R&D Intensity in the Brazilian Industry: Some Distributional Regularities*

Luís Otávio Façanha
Marcelo Resende
Instituto de Economia, Universidade Federal do Rio de Janeiro

Recebido: 5/11/2003 Aprovado: 8/7/2004

ABSTRACT

Previous empirical results on the determinants of R&D intensity have indirectly suggested the relevance of unobserved R&D-related capabilities. Cohen and Klepper (1992) have developed a probabilistic model that is able to conform with well known stylized facts on the R&D intensity distribution such as clustering around zero, unimodality and positive skewness and generates well defined predictions about the correlation between the different distribution moments (mean, variance, coefficient of variation and skewness). The evidence for 3-digits sectors of the manufacturing industry in São Paulo is to a great extent consistent with the referred theoretical model.

KEYWORDS | R&D Intensity; Distribution

JEL CODES | C16; L60; O32

* The authors acknowledge valuable comments from two anonymous referees, but the usual disclaimer applies.

Resumo

Resultados empíricos anteriores sobre os determinantes da intensidade em P&D têm sugerido indiretamente a relevância de competências não observadas relacionadas a P&D. Cohen e Klepper (1992) desenvolveram um modelo probabilístico que é capaz de ser consistente com fatos estilizados bem conhecidos acerca da distribuição de P&D tais como agrupamento em torno de zero, unimodalidade e assimetria positiva e gera previsões bem definidas sobre a correlação entre os diferentes momentos da distribuição (média, variância, coeficiente de variação e assimetria). A evidência para setores a 3-dígitos da indústria de transformação em São Paulo é em grande parte consistente com o referido modelo teórico.

Palavras-Chave | Intensidade de P&D; Distribuição

Códigos JEL | C16; L60; O32

1. Introduction

Inter-industry regularities have gradually reached a lower status in the recent empirical Industrial Organization literature. In fact, one can testify a movement of that research area towards the increasing consideration of industry-specific effects to explain the behavior of various variables in Industrial Organization. Schmalensee (1989), for example, questions the relevance of inter-industry studies in the context of traditional structure-conduct-performance framework, whereas Bresnahan (1989) points out that structural econometric modeling of market power necessarily requires industry case studies to properly approach industry-specific idiosyncrasies. The absence of general patterns appears to be also salient in the context of industry dynamics aspects such as entry and exit rates as indicated by Dunne and Roberts (1988) and Façanha and Resende (2003) for the U.S. and Brazilian manufacturing industries respectively.

The natural question that follows the previous considerations is whether any important regularity can be obtained in Industrial Economics. It turns out, that some within-industry patterns appear to be recurring. Skewed size distributions for firms are often encountered and could be consistent with Gibrat´s law [see *e.g.* Sutton (1997) for an overview]. Other interesting distributional regularities can be motivated for R&D intensity as developed by Cohen and Klepper (1992). In fact, the bulk of the empirical literature on the determinants of R&D often indicate that observable explanatory factors like firm size, cash flow and diversification might have limited role in explaining R&D intensity (see Cohen & Levin, 1989). These results can potentially motivate the role of unobserved factors that were generically denoted by R&D *related capabilities* as in fact has already been recognized by Nelson and Winter (1982). These authors emphasized the importance of capabilities accumulated through experience and capabilities acquired through positive factors arising from the stochastic environment in explaining inter-firm R&D intensity. Cohen and Klepper (1992) provide a simple probabilistic framework generating results that are consistent with some stylized facts pertaining the shape of R&D intensity distributions. The paper aims to partially fill the existing gap in the Brazilian empirical literature. In fact, studies in related topics are scarce in the Brazilian case in part due to data availability restrictions. Exception includes descriptive studies assessing the behavior of R&D, for example in terms of its relation with firm size as provided by Resende and Hasenclever (1998) and Macedo and Albuquerque (1999). The analyses considered by those papers were constrained by data availability. The broader sample utilized in this paper allows to explore distributional patterns of R&D intensity in terms of a theoretical framework that underscores the role of stochastic unobserved factors in determining R&D intensity. We believe that the exercise might be valuable in the context of Brazilian industry despite important differences between Brazil and US regarding technological effort.

The paper is organized as follows. The second section presents the basic probabilistic elements underlying the model of R&D intensity developed by Cohen and Klepper (1992). The third section presents the empirical analysis of R&D intensity in the Brazilian case, and discusses therefore the data source and analyses the empirical distributional regularities obtained for the Brazilian case. The fourth section brings some final comments.

