Modelling credit risk for innovative firms: the role of innovation measures

Chiara Pederzoli♠♣, GridThoma♦♣, Costanza Torricelli★♣

♠ University of Milano-Bicocca
♦ University of Camerino
★ University of Modena and Reggio Emilia
♣ CEFIN, Centro Studi Banca e Finanza, Modena

Corresponding address: grid05@gmail.com

Abstract

Financial constraints are particularly severe for R&D projects of SMEs, which cannot generally rely on equity markets and, in the EU, on a sufficiently developed VC industry. If innovative SMEs have to depend on banks to finance their R&D projects, it is particularly important to develop models able to estimate their probability of default (PD) in consideration of their peculiar features. Based on the signaling value of some innovative assets, the purpose of this paper is to show the importance to include them into models which have proved to be successful for SMEs. To this end, we take a logit model and test it on a unique dataset of innovative SMEs (based on PATSTAT database, EPO BULLETIN and AMADEUS) to estimate a two-year PD with default years 2006-2008. In the regression analysis the innovation-related variables are two in order to account for R&D productivity at the level of the firm and to consider the value of the inventive output. Our analyses first address measurement issues concerning innovation-related variable and then show that, while the accounting variables and the patent value are always significant with the expected sign, the patent number per se reduces the PD only in the presence of an appropriate equity level.

1. Introduction

The advent and the fast growth of the knowledge economy and the parallel development of science-based industries (e.g. biotechnology, software) have been accompanied by the emergence and success of innovative start-ups, which in many instances have outperformed incumbent firms. Examples include Microsoft in operating systems, Google and Yahoo in web applications, Amgen and Genentech in biotechnology, Echelon in automation and many others. This evidence could be directly related with the higher experimentation and innovation propensity of small firms vis-à-vis large ones [6, 5, 31].

Innovative firms, independently of their size, face financial constraints as stressed by a broad literature, which has given special attention to the role of equity finance (Brown et al., 2009). Hall [23] concludes that "… the capital structure of R&D-intensive firms customarily exhibits considerably less leverage than that of other firms". The issue has been extensively surveyed in Hall and Lerner [26], who claimed that financial constraints are fuelled by information asymmetries between inventor/entrepreneur and investor. In particular, these asymmetries regard the fact that an inventor has a better understanding on the potential success and structure of the R&D project, and thus, the marketplace for financing the innovative assets could be characterized by a typical "lemons market problem".

Financial constraints are particularly severe for R&D projects generated by SMEs [23, 26], which - as it is the case also for non innovative firms – do not normally rely on equity markets. However, innovative SMEs encounter a stronger adverse selection in the credit market [31]: since new innovators are corresponded by a financial distress in an early stage, they face comparatively higher interest rates and reduced credit availability, which in turn have an effect on their financial performance. This hampering mechanism is even more pronounced in sectors and/or countries where there is a lack of a sufficiently developed VC industry, such as the case of EU countries [15]. With the exception of UK, the continental EU countries show very low intensity of VC investments relatively to their GDP compared to USA, Israel, Canada and Switzerland [41].

In sum, innovative SMEs add to the well-known financing difficulties of "traditional" SMEs, the above-mentioned problems typical of innovative firms thus encountering peculiar difficulties in financing their activities. If innovative SMEs have to rely on banks to finance their R&D projects, it is particularly important to develop models able to estimate their probability of default (PD) in consideration of their specific features.[1] The empirical literature on SMEs default prediction has proved, over different period and different datasets, the good performance of logit/probit models[2] and, despite some differences among various research works, a convergence emerges on five categories of financial indicators (leverage, liquidity, profitability, coverage, and activity) and the importance of qualitative variables is also recognized [20].

The purpose of this paper is to show that the credit quality of innovative firms cannot be appropriately gauged only on the basis of indicators related to balance-sheet variables. Based on the signaling value of some innovative assets such as patents, we believe it is important to include them into models which have proved to be successful for SMEs. Previous

---

[1]The issue is relevant also in terms of capital regulation, given that Basel II recognizes a different treatment for the exposures towards SMEs, which benefit from a reduction in the capital requirement proportional to their size.

[2] For a broader discussion of the issue, see [4].

research in entrepreneurial finance has claimed that that patents can constitute rich information source for financial investors in assessing the quality of innovative firms.[3] Recently, Hsu and Ziedonis [35] show that patents improve the terms by which new firms access venture capital. In particular, they document that the larger the patent portfolio of start-ups, the bigger the money evaluation by VCs and that this effect is even more pronounced for younger and inexperienced firms. In the same vein, Harhoff et al. [33] demonstrate similar findings and they argued that the granting decision by the patent office does not trigger additional financial evaluation from VCs because this event is fully anticipated thanks to information indicators revealed in the patent application (e.g. such as patent citations).

In this paper we take a logit model and test it on a unique and novel dataset of innovative SMEs (based on PATSTAT database, EPO BULLETIN and AMADEUS[4]) to assess, beside standard accounting variables, the fundamental role played by knowledge-related variables as proxied by patent and patent indicators in predicting a default probability (PD) with default years 2006-2008. In the regression analysis the knowledge- or innovation-related variables are two in order to consider the value of the patent portfolio of companies and to account for R&D productivity at the level of the firm.

The paper is organized as follows. Section 2 describes the dataset. Section 3 illustrates the default prediction model used, while Section 4 discusses the various issues connected with the measurement of the knowledge-related regressors. Section 5 presents the results obtained and the last Section concludes.

2. Data description: the sample of innovative firms

The first issue to be addressed in the construction of the dataset is the definition of innovative firms. To this end, we use patent data: while not all inventions are patented, patenting activities have increased significantly in the last decade in terms of larger company patent portfolios and larger share of firms applying for patents in many different technologies [41]. On the other hand, patents can be considered a highly objective data source over time and they provide very detailed information regarding the invention and its inventors [19].

To define the set of innovative companies in this paper we include all European firms that have filed at least one patent application in the EPO and PCT/WIPO system.[5] We decided to limit our analysis only on these two patent systems in order to take into account only the most relevant patent inventions by a firm and to achieve higher homogeneity across patent measures. In fact, patents document varies significantly in terms of their economic value from one legislation to another. In 2003 for example the average cost of obtaining a standard patent was estimated at EUR 30,530 for EPO, EUR 46,700 for PCT and only EUR 10,250 for USPTO and EUR 5,460 to acquire a JPO grant [41].

The data source is the PATSTAT database (version April 2009) and EPO BULLETIN (version December 2009).[6] In particular, our database covers all patent document publications –applications and grants – since the inception of EPO and PCT/WIPO system up to Dec 31st, 2009. Then, relying on the AMADEUS business directory we integrate the patent owner names with demographic and accounting information, such as sector activity, ownership, balance sheet, profit and loss account. Our methodology builds over a previous contribution by Thoma et al. [48], who developed a complex matching algorithm to merge extensive company information. Our dataset relies on the overall population of patent owners, which allows to overcome any selection bias limitation.

