The impact of introducing economic variables into credit scorecards: an example from invoice discounting.

Jie Zhang[a] and Lyn C Thomas[a,b]

a: University of Southampton, Southampton, UK.

b: Corresponding author
School of Management
University of Southampton
Southampton UK
SO17 1 BJ

Tel : 44-238059-7718
Fax : 44- 238059-3844
Email: l.thomas@soton.ac.uk

Abstract: The paper shows how introducing economic variables into a credit scorecard improves the predictive power of the scorecard. Such a scorecard can forecast default rates accurately even when economic conditions change This means one can develop a single step approach to estimating the Point in Time PDs which are requirements of the Basel Accord banking regulations. A one step approach has several advantages when compared with the more standard approach of estimating scores with no economic variables first and then segmenting the portfolio by score bands and estimating the PD per segment. To build such a scorecard we decompose it into the population odds and the weights of evidence and shows that economic variables model the dynamics of the population odds part of the scorecard and so leads to this improvement in prediction.

The paper then applies this extension to credit scoring to a real problem in invoice discounting. This is when banks lend to small businesses using the invoices that the businesses have issued as collateral. There is a significant volume of such lending, but it is not much addressed in the previous literature. The scorecards used to assess the risk of default of such small businesses are very similar to the behavioural scorecards used to assess default risk in lending to consumers. The results show that modelling the population odds by economic variables is very effective but there is little improvement in the scorecard's performance if one models the dynamics of the weights of evidence by adding interactions between the economic variables and the performance characteristics of the borrowing firm.

Key words: Finance; risk analysis; credit scoring; probability of default; invoice discounting; economic factors; small and medium sized enterprises
JEL codes: C53; G21; G32

1 Introduction

With the advent of the Basel Accords, Basel II (BCBS 2006) and Basel III (2011), credit scoring – the mainstay of credit risk assessment for consumer lending (Anderson 2007, Thomas 2009) for half a century – has become a vital tool in estimating Probability of Default (PD). This is one of the three parameter estimates required on all type of loan portfolios by the Accords. The Accords determine how much regulatory capital banks must keep to deal with the credit risks they incur. The philosophies underlying the Basel definitions of Probability of| Default have been fully discussed in a number of books Englemann and Rauhmeier (2006), Ong (2006) and Van Gestel and Baesens (2009) and many papers . for example those in the literature survey in Carlehed and Petrov (2012). The two ends of the spectrum are Point-in-Time (PIT) conditional PD and Through-the-cycle (TTC) unconditional PD. The latter is sometimes called the long run average PD (LRPD) but the Basel Accord requires estimates of the loan run average of the one year look ahead PD which is not quite the same as either. PIT PD should take into account the current information including the current economic conditions. So it changes as the economy changes. TTC PD on the other hand should stay constant through the business cycle unless there are permanent changes to the borrower's situation which would affect the ability to repay in the long term. There are different positives about each approach and these lead to their different uses. TTC PD is procyclical and so is useful in capital management and its stability is seen as useful in capital adequacy regulations. PIT PD is preferable for risk management as it gives better predictions of the immediate losses and its models are easy to verify by back testing. For corporate loans most models are considered hybrids but rating agency models tend to the TTC end of the spectrum while banks tend to build closer to PIT type models.

For corporate loan portfolios a two step approach was developed by Aguai et al (2006) and Miu and Ozdemir (2008). LRPD was first estimated, often by taking the agencies ratings, and then these results are conditioned on macro-economic variable values to get PIT PD estimates. An alternative two step approach was developed for retail and SME portfolios by first building a scorecard with no economic characteristics in it . The scores were then used to segment the portfolio and the LRPD for each segment of the portfolio is calculated by averaging the default rates over the economic cycle. These LRPD for the segments may need to be adjusted if they do not give rise to the LRPD for the portfolio as a whole, (Bank of England 2013). So scores are only used for discrimination and not for estimating the PD values. What is proposed here is a single step process which gives PIT PDs directly at the individual loan level. Since these include the economic variables in the scorecard one could use simulation to estimate LRPD if required (See (McDonald et al 2010) for an example of this using a mortgage portfolio). One benefit of the single step approach is that one gets discrimination and estimates of PD values from the same scorecard, Since the discrimination is used in operational decisions one can show the Basel "use" test is satisfied. Moreover one can use back testing to check regularly both the discrimination and the probability forecasts of the scorecard.

This paper outlines why credit scorecards are improved for both consumer and small firm lending by including economic effects into the scorecard. We show how the decomposition of the scorecard into two parts allows one to recognise theoretically the two effects that including economic variables will have in a scorecard.

This idea of introducing economic variables is applied to a real problem that arose in invoice discounting, where the bank's lending to small companies is secured against sales invoices which the companies have issued but not yet received payment on. The major bank involved used a logistic regression based scorecard to assess the default risk of the SME companies they lent money to using invoice discounting. In the last financial crisis, their scorecard, which had been built in a more benign period, did not respond well to the change in the macroeconomic environment and predicted too few defaulters. The scorecard though continued to discriminate well in terms of the ranking of the companies' default probability. Exactly the same phenomenon occurred with consumer scorecards during the subprime mortgage crisis. This case study showed that introducing economic variables directly into the scorecard improved the default probability predictions considerably but introducing interaction variables between the economic variables and borrower characteristics made no significant improvement in either probability prediction or discrimination.

