Can Credit Risk Be Rated Through-the-Cycle?

Karlo Kauko[*]

Abstract

Since the introduction of Basel II, it has been argued that the use of internal credit risk models in banks may strengthen the procyclicality of the financial system. This problem could be alleviated by using through-the-cycle (TTC) ratings. A TTC rating ignores cyclical fluctuations of credit risk. The existence of transitory cyclical fluctuations in corporate credit risk at the company level in Moody's KMV data on Finnish publicly listed companies is tested. The evidence is mostly negative. The distance-to-default of a typical company seems to follow a unit root process. In most cases company level credit risk does not follow cycles with a predictable regularity, and companies that suffered most from the previous downturn may not benefit particularly strongly from the following upturn. Due to continuous entry and exit of firms, average credit risk can be stationary even if the distance-to-default of each individual company is a unit root process.

Keywords: Through-the-cycle rating, Credit risk, Distance to default, Procyclicality
JEL classification: E44, G21, G33

[*] Bank of Finland, PO Box 160, FIN-00101 Helsinki, Finland, karlo.kauko@bof.fi
Thanks are due to Jouko Vilmunen, Matti Virén, Esa Jokivuolle and an anonymous referee for several valuable comments.

1 - Introduction

A central feature of both Basel II and future Basel III capital adequacy systems is the use of internal ratings; many banks are allowed to determine the credit risk of each debtor by using their own internal ratings based approach (IRBA) models. It has been suggested that the system might amplify business cycles. During any recession credit risk tends to worsen, leading to higher capital requirements per exposure, which may diminish the supply of loans and possibly lead to a credit crunch, which would worsen the recession. This literature has been reviewed by Gordy and Howells (2006) and Drumond (2009).

A number of potential solutions to the procyclicality problem have been presented, ranging from countercyclical capital buffers to dynamic loan loss provisioning. Unfortunately there seems to be very little in-depth analysis on the feasibility and usefulness of the various proposed remedies. If the problem and its potential solutions are to be taken seriously, detailed analysis on these approaches is needed.

This paper intends to contribute to assessing one of the proposed solutions, namely Through-The-Cycle (TTC) rating methods and their feasibility. There seems to be no consensus on the precise definition of TTC ratings, but Löffler (2004) has phrased and explicitly used a good conceptual framework. Using its analogue, changes in credit risk are hypothesised to consist of permanent structural changes and transitory cycles. TTC ratings are based on the structural component and ignore the cyclical component. If the cyclical component exists, and if it can be measured with acceptable accuracy almost on real time, TTC ratings can be calculated by eliminating the cyclical component from the perceived point-in-time credit risk.

Rating agencies often claim they present TTC ratings. Empirical evidence indicates that these ratings are slow to react to new information. (Altman and Rijken 2005, Löffler 2004) Agencies' assessments are more useful for investors if the mere level of rating is complemented by its momentum (Güttler and Raupach 2010). Otherwise, there seem to be relatively little literature on the time series properties of credit risk indicators at the debtor level, and there seems to be little or no evidence on the existence of transitory cycles in credit risk.

This paper uses a Merton (1974) type credit risk measure. The data have been provided by Moody's KMV. [1] The data are interpreted as point in time (PIT) estimates of corporate credit risk. It is analysed whether these estimates are subject to temporary cyclical variations that could be filtered out in order to get TTC estimates of credit risk. According to the results, companies typically have got no persistent structural levels of credit risk. Companies' Distance-To-Default (DD) indicators tend to be unit root processes. In some cases the behaviour of corporate credit risk during one phase of the cycle predicts its change during a following phase, but cross-company differences in cyclicality seem highly unstable; companies that suffered most from the previous recession may or may not improve their creditworthiness more than others during the following benign period. Cyclicality may even be systematically reversed. The collapse in 2007-2009 and the subsequent recovery are the only good example of relatively persistent cross-company differences in cyclicality, and even this observation may be due to mismeasurement of credit risk. Moreover, credit risk in small-cap firms has a weak tendency to revert to its past level, but no evidence on such regularity is found among large and medium size companies. Even among small-cap companies the tendency to reversion is not due to the hypothesised propensity of some companies to react particularly strongly to every upturn and downturn. Thus, it seems very difficult to implement rating systems that would reliably filter out the temporary cyclical component of credit risk. Evidence on the existence of this component is mixed and too weak.

The long-term properties of aggregate credit risk cannot be directly derived from the credit risk of individual debtors. If the average credit risk in the economy were a unit root process, it could gradually migrate to any unrealistic extreme. However, simulations demonstrate that the average credit risk in the corporate sector may be stationary even though the credit risk of each individual company is a unit root process. This is possible because the population of companies is subject to continuous entry and exit. Exits through bankruptcy eliminate financially weak companies and prevents the average DD from declining indefinitely. Simultaneously, the entry of new companies prevents the average DD from increasing with no limits. If average credit risk is stationary, any deviation from the long term average is temporary, and can be regarded a "cycle".

[1] These data have been relatively seldom used for research purposes, but a suitable data set on Finnish publicly listed companies was readily available for this project, and the data provider granted the authorisation to use it for research purposes.

The second section describes the data. A theoretical framework and econometric results, including unit root tests and various regression analyses, are presented in section three. The fourth section presents simulation results. The fifth section concludes and discusses some of the findings.

2 - Data

The data for the following analyses were provided by Moody's KMV. These data are calculated using a modified, empirically driven version of the option pricing theory and the Merton (1974) model for corporate credit risk. The inputs consist of market capitalization of corporate equity, its historical volatility and various types of corporate debt on the balance sheet. With these inputs it is possible to calculate the so-called distance-to-default (DD), which is simply the number of standard deviations between the value of assets and a critical threshold value related to corporate debt: if asset depreciation is strong enough, the company defaults. Expected Default Frequencies™ (EDF) reported by Moody's KMV are calculated with a three step procedure.

1) Like in the original Merton model, the value and volatility of firm assets are derived from balance sheet and stock market data using option pricing theory: a company with hardly any net assets becomes more valuable if volatility enhances the option nature or shares. Assets consist basically of all the tangible and intangible items valued by the market.
2) The DD is derived from three variables, namely liabilities on the balance sheet, asset volatility and asset value.
3) The EDFs are calculated by using an empirically fitted model that explains the probability of a corporate default as a function of the DD, the industry and the home country of the company. No normal distribution assumption is used in calculating how likely the default is.

The method is described in more detail by Crosbie and Bohn (2003, p. 10-18). The result is a point-in-time (PIT) measure of credit risk; no changes in DD are ignored as transitory and cyclical.

In the following analyses, a transformation is made for each observation. The original data provided by Moody's consists of 12 months expected default frequency estimates (EDF) calculated by the data provider. The EDF cannot become negative and values higher than 100 % are theoretically impossible. Instead, the underlying DD can have any finite value. Not being constrained by theoretical upper and lower boundaries is a highly desirable characteristic because it makes e.g. unit root testing meaningful. Moreover, using the DD makes it meaningful to carry out regression analyses which are based on the assumption that the same coefficients apply in both average and extreme cases, when the creditworthiness of a company is close to its theoretical limits. Unfortunately the data provider does not disclose enough details to enable us to transform the reported EDF into the original DD. However, by assuming that the empirical distribution is relatively similar to the normal distribution we can calculate a proxy (*Dit*) for the distance to default. If the value of firm assets follows the normal distribution, the probability of ending up in a situation where debt exceeds liabilities equals

$$EDFit = \Phi(-Dit) \equiv ½\{1 + \mathrm{erf}[(-D_1 it - \mu)/(\sqrt{[2\sigma^{\dagger}2]})\ ]\}$$

Where Φ is the cumulative standard normal distribution and erf is the Gaussian error function.[2]

Because μ=0 and σ=1, it follows that

$$2EDF_{it} - 1 = \mathrm{erf}\,[-D_{it}/\sqrt{2}]$$

$$\Longleftrightarrow\ Dit = \sqrt{2}\ \Psi(2\ EDF_{it}) \qquad (1)$$

Ψ is the inverse complementary error function[3] of the standard normal distribution and EDFit is the expected default frequency of company i month t reported by Moody's KMV. The resulting *Dit* may not equal the underlying DD. Nevertheless, this proxy suits the purposes of this paper because the difference between the non-disclosed actual DD and its proxy (*Dit*) is probably within reasonable limits. In the following, *Dit* is loosely called the distance-to-default of company i at the moment of time t, even though it is strictly speaking a mere proxy. The transformation presented in formula (1) yields an explained variable not bounded by theoretical minimum and maximum values.

The predictive power of the Merton and KMV methods has been analysed in academic literature. Bharath and Shumway (2008) calculated their own data by using the same methodology. The authors concluded that the Merton model alone is not a perfect measure of credit risk, but its predictive power was clear. Rather similar results have been reported by Reisz and

[2] See e.g. Johnson & al (1994), ch 13
[3] Complementary error function = 1 -erf

Perlich (2007). Li and Miu (2010) concluded that the Merton model is particularly good at predicting bankruptcies among financially weak companies. These previous analyses are now taken as sufficient evidence on the relevance of the data.

The following analyses are carried out with monthly data on annual default probabilities of non-financial companies quoted on the Helsinki Stock Exchange. With the exception of banks and insurance companies, all the 119 firms with at least five years of data between August 1999 and June 2010 were included. Financials have been excluded because the focus of discussions on procyclicality is on the creditworthiness of banks' customers. Companies with less than five years of history would hardly be of any use in the following analyses. Each observation corresponds to the last business day of the month. It might be interesting to extend the analysis to other countries or longer periods of time, but such data were not readily available to the author.

