Variables Affecting High-Tech Cluster Innovation: A Statistical Approach

Mousa Al-kfairy, Prof. Souheil Khaddaj and Dr. Robert B. Mellor

School of Computer Science and Mathematics, Kingston University, London

Abstract: *Regional development and innovation capabilities are topics of prime importance to economists. Prime amongst tools to foster and encourage regional innovation is the "business cluster" particularly high-tech clusters which are believed to foster more innovation. However, the factors leading to this elusive effect are ill-defined. In this paper, we applied statistical modelling methodologies on a longitudinal data set derived from a successful business cluster in its mature stage (Mäjardevi Science Park, Sweden). We examined the relationship between cluster's innovation output using patents and licenses income as indicator, and its input using R&D and social expenses as another indicator. The results show that innovation output is strongly connected with both R&D expenditure and with social networking. Perhaps surprising is the correlation between innovation output and the number of firms in the cluster. Thus, managers of high-tech clusters' can set innovation targets based on the number of firms involved and from which they can calculate what healthy levels of R&D and networking expenses should be.*

Keywords: *innovation capabilities, firms' networks, social expenses, cluster innovation, innovation indicators.*

1. Introduction

Probably the most accepted definition of a business cluster is "a set of interconnected firms in close proximity with a connection to higher education institution, and working in the same or similar industries" [1]. They are assumed to foster innovation and stimulate regional development through knowledge spill-overs and networking within the firms that inhabit the cluster [2]. Because of the importance of regional development (high-tech or otherwise), the subject has been the focus of several studies, the majority of which have concentrated on revealing which metrics or variables can reliably be used as indicators of success - or otherwise. Examples of indicators include R&D expenditure, and indeed reference [3], found a strong correlation between R&D investments and sales growth. Others found R&D expenditure volatility (which is defined as change and variation of R&D investments overtime) to be a potential indicator, as reported in [4]. They used the changes in R&D investments to measure how firms are moving between experimentation stage to the utilization of the knowledge, which involves different amount of R&D investments needs overtime. However, their findings confirm that firms with sequential changes in their R&D investments experience a higher sales growth in large firms, while it has a negative impact on small firms and slow-moving industries [4].

Networking at worker level and managerial networking have also been used for measuring innovation and explaining the growth of corporate performance (e.g. in [5] and [6]) as well as other examples like the value of formal and informal networking channels between firms [7]. Other common indicators include patenting [8] and R&D sustainability [9] which have been tried as potential qualitative metrics to be used for measuring innovation at cluster level [9].

The largest part of the body of existing literature endeavors to link corporate performance with the capabilities of the firm to promote and manage its capability to innovate. A characteristic of the majority of these studies is the assumption that innovation can be achieved through investing more on R&D, in order to develop valuable networks or adding to patent portfolios (or produce patented products under license). However, there is a lack of evidence as to how firms can achieve these goals, especially given a high background "noise". Because

anecdotal evidence suggests that firms or individuals will often submit "blocking patents", never meant to achieve a return for the applicant.

The work presented in this paper differs from previous studies in that, our results differentiate between innovation input and output, where meaningful patents (i.e. those commanding license fees) are an obvious output of research and development and the input is the cost involved in achieving these deliverables. In order to do this, pointers from the existing literature were taken and analyzed statistically by regression analysis on scatter plots, a technique which enables a higher degree of precision than otherwise possible. For example, we assume that networking would involve social activities and will have a cost attached to it, thus social expenses could be another innovation input. Based on that and other metrics, we used documented R&D outlay as well as social expenses (the importance of social inter-connectivity to promote innovation has been described by many, for a recent overview see e.g. [10]) as innovation input, and the innovation output is the value of patents as expressed as income from license agreements.

To achieve meaningful results, we added both the number of firms to our equations, as well as number of employees, and where needed, we combined these to reach a "per capita" result, to come up with a meaningful set of indicators that apparently truly do affect innovation output at cluster level.

2. Purpose of the Study

This study defines and explores the statistical relationship(s) between the innovation input of companies within a mature business cluster and their innovation output. We compare these metrics in order to see which metrics are important to a more or lesser extent in the innovation output of a business cluster.