Section 2 recalls the basic decomposition of a credit scorecard and highlights why there is a need to introduce economic variables so as to deal with the changes in the population default rate over time. Section 3 outlines the literature on building scorecards and Section 4 describes invoice discounting and reviews the literature on it. Section 5 describes the data used in this research while section 6 uses logistic regression to build a credit scorecard for invoice discounting using the data. Section 7 adds economic variables to such a scorecard, which improved the prediction of how many companies default. Section 8 considers using interactions between the economic variables and firm specific behavioural variables to improve predictions. This corresponds to including economic variables in the firm specific term in the scorecard.

Finally Section 9 draws some conclusions on the advantage of using economic variables in the scorecard.

2. Decomposition of scorecards

Credit scoring is a way of estimating which borrowers will default over some future time horizon. These are the" Bads" (B) while the others who have an acceptable performance over that period are the "Goods" (G). A credit score is essentially a sufficient statistic in that for a borrower with characteristics **x** the score s(x) satisfies $P\{G|\mathbf{x}\} = P\{G|s(\mathbf{x})\}$. Most scores, including all produced by logistic regression – the way 95% of scorecards are produced in practice- are log odds scores (Thomas 2009) so that

$$s(\mathbf{x}) = \log\left(\frac{P\{G|s(\mathbf{x})\}}{P\{B|s(\mathbf{x})\}}\right)$$

Such scores can be decomposed into the sum of a population odds score, s_{pop}, and a weights of evidence term, *woe*(**x**), by Bayes Theorem. Formally if p_G and p_B are the proportions of Goods and Bads in the whole borrower population, then

$$s(\mathbf{x}) = \log\left(\frac{P\{G|s(\mathbf{x})\}}{P\{B|s(\mathbf{x})\}}\right) = \log\left(\frac{P\{s(\mathbf{x})|G\}p_G}{P\{s(\mathbf{x})|B\}p_B}\right) = \log\left(\frac{p_G}{p_B}\right) + \log\left(\frac{P\{s(\mathbf{x})|G\}}{P\{s(\mathbf{x})|B\}}\right) = s_{pop} + woe(\mathbf{x})$$

Ideally we should consider the dynamics of a scoring system and define s(**x**,t) to be the score that a borrower with characteristics **x** should have at time t. Hence one would then have the decomposition

$$s(\mathbf{x},t) = s_{pop}(t) + woe(\mathbf{x},t)$$

The problem with a scoring system is that it is static and uses the data available at t_0 when the scorecard is being built.. So the score at time t will still be

$$s(\mathbf{x},t_0) = s_{pop}(t_0) + woe(\mathbf{x},t_0)$$

One needs to include in the scorecard characteristics which change over time and so are able to give the borrower the score s(**x**,t) at time t. $s_{pop}(t)$ is a function of the Population Bad rate at time t and so one could introduce economic variables $e_i(t)$ into the scorecard to model these changes, i.e.,

$$s_{pop}(t) = a_1 e_1(t) + \cdots a_n e_n(t)$$

More problematic is whether the weights of evidence term *woe*(**x**,t) does depend on t. If so can the dynamics of *woe*(**x**,t) be explained by interaction variables between the performance characteristics **x** of the borrower and the economic variables which change with time? If so, one can replace the term $woe(\mathbf{x},t)$ by $woe(\mathbf{x},e(t))$ where e(t) are the economic conditions at time t. We investigate whether both these effects occur and whether one can build a scorecard that allows for them. This is done in the context of modelling the credit risk for invoice discounting which has not been addressed in the open literature previously.

3. Literature review of retail and SME credit scoring and the introduction of economic variables

The literature on credit scoring, especially for consumer lending, has grown substantially in the last decade (Thomas et al 2002, Mays 2004, Anderson 2007, Thomas 2009). The most widely used technique is logistic regression (Thomas, et al (2002). Traditionally consumer credit scorecards ranked potential applicants in terms of default risk and lenders choose a cut-off score of whom to accept and whom to reject using business reasons. With the advent of the Basel Accord requirements, scorecards must now also give good predictions of the default probability as well as accurate rankings.

On the corporate side, Altman (1968) used multiple discriminant analysis (MDA) to build a score based on a firm's financial ratios to estimate the probabilities of the firm defaulting. Altman and Sabato (2007) developed a financial ratio based logistic regression model for small and medium sized enterprises (SME) and showed that it discriminated well. However these models do not use the information on the volume and performance of the invoices a firm is raising which are available to those who score SMEs for invoice discounting, nor do they introduce economic variables. Demirovic and Thomas (2007) while identifying which accounting ratios were significant in estimating large company default rates pointed out that the average of such rates depends on the state of the economy. This is akin to saying $s_{pop}(t)$ is a function of e(t).

Dietsch and Petey (2002) built variants of Creditmetrics and Credit Risk+ which could be applied in the SME context. However both were two stage models depending on an existing credit score which split SMEs into appropriate risk grades. This paper introduces a one stage approach to the problem.

Bellotti and Crook (2009) and Malik and Thomas (2010) introduced economic variables into scorecards but by using the survival analysis approach to scorecard building. This does not lead to the log odds scores, which we look at in this paper and which are produced by the standard logistic regression approach used by most lenders. Breeden (2009) developed dual time dynamics which is a portfolio level model that includes vintage and maturation and calendar effects. The last of these involves economic variables. Crook and Bellotti (2010) reviewed consumer default risk models including ones with time variable covariates. They cover a number of approaches but do not discuss the impact that including economic variables have on the default risk. Sousa et al (2013) develop a two stage model. First they build a conventional scorecard with static characteristics. Secondly they apply linear regression to relate the portfolio level default rate to the economic conditions and then apply this adjustment at the individual default level. John et al (2010) looked at how health care loan delinquencies are affected by macroeconomic variables. This was a portfolio level model and so no scorecard was involved.