Unfortunately, the sample is still relatively short for analysis on cyclical phenomena, but fortunately it describes a highly cyclical economy. The beginning of the sample describes a benign period; the annual growth rate of Finnish real GDP reached 5.3 % in 2000. A moderate slow-down took place in 2001-2003, but growth rates remained positive. A new benign period began in 2004, and growth peaked again at 5.3 % in 2007. In 2009 output collapsed by 8 %.

Since 2006 companies listed on the Helsinki Stock Exchange have been divided into large, mid-cap and small-cap companies. Because stock prices are a highly important driver of changes in default probability estimates in the data, especially in the short run, and because pricing inefficiencies can be particularly severe in the case of small-cap stocks (See Avramov 2002 or Hung et al 2009), some of the analyses are carried out either without small-cap companies or specifically for small-cap companies only. In the following, each company was classified according to its size category in early 2010. Companies not quoted in 2010 (12 cases) were classified by their 2002 market capitalization. If the market cap was less than EUR 150 million, the company was classified as small; this threshold is the same as the one introduced by the Helsinki Stock Exchange in 2006. There are 60 companies in the small-cap category of the sample. Two companies exited the sample because of bankruptcy. Ten companies were delisted during the observation period because of other reasons, mainly because of being target of M&A. No particular industry dominates the Helsinki Stock Exchange, even though the number of companies in electronics and wood processing may be larger than in most other markets.

There seem to be almost no seasonal variation in DDs . When the first difference of *D*s in the monthly panel data (D_{it}-D_{it-1}) is explained with panel OLS with no other explanatory factors than month specific dummy variables, the overall fit of the regression is almost zero (R^2 =0.01). The average distance to default seems to have slight tendency to marginally increase in January and to decrease during the rest of the year, especially in June, but these effects are insignificant.

The most important short-term driver of credit risk in this model is the change in stock quotations because balance sheet data is updated infrequently, only once a year. There is an ongoing debate on whether equity prices follow a pure unit root process, and whether temporary fluctuations can be identified in stock market data. (See e.g. Narayan and Narayan 2007; Bali et al. 2008; Choe et al. 2007). Whatever the best answer to the question on the time series properties of stock quotations is, the result cannot be directly applied to credit risk implied by the KMV model because stock prices are not the sole driver of DDs, especially not in the long term. The Merton model is based on the assumption that firm assets follow a random walk until the predefined future date. This does not necessarily imply the DD must follow a unit root process in any data that can be meaningfully analysed with the Merton method because factors not related to exogenous shocks to the value of corporate assets affect the default probability. Managers and shareholders would typically react to different difficulties and opportunities by, for instance, expanding the undertaking, disposing of business units, issuing more equity capital to strengthen the solvency, adapting the dividend policy etc. For instance, it has been found that firms near credit rating upgrades or downgrades issue less debt relative to equity than firms not near a change in rating (Kisgen 2006) and recently downgraded firms typically reduce leverage (Kisgen 2009). The combination of different drivers of credit risk creates a very complicated process. This highly complex system is now regarded as a black box; the focus is on the typical time series properties of credit risk, not on causal mechanisms behind it.

3 - Empirics

3.1 - A simple model with three empirical predictions

The following analysis is based on the analogue of Löffler (2004), one of the few papers that phrase out an explicit view on the precise nature of credit risk cyclicality. The Löffler decomposition is applied to these data. If there are both genuine cycles and a non-cyclical structural TTC component in the data, the distance to default of company i is determined by the following process, or at least by something very similar.

$$D_{it} = S_{it} + C_{it} \tag{2}$$

where D_{it} is the point-in-time (PIT) proxy for the DD of company i on the last day of the month t. S_{it} is the value of the relatively stable structural through-the-cycle (TTC) component of credit risk of the company i at the moment of time t. The structural component follows a unit root process, is subject to infrequent abrupt shifts or remains constant, but does not undergo transitory cycles. C_{it} is the cyclical component of credit risk. The cyclical component C_{it} is assumed to be stationary with mean zero.

$$C_{it} = \alpha\, C_{it-1} + \beta_i\, C_{macro,t} + \varepsilon_{it} \tag{3}$$

where $0<\alpha<1$. The company specific cyclicality parameter β_i is positive for most firms. The idiosyncratic shock of the company i in period t is denoted ε_{it}. If α were equal to 1, macroeconomic shocks would cause permanent changes in credit risk and there would be no transitory cycles to be filtered out in credit ratings. It would not be possible to distinguish the shock term ε_{it} from idiosyncratic shocks to S. The macro level cyclical shock $C_{macro,t}$ has got mean zero and it is common to all firms, and it causes correlation in firms' credit risks. Both macroeconomic and idiosyncratic shocks may be autocorrelated.

A PIT rating would be based on D_{it}. A TTC rating would ignore the value of C_{it} by dropping off the whole component and focusing exclusively on S_{it}. If the structural component remains constant, the moving average of point-in-time ratings would normally be a satisfactory proxy for S_{it}. Alternatively, one might assign the cyclical component C_{it} a firm specific constant negative value to yield default probabilities under adverse conditions. The choice between these approaches would affect the level of TTC default risk, not its variation over time.

Parameters for this simple model can hardly be estimated. Attempts to apply state-space methods to the data and to treat *C* and *D* as state variables failed. Nevertheless, the model yields three testable hypotheses.

1. Relatively many companies may have got a seemingly stationary D_{it}. Variations are entirely or almost entirely due to transitory fluctuations of C_{it}. The structural component either remains constant or variations in it are too weak to be detected.

2. There is some negative serial correlation in differenced D_{it} ; upturns are followed by downturns. The negative serial correlation is caused by fluctuations of *C*. This parameter is subject to idiosyncratic shocks (ε_{it} in equation 3) and cyclical ones ($C_{macro,t}$). Because *C* is characterised by a tendency to revert towards zero, any shock to it gradually vanishes. After the first month only the fraction $\alpha<1$ of a shock is left, after two months only the fraction α^2 etc. New shocks to both the structural and the cyclical component make the relationship between past and future changes of D_{it} merely statistical, but it should be detectable.

3. Unless the term beta of equation 3 changes frequently, companies' reactions to cycles remain relatively constant. If the D_{it} of a company deteriorates more than average during a recession, it will improve remarkably strongly during the following benign period because the company specific parameter β_i in equation 3 must be large. If it were not, the impact would have been weak. If the cumulative sum of macroeconomic shocks during a boom phase is +1, and the sum of negative macro shocks is -1 during the recession, the expected value of the change in the D_{it} is $+\beta_i$ during the boom and $-\beta_i$ during the recession. If there were no idiosyncratic shocks ($\varepsilon_{it} = 0$ for all i and t), and if structural components remained constant, there would be a perfect negative correlation.

If these hypotheses find no support in the data, the equations 2 and 3 are a very poor description of reality, and the existence of any cyclical component remotely similar to *C* is unlikely. If there is something we could call "cycles", the phenomenon is probably much more complicated.

3.2 - Testing the first empirical prediction

Panel unit root tests typically have got higher power than unit root tests on individual series (See Maddala and Wu 1999 and references therein). However, serial correlation, especially negative one, may seriously bias many unit root test results. (See Schwert 1989, Hlouskova and Wagner 2006) Such negative correlation is found in several series of the sample (see section 3.3). The ADF test can be applied, even though the presence of serial correlation accentuates the importance of lag length selection in unit root tests; the use of standard Akaike and Schwarz criteria would often lead to excessively short lag structures. (Ng and Perron 1995; Lopez 1997). The modified Akaike criterion is used in the following analyses. It takes into account the consequences of the potentially biased sum of autoregressive coefficients. Using the modified criterion significantly improves the reliability of unit root test results (Ng and Perron 2001).

Many panel unit root tests simply evaluate the joint significance of p-values obtained by testing each series separately. The null hypothesis is unit roots in the whole data, the alternative hypothesis being that at least some of the series are stationary. If most of the variation in DDs is cyclical, it would be natural to expect that some subgroup of companies has not undergone any structural changes, or changes are too weak to be detected, and the null hypothesis of unit roots in the whole data would be rejected.

Panel unit root tests were run with the data set on all the D_{it} observations. Three different test statistics are reported. The Fisher ADF-approach applies the chi squared distribution to a function of logarithmic p-values of ADF tests on individual series. Choi (2001) proposed a Z-statistic of the significance of unit root tests. A test statistic based on averaging individual ADF test statistics has been presented by Im et al (2003). As can be seen in table 1, there is no evidence that the proxy for DD of any company in the data would be entirely driven by cyclical fluctuations. If the cyclical component exists, the vast majority of companies, and possibly all of them, have had major changes in the structural component of credit risk.

Table 1 Panel unit root tests

Method	Test statistic	p-value
ADF Fisher Chi-squared, intercept, no trend	232.20	0.594
ADF Choi Z stat, intercept, no trend	-0.767	0.222
Im, Pesaran Shin W-stat, intercept, no trend	-1.035	0.150

Panel unit root tests (levels) for 119 companies; D_{it} in monthly data
(Aug 1990 - Jun 2010)
Modified Akaike criterion in lag-length selection

Some interesting methods cannot be applied to panel data because there is no straightforward way to evaluate the joint significance of tests on individual series. Ng and Perron (2001) recommended a GLS detrending combined with the modified Akaike criterion in lag length selection in the presence of negative serial correlation in the error term. This method was applied to each company separately. It was found that the unit root hypothesis would in some cases be rejected at each significance level, but the number of rejections does not differ much from what one would expect as a random outcome. (See table 2) The results corroborate the findings of panel unit root tests reported in table 1.