Given the focus on SMEs, the second issue concerns the definition adopted to identify this category of firms. The definition given by the European Union refers both to the number of employees and to sales: firms are considered small if they have less than euro 50 million in sales or less than 250 employees.[7] The Basel Committee on Banking Supervision [7], for the purpose of capital requirements, imposes a criterion based on sales only to discriminate between SMEs and corporates: firms with annual sales less than 50 million euros are considered SMEs and this imply for the intermediary a reduction in capital requirement proportional to the firm's size.[8] In our sample, we have included firms with turnover in the range of 1-50 million Euros, whereas the geographical context has regarded EU15 countries, Switzerland and Norway.[9] Thus, consistency with the Basel II definition, allows to use the estimated PDs as input in the Basel II capital requirement formula.

[3] For a survey on developments of entrepreneurial finance see [13].

[4] See Thoma et al. [48] for more details.

[5] EPO is the acronym for European Patent Office, whereas PCT/WIPO for Patent Cooperation Treaty/World International Patent Office. For information on these patent systems see [21].

[6] Both the PATSTAT and BULLETIN database are available to any user under request from the EPO. The data have been managed by with SQL and STATA software toolboxes. For more details on this task see [48].

[7] Commission Recommendation 96/280/EC of April 3, 1996, updated in 2003/361/EC of May 6, 2003. See http://europa.eu/scadplus/leg/en/lvb/n26026.htm.

[8] The reduction applies to the capital function through the correlation, which is reduced by a maximum of 0.04 for the smallest firms. This correction is justified by the assumption that defaults of small firms are less correlated and therefore less risky on the whole for the portfolio.

[9] The turnover is given by the sum of sales and net stocks of the reference year. In the present analysis we use turnover and not sales because the AMADEUS does not report information on sales for some countries such as UK, Ireland and Denmark.

A further important issue is the definition of default to be used in the classification. In order to classify defaulted firms in our sample, we need first of all to adopt a definition of default, since literature does not provide a univocal one. Altman and Hotchkiss [3] stress that four terms - *failure, insolvency, default* and *bankruptcy* - are used interchangeably in the literature but have different meaning and refer to different situations in different countries' bankruptcy law. The BCBS [7] adopts a wide default definition in that "a default" is considered to have occurred with regard to a particular obligor when either or both of the two following events have taken place:

- The bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realizing security (if held).

- The obligor is past due more than 90 days on any material credit obligation to the banking group overdrafts will be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than current outstandings.

Often default definitions for credit risk models concern single loan defaults of a company versus a bank, as also emerges from the above Basel II instructions. This is the case for banks building models based on their portfolio data, that is relying on single loans data which are not public (for example Altman and Sabato [2] develop a logit model for Italian SMEs based on the portfolio of a large Italian bank). However, traditional structural models (i.e. Merton-type models) refer to a firm-based definition of default: a firm defaults when the value of the assets is lower than the value of the liabilities, that is when equity is negative.

In this work, we identify a firm's legal status according to the following taxonomy:

i) Active: if a company is currently performing economic activities;

ii) Inactive: if a company has not been performing economic activities in the last three years;

iii) Bankrupted: (a) unable to pay the creditors; (b) the assets are held by a receiver; (c) assets and property of the company redistributed;

iv) Dissolved: when the legal life of company has come to an end;

v) Merged-demerged-acquired: whether a firm has been merged with another company, acquired or split in more than one other company;

vi) Unknown, firms with unavailable legal status.

Consistently with previous studies [4] we include only firms with a legal status active or bankrupt. The reason for this choice lies in the data availability but it is also motivated by the objective of the paper: our aim is to define a model, based on public and accessible data, that measures the health state of the firms and enables any economic subject interested in a specific firm's health (i.e. suppliers, customers, lenders, etc.) to estimate the probability of a particular firm to get bankrupted.

In line with previous literature we adopted a reduced sample approach with a ratio of bankrupt firms of 6% of the overall sample. This rate is the sample default rate before cleaning and is in line with the one assumed by Altman and Sabato [4]. To build our sample, first we start with all the bankrupt with available information on profit/loss and balance sheet accounts in five macro business activities: low tech process industries (US SIC 10-33), chemicals and pharmaceuticals (US SIC 28), manufacturing (US SIC 34-39), distribution (US SIC 50-60) and services (US SIC 70-99).

Second, we randomly select firms with active legal status up to 94% of the sample and in order to obtain a full independence of the observations we adopt a sampling strategy without replacement.

Finally, we adopt a pooled cross section logit model – as described in the next Section – to estimate a PD with default years 2006-2008 that correspond respectively to fiscal years 2004-2006. The final sample consists of 2,665 firms, whereby 160 are classified as default and 2,505 are active companies according to the AMADEUS business directory.

Table 1 reports the distribution of firms by cohort and macro industry. To be noticed that about 2/3 of our firms originate from manufacturing and low tech industries, whereas Services account for only 13.7%; the age distribution of the firms is relatively old, with about 70% of the firms incorporated before year 1990.

[Table 1 about here]

3. The default prediction model

There is a wide range of default prediction models, i.e. models that assign a probability of failure or a credit score to firms over a given time horizon. The literature on this topic has developed especially in connection with Basel II, which allows banks to set up an internal rating system, that is, a system to assign ratings to the obligors and to quantify the associate PDs [7]. However, some sophisticated models available in the literature can be used only if market data on stocks (structural models) or corporate bonds and asset swaps (reduced-form models) are available. As for SMEs, for which market data are generally not available, either heuristic (e.g. neural network) or statistical models can be applied.

Beaver [8] and Altman [1] first used discriminant analysis (DA) to predict default. In order to overcome the limits inherent in DA (e.g. strong hypotheses on explanatory variables, equal variance-covariance matrix for failed and not failed firms), logit and probit models have been widely adopted. An important advantage of the latter models is the immediate interpretation of the output as a default probability. A seminal paper in this respect is the one by Ohlson [43], who analysed a dataset of US firms over the years 1970-1976 and estimated a logit model with nine financial ratios as regressors. More recently Beaver [9], by analyzing a dataset of US firms over the period 1962-2002, has shown that balance-sheet financial ratios still preserve their predictive ability, even if market-based variables partly encompass accounting data.