4. Invoice discounting

Invoice discounting is a form of short-term borrowing used to improve a small company's working capital and cash flow position. It allows a business to borrow money against its sales invoices before the customer has actually paid (abfa 2014). To do this, the company borrows a percentage of the value of its sales ledger from a bank or financial institution, effectively using the unpaid sales invoices as collateral for the borrowing. Invoice discounting differs from debt factoring, in that invoice discounting only involves two parties, the invoicing company and the finance company or bank. In debt factoring, a company sells its invoices as receivable to the factors (the financial institutions) at a discount. The factors collect the money due in the invoice from those who had received the goods (the debtors).

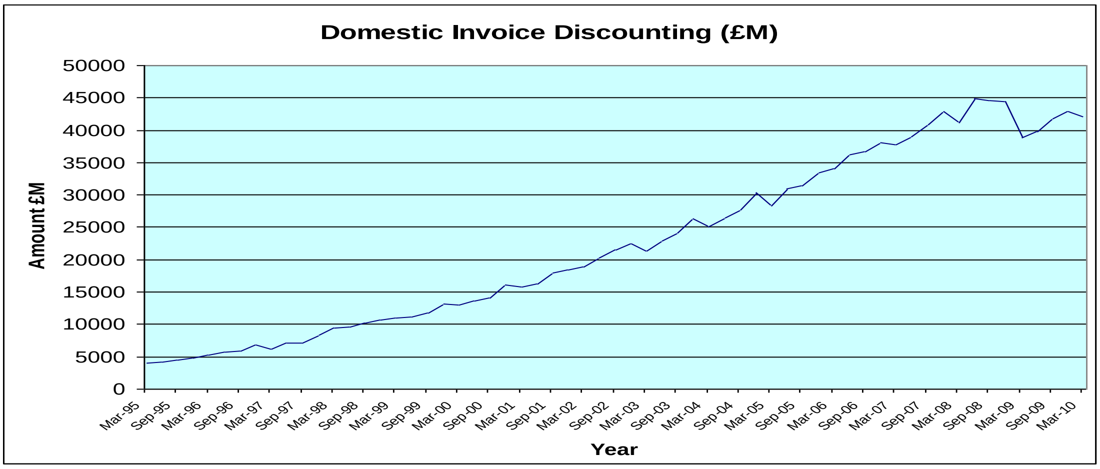

Figure 1: Amount borrowed in UK using invoice discounting (www.abfa.org.uk)

Figure 1 reflects the increasing trend of UK invoice discounting from 1995 to 2010. It only refers to that lent by members of the Asset Based Finance Association and so the total lending is higher. The figure for borrowing worldwide using invoice discounting was put at $1 trillion even as far back as 2001 (Omni Rand 2002) .

There is very little literature on invoicing discounting (ID) and factoring (afba 2014), and what there is concentrates on which firms use invoice discounting and factoring. Smith and Schnucker (1994) examined the structure of the organisation to evaluate the economics of the decision of whether to use invoice discounting and claimed that economies of scale were the major driver in whether to use invoice discounting or not. Summers and Wilson (2000) found evidence of a 'financing demand' explanation for the use of factoring, and they argued that the use of factoring was more related to the demand for asset-based finance from small companies than to the firm's organizational structure. Soufani (2000) profiled which businesses use factoring and invoice discounting in terms of sector, size, age and type of ownership. Soufani (2002) surveyed 3805 SMEs and built a logistic regression model to test hypotheses about which businesses use factoring in terms of their demographic characteristics, their relationship with their banks, their size and the value of their collateral.

Invoice discounting credit scores are developed for small companies, which need bank loans to help with their cash flow. Thus they are much closer to consumer credit scorecards than the ratings models used by rating agencies on large corporate.
In invoice discounting (ID), default means the invoicing company defaults, at which point the bank cannot collect on the invoices. Unlike other corporate lending, the bank or finance company has very up to date information on the state of the firm, by seeing the value of the invoices being issued, and by observing the financial statements being

submitted monthly to the bank. The banks can close down ID accounts at any time or seek further collateral from the company if necessary. So they need to estimate regularly the probability of a firm defaulting and not just how it compares or ranks with other ID firms. The Hosmer-Lemeshow test and the expected versus actual number of defaulters measure how accurate are the probability predictions based on the score. For invoice discounting, these measures are as important as the Kolmogorov-Smirnov statistic (KS) and the Gini coefficient (GINI), which measure how well the scorecard ranks the borrowers. (BCBS 2005)

5 The data and data preparation

The data used in this research has information on each company that is being lent money by the bank under the ID approach. The data records every month how much the companies are invoicing their customers and the state of the invoices. The records start from July 2003 and end at December 2009, and there is information on the invoices sent out by 5826 companies, and their subsequent repayment. 1184 of the companies defaulted during this period. The dependent variable is whether the company defaults within the next 12 months. There are 75 independent variables, some of which are basic information about the company, and some of which relates to the state of the company's invoices and its sales ledger. Examples of these characteristics are

Duration of Account (how long company has had ID account)

Financial rating of company

Ledger Difference

Annual Turnover Trend (average value of sales in last 3 months divided by sales in last month)

Disapproval due to Age (% of invoices not acceptable because they are too old)