Table 2 Unit root testing of company level data

Method	Ng-Perron; Modified Akaike criterion, GLS detrending	Shift in level, abrupt shift dummy	Shift in level, smooth exponential shift
Test statistic	Elliot-Rothenberg-Stock (1996)	Lanne & al (2002)	Lanne & al (2002)
Nr of rejections at the 10 % level	12	11	10
Nr of rejections at the 5 % level	4	7	8
Nr of rejections at the 1 % level	1	0	1

Separate unit root tests (levels) for 119 companies, D_{it} observations in monthly data (Aug 1999 - Jun 2010)
Modified Akaike criterion in lag-length selection
In level shift tests the number of lags has been obtained for each company separately by running a separate ADF test with company specific trends and fixed effects. The number of lags proposed by the modified Akaike criterion was used in level shift tests in determining the break date and unit root testing itself.

Even if there were only one major change in *S*, the methods applied so far would typically accept the null hypothesis of unit roots. Unit root tests that allow for the presence of a structural break were applied to the data. Following Lanne et al (2003), the analysis was begun by optimising the number of lags by running separate ADF analyses with individual intercepts and trends. The number of lags suggested for each company by the modified Akaike criterion was used at the following stages; these preliminary analyses were not used for any other purpose. As a second step, the date of the structural break was determined endogenously. A deterministic component, consisting of constants and the shift caused by a structural break, was deducted from the original series by using a GLS procedure. The structural breaks were identified in two different ways, first by assuming an abrupt shift dummy in level and then by assuming a smoother exponential shift. Finally, the unit root test was run on residuals after deducting the deterministic component. (See Saikkonen and Lütkepohl 2002). Lanne et al (2002) tabulated critical values for the t-value of the lagged non-deterministic component of the original series. Again, each company was tested separately because there is no straightforward way to apply the method to panel data. As can be seen in table 2, the number of rejections is roughly equal to what one would expect as a random outcome. Hence, the number of structural shifts of credit risk seems much higher than one in the vast majority of companies. The unit root component (*S* in equation 2) is a major driver of credit risk in most or possibly all companies, leaving open the question on the existence of the stationary component *C*.

3.3 - Testing the second empirical prediction

The evidence presented in section 3.2 corroborates the hypothesis that unit root processes are major drivers of credit risk, but it does not necessarily prove the hypothesised cyclical component *C* is non-existent. In order to test the existence of component *C* in equation 2, a few panel regressions were run with annual data. The company level DD for each year is the three months simple moving average ($\check{D}_{it}=\{D_{it}+D_{it-1}+D_{it-2}\}/3$) of December. The analyses were carried out with data on the 117 companies for which it was possible to find data on six consecutive fourth quarters. Because the data does not seem stationary, the regression was run in differences. The annual difference of the distance-to-default was regressed on its past values. No weighting was applied. Obviously the limited time dimension makes it pointless to try to identify company specific effects. The results are presented in table 3. The first equation was run with no period or company specific effects. The Hausman test indicated that the random period effects model is more suitable than the fixed period effects model (Chi sq = 0.21), and the 3rd equation could be considered the main model for the whole data. The first, second and fourth lags are statistically significant. Thus, the DD is characterised by an observable tendency to return to its past level.

Interestingly, the tendency to reversion is related to firm size. There is no evidence on the existence of temporary fluctuations and the cyclical component *C* in the data if small-cap firms are excluded from the sample. (Equations 4-5 in table 3). If, instead, the focus is on small-cap firms, the DD clearly tends to return to its past values (Equations 6-7). It is difficult to say whether this is a genuine property of credit risk in small-cap firms or something related to potentially inefficient pricing and illiquidity of small-cap companies on the stock exchange. There is almost no difference in the standard deviation of the explained variable between the two size groups; in both groups it is slightly higher than 0.36.

Table 3 Annual difference of DD as a function of lagged differences

	1	2	3	4
	No year specific effects	Fixed period effects	Random period effects	Small-cap firms excluded; fixed period effects
Constant	0.02	0.01	0.01	-0.02
	(0.2)	(1.3)	(0.1)	(-2.2)**
$\check{D}_{it-12}$-$\check{D}_{it-24}$	-0.07	-0.09	-0.09	-0.13
	(-0.7)	(-0.9)	(-2.1)**	(-1.0)
$\check{D}_{it-24}$-$\check{D}_{it-36}$	-0.04	-0.09	-0.09	0.00
	(-0.6)	(-2.1)**	(-2.2)**	(0.0)
$\check{D}_{it-36}$-$\check{D}_{it-48}$	-0.16	-0.05	-0.05	-0.07
	(-1.9)*	(-1.7)*	(-1.3)	(-1.2)
$\check{D}_{it-48}$-$\check{D}_{it-60}$	-0.14	-0.10	-0.09	-0.06
	(-3.9)***	(-2.6)***	(-2.7)***	(-1.3)
R2	0.04	0.39	0.02	0.51
F	6.592***	44.01***	3.819***	36.81***
N	624	624	624	330

t-values corrected for heteroscedasticity and autocorrelation
*= 10 % signif, ** = 5 % signif, ***=1 % signif
Explained variable $\check{D}_{it}$-$\check{D}_{it-12}$; Decembers only

Table 3 (continued)
Annual difference of DD as a function of lagged differences

	5	6	7
	Small-cap firms excluded; random period effects	Small-cap firms only; fixed period effects	Small-cap firms only; random period effects
Constant	-0.02	0.05	0.05
	(-0.2)	(4.5)***	(0.6)
$\check{D}_{it-12}-\check{D}_{it-24}$	-0.13	-0.10	-0.10
	(-1.0)	(-1.2)	(-1.2)
$\check{D}_{it-24}-\check{D}_{it-36}$	0.00	-0.17	-0.17
	(0.0)	(-4.6)***	(-4.8)***
$\check{D}_{it-36}-\check{D}_{it-48}$	-0.07	-0.04	-0.04
	(-1.3)	(-0.7)	(-0.7)
$\check{D}_{it-48}-\check{D}_{it-60}$	-0.06	-0.14	-0.14
	(-1.3)	(-3.2)***	(-3.2)***
R2	0.02	0.31	0.05
F	1.75	13.857***	3.643***
N	330	294	294

t-values corrected for heteroscedasticity and autocorrelation
*= 10 % signif, ** = 5 % signif, ***=1 % signif
Explained variable $\check{D}_{it}-\check{D}_{it-12}$; Decembers only

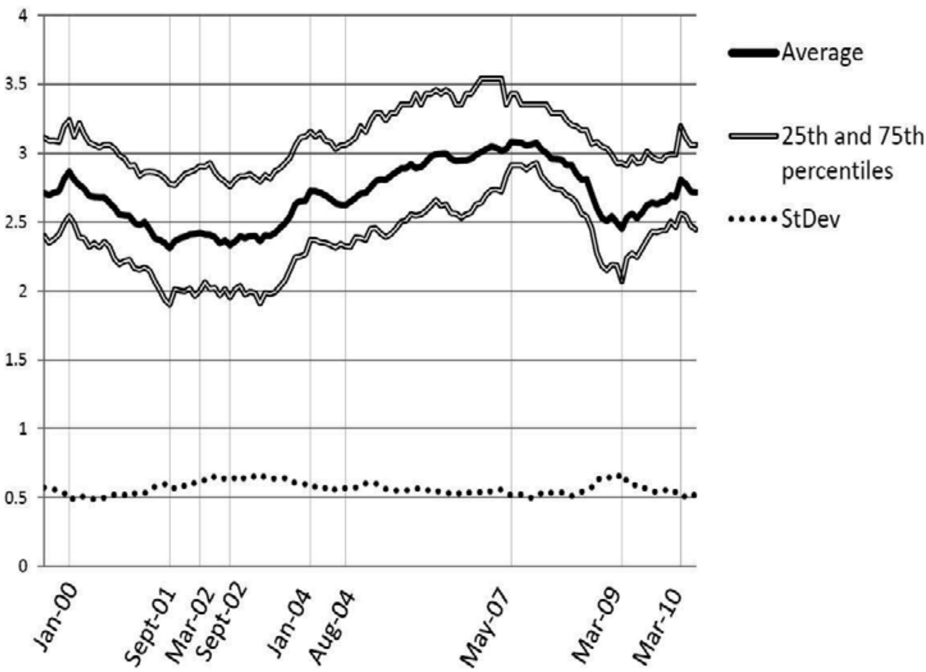
3.4 - Testing the third empirical prediction

In section 3.3 it was found that firm-specific DD has got, at least in small-cap firms, some tendency to return to its past level, but such temporary fluctuations may be due to either macroeconomic shocks or idiosyncratic factors. From the point of view of procyclical capital requirements, idiosyncratic variations are uninteresting. The macro level credit cycle is not directly observable but it affects the average credit risk in the economy. The figure 1 presents the development of the mean of the DD. Its cross-sectional 25th and 75th percentiles and standard deviation are also presented in the chart. If anything, cross-sectional variation across firms has widened when credit quality has deteriorated. Spectral analysis found no satisfactory evidence on the existence of genuine cycles in the average. At least in quarterly data these fluctuations are correlated with the proxy for output gap[4]; the immediate correlation between the output gap and average DD is +0.51 in levels.