2. R&D intensity: some probabilistic aspects

The search for probabilistic foundations for the distribution of R&D intensity is recent in the literature. Cohen and Klepper (1992) advances a simple probabilistic model in order to conform with some apparent stylized facts pertaining the distribution of R&D intensity. These regularities include: a large number of non-performing firms clustering around zero, unimodality and positive skewness. The weak empirical results on the determinants of R&D intensity coupled with some inter-industry regularities raise the possibility that chance ("unobserved R&D-related capabilities") may play a role on the observed patterns of R&D intensity.

Earlier contributions recognized the stochastic nature of R&D (see *e.g.* Evenson & Kislev, 1976; Nelson, 1982; Nelson & Winter, 1982). In particular, emphasize the uncertain outcomes accruing from R&D experimentation whereas the modeling strategy by Cohen and Klepper (1992) attempt to focus on experimentation itself in its relation with randomly acquired expertise.

The starting point of the model refers to randomly acquired expertise that reflects intrinsic characteristics difficult to change in the short-run. The firm engages in experiments of the approach to innovation for which it possesses expertise. Let n_j denote the number of possible approaches available to firm j and z_{ij} refer to the number of experiments in approach i conducted by firm j. The marginal cost is assumed as constant across experiments and firms and equals

$MC(z_{ij}) = c$. Defining also r_{ij} as the total R&D expenditure in approach i by firm j, it follows that $c = r_{ij}/z_{ij}$.

From the revenue point of view, let q_j denote total output of firm j times the average number of units of quality per output. Firms are price-takers in the sense that the R&D expenditures are independently chosen relative to the other firms and also in the sense that the firms receive a price h (price per unit of quality).[1] The resulting marginal revenue (equal to average revenue) $MR(z_{ij}) = q_j h/z_{ij}$ (where the term in the numerator refers to total sales in units of quality – say s_j).

[1] The use of quality-adjusted variables is convenient to simultaneously encompass the possibility of both process and product innovations.

From the standpoint of profit maximization the optimal level of experiments should be consistent with the equality between marginal cost and marginal revenue. However, in order to avoid trivial results and enrich the model, the authors take into consideration parameters that reflect technological opportunities and ability to appropriate rents from innovation. For the former effect, the authors consider a generic positive g whereas for the latter a fraction b, both assumed to be constant across firms. In this sense, one can readily obtain:

$$\frac{r_{ij}}{s_j} = bg \equiv d \qquad (1)$$

In this sense, d represents the R&D intensity that is specific to a particular approach. If firm j adopts n_j different approaches, such that $\sum_i r_{ij} = r_j$ it follows that:

$$\frac{r_j}{s_j} = n_j d \qquad (2)$$

The theoretical prediction has, therefore, implications on the pattern of R&D intensity. The previous hypotheses imply that there is a probability p of a firm being endowed with a given expertise and therefore the search of innovative opportunities within a given approach evolves in accordance with a Bernoulli process (with probability of success equal to p). Hence, it follows that $n_j d$ (and therefore R&D intensity) is distributed as a binomial random variable [b(N,p), where N refer to the total number of approaches adopted by all firms]. In that particular context, we can rely on standard results for the moments of a binomial random variable. Specifically:

$$E(r_j / s_j) \equiv \mu = dnp \qquad (3)$$

$$V(r_j / s_j) \equiv \sigma^2 = d^2 p(1-p) \qquad (4)$$

$$CV(r_j / s_j) \equiv \varpi / \mu = N^{-1/2} p^{-1/2} (1-p)^{1/2} \qquad (5)$$

$$S(r_j / s_j) \equiv \mu^3 / \varpi^3 = (1-2p) N^{-1/2} p^{-1/2} (1-p)^{-1/2} \qquad (6)$$

where E(.), V(.), CV(.) and S(.) stand respectively for mean, variance, coefficient of variation and skewness. It is possible to examine the effects of N, d and p on the previously indicated moments. For that purpose, one can compute the partial derivatives of those moments with respect to the referred parameters. Under the simplifying assumption that $p \leq 1/2$, the analysis becomes more straightforward as the mean and the variance rise as N, d and p rise, whereas the coefficient of variation and skewness decline. Taken together, these results produce some implications in terms of the correlations between the different moments as summarized in Table 1:

TABLE 1
R&D intensity: moments correlations

Moments	$E(r_j/s_j)$	$V(r_j/s_j)$	$S(r_j/s_j)$
$V(r_j/s_j)$	+		
$S(r_j/s_j)$	–	–	
$CV(r_j/s_j)$	–	–	+

The empirical evidence obtained by Cohen and Klepper (1992) for the U.S. provides acceptable support for the model. First of all, the visual inspection of the histograms for the 15 sectors studied consistently indicated that clustering around zero, unimodality and strong positive skewness are salient features. In the context of theoretical implications, the correlations between the sample moments of R&D intensity (as summarized in Table 1) are largely supported. In fact, all the expected signs are obtained with the exception of the small and insignificant correlation between the variance and the skewness. As a complementary test for the validity of the binomial distribution, the authors considered the expressions (3), (4) and (5) that were solved for d, N and p. Since d (the R&D intensity related to particular approach) must be positive, the equations for the first three moments simultaneously imply $T_1 \equiv 2\sigma - S\mu > 0$. Additionally as $N > 0$ and $0 < p < 1$, the condition $T_2 \equiv \sigma - S\mu > 0$ should hold. Cohen and Klepper found that the first condition was satisfied in 88.89% of the sectors studied whereas the second condition was satisfied in only 26.26% of the cases.

This line of research was further pursued by Lee (2002). At a descriptive level, the author investigates the distribution of R&D intensity in seven industries and in six different countries for R&D performing firms (Canada, Japan, Korea, Taiwan, India and China). The distribution patterns for those countries are less homogeneous than those found by Cohen and Klepper for the U.S. but the regularities are still present. In particular, strong positive skewness is observed. This result was further investigated by means of (log) normality tests. The evidence strongly supported the (log) normality of R&D intensity distribution. An extension developed by the author considered a variable called firms´ "technological capability" that is approximated in terms of a categorical variable that reflects the firm´s perception of capability relative to the sector leader. This information provided a starting point to devise a proxy for "unobserved R&D-related capabilities" and the resulting distribution is consistent with log-normality.

3. Empirical Analysis

3.1. Data Construction

The paper relied on a unique data set from a survey conducted by Fundação SEADE for the state of São Paulo in 1996 (Pesquisa de Atividade Econômica Paulista –PAEP). This extensive survey provided not only some basic accounting information as well as information on innovative activity and organizational practices. For the present study, we considered the industry questionnaire and constructed an R&D intensity variable defined as the number of employees allocated to R&D activities divided by the total number of employees. Even though, the available data refers to São Paulo, it can be considered representative as a substantial portion of the manufacturing industry is located in that state. Information on R&D activity was available for firms with more than 100 employees. We further restricted our attention to 3-digits sectors that had 30 or more firms. In that sense, the final sample included 1,341 observations distributed over 23 sectors.[2] In comparison wich Cohen and Klepper's sample for the U.S., one can identify some contrasts. The authors consider data for 2,491 business units operated by 352 firms (obtained from Fortune´s top 1,000 list), where business units are spread over 254 sectors (defined by the

[2] See the appendix for a list of the 3-digits sectors considered in the present study.

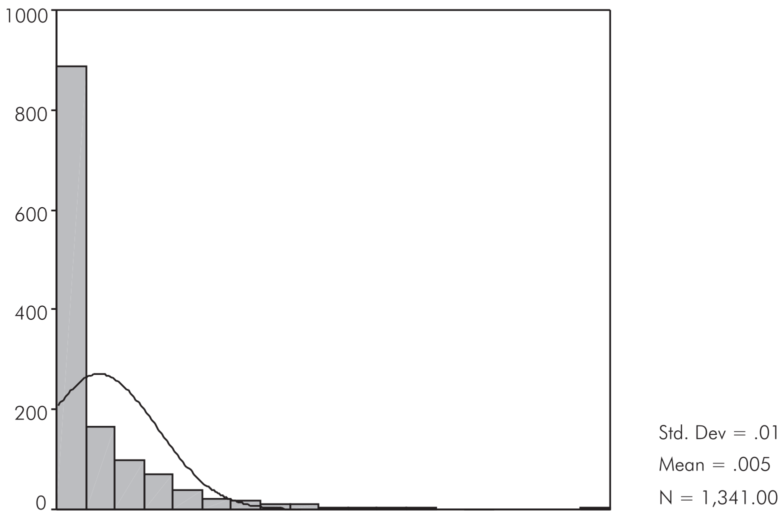
Federal Trade Commission´s Line of Business Program). It is important to point out that the correlation across moments and the associated tests were carried out for 99 lines of business comprising a minimum of ten business units each.