3. Methodology

3.1. Sample Selection

A longitudinal data-set from 2006 to 2016 was obtained from the Swedish 'Ratsit' database for companies using industrial code '62' (programming and related industries). The over 300 companies were divided into two groups, those inhabiting the Linköping science park development 'Mjardevi Science Park' were called 'on-cluster' firms, while those in in Linköping municipality but not in Mjardevi were designated 'off-cluster' firms.

The data was then smoothed by dividing all the values by 1000 (one thousand) to simplify the calculations, thus all data reported in MSEK (Million Swedish Krona, original data was in KSEK(thousand Swedish Krona)), except for that concerning the number of employees.

3.2. Statistical Approach

The data was then cleaned for each category (field), and summed up using the following function, where *v(c,j) is the value of category (c) for company (j) e.g the number of employees in firm number two will have the function v(number of employees, 2)*:

$$V(i) = \sum_{j=1}^{n} v(c,j), j = 1..n \text{ firms in } (i) \text{ year} \tag{1}$$

As discussed earlier, the literature indicates that outlay on social expenses (networking) and R&D expenditure are the innovation input, and that the patents and license income is the innovation output. Thus equation (1) was calculated using the cleaned data for patent and license income, R&D, social expenses. The total number of employees, and total number of firms were added to the model to show how innovation output can be mapped with the different development stages of high-tech cluster, and the simulation and run as linear and non-linear regression analyses using Minitab statistical software version 17.

4. Results

Table 1 shows a summary of the relevant data set for cluster firms, the data represent the years 2007 – 2015, since the data for 2006 and 2016 were not complete (all values are of million Swedish Krona (MSEK), except for the number of firms and number of employees). Then, we apply Anderson-Darling normality test on R&D, P&L and social expenses. The test reported P-values of 0.792, 0.136 and 0.253 for the three variables respectively, which means we fail to reject the null hypothesis, concluding that linear regression can be applied. So, we applied linear regression as shown in (2) [11], where patents and licenses income can be predicted using

social expenses or R&D multiplied by a coefficient value α, which means changing one unit in social expenses will change α in P&L income (see fig.1 and fig.2 for the actual values α). On the other hand, β represents the y-intercept.

TABLE I: Summary of Data Set

Number of firms	R&D (MSEK)	P&L (MSEK)	Number of employees	social expenses (MSEK)
55	487.867	1.801	1203	269.757
62	491.717	3.005	1281	281.753
69	559.346	3.336	1353	286.929
69	608.261	5.191	1340	329.112
71	644.63	38.222	1468	332.656
74	727.021	62.16	1610	359.42
74	773.372	46.78	1630	458.286
83	814.879	27.804	1470	371.512
83	884.404	67.697	1463	497.654

R&D is summed annual outlay on R&D.
P&L is summed annual income from patents and licenses.

$$P\&L = \alpha \cdot X + \beta,\ X \text{ is the value of either social expenses or R\&D after smoothing} \tag{2}$$

Using equation (2) all the variables in Table 1 could be plotted against each other using Minitab. The results are shown in Fig. 1 and Fig.2 reveal that that P&L can be predicted from social expenses and R&D.
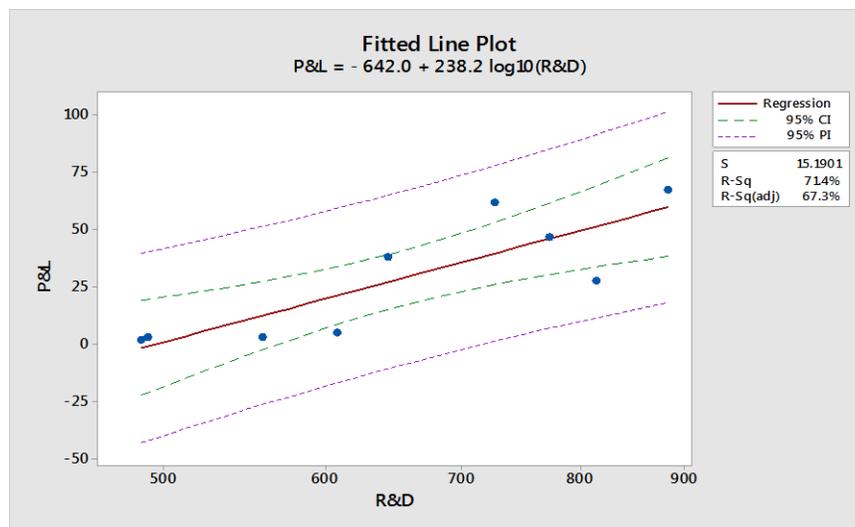

Fig.1: P&L Regression with R&D INPUT.