Focusing on SMEs, a few recent works use logit/probit models, or some evolution of the same, for the PD estimation: Altman and Sabato [4] use a dataset of US SMEs, Altman and Sabato [2] analyse separately US, Australian and Italian SMEs, Behr and Güttler [10] and Fantazzini and Figini [16] analyse German data, Fidrmuc and Heinz [17] use data from Slovakia, and Pederzoli and Torricelli [45] focus on the Italian case. Despite some differences among these analyses, a convergence emerges on some types of financial indicators, which can be grouped into five categories: leverage, liquidity, profitability, coverage, and activity.

Thus, in line with most of the literature, we use a binary logistic regression model where the default probability is estimated by following equation:

$$PD_i = P(Y_{i,t+1} = 1) = \frac{\exp(\alpha + \sum_{k=1}^{R} \beta_k X_{ik,t})}{1 + \exp(\alpha + \sum_{k=1}^{R} \beta_k X_{ik,t})}$$

where:

$$Y_{i,t+1 \quad i=1,\dots,n} = \begin{cases} 1 \; if \; obligor \; i \; defaults \; in \; t+1 \\ 0 \; if \; obligor \; i \; does \; not \; defaults \; in \; t+1 \end{cases}$$

$$X_{ik,t \quad i=1,\dots,n} = k^{th} \; regressor \; for \; obligor \; i \; in \; t$$

We quantify the dependent variable according to the definition of default given in Section 2, while we consider both balance sheet variables and knowledge related variables as regressors.

In the case under investigation in this paper, i.e. innovative SMEs, one issue is still the selection of appropriate and informative balance sheet variables, but the main one is the definition and the measurement of the knowledge related ones. While the former is tackled by means of a backward elimination procedure based on the Schwartz Information Criterion (SIC) and is illustrated in Section 5, the latter requires the discussion of many issues as outlined in the following Section.

4. Innovation-related variables

An increasing number of studies have used patent counts and patent-related indicators to measure the quantity and the 'value' of inventive output. Several studies have shown that patent counts are strongly correlated to size of innovative investments typically measured by R&D ([19],[44]). However, crude patent counts are a biased indicator of inventive output because they do not account for differences in the value of patented inventions. This is the reason why innovation scholars have introduced four main patent-related indicators as a measure of the 'value' of the inventive output ([33],[27]).

First, the number of inventors of a patent has been associated to the economic and technological value of patents. In fact, the technical value of an invention is related to the research cost of the underlining R&D project, which is made up in large part of wage bills for the human resources involved in the project [34]. In this direction, the more inventors in a patent, the more research-intensive and expensive the R&D project ([22], [18]).

A second indicator of the value of patents is given by the geographical scope of patent protection; that is, with the number of national and international offices in which a patent document has been applied. Typically, the international patent protection requires additional filling costs and this decision by the owner signals a higher expectation of economic value related to the invention ([38], [29]).

Third, the number of citations received (henceforth also forward citations) is widely used as indicator of patent value ([32], [49]). So far the previous literature has provided two main explanations. On the one hand they demonstrate the cumulativeness of a given technology, suggesting additional R&D being performed and hence market potential. On the other hand, since citations reveal a knowledge transfer process, they show that a technology is being used and hence it is valuable.

Fourth, the number of technological classes has been shown to be an indicator of technological "value" similar to the number of citations by Lerner (1994). In particular the number of International Patent Classifications (IPC) codes can be viewed as a measure of technological scope or generality of the patent. To guarantee a reasonable level of precision, we use the number of eight-digits IPC classification codes reported in the patent document.

In the empirical analysis of the present paper, the innovation-related variables are two. First, we account for R&D productivity at the level of the firm by computing the ratio of patents divided by number of distinct inventors employed by the firm over a five years period. Second, to consider the value of the inventive output we built a multidimensional factor index according to the methodology explained in the next section.

4.1. Patent value factor index

Three indicators – family size, citation and IPC technical classes– were combined into a composite index of patent 'value' derived from a common factor model based on an approach similar to Lanjouw and Schankerman (2004) and Hall et al. (2007). In factor models each series of data (quality indicator in this case) is broken down into a common component and an idiosyncratic component. Given N indicators, the common component is driven by only a few common shocks, Q with $Q < N$. In a static factor model, the common shocks affect the indicators only contemporaneously. The basic model is given by:

$$\boldsymbol{X} = \boldsymbol{UB} + \boldsymbol{E} = \boldsymbol{K} + \boldsymbol{E}$$

where $\boldsymbol{X}$ is the $(T \times N)$ matrix of observations on the N indicators whose columns are series of length T-normalized to have mean 0 and variance 1, $\boldsymbol{U}$ is the $(T \times Q)$ matrix of Q common shocks and $\boldsymbol{B}$ is the $(Q \times N)$ matrix of factor loadings, which determines the impact of common shock q on series n. The common shocks and the factor loadings together make up the common component $\boldsymbol{K}$. After the influence of common shocks has been removed, only the idiosyncratic component $\boldsymbol{E}$ remains. To estimate the common component we have to find a linear combination of the indicators in $\boldsymbol{X}$ that explains as much as possible the total variance of each indicator, minimizing the idiosyncratic component (for a technical discussion of factor models see Jolliffe, 2002).

The parallel with least squares estimation is clear from this formulation, but the fact that the common shocks are unobserved complicates the problem and the standard way to extract the common component in the static case is principal component analysis. In principal component analysis the first Q eigenvalues and eigenvectors are calculated from the variance-covariance matrix of the dataset $\boldsymbol{X}$. The common component is then defined as: $\boldsymbol{K= XVV'}$, with $V = [p_1,\ldots,p_Q]$ and where p_i is the eigenvector corresponding to the ith largest ($i = 1 \ldots Q$) eigenvalue of the covariance matrix of $\boldsymbol{X}$. This method does not ensure a unique solution. A further problem is that *ex ante* it is not known how many common shocks Q affect the series in $\boldsymbol{X}$. Following the approach suggested by Lanjouw and Schankerman [38], we use a multiple-indicator model with an unobserved common factor:

$$y_{k\,i}= \lambda_k q_i + \beta'\boldsymbol{X} + e_{ki}$$

where y_{ki} indicates the value of the kth patent indicator for the ith patent; q is the common factor with factor loadings λ_k and normal distribution, and $\boldsymbol{X}$ is a set of controls. The main underlying assumption is that the variability of each patent indicator in the sample may be generated by the variability of a common factor across all the indicators and an idiosyncratic part e_k not related to the other indicators with distribution $N(0,\sigma^2_k)$.

In our setting, the common factor is the unobserved characteristic of a patent that positively influences three "value" indicators. Estimation of common quality index q is based on information extrapolated from the covariance matrix of our four indicators. By assuming the normality of q_i and e_k we can estimate by maximum likelihood, which ensures a unique solution. Once the estimates of λ_k are obtained, the model is inverted to calculate q.