Any observation higher or lower than any of the following or preceding five observations is now classified a local extreme, and the transition from one local extreme to the next one is called a phase. Local maxima of the average DD can be identified in January 2000, March 2002, January 2004 and May 2007. Local minima are found in September 2001, September 2002, August 2004, and March 2009. These turning points are marked in chart 1 by vertical lines. This set passes one compliance test proposed by Bry and Boschan (1971, p 16-29): peaks and troughs alternate. Using the method of Bry and Boschan, cyclical peaks and troughs in the mean DD were identified. There were only three cyclical turning points that satisfied all the Bry-Boschan criteria, namely September 2001, May 2007 and March 2009. Moreover, another criterion of a cycle was used; phases where the value of the average of DDs changed more than by 0.3 from one local extreme to the next one are classified as cyclical irrespective of the duration of the phase.

[4] The output gap was calculated as the Hodrick-Filter residual ($\lambda = 1600$) run on logarithmic real GDP data for the period Q1/1995 - Q4 /2013; Data after the second quarter of 2010 were from the latest Bank of Finland forecast.

Figure 1 Average proxy for DD and its cross-sectional variation



If the business cycle sensitivity of each firm, the parameter βi in equation 3, remains broadly constant for lengthy periods of time, the development of credit risk during the previous phases of the cycle should be relevant to changes of credit risk during any upturn or downturn. The drastic deterioration of credit quality between May 2007 and March 2009 should have been stronger among cyclical companies that experienced particularly strong improvements of creditworthiness during the preceding benign phase. The change of DD during the upturn in September 2001 – May 2007 should obtain a negative regression coefficient as a correlate of the change in DD in May 2007 – March 2009. OLS results of cross-sectional analyses are presented in table 4. Results for the whole sample, small-cap companies and large to mid-cap companies are presented separately. These results are completely inconsistent with the hypothesis that the credit quality of each company would react to macroeconomic factors as described by equation 3. The slight tendency to reversion among small-cap companies observed in section 3.3 is not reflected in cyclical variation. If there are company specific differences in the sensitivity to cycles, these sensitivities vary so frequently that information on past cyclicality is of virtually no use in predicting future cyclicality.

Table 4 Change of distance to default in 2007-2009

	Change in average D	C	Change in D in Sept 2001 - May 2007	R2	F
Whole sample N=101	-0.656	-0.741 (-6.8)***	0.107 (0.9)	0.01	0.79
Small-cap N = 46	-0.504	-0.483 (-4.3)***	-0.027 (-0.2)	0.00	3.43
Large and mid cap, N = 55	-0.783	-0.984 (-5.9)***	0.255 (1.5)	0.04	1.97

Cross-sectional OLS analyses with identifiable cycles
Change of *D* in May 2007-March 2009 as a function of its past change
t-values in parentheses, corrected for heteroscedasticity (White)

The second method to differentiate cycles from insignificant random variation identifies more cycles to be analysed. In the cross-sectional OLS analyses presented in table 5, it is tested whether firm specific changes in DDs during the strongest swings are related to changes during previous accentuated phases. Phases where the average of DDs has changed by less than 0.3 are excluded as both explanatory and explained variables. As hypothesised, the upturns in September 2002 - January 2004 and August 2004 - May 2007 were stronger in companies that were affected particularly severely by the downturn in January 2000 - September 2001. Instead, the statistical relationship between the upturns in September 2002 - January 2004 and August 2004 - May 2007 is inconsistent with the permanent cyclicality hypothesis; companies that benefited particularly strongly from the first upturn benefited less than others from the second one. The most dramatic phase of the sample, namely the collapse in 2007-2009, is not related to previous strong phases. Instead, the permanent cyclicality hypothesis works as expected in the recovery in 2009-2010; companies that were affected particularly strongly by the crisis recovered more than others. Hence, the evidence is mixed.

Table 5 Change of DD during strong phases

		Mar09-Mar10	May07-Mar09	Aug04-May07	Sept02-Jan04
		Upturn	Downturn	Upturn	Upturn
		12 months	22 months	33 months	16 months
	Mean of explained variable	0.356	-0.687	0.428	0.394
	Constant	-0.05 (-0.7)	-0.76 (-5.8)***	0.50 (9.1)***	0.28 (8.0)***
Downturn	Change of *Di* in May07-Mar09	-0.51 (-7.6)***			
Upturn	Change of *Di* in Aug04-May07	<u>-0.14</u> (-1.1)	<u>0.06</u> (0.4)		
Upturn	Change of *Di* in Sept02-Jan04	0.19 (1.7)*	-0.02 (-0.1)	<u>-0.42</u> (-3.7)***	
Downturn	Change of *Di* in Jan00-Sept01	-0.05 (-0.7)	-0.09 (-0.9)	-0.17 (-2.0)**	-0.22 (-4.0)***
	R2	0.51	0.01	0.14	0.15
	F	21.7***	0.37	7.36***	17.4***

Explained variable *Di end-of-phase* - *Di* end of previous phase
t-values corrected for heteroscedasticity (White)
Averages of the explained variable may differ from table 4 because samples may differ.
* = 10 % significance, ** = 5 % significance, *** = 1 % significance
Coefficients with "wrong" sign underlined

Cycles of different companies might be imperfectly synchronised; each company may react to the same macroeconomic factors but some of them may react faster than others. The most extreme phase of the whole data, i.e. the financial crisis of 2007-2009, and the preceding lengthy benign period were analyzed in order to test this hypothesis. Each extreme value is the highest or lowest monthly value observed during a certain period of time. The extremes of DD for each company i are defined as follows.

- $D_{i\ crisis} = \mathrm{Min}[D_{i\ Nov\ 2008} \ldots D_{i\ Sept\ 2009}]$
- $D_{i\ peak} = \mathrm{Max}[D_{i\ Jan2007} \ldots D_{i\ Nov\ 2007}]$
- $D_{i\ pre\text{-}peak} = \mathrm{Min}[D_{i\ Jan\ 2004} \ldots D_{i\ Dec\ 2006}]$

These data were used to calculate differences $D_{i\ crisis}$-$D_{i\ peak}$, and $D_{i\ peak}$ -$D_{i\ pre\text{-}peak}$. One would expect that there must be a highly negative correlation between these differences because cyclical companies experienced the strongest improvement before the crisis and the worst collapse during it. The correlation between these variables turns out to be non-significant and positive (+0.12, N=108).

4 - Stationarity at the aggregate level - simulations

The section 3.1 demonstrated that instead of being stationary, firm-specific credit risk has probably got a unit root. If the distance-to-default of every company follows a unit root process, we might draw the conclusion that the average credit risk in the economy should also be a unit root process. Companies' average annual default probability might gradually migrate to 99.99 percent, or the economy might end up in a situation where corporate bankruptcies are unheard of. These extreme alternatives do not seem realistic; to be credible, the unit root hypothesis should be consistent with reasonable assumptions on average credit risk in the economy.

In principle, the average of a large number of unit root processes could be stationary in the presence of a suitable cointegration. This possibility is not realistic in this case; the correlation between DDs should be negative in the long term. It is difficult to see which forces would cause this negative relationship. A more plausible explanation is related to continuous entry and exit. Only those companies that exist at a given moment of time are included in the average.

In order to draw some conclusions on the implications of this way of thinking, a number of simulations were run. The oldest company of the sample in section 3 was 360 years old in 2009, the last full year of the sample.

Taking this as the starting point in model calibration, simulations were run for 360 consecutive "years". During each period t a number of new companies were established. The number of start-ups (*Nt*) grows in a growing economy. The number of new start-ups in period t is determined by a very simple function.

$$Nt = 5 \text{ Round}(1.025^{t-1}) \tag{4}$$

In total, 1 450 645 companies were established in each simulation, about 91.5 % of them during the last 100 "years". The distance-to-default develops as a unit root process, unless the firm fails.

$$D_{it} = D_{it-1} + \varepsilon_{it} \tag{5}$$

where ε_{it} is an iid normally distributed random variable with mean zero and standard deviation 0.198. This standard deviation was calibrated to produce the cross-sectional standard deviation of the ten year difference in section 3 data. In order to save computing capacity, it was assumed that firms of the same cohort can be divided into 5 groups of equal size. Companies in the same group are always identical and they are assigned the same distance-to-default. Each group *i* of start-ups is assigned the distance-to-default $1.33+\varepsilon_{it}$. The parameter value 1.33 was calibrated to produce the observed five-year survival rate of newly established Finnish firms, which is reported by Nurmi (2004, p.40) to be slightly less than 60 %. Firms' DDs change already before the first possible moment of bankruptcy at the end of the entry period.

Firms do not exit in any other way than through bankruptcy. Each year t the firms of the group i exit iff

$$\Phi(D_{it}) < R_{it} \tag{6}$$

where Φ is the cumulative standard normal distribution and R_{it} is an evenly iid distributed random variable between 0 and 1. A firm may exit even during the same year it is established.

The first 260 "years" were discarded and nothing but the last 100 moments of time were used in subsequent analysis. The average distance-to-default was calculated for each year taking into account nothing but firms that had been established but not failed by the end of the year. This simulation was repeated a hundred times, producing a panel of ten thousand observations on average DD in the artificial economy. These time series converge towards an equilibrium where the average DD seems constant. The mean DD was 1.86 for the first observation (t=261) and 1.87 for the last observation (t=360).

