Predicting Business Performance through Patent Applications

Daniel Müller, Yiea-Funk Te, Pratiksha Jain
Department of Management, Technology and Economics
ETH Zurich
Zurich, Switzerland
danielmueller@ethz.ch, fte@ethz.ch, 25pratiksha@gmail.com

Abstract— Small and medium sized enterprises represent most businesses and are a crucial part of the economy of all countries. Therefore, the study of factors responsible for the success and the failure of SME is important. In this study, we investigate the role of innovation, a major driver of business success, measured by patent applications. Applying a random forest model to a large sample of Swiss companies, we find a high feature importance of attributes derived from patent applications to predict business performance. We show that patent applications, which are publically available, have a high prediction power in estimating the business performance of small and medium sized enterprises.

Keywords- business performance, public data, random forest, patent application, innovation

I. INTRODUCTION

Small and medium sized enterprises (SMEs) play a crucial role in strengthening the economy [1]–[3]. In Europe, SMEs account for 98% of all enterprises, 58% of GDP and 67% of total employment [4]. From bakeries to precision parts manufacturers operating on the international market, SMEs are vital elements of the economy [2], [5]–[7]. There is broad consensus that SMEs are job creators for any economy [8]–[10]. Thus, SMEs are closely linked to economic well-being of any country [11]. Therefore, the study of factors affecting growth of SMEs becomes the topic of worldwide interest and is also important for the financial stability and robustness of society [11]. Scholars have been trying to develop methods and algorithms for prediction of growth of companies since 1960s [12], but majority of them were focused on large companies (eg. [13], [14], [15]). Only a small number of SME specific extensive models, for example "Lussier model" [16] are known. In the last decades, researchers have tried to prove the correctness of this model [17]. Many other studies for success and failure prediction of SMEs are mainly based on survey data [17]–[19]. Hence, their findings are based on small set of companies due to lower turn around and dependent upon the survey answers which might not be the factual numbers.

Researchers have also shown that SMEs are pillars of innovation [20] and have capability of structural change and market disruption in any economy. From other studies [5], it has been revealed that small firms in the US spend almost twice as much of their R&D dollars in fundamental research compared to the large firms. Innovation in companies is difficult to measure, but one way to examine it, is through patenting behavior [21]. Innovation leads to urge of acquiring intellectual property rights of the invention to benefit from the investment in knowledge creation. Patents grant temporary monopoly for exploitation of knowledge. Through this monopoly, commercial use of invention can be prevented, thereby providing an opportunity of market dominance by selling the invention. Therefore, the impact of patent system upon SMEs is of particular importance.

Researchers [22] have claimed that evaluating patterns of innovation in respect to sales growth is hard at empirical and theoretical level. The research in the field is not extensive, but practitioners and academics alike have in their studies identified growth of revenue as meaningful indicator of post innovation performance [13], [23]–[26]. These studies are mainly done on datasets of large scale and global companies and are industry specific [24], [27]. Such studies are limited for SMEs as their revenue figures are not public and hence not easily available. Empirical researches have not been able to find any strong links between innovation and growth of revenue for SMEs. We aim to address this research gap by capturing the effect of innovation on growth of revenue for small and medium sized enterprises.

We will address these topics with this study of patent and non-patent filing SMEs, all operating in Switzerland. We aim to apply data analysis and machine learning on a Swiss SME dataset to explain the growth of companies with patent data and derived gender data of patent applications. If current growth numbers could be explained by the patent data or are correlated to patent data, future growth will eventually also have similar dependence on patents. By explaining the historic growth of revenues, we want to investigate, whether patent data could be used as a feature for the prediction of future growth of companies.

II. RELATED WORK

Many researchers have investigated the relation between innovation and business performance [18], [28]–[31]. It is also widely shown that innovative companies respond better and faster to challenges and changes, and compete most effectively in the market [32]. They can exploit new products and market opportunities better than non-innovating
companies [32]. Executives overwhelmingly say that innovation is what their companies “need most for growth.” [33]. Further, authors of [18] showed that organizational size strengthens the positive relationship between innovation and performance within firms. Baker and Sinkula [34] showed the effect of market orientation and business performance. They explain how capabilities lead to superior products through innovation capabilities. Mowery [24] used a dataset of US manufacturing firms and found that R&D expenditures have positive impact on firm’s growth. Law [27] researched with the help of a dataset of steel and petroleum industry over a 40 year period, that innovators grew quickly. Freel [25], in its study of 209 leading UK firms, showed that innovating companies grew faster and have larger profits. However, these studies either exclusively address large companies or are specific to a particular industry. Previous literature investigating innovation and growth for SMEs rely on survey data and research in terms of process and product release [17]–[19].

We find that the literature addressing the role of patents in growth of small and medium companies is underdeveloped, mainly due to the paucity of factual datasets and requires more research.

In our study, we investigate the role of patents in SME growth. To study the impact of patents on SME business performance, we answer in the following two research questions:

R1: Is there any difference in the growth rates of SMEs with and without patents?

R2: Which patents application attributes are useful to predict business performance of SMEs?