4.2. The depreciation problem

One key aspect of knowledge is its cumulativeness. This means that the knowledge assets of a firm strongly depend on previous vintages of other knowledge. However, knowledge depreciates too. As time goes by new technologies are invented that improve the pre-existent ones or in some cases supply radical new technical solutions that replace the others. The pace of this process is more fierce in some areas than others. For example, in the last years the rate of technical change has been considered very fast in software and other ICTs related industries.

In the literature to account for time dimension of the knowledge accumulation process previous contributions have adopted conventional declining balance formula using a directly comparable relation with ordinary investment and capital:

$$K_t = R_t + (1-\delta)K_{t-1}$$

Where K_t is stock of knowledge at time t, R_t the production of knowledge between t-1 and t, and δ is the depreciation rate. Although a variety of choices for the depreciation rate have been explored in the past, the choice makes little difference for estimation, and most of previous works use the 15 per cent (see for a survey [25] and [28]).

For R&D investments or personnel, typically the starting stock is calculated for each firm at the first available R&D observation year as $K_o= R_o/(\delta+g)$, where g is a conventional growth rate and approximated with 8 per cent. This assumes that real R&D has been growing at a constant annual growth prior to the sample.

Similarly, patent related variables are obtained using the same method. However, given the longer pre-sample history of patenting (back to 1970s) than for R&D the impact of the initial stock is minimal and thus the initial available patent counts are often not discounted to obtain an initial capital stock.

4.3. Data censoring and other measurement issues

Patent data suffer several truncation issues. First, EPO/PCT patent application information are available only with a delay in time. A patent application is generally published 18 months after it has been filed, whereas the time lag between filing and grant or refusal of patents is not fixed. In our analysis to overcome this end of sample bias we considered all the patent applications and not just grants.

Second, the filling date cannot always be defined as the closest recorded date to the invention activity if the EPO/PCT patent application is secondary filling of a priority patent from a national office – and typically this is the case. Hence, we considered as reference year for the patent information the priority year rather the application year.

A third censoring problem regards the patent value indicators. In particular, forward citations to a patent take place over a very long period of time. For example Hall et al. (2007) have documented that – over a reference period of three decades – on average 2/3 of the all citations received by a patent take place within the first decade whereas in the first three and five years only about 20% and 40% respectively. Based on this evidence and given that our firm sample regards fiscal year 2004-2006, we opted to count the forward citations only those taken place after three years from the priority date in order to achieve a homogenous measure across years.

Another measurement problem of the patent value indicators concerns the different statistical structure across technologies. For example the citations cumulate more slowly in Chemicals rather than Electronics, because the pace of technical change is faster in the latter technologies. In turn family size in globalized industries such as Pharmaceuticals is higher than Mechanicals. Similarly, number of IPC classes is more numerous in general purpose technologies such as ICTs rather than in Consumer goods. In the literature to correct for this bias several statistical procedures have been advanced (Hall et al., 2001), whereby one of the most frequent approach is the detrending by time and technology fixed effects. In this work we scaled our three indicators by the geometric averages computed by reference year and technology groups.[10] The future developments of this project will consider also other approaches to correct for time and technology fixed effects such as the semistructrural model discussed in Hall et al. [25].

5. Model estimation and results

The financial variables included in the model have been selected according to a back stepwise procedure. We select among the same candidate predictors proposed in Altman and Sabato [4] by means of a backward elimination procedure based on the Schwartz Information Criterion (SIC).

Table 2 summarizes the list of variables selected. As for the accounting variables, they are meant to describe the five main features of a firm's profile as recalled in Section 3, i.e. leverage (EQ_RAT), liquidity (LIQ_RAT), profitability (PROF_RAT and EX_RAT), coverage (COV_RAT) and activity (SALES). The sign expected from the logistic regression is naturally negative for all.

Regarding the patent variables we adopt a specification similar to Hall et al [27] who proxy the knowledge assets of a firm according to two dimensions as explained in Section 4: the patent value yield given by the patent citations over the patent stocks and the patent productivity as measured by the capitalized patent counts standardized by R&D stocks. While a higher patent value should decrease the PD so that the its coefficient in the logistic regression is expected to be negative, the patent productivity alone may imply a higher firm's riskiness and thus there is no strong a-priori about the sign of its coefficient.

Due to the lack of information on the R&D expenditures on SMEs we measured the patent productivity as the ratio of patents counts divided by the number of active inventors over a 5 years window. Previous research [34] has shown that R&D investments are made up of for about 70% of labor costs – typically wage bills for the R&D personnel– and the remaining part is highly correlated with the size of the R&D personnel. Moreover, we think that for the SME case this measure is more suitable than the one based on R&D investments because even when R&D investment is reported in the P&L account it may underestimate the actual intensity of innovation activities. Indeed, in the case of SMEs, R&D activities are not formalized in structured labs and typically R&D costs are mixed with labor costs and/or with other fixed costs when R&D is outsourced.

As for the patent value yield, we complement the citation counts with the patent value factor index described in Section 4.1. Again this indicator is a ratio of the capitalized factor index with capitalized patent counts. Both the patent productivity and patent value yield have been aggregated by year at the firm level. Table 2 summarizes the variables of the econometric model.

Then we analyzed the descriptive statistics of the variables for the whole sample, by industry and cohort. On the whole the data highlight that profitability is highest in low tech industries and manufacturing which are characterized by a lower level of equity and higher bank financing compared to services. Since services typically have no or very limited collateral value, external financing is low, and service firms are additionally characterized by a stronger liquidity also because the payment due dates are shorter. For the same reason we observe a similar pattern when we compared young vis-à-vis old firms. Regarding the innovative variables, as expected, we can notice that chemicals and pharmaceuticals are the industries with the bigger patent value, whereas the patent productivity is lower than other industries. This evidence is consistent with previous findings [18] who have claimed that the extreme tale of the patent value distribution is populated by chemicals and pharmaceuticals patents. On the other hand patent productivity in this industry has been declining since the begging of the 1990s and rising significantly in manufacturing [38]. However, the bigger patent productivity in manufacturing has not been attributed to a real burst of innovative activities but to legislative changes in the patent system [50] such as: the introduction of the Court of Appeals of Federal Circuit in the US; the extension of patenting to new subject areas such as biotechnology, software, business methods at the USPTO which in turn influenced directly or indirectly filings at the EPO and PCT; and the introduction of the TRIPS agreements in the World Trade Organization member countries.

[Table 2 about here]

In order to perform out-of-sample analysis, the sample has been divided so that the estimation is done over 2/3 of the full dataset while the remaining 1/3 is left for out-of-sample cheks[11]. Table 3 reports the results of the logistic regression analysis for the in-sample dataset. In order to prove the importance of including patent-related variables, we estimate different variants of the model. In Model 1 we take only accounting variables as regressors, while in Model 2 and 3 we include also patent-related variables, whereby the interaction terms in Model 3 aims to highlight a joint consideration of innovation and capital structure in assessing the credit quality of innovative firms.