The panel unit root test proposed by Levin et al (2002) suits the setting particularly well. The method assumes that series of the sample are identical with respect to the first-order partial autocorrelation, but other parameters can vary across units. The null hypothesis is that each time series of the panel has got a unit root, the alternative hypothesis being that none of them have. As pointed out by the original authors (p 18), the method is of limited use unless the time series are free from contemporaneous correlation and identical with respect to the presence of a unit root. Because of the nature of the data generating algorithm these criteria should be satisfied in the artificial sample. The model applied in the following is numbered 2 by Levin et al (LLC); each time series has got its own intercept, but no trends are allowed in the data. It can be essentially important to choose the right bandwidth, a parameter value used in correcting test statistics for serial correlation. Westerlund (2009) tested with Monte Carlo simulations different details concerning the use of the LLC test. He strongly recommended the use of the bandwidth selection criterion $Ki = 3.21T^{(1/3)}$ whereas the lag length selection criterion proved to be of secondary importance. The Schwarz-Bayesian lag length selection criterion was one of the methods considered by Westerlund whereas the modified Akaike criterion was not; results based on both criteria are reported in table 6. Moreover, the results of panel ADF were calculated for comparison. These methods unanimously reject the unit root hypothesis.

Table 6 Unit root testing of artificial panel data

	Stat	P-value	Lag selection criterion
Levin-Lin-Chu 2002 (LLC)	-15.09	0.000	Schwartz-Bayesian
Levin-Lin-Chu 2002 (LLC)	-7.56	0.000	Modified Akaike
Panel-ADF; Fisher chi squared	511.82	0.000	Modified Akaike
Panel-ADF; Choi Z-stat	-11.47	0.000	Modified Akaike

The LLC test was run with the Bartlett kernel, the bandwidth criterion $K_i = 3.21T^{(1/3)}$; No trends are included
N=100; T=100

The stability of the average DD is due to the balance between two counteracting forces. The average distance-to-default among surviving firms of each cohort increases over time when disproportionate exit takes place among firms whose DD has deteriorated, or not improved enough to promise a long life. Simultaneously, the relative weight of each cohort steadily declines even if no exit takes place because increasing numbers of new firms enter. In real life exit may take place even without bankruptcy and default, for instance because of voluntary closures due to M&A or non-economic factors, such as retirement of the entrepreneur. Hence, in real life the long-term forces that would force the average DD upwards are counteracted by multiple factors.

This way of thinking offers a possible explanation to the observed propensity of newly established firms to fail at a higher probability than seemingly similar older firms. An old firm must have survived for many years to be observed, which is unlikely unless the firm has developed a low probability of default soon after entry. This reasoning is essentially the same as the one presented by Thompson (2005) who argues that high exit rates are commonplace among new firms because this group includes a lot of companies that were established with an inherently high failure probability, which is unlikely in the case of firms that have managed to survive several decades.

5 - Conclusion

It has been proposed that the presumed procyclicality problem of capital adequacy regulations could be alleviated by using through-the-cycle (TTC) ratings in banks' IRBA models. Many discussions on this possibility are not based on any explicit assumptions on the nature of business and credit cycles, let alone their mutual interactions. However, the feasibility of TTC ratings depends on the time series properties of credit risk at the debtor level. This paper has presented some empirical evidence on this issue. The original KMV data provided by Moody's was used to calculate monthly proxies for distance-to-default (DD) of 119 Finnish publicly listed companies. A simple model based on the decomposition of Löffler (2004) was presented. The model was based on the assumption that companies' DDs contain a stationary cyclical component. The model had three major empirical predictions to be tested, but these predictions did not describe the data particularly well. It was found that few if any companies are characterised stationary DDs. DDs of small-cap companies have got some tendency to revert to past levels. However, no such phenomenon can be observed in a sample consisting of large and mid-cap companies. In some cases the development of DD during a cyclical phase predicts developments in the following phase, but this is not always the case, and the cyclicality of a typical company seems highly unstable.

The co-existence of a cyclical component and a structural component finds little or no support in the data. If there are 'cycles' in company level credit risk, they are so different from hypothesised temporary fluctuations driven by macro level phenomena that it is difficult to imagine what they would be. Therefore, TTC rating philosophies based on the idea that transitory cycles are filtered out and the rating should be based on the structural component do not seem fully feasible because such cycles do not seem to exist. Even among small-cap firms the negative serial correlation is of little use in eliminating cyclical variations because the reverting component accounts for a relatively insignificant part of total variation of DDs, and even among them the cyclicality of a company can be suddenly reversed.

In one particular case there seemed to be some evidence on the existence of hypothesised transitory cycles. It was found that the collapse of credit quality during the crisis in 2007-2009 was strongly correlated with the subsequent recovery. The relatively clear cyclical component in 2007-2010 could be due to measurement problems. Possibly fire sales during the worst crisis period affected the prices of illiquid stocks pushing them below their fundamental values. This mispricing may have been temporary, which would explain the recovery of perceived credit risk in 2009-2010. However, both the deterioration during the crisis and the recovery in 2009 were, on average, somewhat weaker among less liquid and potentially mispriced small-cap companies. This may not necessarily prove that the observed cyclical component of credit risk in 2007-2009 was not due to mispricing. It could simply be due to the limited number of regulated institutional investors investing in Finnish small-cap companies. Fire sales induced by solvency requirements cannot take place in absence of such regulated institutional investors.

Another often used definition of TTC ratings combines the current credit risk with its perceived sensitivity to the macroeconomic environment. If two debtors are characterised by equal probabilities of default at the moment, the company with a higher sensitivity to business cycles would be given a weaker credit rating. Emphasising this vulnerability in credit quality should be possible, provided it is possible to estimate the sensitivity of different debtors. Unfortunately, it was found in section 3.4 that the cyclicality of a typical corporate debtor undergoes frequent and fundamental changes. Firms that were strongly affected by the previous stage of business cycles can suddenly become particularly insensitive to the macroeconomic environment, and vice versa. At least historic correlations are of little use in assessing business cycle sensitivity in TTC ratings.

Some of the findings in section 3 could be consistent with the hypothesis of multiple credit risk drivers instead of one general macroeconomic factor. Some earlier contributions (Koopman et al 2009; Jiménez & Mencía 2009) have found evidence in favour of this hypothesis but a simple factor analysis[5] on section three data found no evidence to support it. This analysis was by no means exhaustive. Applying sophisticated factor analysis techniques and different rotations to the same data set might be an interesting way to expand the analysis; a satisfactory explanation to the irregular changes of companies' cyclicalities could be found. However, the Basel II and III frameworks are based on the Asymptotic Single Risk Factor approach. Even if these hypothetical multiple factors could be identified, it may not be obvious how to eliminate the impact of any truly cyclical factor without abandoning the basic assumptions of the capital adequacy framework itself.

[5] Details are available from the author upon request.

Even though companies' credit risks seem unit root processes, simulations of section four demonstrate that the average credit risk of a representative loan portfolio may be stationary and therefore subject to transitory fluctuations. This paradox is explained by the entry and exit of debtors. Hence, it would be possible to make a cyclical adjustment to the whole loan portfolio, at least if the loan portfolio is subject to a same kind of entry and exit of firms as the artificial economy of section four. On the other hand, it is not obvious why it would be useful to calculate the credit risk of each debtor by using highly sophisticated methods, and then to apply a coefficient that prevents the variation of the average capital requirement. Choosing a suitable constant risk weight for the whole portfolio would yield the same capital requirement with much less work. A somewhat more sophisticated way to implement TTC ratings is to apply smoothing at the rating category level. Potential debtors can be ranked according to their PIT credit risk, and they can be assigned ratings according to the credit risk relative to other debtors. The number of debtors in each category is held broadly constant by tightening criteria during cyclical upturns and loosening them during downturns. The risk weight of each rating category is held constant irrespective of how the actual default probability of the category develops. This approach was tested in the simulations by Gordy and Howells (2006); they concluded that the cyclical behaviour of capital requirements under this TTC rating system depends on the cyclical development of banks' policies concerning new lending. However, it can be argued that even under this more sophisticated smoothing one runs the risk of making complicated calculations in order to obtain a predefined, exogenously given result. If half of the loan portfolio is always assigned the weight 1.2, and the other half the weight 0.8, the average risk weight obviously cannot differ from 1. It would be more meaningful to use all would-be borrowers in the economy as benchmark when debtors are assigned ratings, if the bank has got access to the required information set.

Previous empirical literature has not reached a clear consensus view on the time series properties of the development of real GDP. Nelson and Plosser (1982) questioned the traditional view of GDP growth as a trend stationary process. Some authors claim that output grows as a unit root process whereas some others have reached the conclusion that long-term economic growth is a trend-stationary process (see Beechey and Österholm 2008). Moreover, there may be structural breaks in the development of GDP (Papell and Prodan 2004). The answer to the question on trend stationarity vs. unit root may depend on the era and the country (Gaffeo et al 2005).

Interestingly, this literature is not often referred to in analyses on through-the-cycle ratings even though the credit cycle is almost by definition related to macroeconomics. Even though the statistical relationship between credit risks and macroeconomic variables appears weak (Koopman et al 2009) or very complicated (Figlewski et al 2012), there must be some connection between aggregate output and the credit risk of the corporate sector, simply because private companies account for most of the GDP.