We aim to study the role of patents as a public data source to predict the future business performance of their SME client’s. Answering these research questions holds importance for business to business companies, especially though for financial institutions that benefit from understanding which of their business customers will outperform the others, hence eventually directly or indirectly benefit them. This is due to the fact, that most financial institutions (i.e. Banks and insurances) own revenues correlate to some extent with the volume of the business of their customers.

III. METHODOLOGY

A. Data set

We created our population sample from a set of general liability insurance policy data. A Swiss insurance company that sells insurance policies all over Switzerland, provided the dataset used in this research and includes policy information from 2010-2016. The dataset of SMEs includes information such as company name, unique id, industry type, annual revenue and incorporation year. The dataset is very diverse in size, group, ownership, industry type, location and is representative for Switzerland. For integrity of the data, we have excluded the firms which have multiple policies. We have removed the outliers in revenue numbers by using the quantile function to truncate values at 1st and 99th percentile in the data [28]. We excluded the revenue numbers of 2016 to use them at a later point as label for growth prediction. To compare the revenues of each firm from the scale of 1, we divided the revenues of each firm by their corresponding revenue of the year 2010. We then combined

Therefore, patents should capture continuous innovation and should reflect its existence through growth of SMEs.
this data with federal statistic office datasets of Switzerland (SHAB data, BFS data [45]) to add more features.

We matched this data set with a patent database that covers all patents granted in Switzerland (the Lens [46], our data set two). As the linking attribute, we used the policy holder name from dataset one and matched it with the applicant name from data set two. By matching, we identified those Swiss SMEs which have filed at least one patent after 2009.

**B. Data preprocessing**

We have categorized the industries of SMEs in our dataset in four broad categories that are Manufacturing, Service, Trade and Real Estate. These features will later be used for our prediction model. Figure 1 shows the distribution of each of the four industries in dataset. 1,625 SMEs belong to the service industry, 1,067 to real estate, 530 to trade and 881 to manufacturing industry.

In order to investigate research question 1, we separated those companies with patents from those without. For the two different groups. Further, we classified the SME into 3 classes:

- SME with 0 patents as class zero
- 1-2 patents as class one
- >2 patents as class two

Our final dataset has 4,104 companies in which 3,618 belong to class 0, 376 belong to class 1 and 109 belong to class 2. We call this field SME class. We further calculate patent count, number of cites, count of male and female inventors as features from patent data to investigate their significance in predicting business growth between 2010 to 2016.

Additionally we assigned the attribute gender to the first names of the patents applicants and counted the males and females names for each applications. Our first name to gender mapping algorithm, previously performed with a 96.10% of F1-score on test set of 20,000 instances on another set of Swiss names and was trained on more than 1.6 million Swiss names. Applying the first name to gender mapping, we found that 79.2% of all patents are filed by males only. 8% of them are filed by females only and 11% of them are filed by mixed teams. These numbers are identical to the findings published by Statista Switzerland [47].

**C. Prediction and Evaluation**

We have divided our dataset in a training set (75%) and a test set (25%). In order to predict business growths, we report the result of a random forest with 500 trees, a maximum depth of the trees of 10 and the parameter $r=\sqrt{n}$, with $n =$ Total features available. We compute the performance of our model by computing RMSE and R-squared values on the test set.

**IV. ANALYSIS AND RESULTS**

In order to investigate the importance of patents, we show the change in revenue, indexed to 1 for the year 2010 of SMEs and the consecutive 5 years.

Figure 2 shows that the growth of SME with and without patents, as well as the average growth of all SMEs (Grand Total in color blue) experienced on average 60% and 48% growth accumulated over 5 years. The groups zero, representing SMEs without patent applications on the other hand experience almost no growth (3% accumulated over 5 years, indicated by flat yellow line). In order to validate the statistical significance of the average growth values of the 3 classes, we performed a t-test on all three possible combination as well as an analysis of variance (ANOVA). ANOVA is a hypothesis testing statistical technique used to compare the means of more than 2 groups. We have used a confidence interval of 0.05 for the test. Both tests show a significant difference between the patent groups and the non-patent groups. Further applying a t-test on all three possible combination of SME classes, we find that the t-tests between class zero and class two show a p value less than 0.05.

Further investigating the role of patent application, we group SMEs with patents by gender composition of the inventors mentioned on the application. We find that SMEs with mixed inventor teams (both males and females) and female only inventor teams outperform SMEs with male only inventors. We show the indexed performance of male only, female only, mixed teams as well as the grand total of all SMSs in our sample between 2010-2015 in Figure 3.
We validated our findings for statistical significance of the three groups over all the years and found that there was no significant difference between the female inventor companies and mixed gender companies in any year. There was significant difference between mixed gender and male inventor companies in the years 2014 (p value=0.0381 T value=2.764) and 2015 (p value=0.0491 T value=2.034). However, there was also no significant difference in the growth of male only and female only patent inventors.