[10] We followed the technology grouping proposed by the Observatoire de Science et the Techniques which is articulated in 30 categories. In particular: 1 Electrical devices - electrical engineering; 2 Audiovisual technology; 3 Telecommunications; 4 Information technology; 5 Semiconductors; 6 Optics; 7 Analysis, measurement, control; 8 Medical engineering; 9 Nuclear engineering; 10 Organic fine chemicals; 11 Macromolecular chemistry, polymers; 12 Basic chemical processing, petrol; 13 Surfaces, coatings; 14 Materials, metallurgy; 15 Biotechnology; 16 Pharmaceuticals, cosmetics; 17 Agriculture, food; 18 General processes; 19 Handling, printing; 20 Material processing; 21 Agriculture & food machinery; 22 Environment, pollution; 23 Mechanical tools; 24 Engines, pumps, turbines; 25 Thermal techniques; 26 Mechanical elements; 27 Transport; 28 Space technology, weapons; 29 Consumer goods & equipment; 30 Civil engineering, building, mining.

[11] See [47] for the selection of the out-of-sample dataset.

[Table 3 about here]

Overall, the size of the coefficients is stable and robust across all specifications: the variable with highest impact on the probability of default event is liquidity (LIQ_RAT) followed by the two profitability ones (PROF_RAT and EX_RAT). In particular, one standard deviation increase in the liquidity reduces the PD by 2.8% whereas in terms of profitability by 1.4%. We find that the PD moderately decreases with firm size as approximated by sales.

Model 1 shows that the accounting variables are always significant and have the expected sign as in the literature on traditional SMEs. In Model 2, the coefficient of the value of the innovative assets (VAL_RAT) is, as expected, negative and statistically significant. By contrast, patent productivity (INV_RAT) turns out to have a positive coefficient thus indicating that it contributes to increase the PD, which might seem at first counterintuitive. In order to solve this apparent puzzle, in Model 3 we further investigate the role of this variable in connection with the capital structure of the firm. To this end, we interact the patent productivity with the equity over debt ratio. This model shows that, while the patent value is always significant, the patent productivity alone loses explanatory power, but when it is interacted with the EQ_RAT it turns to be a very strong predictor of the default event with an elasticity of 4.4%.

In sum, by comparative inspection of Model 2 and 3, our results show that the patent value always reduces the PD as expected, while the patent number per se does reduces the PD only if supported by an appropriate equity level.

This finding is consistent with theory on the existence of financial constraints for innovative SMEs discussed previously. On the one hand, SMEs use more equity to finance the innovation activities because of the existence of the investor's information asymmetries on the quality of their assets. On the other, because more innovative SMEs face relatively tighter adverse selection in the credit market only a few of them, which have appropriate equity, will develop those assets and hence the real effect of financial constraints could be plausibly even more severe than the evidence suggested by results in Table 3.

The overall goodness of fit is more than 20%. This is not a small percentage given the limited number of variables of the model and the fact that our sample is made up of the only cross-section dimension and not a time series one. The innovative variables increase by 8.45% the explanatory power of the estimated model.

In order to further assess the validity of the model, we perform additional goodness of fit tests. The first one is the Cumulative Accuracy Profile (CAP) that measures simultaneously Type I and Type II errors. In the CAP analysis companies are ranked by fitted values of the PD event. For a given percentage of the observations x, a CAP curve is built by computing the percentage actual default events with the risk score equal to or lower than x (for a more detailed illustration see e.g. [46]).

In Figure 1 the thick curve shows the goodness of the estimated model. It depicts the percentage of actual defaults events (vertical axis) versus the predicted defaults by the model (horizontal axis). The diagonal line represents the case of non-informative model, whereas the upper line the perfectly predicting model. In our case the model shows a high predictive power estimating about 50% of the defaulters within only 6% of the observation. A more synthetic measure is the Accuracy Ratio which graphically equals the area predicted by the CAP divided by the area of the perfectly predicting model (see Figure 2).

The model performs well also in out-sample dataset: the CAP-out-of sample follows closely the dynamics of that in-sample though the Accuracy Ratio is smaller (Figure 3). Also the two types of error of the in-sample and out-sample dataset closely co-evolve which again strongly support the validity of the model estimated in Table 3 (Figure 4).

In order to gauge the increased accuracy obtained by the inclusion of the patent-related variable, we now directly the baseline model (Model 1) with the one proposed in this paper (Model 3).

[Table 4 about here]

In terms of Accuracy Ratio, Table 4 highlights that the model we propose perform better both in-sample and out-of-sample. Moreover, the relative reduction of the AR in the out-sample dataset for the models with innovative variables is lower (12.1%) than in the case of the Model 1 (15.1%) thus indicating that our model is more accurate in default prediction.

Figure 5a and 5b compare the CAP curves for Model 1 and Model 3 respectively in- and out-sample. The model with innovative variables shows a higher explanatory power for the medium risk firms (percentiles 15-35 percentiles for the in-sample dataset), whereas for the high risk firms (top 10% of the distribution) the models works similarly. Put differently, the financial distress of high risky firms dwarfs any additional effect of the knowledge variables as determinants on the PD.

6. Conclusions

In this paper we developed a parsimonious logit regression model to estimate a PD with default years 2006-2008 of innovative SMEs in EU15 countries. To the best of our knowledge this is among the first attempt which combines accounting and innovation related variables to predict the default event. Based on the signaling value of patents ([35], [32]), we have included their consideration in an econometric model for the prediction of innovative SMEs, which builds a widespread model based on accounting data.

In the regression analysis the innovation-related variables are two in order to consider both the value of the inventive output and the R&D productivity at the level of the firm and. Our analyses show that, while the accounting variables and the patent value are always significant with the expected sign, the patent productivity per se has a lower significance and points to an increased riskiness of the firm. However, when considered in connection with the capital structure of the firm, it reduces the PD only if equity is appropriate.

This finding can be reconciled with the literature on the financial constraints faced by innovative SMEs ([23] and [26]), which, cannot rely on equity markets and on a sufficiently developed VC industry and hence have to rely on credit to finance their projects. The model proposed in this paper to predict the PD of innovative firms can be useful for banks to gauge the credit quality of this type of firms in consideration of their peculiar feature, which include, beside standard balance-sheet ratios, other measures to account for their innovative value and potential.

Finally, it is noteworthy that patents are not the only information trail to reveal the technological and commercial potential of a start-up. Other studies have claimed that web newswires could constitute an additional information sources for financial investors [36]. We aim to develop similar measures in a later stage of this project.