References

Altman, E. I. and H. A. Rijken, 2005. The impact of rating agencies' through-the-cycle methodology on rating dynamics*, Economic Notes by Banca Monte dei Paschi de Siena* 34, 127-154

Andersen, H., 2011. Procyclical implications of Basel II: Can the cyclicality of capital requirements be contained? *Journal of Financial Stability* 7, 138-154

Avramov, D.,2002. Stock return predictability and model uncertainty; *Journal of Financial Economics* 64, 423-458

Bali, T. G, K. O. Demirtas and H. Levy, 2008. Nonlinear mean reversion in stock prices, *Journal of Banking and Finance* 32, 767-782

Basel Committee, 2005. An Explanatory Note on the Basel II IRB Risk Weight Functions, Bank for International Settlements, July 2005

Beechey, M. and P. Österholm, 2008. Revisiting the Uncertain Unit Root in GDP and CPI: Testing for Non-Linear Trend Reversion, *Economics Letters* 100, 221-223

Bharath, S. and T. Shumway, 2008. Forecasting default with the Merton distance to default model, *The Review of Financial Studies* 21, pp. 1339-1369

Bry, G. and C. Boschan, 1971. *Cyclical Analysis of Time Series: Procedures and Computer Programs*. (NBER)

Choe, K.-I., K. Nam and F. Vahid, 2007. Necessity of negative serial correlation for mean-reversion of stock prices, *Quarterly Review of Economics and Finance* 47, 576-583

Choi , I., 2001. Unit root tests for panel data; *Journal of International Money and Finance 20*, 249-272

Crosbie, P. and J. Bohn, 2003. *Modeling Default Risk - Modeling methodology* (Moody's KMV)

Drumond, I., 2009. Bank capital requirements, business cycle fluctuations and the Basel accords: a synthesis, *Journal of Economic Surveys* 23, 798-830

Figlewski, S., H. Frydman and W. Liang, 2012. Modeling the effect of macroeconomic factors on corporate default and credit rating transitions, *International Review of Economics and Finance* 21, 87-105

Gaffeo, E., M. Gallegati and M. Gallegati, 2005. Requiem for the unit root in per capita real GDP? Additional evidence from historical data, *Empirical Economics* 30, 37-63

Gordy, M. B. and B. Howells, 2006. Procyclicality in Basel II: Can we treat the disease without killing the patient? *Journal of Financial Intermediation* 15, 395-417

Güttler, A. and P. Raupach, 2010. The impact of downward rating momentum. *Journal of Financial Services Research* 37; 1-23

Hlouskova, J. and M. Wagner, 2006. The performance of panel unit root and stationarity tests: results from a large scale simulation study, *Econometric Reviews* 25, 85-116

Hung, J.-C., Y.-H Lee and T.-Y. Pai, 2009. Examining Market Efficiency for Large- and Small-Capitalization of TOPIX and FTSE Stock Indices, *Applied Financial Economics* 19, 735-744

Im, K.S., M H. Pesaran and Y. Shin, 2003. Testing for unit roots in heterogeneous panels, *Journal of Econometrics* 115, 53-74

Jiménez, G. and J. Mencía, 2009. Modelling the distribution of credit losses with observable and latent factors, *Journal of Empirical Finance* 16, 235-253

Johnson, N. L., S. Kotz and N Balakrishnan, 1994. *Continuous univariate distributions*, Vol 1, (John Wiley & Sons; Wiley series in probability and mathematical statistics)

Kisgen, D. J., 2006. Credit Ratings and Capital Structure, *The Journal of Finance* 61, pp. 1035-1072

Kisgen, D. J., 2009. Do Firms Target Credit Ratings or Leverage Levels, *Journal of Financial and Quantitative Analysis* 44, 1323-1344

Koopman, S. J., R. Kräussl, A. Lucas and A. B. Monteiro, 2009. Credit cycles and macro fundamentals, *Journal of Empirical Finance* 16, 42-54

Lanne, M, H. Lütkepohl and P. Saikkonen, 2002. Comparison of unit root tests for time series with level shifts, *Journal of time series analysis* 23, 667-685

Lanne, M, H. Lütkepohl and P. Saikkonen, 2003. Test procedures for unit roots in time series with level shifts at unknown time, *Oxford Bulletin of Economics and Statistics* 65, 91-115

Levin, A., C.-F. Lin and C.-S. J. Chu, 2002. Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of Econometrics* 108; 1-24

Li, M.-Y. L. and P. Miu, 2010. A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach, *Journal of Empirical Finance* 17, 818-833

Lopez, J. H., 1997. The power of the ADF test, *Economics Letters* 57, 5-10

Löffler, G., 2004. An anatomy of rating through the cycle, *Journal of Banking and Finance* 28, 695-720

Maddala, G. S. and S. Wu, 1999. A comparative study of unit root tests with panel data and a new simple test, *Oxford Bulletin of Economics and Statistics* 61, Special Issue, 631-652

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates, The Journal of Finance 29, 449-470

Narayan, P. K. and S. Narayan, 2007. Mean reversion in stock prices: new evidence from panel unit root tests, *Studies in Economics and Finance* 24, 233-244

Nelson, C. R and C. I. Plosser, 1982. Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics 10*, 139-162

Ng, S. and P. Perron, 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association* 90, 268-281

Ng, S. and P. Perron, 2001. Lag length selection and the construction of unit root tests with good size and power, *Econometrica* 69, 1519-1554

Nurmi, S., 2004. The determinants of plant survival in turbulent macroeconomic conditions, In *Essays on Plant size, Employment Dynamics and Survival*, (Helsinki School of Economics A-230)

Papell, D. H and R. Prodan, 2004. The uncertain unit root in U.S. real GDP: evidence with restricted and unrestricted structural change, *Journal of Money, Credit and Banking* 36, 423-427

Reisz, A. S and C. Perlich, 2007. A market-based framework for bankruptcy prediction, *Journal of Financial Stability* 3, 85-131

Saikkonen, P. and H. Lütkepohl, 2002. Testing for a unit root in a time series with a level shift at unknown time, *Econometric Theory* 18, 313-348

Schwert, G. W., 1989. Tests for unit roots: A Monte Carlo investigation, *Journal of Business & Economic Statistics* 7, 147-159

Thompson, P., 2005. Selection and firm survival: Evidence from the shipbuilding industry 1825-1914, *The Review of Economics and Statistics* 87, 26-36

Westerlund, J., 2009. A note on the use of the LLC panel unit root test. *Empirical Economics* 37, 517-531

A Probit Model for Insolvency Risk among Insurance Companies

Leo de Haan*

Jan Kakes†

Abstract

We estimate a probit model of insolvency risk, using a dataset of about 400 Dutch insurance companies during the period 1995-2005. The results suggest that surplus capital, company size, profitability, long-tailed business and being a mutual insurer reduce the risk of insolvency. The model can be used to identify insurers with high insolvency risk one year ahead. It is shown that the choice of the threshold above which an insurer is classified as having high insolvency risk, is an important determinant of the relative occurrence of type I and type II prediction errors. We use a loss function to find the optimal threshold given the supervisor's relative preference with respect to missing insolvencies and false alarms.

Keywords: Probit model, Insolvency risk, Insurance
JEL Classification: G22, G32

* Corresponding author. E-mail: l.de.haan@dnb.nl. Phone: +31 20 5243539. Fax: +31 20 5242514. De Nederlandsche Bank, Economics and Research Division, Amsterdam, The Netherlands.

† De Nederlandsche Bank, Financial Stability Division, Amsterdam, The Netherlands.

The authors wish to thank an anonymous referee, as well as Jakob de Haan and participants of the conference on Modelling and Analysis of Safety and Risk in Complex Systems (Saint Petersburg, 2009) for useful comments on earlier versions. The views expressed are those of the authors and do not necessarily reflect official positions of De Nederlandsche Bank.

1 - Introduction

For a supervisor of the insurance sector, it may be useful to have a statistical tool to assess the insolvency risk of insurance companies. This paper presents a probit model to investigate the determinants of insolvency.

Earlier applications of classification techniques for financial institutions usually distinguish bankrupt from non-bankrupt companies (see, e.g., BarNiv and McDonald, 1992, for a survey). Most of these studies use US data, where bankruptcies occur frequently. Since bankruptcies among Dutch insurance companies are very rare, other classification criteria have to be used for a sensible statistical analysis. Instead of actual defaults, we therefore use a broader concept that also includes near insolvencies. To identify these, we use an objective classification criterion based on the solvency ratio, which is defined as the actual capital-to-assets ratio of an insurance company divided by the required capital-to-assets ratio that is prescribed by the supervisor.

Besides the fact that we focus on a European country, an important value-added of our paper is that we have a comprehensive dataset. Our dataset comprises the entire Dutch insurance industry – about 400 Dutch life and non-life insurers – and covers a long period (1995-2005) including a full business cycle. So, peculiarities due to specific insurers or a specific year are unlikely to drive our results.

The estimation results suggest that surplus capital, company size, profitability, long-tailed business and being a mutual insurer reduce the risk of insolvency. The model can be used to identify insurers with high risk of insolvency one year ahead. It is shown that the choice of the threshold above which an insurer is classified as having high insolvency risk, is an important determinant of the relative occurrence of type I and type II prediction errors. We use a loss function to find the optimal threshold given the supervisor's relative preference with respect to missing insolvencies and false alarms.

The outline of the paper is as follows. After a literature review in Section 2 to put our contribution in perspective, the model is presented in Section 3. Section 4 discusses the data. The estimation results are discussed in Section 5, after which Section 6 shows how the model can be used for predicting insolvencies. Section 7 concludes.

