of leather footwear
	070601000000	Full leather shoes
	070602000000	Canvas shoes
	070603000000	Full shoes or boots
	070604000000	Slippers
	070605000000	Plastic footwear
	070606000000	Footcare products
	070607000000	Shoe uppers
	070608000000	Shoe soles/heels

Source: Prowess database

We remove from the sample products defined only at 2 digit, as they identify broad categories of products which might display a high degree of variability. We also remove the products produced in the

first year of the database since we do not have information on which year the firm started producing the product. Table 10 gives the share of products at different levels of disaggregation. From the table you can see that removing the first 2 digits of disaggregation would remove only 3.32 percent of total sales reported.

Table 10: Share of products and firms producing products at various levels of disaggregation

	2 digit	4 digit	6 digit	8 digit	Total
Share of products	6.68	60.70	31.40	0.03	100
Share in total sales	3.32	75.66	21.00	0.003	100

B Relation between price and unit production cost

Here, we present the results from a fixed effect regression of equation 1. Column 1 reports the results from a fixed effects regression of price on unit cost.

Table 11: Relation between cost of production and price of product

	(1)
Unit cost	0.4305*** (0.0045)
Time Dummies	Yes
Observations	19669
R^2 within	0.364
R^2 between	0.899
R^2 overall	0.869
Number of firms	2581

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$