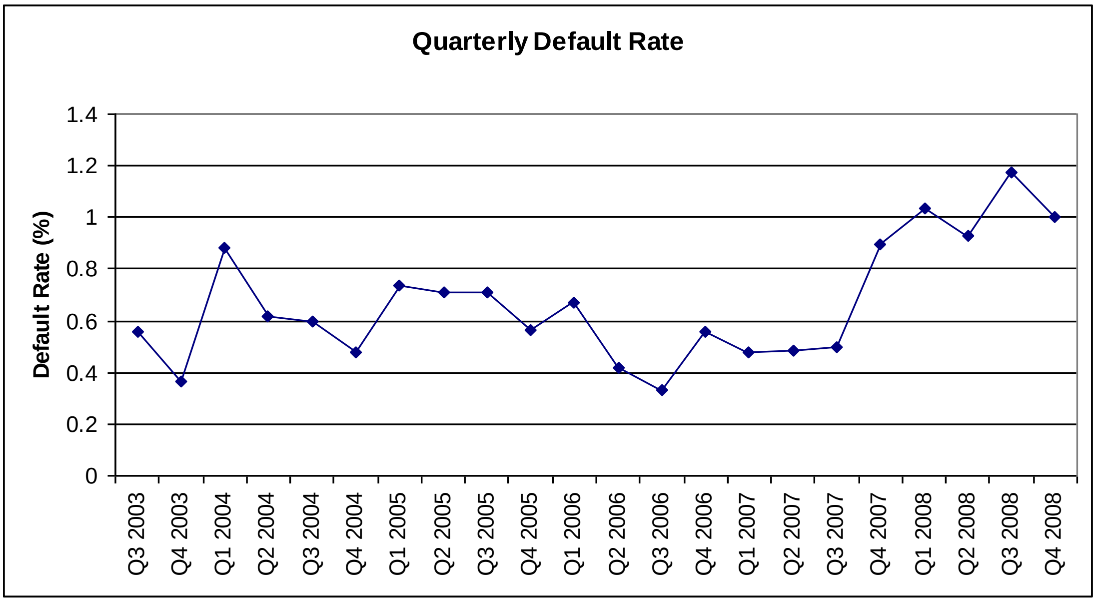

Figure 2: Quarterly default rate of invoicing companies

The actual default rate is reported for each quarter in Figure 2. The reported default rate leads the actual default rate by 12 months, because the dependent variable describes the percentage of defaults in the next 12 months. The default rate rises in the fourth quarter of 2007, because of the global financial crisis that began in the second half of 2008.

The total number of company/month observations in the data set is 173,542. Deleting observations with significant amounts of missing data or with outliers (extremely small or larger values), cuts the sample to 137,271. These are the cases used to build and to test the model.

We split the whole population into three parts: a training sample, an in-time test sample and an out-of-time test sample used to test the model. Data on companies from July 2003 to June 2008 is randomly split into two parts: 2/3 of the firms are used as the training population, and 1/3 of the firms are an in-time test population. In both the training population and the in-time test population, the bad observations are only 6 percent of the observations, since each default gives rise to twelve Bad observations in the twelve months preceding default. So the ratio of ‘good’ (non-default) to ‘bad’ (default) is about 15:1. Scorecards built on data with such a low bad rate are not that robust (Vinciotti et al (2002), Anderson (2009)), so we took a random sample of the good observations and kept all the bad observations to make the ratio of good to bad to be 3:1. For the out-of-time test sample, observations from July 2008 to December 2008 (with observation window ended at December 2009) are kept untouched, and no sampling is made.

6. Building the scorecard

The scorecard is built using the usual approaches to the preparation and selection of the variables. The attributes of a categorical variable are combined if they have similar bad rates and are similar in their detail. The values of continuous variables are coarse classified into a number of indicator variables for suitable intervals of the values. Two continuous variables – average utilization of available loan in the last 90 days and Debt Turn, which is how long it takes the payments of the invoices to cover the sales outstanding at the beginning of the month - are kept as ordinal variables, since both produce a monotonic trend in the bad rate. Correlations between the independent variables are calculated and only one variable from each group of strongly correlated variables is selected for model building.

Logistic regression, the most popular approach in building credit scorecards, is used to predict the default of the invoicing companies. A logistic regression model assumes

$$\log\left[\frac{1-p}{p}\right]=\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_k x_k+\varepsilon \qquad (1)$$

where, p is the probability of default, $x_1, x_{2,} ..., x_k$ are independent variables which describe characteristics of the invoicing firm and its invoices . $\beta_0, \beta_1, ...\beta_k$ are unknown parameters and ε is a random error term.

From equation (1), we can derive *p*, the probability of default

$$p = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k)} \qquad (2)$$

As only a fraction α of good observations were used in building the model, then one needs to adjust the relationship between *p* and the independent variables to

$$\log\left[\frac{\alpha(1-p)}{p}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon \qquad (3)$$

or $$p = \frac{1}{1 + \frac{1}{\alpha}\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k)} \qquad (4)$$

In such cases if *p* is the apparent probability of default in (2) then the actual probability of default *p'* in the out-of-time test sample is

$$p' = \frac{1}{1 + \frac{1}{\alpha}(\frac{1}{p} - 1)} \qquad (5)$$

Logistic regression is undertaken using the stepwise approach but with only the twenty most significant variables from the univariate analysis kept in the model. No economic variables are included. The variables selected to enter the scorecard (Model 1) are shown in Table 1. The information on the invoices sent out by the company and the subsequent payment of the invoices are very strong indicators of the company's chance of default. They are more specific than micro economic indicators of the company sector or geographical area. The latter were included in the characteristics considered for inclusion in the scorecard but were not selected as significant by the stepwise regression.