2 - Literature review

The modeling of insolvencies dates back to the early literature on corporate bankruptcy, including the seminal work by Altman (1968) who used discriminant analysis. Later on, logit or probit models were found to be superior (e.g. Ohlson, 1980; Lennox, 1999). These models assume that the probability of bankruptcy can be predicted by a set of company characteristics and other variables. Discriminant analysis is criticized for its restrictive distributional assumptions, specifically that the variables' distributions are normal and that their variance-covariance matrices are the same for both non-bankrupt and bankrupt firms. Logit and probit models do not impose such assumptions. Moreover, logit and probit models directly give a probabilistic prediction whereas discriminant analysis produces a score, such as the Z-score or the Taffler-score, which has to be converted into a probabilistic measure, being an additional source of error (Ohlson, 1980). A different approach is the Merton (1974) model, which imposes assumptions on the value of firms' underlying assets and capital structure. A firm's default is determined by the market value of its assets relative to its liabilities. While attractive because of its strong theoretical underpinning, the Merton model has been criticized because of the underlying assumptions of perfectly liquid markets, absence of transaction costs and irrelevance of a firm's financial structure.[1] To take into account these weaknesses, the basic Merton model has been extended and modified, while as an alternative reduced form models have been developed, initiated by Jarrow and Turnbull (1995).

These different methodologies to model insolvencies have also been applied to insurance companies, which have a number of special characteristics compared to other firms. Because of the nature of their activities, insurers tend to be sensitive to financial market sentiment, whereas they use specific tools – like re-insurance – to mitigate their risk. Furthermore, like banks, insurers are subject to prudential regulation and supervision. Most studies on insurers' financial structure and the prediction of failures are based on US data. Several of these studies examine the performance of the US regulator's risk-based solvency rules, which were introduced in the early

[1] Furthermore, the estimation of the Merton model requires high-frequency data on share prices for the assessment of market volatility to deduce default probabilities. The outcomes may therefore be largely driven by excessive price fluctuations rather than market fundamentals. In addition, this information is not available for our sample of mostly non-listed insurance companies.

1990s. BarnNiv and McDonald (1992), Cummins et al. (1995, 1999), Grace et al. (1998) and Pottier and Sommer (2000) analyse insolvencies in the property-liability insurance industry. BarNiv and Hershbarger (1990) investigate different models to predict failures of US life insurers. A general finding of these studies is that the risk-based capital ratio used by US regulators can be improved as a leading indicator of failures. Cummins et al. (1999) conclude that cash-flow variables enhance the explanatory power of this static ratio, while Pottier and Sommer (2000) find that ratings by a private agency outperform the risk-based ratio.

Another strand of empirical literature focuses on the actual capital-to-assets ratio of insurers instead of insolvency risk. See, for example, Harrington and Niehaus (2002), Cummins and Nini (2002) and Klein et al. (2002) for the US. De Haan and Kakes (2010) find that Dutch insurers' capital-to-asset ratios can be related to risk characteristics and tend to be significantly higher than the regulatory minimum.

Analyses on solvency of European insurers are scarce, exceptions being Kramer (1996, 1997) for Dutch non-life insurers and Schmeiser (2004) who uses data of a single German insurance company. These studies typically look at a subset of insurers and consider only a brief period. Apart from data limitations, it is difficult to test the performance of indicators of insurer failures in Europe, for the simple reason that insurer defaults hardly ever occur in most European countries including the Netherlands. Because of the lack of bankruptcies in the Netherlands, Kramer (1996, 1997) investigates under what circumstances Dutch insurers are classified as 'strong', 'moderate' or 'weak' according to assessments by the supervisor. His dataset comprises about 200 Dutch non-life insurance companies in a single year (1992). Using an ordered logit analysis, as well as artificial intelligence techniques (neural network, expert system), his classification matches the supervisor's own qualitative assessments in about 90 percent of the cases. Some important explanatory variables – in particular insurers' solvency and profitability – are in line with our results. By contrast, he does not find that organisation type (mutual or stock) is a significant factor while we conclude it is relevant.

3 - The model

We estimate a probit model that classifies insurers into two groups, one with high and the other with low insolvency risk, respectively. Contrary to Kramer (1996, 1997), our classification criterion is based on an objective measure: the solvency ratio, which is defined as the actual capital-to-assets ratio of an insurance company divided by the required capital-to-assets ratio that is prescribed by the supervisor.[2] Our dataset is also much larger, comprising 400 Dutch life and non-life insurers during the period 1995-2005.

The probit model, as introduced by Bliss (1935), assumes that

$$P(Y = 1 \mid X = x) = \Phi(x'\beta), \tag{1}$$

where P denotes the probability, $Y \in (0,1)$ is a binary outcome variable which – in the present study – is 1 if an insurer runs high insolvency risk and 0 if insolvency risk is low, X is a vector of explanatory variables, Φ is the cumulative distribution function of the standard normal distribution, and β is a vector of parameters. Hence, the probability of high insolvency, given certain values x for the explanatory variables, is a function of these values multiplied by a set of parameters. The parameters β are typically estimated by maximum likelihood. This will be done in Section 4. But first, we discuss the data and define the variables that will be used.

4 - Data

We use confidential data on individual insurers for the period 1995-2005, gathered for supervisory purposes and published only in aggregated form in the *Statistical Bulletin* of De Nederlandsche Bank.[3] The unbalanced panel consists of around 400 insurance companies from all lines of business.

---

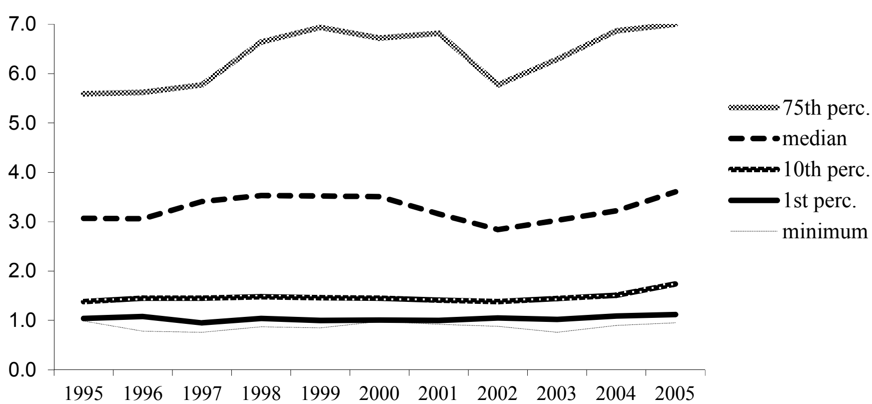
[2] The required solvency ratio is based on an insurer's performance and activities in recent years. For life insurers, the required solvency ratio is largely based on gross technical provisions. For non-life insurers, the required solvency ratio is largely based on premium income and incurred losses over recent years. See De Haan and Kakes (2010) for a more extended discussion of solvency requirements for Dutch insurers.

[3] The dataset used in this study was discontinued after 2005 because of a change in the supervisory reporting framework.

Insurers' financial soundness is often measured by the capital-to-assets ratio. However, the actual capital-to-assets ratio gives only limited information on the financial position of insurers as it does not take account of the risk profiles of their insurance obligations. For example, if a high-risk company and a low-risk company have identical capital-to-assets ratios, the former may be undercapitalized while the latter is overcapitalized. Therefore, insurers' solvency positions are usually presented in terms of the solvency ratio, defined as the actual capital-to-assets ratio of an insurer divided by the required capital-to-assets ratio that is prescribed by the supervisor. The availability of supervisory data on the required capital-to-assets ratio of individual insurers makes our dataset unique and enables us to use the solvency ratio.

Figure 1 shows the percentile distribution of the solvency ratio for each year, including the minimum solvency ratio observed. The median indicates that half of the population has at least thrice the required amount of capital; for the third quartile this is five times or even more. De Haan and Kakes (2010) find that Dutch insurers hold more capital than required because regulatory requirements do not sufficiently take risk factors into account.

Figure 1. Percentile distribution of solvency ratio



Note: The solvency ratio is the actual capital-to-assets ratio divided by the required capital-to-assets ratio.

The dependent variable in our model is the binary outcome variable 'insolvency', which we set equal to 0 if the solvency ratio is greater than 1.5

and equal to 1 if the solvency ratio is less than or equal to 1.5. The frequency distribution of insolvency is given in Table 1. Around 12% of the sample runs high insolvency risk according to this criterion. The hurdle rate of 1.5 reflects supervisory practice at the time and ensures that sufficient observations of high insolvency risk are included, which obviously is essential for a statistically sensible analysis.[4]

Table 1. Frequency distribution of insolvency, by year

	Insolvency = 0		Insolvency = 1		Total number of observations
	Number of observations	Percent of total	Number of observations	Percent of total	
1995	326	86.0	53	14.0	379
1996	345	88.0	47	12.0	392
1997	340	86.1	55	13.9	395
1998	357	87.9	49	12.1	406
1999	354	88.3	47	11.7	401
2000	324	88.3	43	11.7	367
2001	305	86.4	48	13.6	353
2002	289	84.8	52	15.2	341
2003	292	87.4	42	12.6	334
2004	283	89.3	34	10.7	317
2005	278	93.0	21	7.0	299
Total	3493	87.7	491	12.3	3984

The following explanatory variables are considered:

[4] The main results of the probit analysis proved to be not particularly sensitive to this particular choice of hurdle rate. Lower hurdle rates gave less significant outcomes though, due to the fact that there remain too few observations in the insolvent group, which diminishes the discriminatory power of this type of model.

(1) *Surplus capital*, defined as the actual capital-to-assets ratio minus the required capital-to-assets ratio. We expect the probability of insolvency to be lower, the higher the surplus in the preceding year.

(2) *Company size*. Presumably, as large insurers have more scope for diversification than small insurers, their total losses are more predictable. Hence, large firms probably run less risk of insolvency. The size of the company is measured by the natural logarithm of total assets.

(3) *Profitability*. A profitable firm has more internal funds at its disposal that can be hoarded as a buffer, so higher profitability will lead to higher solvency ratios. This is consistent with Kramer's (1996) finding for Dutch insurers that high profitability is one of the significant variables that reduce the probability of insolvency. Profitability is measured by annual return on assets.