All values except the response were smoothed using logarithm to the base 10. The results shown in Fig. 1 indicate that the factor that influences innovation output is and R&D expenditure. Fig. 2 investigates networking within the business cluster, through socializing between colleagues and firms. Comparing Fig. 1 and 2, it can be seen that social expenses have a little more impact on the innovation output than R&D (compare fig. 1 and fig. 2).
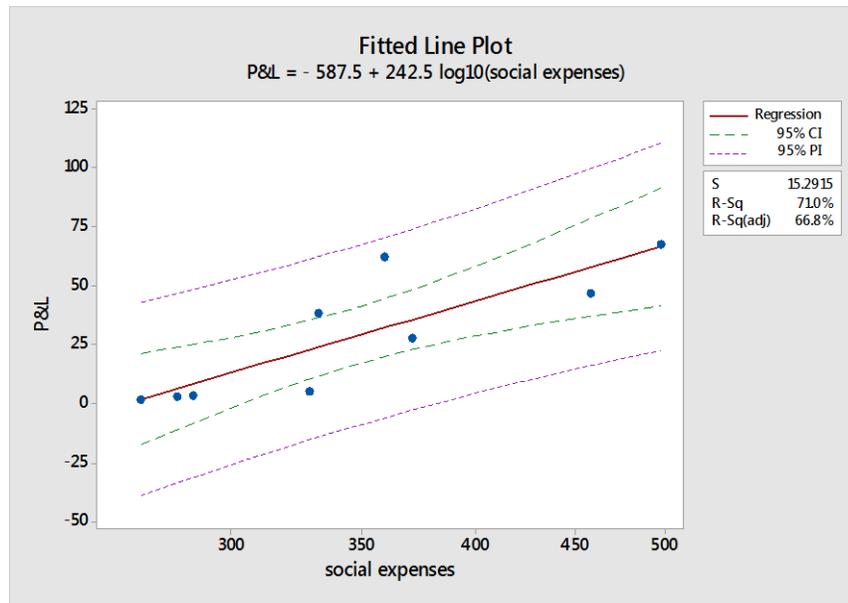
Fig. 2: P&L Regression Using Social Expenses as Input.

These results were built on to model both R&D and social expenses using the formulas shown below as equation 3 and equation 4.

$R\&D = \alpha \cdot X^3 + \beta \cdot X^2 + \gamma \cdot X + c$, Where X is the number of firms, after transformed with Log_{10} (1)

$Social\ Expenses = \alpha \cdot X^3 + \beta \cdot X^2 + \gamma \cdot X + c$, Where X is the number of firms, after transformed with Log_{10} (2)

The results show that both variables are dependent on the total number of firms. However, R&D can be predicted with a cubic function, while social expenses are predicted using linear regression, which explains why the coefficients β and γ are zeros. On the other hand, we can notice different signs for all coefficient when predicting R&D, this indicates that higher number of firms does not necessarily mean a hug R&D investment is needed. while table 2 lists the values of the coefficients:

TABLE II: Coefficient Values for Equations 3 & 4

	A	**β**	**γ**	**C**
R&D	-208854	1158854	-2139427	1314797
Social expenses	1277	0	0	-3670

Finally, to see if both R&D and social expenses can statistically be used for predicting innovation output, the relationship between the two variables was checked using Pearson correlation. A very strong statistical correlation was found, with correlation coefficient =0.914 and P-value = 0.001 (while Spearman correlation rank = 0.983 and P-value of less than 0.001). Based on that, we concluded that these variables can largely replace each other when predicting P&L income.