Characteristic	Estimate	Standard Error	P-value	Description
Intercept	-4.640	0.087	<0.0001	
Acc_Dur_2	-0.343	0.052	<0.0001	Account Duration 1235-2081 days
Acc_Dur_3	-0.509	0.051	<0.0001	Account Duration >2081 days
Add_ABL	0.532	0.064	<0.0001	Has a supplementary product
ATTrend3m_2	0.212	0.052	<0.0001	Annualised Turnover 3 month trend 1.01-1.04
ATTrend3m_3	0.415	0.062	<0.0001	Annualised Turnover 3 month trend 1.05-1.08
ATTrend3m_4	0.622	0.089	<0.0001	Annualised Turnover 3 month trend 1.09-1.11
ATTrend3m_5	0.747	0.061	<0.0001	Annualised Turnover 3 month trend >1.11
BandName_1	-0.410	0.083	<0.0001	Band Name: 'Global'
Bank_1	1.286	0.050	<0.0001	Not banking with this bank
Disapp_Age_1	0.211	0.041	<0.0001	Disapprovals due to 'age' >5.61
Ent_Avg_Tr3m_1	0.253	0.057	<0.0001	Entitlement Average 3 month trend >1.13
Rating_2	0.286	0.045	<0.0001	Financial Rating '6'
Rating_3	0.525	0.051	<0.0001	Financial Rating '7' '8' '9'
LedgerDiffer_1	0.230	0.042	<0.0001	Ledger Difference <=0.97
PayColRatio_1	0.412	0.084	<0.0001	Payment Collection Ratio <=0.53
Top5PcDebtors_1	-0.582	0.082	<0.0001	Top 5 biggest debtors' percentage of the current Sales Ledger <=22.22
Top5PcDebtors_2	-0.274	0.062	<0.0001	Top 5 biggest debtors' percentage of the current Sales Ledger 22.23-34.96
Top5PcDebtors_4	0.406	0.044	<0.0001	Top 5 biggest debtors' percentage of the current Sales Ledger >=62.65
UtiliAve90	0.150	0.005	<0.0001	Utilisation average in last 3 months; ordinal variable, 1-18
DebtTurnClient	0.034	0.005	<0.0001	How many days to cover closing sales ledger; ordinal variable, 1-16

Table 1: Characteristics and their coefficients that appear in Model 1

Table 2 reports the Gini, KS, and Hosmer-Lemeshow test results for Model (1) on the three sets. The Gini and KS values measure the discrimination of the scorecards. InTable 2 the model discrimination is roughly the same in the training, in-time test, and out-of-time test samples. So there is no obvious overfitting in the model. However, from the Hosmer-Lemeshow test, we can see the Chi-square value of the out-of-time test sample is much higher (1470.94) than that in the training sample and the in-time test sample. The reason is that the expected number of defaulters (the sum of predicted probability of default of every observation) given by the model prediction, 605, is much lower than the actual number of defaulters in the out-of-time test sample which is 1409.

	Training	In-time Test	Out-of-time Test
Gini	0.62	0.63	0.60
KS	0.46	0.49	0.46
Hosmer-Lemeshow test (Chi square)	43.16	26.93	1470.94
Actual bad's	4666	2247	1409
Expected bad's	4666	2201	605

Table 2: Model 1 measurement results

Model (1) displays the same problem that the bank reported in practice: In the out-of-time sample, the scorecard continued to discriminate well but the number of predicted defaulters is much less than the actual number of defaulters. In the next section we deal with this problem by adding macroeconomic variables to the scorecard.

7 Adding macroeconomic variables

Introducing economic variables in the scorecard allows the dynamics of $s_{pop}(t)$ to be modelled accurately with a corresponding improvement in the predicted number of defaults. Authors have reported that macroeconomic conditions do affect default risk. Most of the analysis has been done on lending to large companies (Figlewski et al 2007). There has been little analysis of what economic factors affect consumer and small company default risks (Liu and Xu 2003)..

The economic variables considered are:

GDP Growth: change in quarter on quarter in previous year, seasonally adjusted

RPI Growth: Retail Price Index, percentage change over last twelve months

Unemployment Rate: percentage, seasonally adjusted

Interest Rate: Bank of England Libor rate

Production Index: Production Index for manufacturing industries, seasonally adjusted

Business Confidence Index: From the Institute of Chartered Accountants

FTSE All-share: The highest value of FTSE All-Share index in the relevant month

These economic variables are easily obtained (most of them are from the website of Office for National Statistics) and have been previously used in academic research. ‘GDP’, ‘RPI’, ‘Unemployment Rate’, and ‘FTSE’ are suggested as important in consumer finance by Tang et al (2007), and Liu et al (2003). ‘GDP’, ‘RPI’, ‘Unemployment Rate’, ‘Interest Rate’, and ‘FTSE’ are investigated by Figlewski et al (2007) in corporate credit risk, and are shown to influence corporate credit risk. We add two further economic variables, ‘Production Index’, and ‘Business Confidence Index’. These are relevant for invoice discounting since the firms who use it are mainly SMEs in the manufacturing sector. ‘Production Index’ gives an insight into how the manufacturing sector is performing. ‘Business Confidence Index’ describes what managers think the future of their organisation may be. The economic variables are correlated and Table 3 lists the correlation coefficients between them.

	Unemp rate	RPI	GDP	Interest	Produc Index	Business Confidence Index	FTSE
Unemp rate	1						
RPI	-0.71	1					
GDP	-0.93	0.76	1				
Interest	-0.80	0.94	0.85	1			
Produc Index	-0.84	0.88	0.92	0.95	1		
Confidence Index	-0.83	0.51	0.89	0.61	0.74	1	
FTSE	-0.08	0.53	0.29	0.60	0.55	0.04	1

Table 3: Correlation coefficients between economic variables

From Table 3, we can see that except for ‘FTSE’, all the other economic variables have strong correlations with each other. To ensure the coefficients of the variables in the logistic regressions make sense, we do not include strongly correlated variables together in the model.