Acknowledgments
We would like to thank for comments and suggestions Francesco Pattarin and participants at the Cefin Workshop (Modena, November 2010) and 4th CSDA International Conference on Computational and Financial Econometrics (CFE'10, London). Usual caveat apply. Chiara Pederzoli and Costanza Torricelli also acknowledge financial support from MIUR-PRIN 2007.

References

1. Altman, E.I. ”Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy”, *Journal of Finance*, 13, (1968), pp. 589–609.
2. Altman, E.I., Sabato, G. “Effects of the New Basel Capital Accord on Bank Capital Requirements for SMEs” *Journal of Financial Services Research*, 28, (2005), pp. 15–42.
3. Altman, E.I. Hotchkiss, E. Corporate, Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyse and Invest in Distressed Debt, 3rd Edition, Hardcover: Wiley Finance, (New York, 2006).
4. Altman, E.I., Sabato, G. “Modelling Credit Risk for SMEs: Evidence from the U.S. Market”, *Abacus*, 43 (3), (2007), pp. 332-357.
5. Arora A, Fosfuri A, and Gambardella, A., Markets for Technology: The Economics of Innovation and Corporate Strategy, MIT Press (Cambridge, MA, 2001).
6. Arrow KJ, Vertical Integration and Communication, *Bell Journal of Economics* 6 (1975), pp. 173-183.
7. Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version, Bank for International Settlements, (2006).
8. Beaver V., “Financial Ratios as Predictors of Failure”, *Journal of Accounting Research*, 5, (1966), pp. 71-111.
9. Beaver V., “Have Financial Statements Become Less informative? Evidence from the Ability of Financial Ratios to predict Bankruptcy”, *Review of Accounting Studies*, 10, (2005), pp. 93-122.
10. Behr, P Güttler, A. “Credit Risk Assessment and Relationship Lending: an Empirical Analysis of German Small and Medium-sized Enterprises”, *Journal of Small Business Management*, 45(2), (2007), pp. 194–213.
11. Bonaccorsi A. and Thoma G., Institutional Complementarity and Inventive Performance in Nanotechnology, *Research Policy*, 36 (6), (2007), pp. 813-831
12. Brown, J. R., S. M. Fazzari and B. C. Petersen, “Financing Innovation and Growth: Cash Flow, External Equity and the 1990s R&D Boom”, *Journal of Finance*, 64 (1), (2009), pp. 151-185.
13. Denis, D. “Entrepreneurial Finance: an Overview of the Issues and Evidence”, *Journal of Corporate Finance*, 10, (2004), pp. 301– 326.
14. Engelmann, B. Hayden, E. Tasche, D. “Measuring the Discriminative Power of Rating Systems”, Deutsche Bundesbank, Discussion Paper Series 2: Banking and Financial Supervision No 01/2003.
15. EIB, R&D and the Financing of Innovation in Europe, European Investment Bank. 14(2) (2009).
16. Fantazzini D., Figini S. “Default Forecasting for Small-medium Enterprises: Does Heterogeneity Matters?”, *International Journal of Risk Assessment and Management*, 11 (1-2), (2009), pp. 38-49.
17. Fidrmuc J., Heinz C., “Default Rates in the Loan Market for SMEs: Evidence from Slovakia”, University of Munich, IFO Working Paper N. 72/2009.
18. Gambardella, A., D. Harhoff, and B. Verspagen, “The Value of European Patents”, *European Management Review*, 5 (2), (2008), pp. 69-84
19. Griliches, Z. “Patent Statistics as Economic Indicators: A Survey”. *Journal of Economic Literature* XXVIII Dec., (1990), pp. 1661-1707.
20. Grunert, J., Norden, L. and M. Weber, The Role of Non-financial Factors in Internal Credit Ratings, *Journal of Banking & Finance*, 29(2), (2005), pp. 509–531.
21. Guellec, D. and B. Van Pottelsberghe de la Potterie, “Applications, Grants, and the Value of Patent,” *Economic Letters* 69 (1): (2000), pp. 109-114.
22. Guellec, D., and B. van Pottelsberghe de la Potterie, The Economics of the European Patent System: IP Policy for Innovation and Competition, Oxford University Press, (New York, 2007).
23. Hall, B.H., The Financing of Research and Development, NBER Working Paper No. 8773, 2002, Cambridge, Mass.
24. Hall, B. H., Jaffe, A. B., and M. Trajtenberg, The NBER Patent Citations Data File: Lessons, Insights,

and Methodological Tools, NBER Working Paper No. 7741, 2001, Cambridge, Mass.