5. Conclusion and Discussion

Our initial analysis of the data available data indicates strongly that innovation output in a mature high-tech cluster depends both on R&D expenditure as well as social networking within the cluster. It is perhaps surprizing that investing in socializing within a business cluster is almost as important as investing in R&D. These results correlate well with previous theory which modelled the strength of organizations through network links between employees and department [11].

It is interesting to see that amongst the variables used, one can see that the number of firms can be used to predict both variables, and thus can conclude that in a successful business cluster, one hidden variable influencing the innovation output is how many firms there are in a cluster. Clearly this has implications in clusters with large numbers of firms because as more firms inhabit the cluster the 'transaction costs' for communication around networking will increase near-exponentially, meaning that the organizational structure of a cluster can be ad-hoc when it is small, but will become a significant factor as the cluster reaches maturity.

6. Further Research Suggestions

This research can go forward in at least 3 directions. First, it can be proceeded by checking if these results can be reproduced in other clusters, perhaps using "Big Data" techniques. This would also enable more parameters (variables) to be added in order to experiment in more dimensions and indeed a comprehensive multi-dimensional model similar to what was proposed in [12] will help in better understanding how each variable is connected. Second, model the best and most efficient organizational structure(s) for facilitating clusters through their life-cycle. Finally, compare out-cluster firms with in-cluster firms to see if the same influences are present or which others predominate.

7. References

[1] M. E. Porter and M. P. Porter, "Location, Clusters, and the" New" Microeconomics of Competition," *Business Economics,* pp. 7-13, 1998.

[2] A. Cojocaru and S. Ionescu, "The Advantages of Business Clusters*," FAIMA Business & Management Journal*, vol. 4, (2), pp. 31-47, 2016.

[3] J. V. García-Manjón and M. E. Romero-Merino, "Research, development, and firm growth. Empirical evidence from European top R&D spending firms*," Research Policy*, vol. 41, (6), pp. 1084-1092, 2012.

https://doi.org/10.1016/j.respol.2012.03.017

[4] R. Mudambi and T. Swift, "Proactive R&D management and firm growth: a punctuated equilibrium model*," Research Policy,* vol. 40, (3), pp. 429-440, 2011.

https://doi.org/10.1016/j.respol.2010.10.014

[5] R. B. Mellor, "Knowledge valley theory*," International Journal of Knowledge-Based Development,* vol. 2, (5), pp. 5-16, 2014.

https://doi.org/10.1504/IJKBD.2014.059804

[6] R. B. Mellor, "The use of knowledge assets: modelling the potential effect of adding innovators to low-innovation and high-innovation SMEs*," International Journal of Knowledge-Based Development*, vol. 5, (4), pp. 367-380, 2014.

https://doi.org/10.1504/IJKBD.2014.068049

[7] G. G. Bell, "Clusters, networks, and firm innovativeness*," Strategic Manage*. J., vol. 26, (3), pp. 287-295, 2005.

https://doi.org/10.1002/smj.448

[8] M. Porter, "The Economic Performance of Regions*," Reg. Stud*., vol. 37, (6-7), pp. 549-578, 2003.

https://doi.org/10.1080/0034340032000108688

[9] F. Lamperti, R. Mavilia and S. Castellini, "The role of Science Parks: a puzzle of growth, innovation and R&D investments*," Journal of Technology Transfer*, vol. 42, (1), pp. 158-183, 2015.

https://doi.org/10.1007/s10961-015-9455-2

[10] R. B. Mellor, "Modelling the value of external networks for knowledge realisation, innovation, organisational development and efficiency in SMEs," *International Journal of Knowledge-Based Development,* vol. 6, (1), pp. 3-14, 2015.

https://doi.org/10.1504/IJKBD.2015.069447

[11] D. Rumsey, *Statistics ii for dummies*. Hoboken, N.J.: John Wiley & Sons, 2013, pp. 55-83.

[12] R. B. Mellor, *Knowledge Management and Information Systems: Strategies for Growing Organizations*. Basingstoke: Palgrave Macmillan, 2011.

https://doi.org/10.1007/978-1-137-28576-8