Finally, it is worth mentioning that unlike Cohen and Klepper´s study we do not have access to plant-level data. It is reasonable, however, to conjecture that in the Brazilian case a smaller number of plants from a given firm pursue R&D efforts. The use of aggregate R&D intensity data in the present study can render heterogeneities less evident. In any case, it is relevant to investigate distributional regularities in R&D intensity at the firm-level in the Brazilian case. Further comments on the contrasts between the two kinds of data will be discussed later in the text.

3.2. Empirical Results

This section presents the empirical results of the application of Cohen and Klepper´s approach to the Brazilian case. First, at a more descriptive level, we can characterize the frequency distribution of R&D intensity for the totality of the sample as indicated by the histogram presented below.

FIGURE 1
R&D intensity (RDI)



The shape of the distribution is similar to the one encountered by Cohen and Klepper and therefore further motivates the consideration of the correlation between the different moments and the implementation of the related tests for the Brazilian case as the binomial distribution cannot be obviously discarded. In particular, it is worth mentioning the strong clustering around zero that would be in fact reinforced if one had data on smaller firms as typically there is high incidence of non-performing firms.[3]

Secondly, Table 2 displays the sample correlation between the moments of the R&D intensity distribution.

TABLE 2
R&D intensity: moments empirical correlations

Moments	$E(r_j/s_j)$	$V(r_j/s_j)$	$S(r_j/s_j)$
$V(r_j/s_j)$	0,881 (0,000)		
$S(r_j/s_j)$	-0,382 (0,072)	-0,029 (0,895)	
$CV(r_j/s_j)$	-0,691 (0,000)	-0,350 (0,102)	0,862 (0,000)

Note: p-value in parenthesis

The results are broadly consistent with the implications derived from Cohen and Klepper´s model. In fact, the only unexpected results referred to the correlation between skewness and variance that exhibited the correct negative sign but was non-significant and a negative but marginally significant correlation between the variance and the coefficient of variation. The results are similar to those obtained by Cohen and Klepper not only in terms of the signs of the correlations but also in terms of their magnitudes in most cases with the exception of the correlation between the two measures of variability. In that case there is some peculiarity in the relationship between mean in variance in the

3 Moreover it is worth mentioning that the clustering around zero is substantial even for large firms. In fact, in the case of multinational firms the R&D efforts are usually carried out in the subsidiaries.

Brazilian case that might deserve additional investigation. A possible intuition for this result is suggested by Table 3, where is becomes evident that one faces a reduced amount of dispersion and therefore one has relatively homogeneous sectors in Brazil with respect to R&D effort whereas one observes large variability in the coefficient of variation. In order to obtain a large negative correlation coefficient one needs that above-mean values for one variable is strongly associated with below-mean values for the other variable. Given the mentioned patterns it would be unlikely to obtain a strong negative correlation between the coefficient of variation and the variance.

As already suggested in the previous section, complementary tests on the validity of the binomial distribution were carried out through the investigation of the conditions T_1 and T_2. The former condition was satisfied in 100% of the cases whereas the latter was satisfied in 13.04% of the cases (as compared to 89% and 26% that were respectively obtained by Cohen and Klepper). Once more, the results for the Brazilian industry are somewhat similar to the evidence obtained for the U.S. It is worth mentioning, that the systematic violation of T_2 is associated with especially large skewnesses that to a great extent follow from the substantial clustering around zero, since unlike Cohen and Klepper case, the mean is low in the Brazilian case and therefore the dominant role in the referred violation is associated with very high skewness. In order to verify the robustness of the results to more extreme observation, we excluded

TABLE 3
Moments of the R&D intensity distribution-summary statistics

	Mean	Std. Deviation	Minimum	Maximum
Mean	4,876E-03	2,881E-03	0	0,010
Mode	0	0	0	0
Std. deviation	8,968E-03	4,250E-03	0	0,020
Coeff. variation	2,098	0,6633	1,410	3,470
Skewness	2,712	0,866	1,320	4,250
Minimum	0	0	0	0
Maximum	0,043	0,023	0,010	0,090

Note: the columns present summary statistics for different sample moments across 23 sectors (at the 3-digits level) firm with R&D intensity equal or above 0,06 (7 firms). For the sample as a whole, T_2 declined from -0,006 to -0,003.

As a motivation for the previous point, we present in Table 3, summary statistics of the sample moments of the different sectors.