To choose the best economic variables, each economic variable is added on its own to the 20 variables originally considered in the Model in section 5. The results were that ‘GDP’ and ‘Business Confidence Index’ had coefficients with negative signs, and small p-values (smaller than 0.0001). These negative signs are reasonable since a positive economic environment leads to high GDP growth and high Business Confidence and also to low default rate. ‘Production Index’ and ‘FTSE’ have negative signs but the p-values are larger than 0.0001. When, either ‘Unemployment Rate’, or ‘Interest Rate’ were added to the original model they were not selected as being significant while ‘Retail Price Index’ was selected but with a low significance value. Secondly, all the seven economic variables were added to the 20 variables in the model of section 5. A stepwise approach to logistic regression was also applied. ‘FTSE’ and ‘Business Confidence’ were both significant with negative signs and extremely small p-Value; ‘Production Index’ and ‘GDP’ were selected in the model but their coefficients were not highly significant. In the light of this, two versions of models with economic variables are used; one version having ‘Business Confidence Index’ and ‘FTSE’ together, and the other version only including ‘GDP’.

	Characteristic	Estimate	Standard Error	P-value	Description
Version 1: (Model 2)	Confidence	-0.026	0.002	<0.0001	Confidence Index
	FTSE	-0.151	0.025	<0.0001	FTSE All shares
Version 2: (Model 3)	GDP	-0.290	0.038	<0.0001	GDP Growth

Table 4: Coefficients of economic variables in the model (Model 2 and Model 3)

Version 1	Training	In-time Test	Out-of-time Test
Gini	0.63	0.63	0.59

KS	0.47	0.49	0.49
Hosmer-Lemeshow test (Chi square)	32.51	29.95	63.81
Actual 'bad's	4666	2247	1409
Expected 'bad's	4666	2202	1306

Table 5: Model 2 results with 'Business Confidence Index' and 'FTSE' in the model

Version 2	Training	In-time Test	Out-of-time Test
Gini	0.63	0.63	0.59
KS	0.47	0.49	0.44
Hosmer-Lemeshow test (Chi square)	36.38	35.32	66.44
Actual 'bad's	4666	2247	1409
Expected 'bad's	4666	2207	1321

Table 6: Model 3 results with 'GDP' in the model

Table 4 gives the coefficients of the economic variables in the two versions. All the 20 variables selected in Model (1) are also in the Model (2) and (3), and their coefficients are very similar to those in Model (1). 'GDP', 'Business Confidence Index' take their original values while 'FTSE' is in standardised form. All three economic variables have negative signs. This suggests that higher 'GDP', 'Business Confidence Index', and 'FTSE' all lead to lower default risk, which is what one would expect. These two models achieve very similar performance in terms of the measures in Table 5 and Table 6. The KS values and the Gini coefficients in Model (2) and (3) are almost the same as those in Model (1), which has no economic variables. On the out-of-time test sample the discrimination values (KS and Gini) in Model (2) are almost the same as Model 1, while the "GDP" model (Model 3) displays a small drop in KS compared with Model (1). In terms of predicting the number of defaults in the two new models the results are startling. The Hosmer-Lemeshow test value in the out-of-time test sample has improved considerably, dropping from 1470 in Model (1) to 63 in Model (2) and 66 in Model (3). This is because of the impressive improvement in the predictions of the number of defaults in the out-of-time test sample. Model (1) without the economic variables only predicts 605 defaulters, but

Model (2) and Model (3) with economic variables predict more than 1300 defaulters, which is much closer to the actual 1409 of defaults. Thus including economic variables has improved the predicted default rate considerably without affecting the discrimination of the scorecard.

One question is whether such a scorecard operates well if the economic conditions are outside those found in the data set. If this is the case then it means the one step approach can be used to estimate LRPD even though there is only data for some of the economic cycle. One could simulate or forecast the economic data for the rest of the cycle. Data for the rest of the cycle was not available in this case to check the results but a comparable situation was found in McDonald et al (2010). There it was clear the regression worked outside the data available provided an appropriate function of the economic variable was used. The choice of function depends on assumptions about what the default rate should be in extremely stressed economic situations.

8 Interactions between economic variables and other variables

In the last section, we saw the macroeconomic conditions did affect the default risk of invoicing companies by changing the population default rate embedded in $s_{pop}(t)$. In this section we investigate whether the interactions between economic variables and other variables has an impact via the $woe(\mathbf{x},t)$ term.

Interaction variables are constructed by multiplying the dummy variables which are indicator variables of the attributes of invoice performance by the economic variables. For example, the variable, 'Account Duration', has been split in the Model (1) scorecard into 4 ranges and so gives rise to 3 dummy variables. (One of them –

Account Duration1- does not appear in the final scorecard).The interaction of 'Account Duration' with 'GDP' is modelled by multiplying the 3 dummy variables by the 'GDP' value, and by including GDP as a separate variable. A similar approach is undertaken for the other non-economic variables selected in Model (1) and the economic variables ('Business Confidence Index', 'FTSE', and 'GDP').

Using Model (2) (including 'Business Confidence Index' and 'FTSE-all share') as the basic model, the interaction variables of 'Business Confidence Index' and 'FTSE-all share' are added, and a stepwise approach is used to select the significant variables. Four interaction variables are selected in the resulting model, Model (4). Table 7 shows the details of economic variables and interaction variables which remain in the final model, where AD, LD6, FR1 and UA2 are invoice performance variables.