A. Prediction results

To evaluate the predictive power of patents, we train a prediction model on 3,104 companies with and without the patent application derived features. First, we build a random forest model capturing the following features:

- Industry
- Area income
- City to rural index of area
- Category of area (small, large)
- Type of company (AG/GmbH)
- Nominal capital of company
- Category of company according to age (new, old, medium aged)
- Number of years since foundation

In order to investigate the power of the patent application features, we build a second random forest model by adding patent features such as Patent count, Number of male inventors, Number of female inventors, Number of citations, Number of male inventors and Number of female inventors.

Both random forests had same hyper parameters except for the number of features. We then test our models on test set of 1,000 companies. In order to compare the models performance, we show their R-square and RMSE values on the test set, which can be found below in Table I.

To evaluate which features from the patent applications contribute most to the models performance, we further plot the features importance of our (improved second) random forest model. We find that the IncNodePurity in random forests, which is the total decrease in node impurities from splitting on the variable, averaged over all trees has a group like structure of features. From the patents related features, we find that Cited_Count, Patent_Count, Male_Inventors, Female_Inventors and Zero (SMEs with 0 patents) to be important. Due to brevity we do not comment whether the value of the features for the trees has a positive or negative impact. However we rank the features. The IncNodePurity of features from model with and without patents can be found in Figure 4 and Figure 5.

### TABLE I. PERFORMANCE OF RANDOM FOREST MODELS

<table>
<thead>
<tr>
<th>Results</th>
<th>Random Forest (without patent features)</th>
<th>Random Forest (with patent features)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
<td>.527</td>
<td>.838</td>
</tr>
<tr>
<td>RMSE</td>
<td>.26</td>
<td>.18</td>
</tr>
</tbody>
</table>

**Figure 3. SME growth by patent application gender**

**Figure 4. IncNodePurity of model without patents data**

**Figure 5. IncNodePurity of model with patents data**
V. DISCUSSION AND IMPLICATIONS

Validated by our statistical test, we can see that SMEs with one or more patents perform better in terms of growth than SMEs with 0 patents. We could also infer that there is no significant difference in growth between SMEs with one patent application or several patent applications. Further, based on the applicant’s name derived data points, which we examined in Section IV, we see that mixed gender inventing SMEs and female only inventing SMEs have higher growth in revenues as compared to male only inventing SMEs. Examining the different groups over a period of 5 years, we found that after 4 years of operations (after a patent application), the groups start to grow at different rates. Comparing the group’s index growth performance, we observe a significant difference in growth for the mixed gender teams (higher growth) compared to the male only teams. From this, we imply SMEs with protectable IP in Switzerland, over the long run, grow faster than their peers which do not file patent applications.

To further evaluate the prediction power of our patent application features, we apply two random forests, trained on 75% of the data and tested on the remaining 25% of the sample. We find that adding patent features greatly improves model performance by increasing the R-square by 0.311.

VI. CONCLUSION

We have studied the importance of patents for small and medium sized companies. We took the data of Swiss SMEs for our study. We identified 496 SMEs which have patents from our dataset of over 106,000 companies. This indicates that a large population of SMEs do have IP, which required protection of decide against patents as the only strategy for appropriation. However, in our study, we investigated the importance of patents for growth of revenues. We compared the averages of growth of revenues in span of 6 years (2010-2015) for three groups of companies that are companies without patents, companies with one or two patents and companies with more than two patents. It has been evident that overall companies with patents perform better in terms of growth of revenue than companies without patents.

Even though not reported in detail, we further mentioned the role of attributes such as Industry, Category of area (small, large), Legal form (AG, GmbH) and Company age category. We were not able to explain the changes in growth through these variables and assume no moderating effect between the patent application attributes and business growth, which we however did not explicitly test.

Further, we addressed the feature importance of patent data for predicting the growth of companies. We applied supervised learning techniques of machine learning such as, random forests to see the importance of patent data. We predicted change in revenue of 2016 from our dataset. We have not used this index revenue of 2016 throughout our analysis of growth behavior so that we do not gain any prior knowledge of this label. We see that patent features are important for growth prediction of companies, which further indicates and strengths the importance of patents for growth of small and medium sized companies.

Our study results bring value by contributing to the discussion about which publically available data sources can be gathered and analyzed to categorize companies by expected future growth. Further, our study findings are of interest for executives of business to business companies such as financial institutions, which benefit by knowing which of their customers are high growing companies, low growing companies and not growing companies. Correctly estimating expected growth rates, could eventually help them prioritize among their customers and divert resources to the most promising ones, i.e. the ones with most growth.

Since our results are based on only Swiss SMEs, we propose similar studies across other countries to validate if our results hold for SMEs across nations.

Despite having interesting results, our study has few limitations. Our study is not able to explore direct impact of patents on growth numbers as there is large time lag in converting of patent into a commercially valuable product. Also, there is considerable time lag between product release and product success which our study could not explore. Additionally, patents are not sole means of innovation. SME research and development expenditures could be important feature for determining the importance of innovation on revenue of SMEs. Our dataset has not taken research and development expenditures of SMEs into account. In respect to the gender attribute, we only had access to a small and possibly biased dataset of 41 female only SMEs and 30 mixed team inventing.

We therefore encourage other researchers to replication our study using a larger dataset for other countries and perform a correlation analysis of the time of patent application and the commercial success of a product or service related to the application.

REFERENCES