It becomes evident that non-performing firms constitute a substantial portion of the sample and typically one observes low R&D intensity mean values and variability. These results combined with strong positive skewness (few moderate values are enough to produce large skewnesses given the clustering around zero). Nevertheless, the overall evidence favors the binomial characterization of the distribution of R&D intensity.

4. Final comments

The paper attempted to pinpoint distributional patterns for R&D intensity that could be associated with unobserved factors that could be governed by chance. The modeling approach advanced by Cohen and Klepper (1992) legitimates a binomial distribution for R&D intensity. The implied moment correlations implications were supported in terms of the empirical evidence. A possible interpretation is that the distributional regularities observed across different sectors would reflect the presence of common unobserved factors that were denoted by the authors as "unobserved R&D-related capabilities".

Interesting avenues for future research include a better understanding of those unobserved forces as was suggested by Lee (2002). A further development in that direction would be the consideration of the adoption of modern organizational practices (relative to the sector "leader") as those are likely to be complementary (in the strategic sense) relative to R&D efforts. In fact, Bresnahan *et al.* (2002), in a similar vein, had already emphasized the possible positive interplay between information technology and workplace organization in conditioning superior firm performance. In the current stage, data availability seriously constrain those endeavors.

Appendix

3-Digits sectors considered in the study

Sector	Number of firms
Sugar Manufacturing	41
Other Food Products	65
Other Food Products	33
Beverages	35
Textile mill products	118
Clothing	55
Footwear	33
Paper and cardboard packing	71
Printing (newspaper, magazines, books, etc.)	37
Alcohol	51
Pharmaceutical	46
Other chemical products	49
Rubber artifacts	129
Plastic products	50
Ceramics	40
Forgery	56
Metallurgy	30
Cutlery	89
Motors and compressors	40
General purpose machines	67
Other specific use machines	52
Vehicle parts	114
Furniture	40

References

Bresnahan, T.F., "Empirical Studies of Industries with Market Power", *in* Schmalensee, R.; Willig, R. (eds.), *Handbook of Industrial Organization*, Amsterdã: North-Holland, 1.011-1.057, 1989.

Bresnahan, T.F.; Brynjolfsson, E.; Hitt, L.M., "Information Technology, Workplace Organization, and the Demand Skilled Labor: Firm-Level Evidence", *in Quarterly Journal of Economics*, 117, 339-376, 2000.

Cohen, W.M.; Klepper, S., "The Anatomy of Industry R&D Intensity Distributions", *in American Economic Review*, 82, 775-799, 1992.

__________; Levin, R.C., "Empirical Studies of Innovation and Market Structure", *in* Schmalensee, R.; Willig, R. (eds.), *Handbook of Industrial Organization*, Amsterdam: North-Holland, 1.059-1.107, 1989.

Dunne, T.; Roberts, M., "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries", *in RAND Journal of Economics*, 19, 495-515, 1998.

Evenson, R.E; Kislev, Y., "Stochastic Model of Applied Research", *in Journal of Political Economy*, 84, 265-281, 1976.

Façanha, L. O.; Resende, M., "Dinâmica Multi-Setorial da Indústria Brasileira: um Estudo Empírico", Rio de Janeiro: IE-UFRJ, mimeo, 2003.

Lee, C.Y., "Industry R&D Intensity Distributions: Regularities and Underlying Determinants", *in Journal of Evolutionary Economics*, 12, 307-341, 2002.

Macedo, P.B.; Albuquerque, E.M., "P&D e Tamanho da Empresa: Evidência Empírica sobre a Indústria Brasileira", *in Estudos Econômicos*, 29, 343-365, 1999.

Nelson, R.R., "The Role of Knowledge and R&D Efficiency", *in Quarterly Journal of Economics*, 97, 453-470, 1982.

__________; Winter, S., *An Evolutionary Theory of Economic Change*, Cambridge: Belnaps, 1982.

Resende, M., Hasenclever, L., "Intensidade em Pesquisa e Desenvolvimento e Tamanho da Firma: uma Análise Exploratória do Caso Brasileiro", *in Estudos Econômicos*, 28, 601-618, 1998.

Schmalensee, R., "Inter-Industry Studies of Structure and Performance", *in* Schmalensee, R.; Willig, R. (eds.), *Handbook of Industrial Organization*, Amsterdam: North-Holland, 951-1.009, 1989.

Sutton, J. ,"Gibrat's Legacy", *in Journal of Economic Literature*, 35, p.40-59, 1997.