Characteristic	Estimate	Standard Error	P-value	Description
Confidence	-0.035	0.003	<0.0001	Confidence Index
FTSE	-0.131	0.026	<0.0001	FTSE All shares
AD_Conf	0.017	0.004	<0.0001	Confidence Index * Indicator variable (Account Duration < 829 days)
LD6_Conf	-0.033	0.009	0.0002	Confidence Index * Indicator variable (Ledger Difference 1.3 - 1.55)
FR1_FTSE	-0.283	0.083	0.0006	FTSE All shares * Indicator variable (Financial Rating '1' or '2')
UA2_Conf	0.021	0.005	<0.0001	Confidence Index * Indicator variable (Utilisation average in last 3 months 0.45-0.77)

Table 7: Coefficients of economic variables and interaction variables in Model 4

	Training	In time Test	Out of time Test
Gini	0.63	0.63	0.57
KS	0.47	0.49	0.44
Hosmer-Lemeshow test (Chi square)	31.75	21.49	81.69
Actual 'bad's	4666	2247	1409
Expected 'bad's	4666	2204	1383

Table 8: Measurement results of the Model 4

In Model (4), the 20 non-economic variables and the two economic variables in Model (2) are still in the model. Their signs are unchanged and their values are only slightly different from their values in Model (2). From Table 8, we can see that in the out-of-time test sample, the Gini coefficient goes down to 0.57 compared with 0.59 in Model (2), which suggests the ranking ability worsens a little when interaction variables are introduced. The good thing is the predicted number of defaulters increases to 1383, which is slightly closer to the actual number.

Using Model (3) (including 'GDP') as the basic model, a similar process of adding interaction variables is undertaken and the results of the resulting model – Model (5) - are displayed in Tables 9 and 10.

Characteristic	Estimate	Standard Error	P-value	Description
ATT_GDP	-0.078	0.019	<0.0001	GDP * Indicator variable (Annualized Turnover 3 month trend 0.86-1)
DTC5_GDP	0.136	0.027	<0.0001	GDP * Indicator variable (DebtTurnClient >= 92 days)
EAT_GDP	-0.105	0.021	<0.0001	GDP * Indicator variable (Entitlement Average 3 month trend =<1.13)
DA1_GDP	-0.309	0.055	<0.0001	GDP * Indicator variable (Disapprovals due to 'age' >5.61)

Table 9: Coefficients of economic variables and interaction variables in Model 5

Again in Model 5, the basic 20 variables are still in the model, but now 'GDP' is not selected even though four interaction variables are selected in the model (see Table 9). From Table 10, we can see that the Gini coefficient and the KS value in the training and the in-time test sample are similar to Model 3, but in the out-of-time test sample the Gini coefficient goes down further to 0.54 and the KS value drops to 0.39. This suggests the interaction variables weaken the discrimination a little. They also do not improve the prediction of the number of defaults compared with Model (3) which had

GDP but with no interaction terms. Also the discrimination drops away over time as the poor out-of-time Gini and KS results show.

	Training	In time Test	Out of time Test
Gini	0.63	0.63	0.54
KS	0.47	0.49	0.39
Hosmer-Lemeshow test (Chi square)	43.64	24.27	146.0
Actual ‘bad’s	4666	2247	1409
Expected ‘bad’s	4666	2212	1309

Table 10: Measurement results of Model 5

9 Conclusions

This paper shows how the decomposition of a log odds scorecard, is a useful way of considering why the introduction of economic and market factors into a default scorecard based assessment systems works so well. Introducing the economic variables as they are is a way of modelling the dynamics of the population default rate namely $s_{pop}(t)$. This leads to a one-step approach to estimating PIT PDs and a substantial improvement in PD estimation. Introducing interaction variables between the economic factors and demographic and performance variables allows one to model the dynamics of the weights of evidence terms $woe(\mathbf{x},t)$ which would affect the ranking of the different borrowers. In the real case study, estimating $s_{pop}(t)$ .in this way works extremely well and overcomes the problem of the poor default rate predictions. However introducing interaction variables between economic factors and borrow characteristics has little effect and if anything is counter-productive.

Thus introducing economic variables into the scorecards deals with the problems identified in the US Security and Exchange Commission report (2008) which reviewed the causes of the subprime mortgage crisis. They pointed out that the scorecards which were built in the early 2000s produced incorrect estimates of the default risks when the economic conditions worsened around 2006-7. This was confirmed by the work of Rona-Tass and Hiss (2008) and Demyanyk and van Hemert (2008) who showed that the scorecard used by US lenders did not change over the period from 2001 to 2008 while the default rate on mortgage loans changed considerably over that period. If the economic variables had been included in such a scorecard it would have dealt with this change in default rate due to the changes in the economic environment.

The case study in this paper is a scorecard applied to invoice discounting – a form of lending to SME companies which is little addressed in the literature despite being a trillion dollar business. The results confirmed that introducing economic variables directly to estimate $s_{pop}(t)$ improves the out-of-time sample prediction accuracy of the models substantially without diminishing the discrimination of the scorecard.

In Invoice Discounting as accounts receivable decreases, the probability of default increases and Loss Given Default is likely to increase since the invoices are collateral on the loan. Having variables which are strong drivers of both PD and LGD suggests there will be strong correlations between these two measures. Building such joint PD and LGD models for ID is an obvious extension of this work.

The failure of consumer and small firm credit scorecards to cope with the financial crisis of 2007-9 can be attributed in part to their inability to deal with the changes in economic conditions. Through our decomposition of the scorecard we show how economic variables can deal with the two parts of the scorecard in different ways. The actual case study then shows that using a simple addition of economic variables introduces a one-step approach to PD estimation which improves the default predictions of the scorecard significantly without affecting its discrimination.

Acknowledgements

We are grateful to the referees for their helpful suggestions in positioning the paper.