25. Hall B. H. “Measuring the Returns to R&D: The Depreciation Problem”, *Annales d’Economie et de Statistique* 79/80 (2005): pp. 341-381
26. Hall B. H. and J. Lerner, Financing R&D and Innovation, in the Handbook of the Economics of Innovation, B. H. Hall and N. Rosenberg (eds.), Elsevier, (Amsterdam, 2010).
27. Hall, B. H., Jaffe, A. B., and M. Trajtenberg, Market Value and Patent Citations. *RAND Journal of Economics* 36, (2005) pp. 16-38.
28. Hall, B.H., Mairesse J. and P. Mohnen, Using Innovation Surveys for Econometric Analysis, in the Handbook of the Economics of Innovation, B. H. Hall and N. Rosenberg (eds.), Elsevier, (Amsterdam, 2010).
29. Hall, B. H., Thoma G. and Torrisi S., “The Market Value of Patents and R&D: Evidence from European firms”, NBER Working paper No. 13426, 2007, Cambridge, Mass.
30. Harhoff, D. “Strategic Spillovers and Incentives for Research and Development,” *Management Science*, 42 (6), (1996), pp. 907-925.
31. Harhoff, D. and T. Körting, “Lending Relationships in Germany: Empirical Results from Survey Data,” *Journal of Banking and Finance*, 22, (1998), pp. 1317-1353.
32. Harhoff, D., F. M. Scherer and K. Vopel, “Citations, Family Size, Opposition and the Value of Patent Rights - Evidence from Germany”, *Research Policy*, 32, (2003), pp. 1343-1363.
33. Harhoff D., C. Häussler, and E. Müller "To be Financed or Not: The Role of Patents for VC Financing", CEPR Discussion Paper 7115, 2009.
34. Harhoff D. and G. Thoma Inventor Location and the Globalization of R&D, Paper presented at the Conference on Patent Statistics on Decision Makers, 17-18 November 2010, Vienna, Austria.
35. Hsu D. and R. Ziedonis Patents as Quality Signals for Entrepreneurial Ventures, *Academy of Management Best Paper Procedings*, Anaheim California, 2008.
36. Kerr, W. R., J. Lerner, and A. Schoar, "The Consequences of Entrepreneurial Finance: A Regression Discontinuity Analysis", NBER Working Paper No. 15831, 2010, Cambridge, Mass.
37. Jolliffe, I T. Principal Component Analysis, Second Edition, Springer-Verlag, (New York, 2002).
38. Lanjouw, J. O., and M. Schankerman, “Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators.” *Economic Journal* 114, (2004), pp. 441-465.
39. Lennox, C. “Identifying Failing Companies: A Reevaluation of the Logit, Probit and DA Approaches”, *Journal of Economics and Business*, 51, (1999), pp. 347–364.
40. Lerner, J., “The Importance of Patent Scope: An Empirical Analysis,” *Rand Journal of Economics* 25 (2), (1994), pp. 319-33.
41. OECD, Patent Statistics Manual, OECD, (Paris, 2009).
42. OECD, Financing High Growth and Innovative Start-ups and SMEs: Data and Measurement Issues, OECD, (Paris, 2010).
43. Ohlson J., “Financial Ratios and the Probabilistic Prediction of Bankruptcy”, *Journal of Accounting Research*, 18, (1980), pp.109-131.
44. Pavitt, K., Uses and abuses of patent statistics. In van Raan, A. (ed.), Handbook of Quantitative Studies of Science Policy, North Holland, (Amsterdam, 1988).
45. Pederzoli, C. and C. Torricelli “A Parsimonious Default Prediction Model for Italian SMEs”, *Banks and Bank Systems*, 5(4), (2010), pp. 28-32.
46. Sobehart, J. Keenan, S. Stein, R. “Benchmarking Quantitative Default Risk Models: A Validation Methodology”, Algo Research Quarterly, 4(1),(2001), pp. 57-72.
47. Stein R. M., “Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation”, *Journal of Risk Model Validation*, 1(1), (2007), pp. 77-113.
48. Thoma G., S. Torrisi, A. Gambardella, D. Guellec, B. H. Hall, and D. Harhoff, “Harmonizing and Combining Large Datasets – An Application to Firm-Level Patent and Accounting Data”, NBER Working Paper No. 15851, 2010, Cambridge, Mass.
49. Trajtenberg, M. A Penny for Your Quotes: Patent Citations and the Value of Innovations. *Rand Journal of Economics* 21, (1990), pp. 172-187.
50. Hall B. H and R. Ziedonis, “The Determinants of Patenting in the U.S. Semiconductor Industry, 1980-1994”, *Rand Journal of Economics* 32, (2001), pp. 101-128.

Table 1. Firms in the sample

	Cohort							
Industries	*pre-1970*	*1971-80*	*1981-90*	*1991-95*	*1996-2000*	*post-2000*	*Overall*	*%*
Chem.&Phar.	62	27	39	11	17	3	159	6.0%
Low tech	275	140	218	77	88	15	813	30.5%
Manufact.	268	175	231	120	122	36	952	35.7%
Retail Distr.	73	54	113	70	50	17	377	14.1%
Services	35	27	91	75	93	43	364	13.7%
Overall	713	423	692	353	370	114	2,665	100%
Overall %	26.8%	15.9%	26.0%	13.2%	13.9%	4.3%	100%	

Note: Low tech industries include US SIC code 10-33, such as agriculture, forestry, fishing, mining, food and wood products, and textiles, but not chemicals and pharmaceuticals.

Table 2. Variables included in the prediction model

Variable name	*Variable description*
Dependent variable	
DEFAULT	Dummy variable whether a firm is bankrupted: (a) unable to pay the creditors; (b) the assets are held by a receiver; (c) assets and property of the company redistributed.
Independent variables	
Accounting	
EQ_RAT	The equity ratio of the firm equals Equity / Total Debt.
LIQ_RAT	The liquidity ratio is given by Cash/Sales
PROF_RAT	The profit ratio is given by Net Earnings / Total Assets.
EX_RAT	The ratio given by Retained Earnings/Total Assets
COV_RAT	The coverage is given by EBITDA/Interest expenses
SALES	The sales variable is measured by the log of Operative Turnover of the firm
Knowledge-related variables	
VAL_RAT	The patent value ratio equals Capitalized Patent Value Stock / Capitalized Patents Stock. We include three measures of patent value: i) forward citations; ii) size of the patent family; iii) number of patent classes.
INV_RAT	The human capital productivity ratio equals Capitalized Patent Stock / Capitalized R&D personnel.
Control variables	
Quoted dummy	The dummy takes value 1 if the capital of a firm is traded in the stock market.
Country dummies	Macro areas: Central Europe (AT, CH, and DE); Benelux (BE, LU and NL); Nordic countries (DK, FI, IS, NO, and SE); Central-South Europe (ES, FR, GR, IT and PT); and others (GB and IE).
Cohort dummies	pre 1970, 1970s, 1980s, 1991-1995, 1996-2000, post 2000 and residual category “unknown”.
Sectorial dummies	process industries, manufacturing, chemicals and pharmaceuticals, utilities, distribution and retail, services.
Year dummies	2006, 2007, 2008

Table 3. Multivariate logistic regression: results

	Model 1	*Model 2*	*Model 3*	*Model 3 Elasticity*
SALES	-0.298 (0.090)***	-0.287 (0.090)***	-0.287 (0.090)***	-0.002
EQ_RAT	-0.665 (0.265)***	-0.656 (0.266)***	-0.703 (0.239)***	-0.006
LIQ_RAT	-3.320 (1.317)***	-3.241 (1.332)**	-3.254 (1.330)***	-0.028
PROF_RAT	-0.651 (0.297)**	-0.704 (0.294)**	-0.774 (0.284)***	-0.007
EX_RAT	-1.404 (0.690)**	-1.493 (0.737)**	-1.654 (0 .714)**	-0.014
COV_RAT	-0.027 (0.006)***	-0.027 (0.006)***	-0.026 (0.006)***	-0.022
VAL_RAT		-0.627 (0.312)**	-0.644 (0.313)**	-0.005
INV_RAT		1.556 (0.869)*	-1.615 (1.605)	
INV_RAT * EQ_RAT			-5.157 (2.012)***	-0.044
McFadden R squared	0.213	0.223	0.231	
Hosmer-Lemeshow Test	98.39	100.74	77.22	

Notes: Each regression includes the dummies listed in Table 2; * significant at 10%; ** significant at 5%; *** significant at 1%; SE in parenthesis.