References

Altman, E. I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, Vol.23, No.4

Altman, E. I. & Sabato, G. (2007) Modelling Credit Risk for SMEs: Evidence from the U.S. Market, *Abacus*, 43(3), 332-357

Anderson, R. (2007) The Credit Scoring Toolkit, Theory and Practice for Retail Credit Risk Management and Decision Automation, Oxford University Press, Oxford

ABFA (Asset Backed financial Association (2014), Invoice Discounting, Https;//www.abfa.org.uk/public/InvoiceDiscounting. Asp

Aguai S., Forest Jr. L.R., King L., Lennon M.C., Lordkipanidze, B. (2008). Designing and implementing a Basel II compliant PIT–TTC ratings framework. In The Basel Handbook, ed by Ong M.K., pp. 267–297. Risk Books, London

Bank of England (2013), Credit risk; Internal Ratings Based Approach, Supervisory Statement SS1/13.

Basel Committee on Banking Supervision. (2006), Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version, BIS, Basel

Basel Committee on Banking Supervision, (2005) Studies on the Validation of Internal Rating Systems, Working Paper No. 14, BIS, Basel

Basel Committee on Banking Supervision, (2011) Basel III: a global framework for more resilient banks and banking systems: revised version June 2011, BIS, Basel,

Bellotti T., Crook J. (2009) Credit scoring with economic variables using survival analysis. Journal of the Operational Research Society 60, 1699-1707.

Breeden J.L. (2009), Modelling data with multiple time dimensions, Computational Statistics and Data Analysis 51, 4761-4785.

Carlehed M., Petrov A., (2012), A methodology for point-in-time-through-the –cycle probability of default decomposition in risk classification systems, Journal of Risk Model Validation 6, no 3, 3-25.

Crook J, Bellotti T., (2010) Time varying and dynamic models for default risk in consumer loans, Journal of the Royal Statistical Society, Series A, 173, 283,-305.

Demyanyk, Y. & van Hemert, O. (2008) Understanding the subprime mortgage crisis. Proceedings of Federal Reserve Bank of Chicago May 2008, 171-192

Demirovic A., Thomas D.C., (2007) The relevance of accounting data in the measurement of credit risk, European journal of Finance 13, 253-268.

Dietsch, M. & Petey, J. (2002) The credit risk in SME loans portfolios: modelling issues, pricing and capital requirements, Journal of Banking and Finance 26, 303-322

Engelmann B., Rauhmeier R., (2006), The Basel II Risk Parameters, Springer, , Berlin.

Figlewski, S., Frydman, H. & Liang, W. (2007) Modelling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions. Working Paper No. FIN-06-007, NYU Stern School of Business

John S., Vijayans S., Priyak K.Z., (2010) Effect of microeconomic variables on health care loan/lease portfolio delinquency rate, 2010 Enterprise Risk management Symposium, Society of Actuaries.

Liu, J. & Xu, X. E. (2003) The predictive power of economic indicators in consumer credit risk management, RMA Journal , September 2008

Malik M., Thomas L.C. (2010) Modelling credit risk of portfolios of consumer loans. Journal of the Operational Research Society 61, 411-420.

Mays, E. (2004) Credit Scoring for Risk Managers, The Handbook for Lenders, Thomson South Western, Mason, Ohio

McDonald R., Matuszyk A., Thomas L.C. (2010), Application of survival analysis to cash flow modelling for mortgage products, OR Insight. 23 (1), 1-14.

Miu P., Ozdemir B., (2008), Estimating and validating long run probability of default with repect to basell II requirements, Journal of Risk Model Validation 2, no 2, 1-39.

Omni Rand (2002) Invoice Discounting, http://docs.google.com/viewer?a=v&q=cache:HkJy8CufqVoJ:www.omni-rand.ca/dloads/Financing_thru_Factoring_show.pps+worldwide+size+invoice+discounting&

Ong M.K., (2006), The Basel handbook: a Guide for Financial Modelers, Risk Books, London

Rona–Tas, A. & Hiss, S. (2008) Consumer and Corporate Credit Ratings and the Subprime mortgage crisis in the U.S. with some lessons for Germany, SCHUFA, Wiesbaden.

Smith, J. K. & Schnucker, C. (1994) An empirical examination of organizational structure: the economics of the factoring decision. Journal of Corporate Finance, 1, 119–138

Soufani, K. (2000) The role of factoring in financing UK SMEs: a supply side analysis. Journal of Small Business and Enterprise Development, Volume 8, number 1, 37-46

Soufani, K. (2002) On the determinants of factoring as a financing choice: evidence from the UK. *Journal of Economics and Business* 54 (2002) 239-252

Sousa M.R., Gama J., Brandao E., (2013). Introducing time-changing economics into Credit Scoring. FEP Working Paper, FEP-UP School of Economics and Management, University of Porto

Summers, B. & Wilson, N. (2000) Trade Credit Management and the Decision to Use Factoring: An Empirical Study, Journal of Business Finance & Accounting, 27(1) & (2), 37-68

Thomas, L. C. (2009) Consumer Credit Models: Pricing, Profit and Portfolios, Oxford University Press, Oxford

Thomas, L. C., Edelman, D. B. & Crook, J. N. (2002) *Credit Scoring and its Applications*, SIAM, Philadelphia

US Securities and Exchange Commission (2008) Summary report on issues identified in the Commission Staff's Examination of Select Credit Rating Agencies by the Staff of the Securities and Exchange Commission, July 2008

Van Gestel T., Baesens B., (2009), credit Risk management, Oxford University Press, Oxford.

Vinciotti, V. & Hand, D. J. (2002) Scorecard construction with unbalanced class sizes, Journal of the Iranian Statistical Society, 2, 189-205