Table 4. A comparison of the accuracy ratio

Accuracy Ratio (AR)	*Model 1*	*Model 3*
AR in-sample	0.689	0.702
AR out-of-sample	0.585	0.617

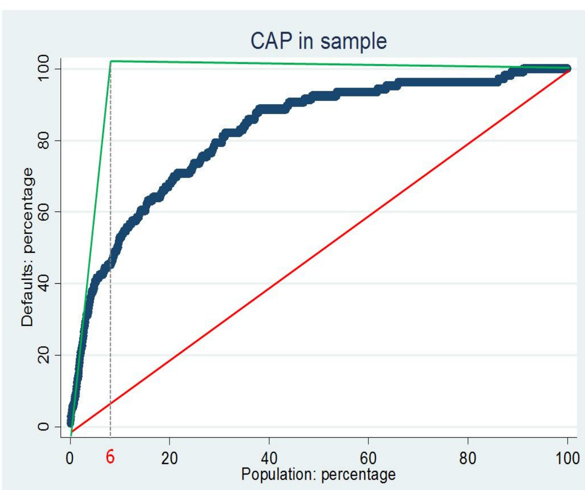

Figure 1 In-sample Cumulative Accuracy Profile

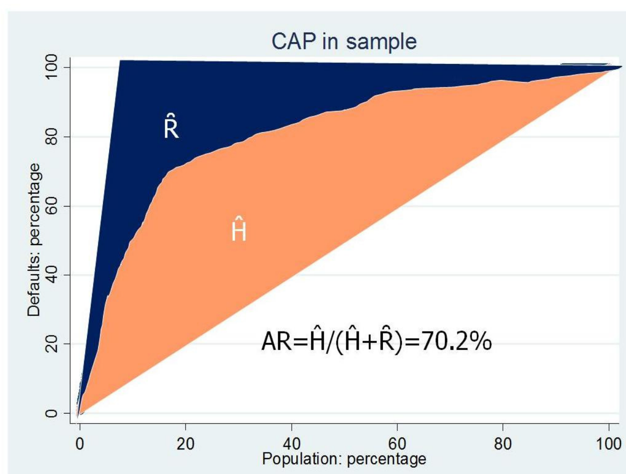

Figure 2 In-sample Accuracy Ratio

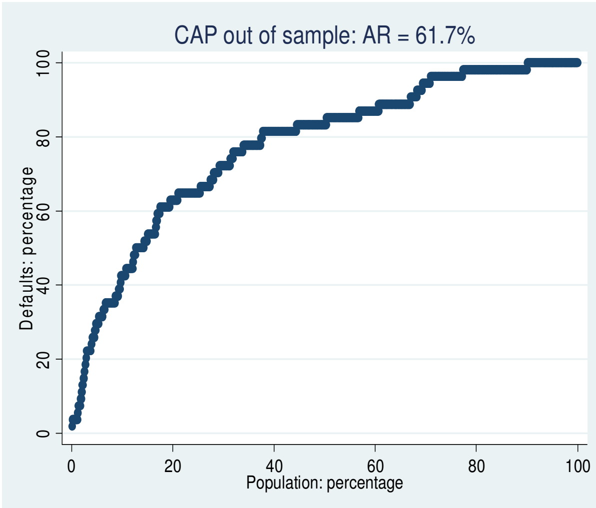

Figure 3 Out-of-sample Cumulative Accuracy Profile

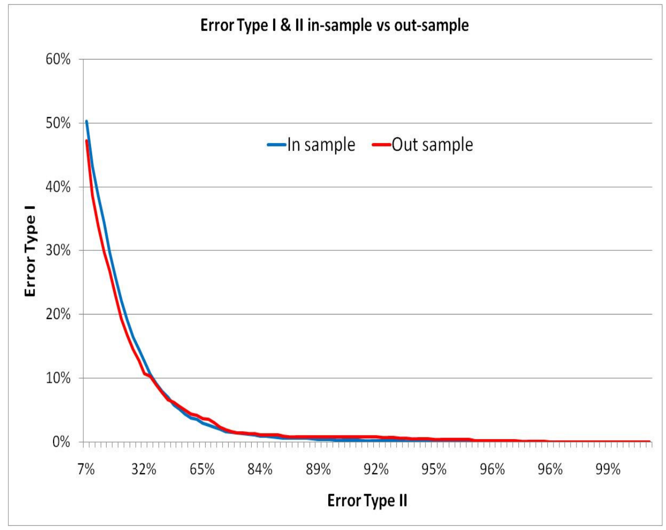

Figure 4 Comparison of Prediction Errors

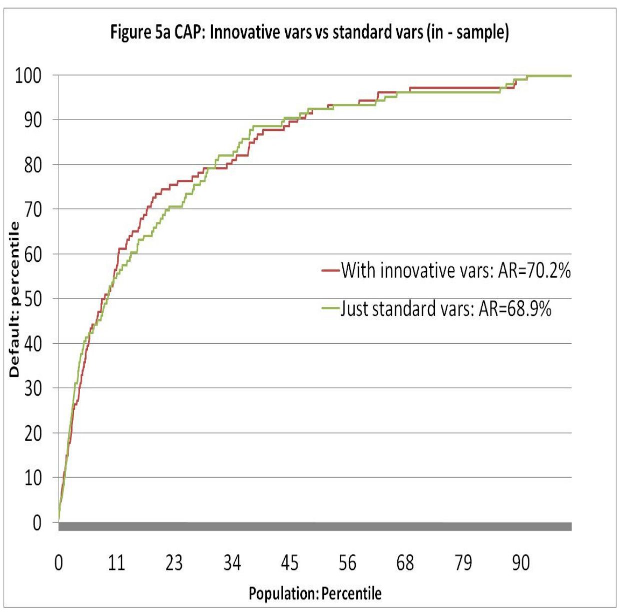

Figure 5a Comparison of Model 1-3 in sample

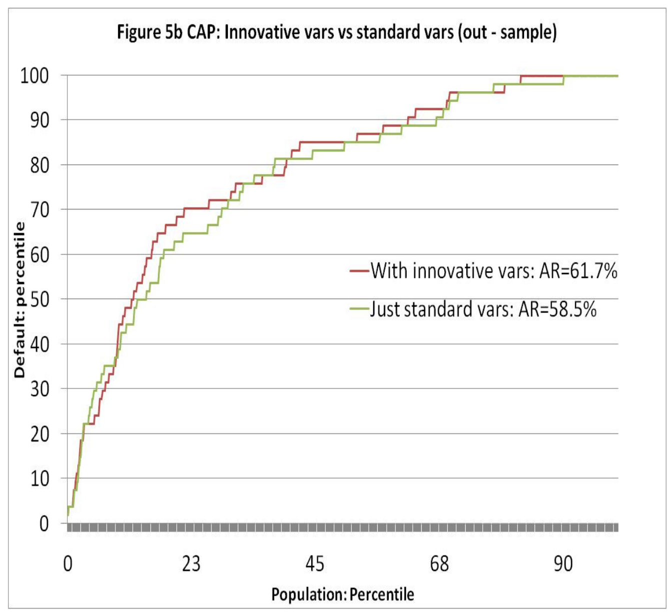

Figure 5b Comparison of Model 1-3 out sample