(4) *Reinsurance*. If a company reinsures the bulk of its risks elsewhere, lower capitalization is required to achieve a given level of insolvency risk. Hence, we would expect a negative relationship between the use of reinsurance and capital. However, it is also conceivable that reinsurers demand a certain minimum level of capitalization from their clients, leading to a positive relationship between the use of reinsurance and capital. The effect of reinsurance on insolvency risk is therefore ambiguous. The use of reinsurance is measured by the proportion of reinsurance premiums paid in total premiums earned.

(5) *Portfolio risk*. The proportion of equity in the investment portfolio is a standard measure for the risk profile of the insurer's asset portfolio. The more volatile assets are held by the firm, for a given solvency ratio, the more vulnerable it is for asset price fluctuations that could lead to insolvency. Indeed, Kramer (1996) finds that for Dutch insurers, a higher proportion of equity increases the probability of insolvency.

(6) *Long-tailed business*. The ratio of loss reserves over incurred losses is a proxy for the time lag between policy issuance and the payment of the claims, with higher ratios indicating longer tailed business. As this ratio increases, the insurer's management is in control of the policyholders' funds for a longer time, which increases liquidity and thus diminishes the risk of insolvency.

(7) *Underwriting risk*. Underwriting risk is measured by the standard deviation of the loss ratio per firm (the ratio of losses incurred to premiums earned), which is frequently used in the insurance literature (e.g. Meyers, 1989; Guo and Winter, 1997; Lamm-Tennant and Starks,

1993). This measure captures the risk on the insurance portfolio, while portfolio risk (introduced above) measures the risk on the investment portfolio. We expect a positive relation with insolvency risk.

(8) *Herfindahl index*. This index measures the degree to which an insurer is diversified across different lines of business (as measured by the written premiums). Considered lines of business are health, car, transport, fire, other non-life, mixed non-life, and life insurance. Insurers with more diversified underwriting portfolios are expected to run less insolvency risk (e.g. Klein et al., 2002; Cummins and Nini, 2002). Lower Herfindahl indices imply higher diversification and, consequently, we expect a positive relation with insolvency risk.

(9) *Mutual*. This is a binary variable taking a value of 1 for mutual and 0 for stock companies. A priori, this variable's impact on a firm's solvency is ambiguous. Agency theory (Jensen and Meckling, 1976) predicts lower capitalization needs for mutual insurers because of the elimination of the owner-policyholder conflict. However, according to the pecking order hypothesis (Myers and Majluf, 1984) mutual insurers have a greater tendency than stock insurers to hoard capital because of their limited ability to access capital markets.[5]

(10) *Line of business*. A binary variable for each line of business taking the value of 1 if the company in that year is predominantly (i.e. more than half of its written premiums) in that particular line of business and 0 if it is not. This dummy variable should capture any remaining unobservable, time-invariant effects related to the line of business of an insurer.

(11) *Year*. A year dummy for each sample year. This dummy variable captures the effects of macro-economic trends, which are common to all insurers.

Table 2 presents summary statistics for the continuous variables.

[5] See De Haan and Kakes (2010) for a more detailed explanation of the underlying theories.

Table 2. Summary statistics, 1995-2005

	Median	Mean	Standard deviation	Number of observations
Surplus capital	0.228	0.288	0.294	3985
Company size	10.210	10.443	2.354	3915
Profitability	0.020	0.028	0.051	3912
Reinsurance	0.135	0.235	0.266	3284
Portfolio risk	0.129	0.185	0.197	2688
Long-tailed business	2.895	8.563	12.240	3005
Underwriting risk	0.116	0.972	4.565	3808
Herfindahl	1.000	0.881	0.237	3993

5 - Estimation results

The estimation results of the probit model including all explanatory variables are presented as 'model 1' in Table 3. Several variables are not statistically significant. Moreover, many of these insignificant variables have missing values for various company-years, so that they limit the number of observations that can be used in the model estimation. Therefore, the model is re-estimated dropping the statistically insignificant variables one by one in order of significance. This backward elimination procedure results in a model with seven statistically significant variables (model 2).[6] A positive side effect of dropping insignificant variables is that it increases the sample size substantially, which makes the results more representative for the population of insurance companies. Despite the dropping of explanatory variables, the model fit, as measured by the pseudo-R^2, improves, from 0.374 to 0.390. Moreover, all coefficients of the selected variables keep their sign and do not substantially alter in magnitude, and some of them become statistically more

[6] Backward elimination involves starting with all candidate variables and than deleting them one by one in order of statistical significance. This is one approach in stepwise regression; other approaches are forward selection or combinations of the two. See, e.g., Hocking (1976).

Table 3. Estimation results. Dependent variable is Insolvency risk

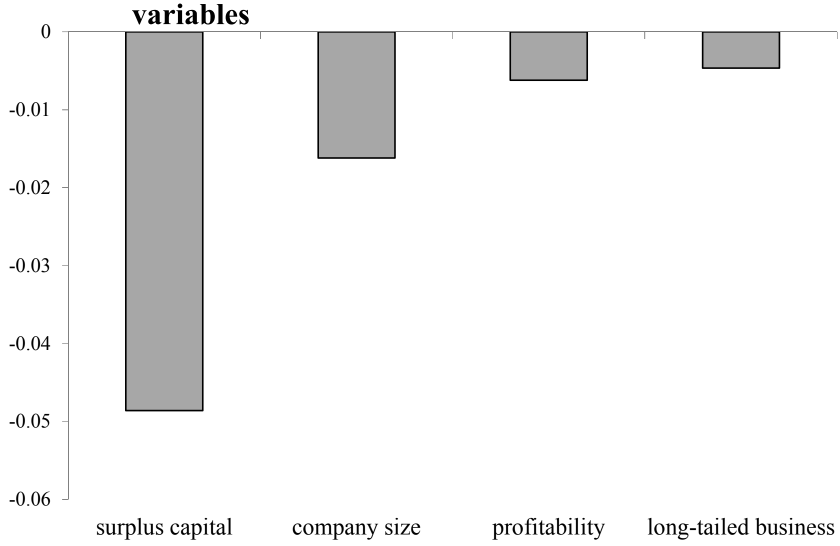
	Model 1			Model 2		
	Coefficient	Standard error	Effect	Coefficient	Standard error	Effect
Continuous variables:						
Surplus capital $_{t-1}$	-7.684**	2.600	-0.112	-7.837**	1.444	-0.177
Company size $_{t-1}$	-0.303**	0.057	-0.004	-0.317**	0.040	-0.007
Profitability$_{t-1}$	-9.210**	1.722	-0.134	-5.759**	1.211	-0.130
Reinsurance $_{t-1}$	-0.162	0.465	-0.002			
Portfolio risk $_{t-1}$	0.171	0.416	0.002			
Long-tailed business $_{t-1}$	-0.017	0.010	-0.0002	-0.017**	0.006	-0.0004
Underwriting risk $_{t-1}$	0.025	0.022	0.0004			
Herfindahl $_{t-1}$	0.050	0.660	0.0007			
Dummy variables:						
Mutual company	-0.510	0.291	-0.006	-0.619**	0.181	-0.012
Health insurance	0.386	0.346	0.008	0.542**	0.169	0.020
Car insurance	0.306	0.441	0.006			
Transport insurance	-0.676	0.392	-0.005			
Fire insurance	-0.167	0.455	-0.002			
Mixed non-life ins.	-0.335	0.539	-0.004	-0.434*	0.182	-0.007
Pseudo R^2	0.374			0.390		
Number of observations	1649			2476		
Number of insurers	277			401		

Explanatory note: Reported effects are marginal effects for continuous variables and fixed effects for dummy variables. Standard errors of the coefficients are calculated using the Huber/White/sandwich estimator and allowing for correlation of observations for the same company. Suffixes * and ** indicate statistical significance at the 5% and 1% level, respectively.

significant. Hence, the estimation results are fairly robust to the selection of variables.

We now turn to the economic interpretation of our selected model. The ('fixed') effect of mutual is -0.012, indicating that a mutual insurer's probability of high insolvency risk is 1.2 percentage points lower than a stock insurer's. The effects of line of business are in the same order of magnitude (+2 percentage points for health insurers, -0.7 for mixed non-life insurers). The economic interpretation of the ('marginal') effects for the continuous variables is less straightforward, as these represent the effects for an infinitesimal change in the explanatory variables. For ease of interpretation, Figure 2 therefore plots the probability effects for these variables. The bars give the partial effects on the probability of high insolvency risk of a one-standard-deviation increase for each explanatory variable. The solvency surplus in the previous year is the most important probability reducer, followed by company size, profitability, and long-tailed business. These outcomes are consistent with the expectations expressed in Section 3.

Figure 2. Probability effects of selected explanatory variables



Note: Partial effect of one-standard-deviation increase in explanatory variable on probability of high insolvency risk.

6 - Model use

A prudential supervisor may wish to use the model to predict, one year ahead, whether or not an insurer runs high risk of falling into the class of companies with high insolvency risk. The performance of a probit model, or any other model or indicator used to predict a dichotomous variable, can be considered in terms of a two by two matrix, as shown in Table 4.[7] A signal is considered *good* (cell A) if it is followed by insolvency and *bad* (cell B) if it remains solvent. If there is no signal and insolvency follows, this is denoted as a *missed signal* (cell C) while a situation in which no signal is followed by no insolvency is a *good silent signal* (cell D).

In standard statistical terms, using type I and type II errors, let the null hypothesis (H0) be equal to insolvency and the alternative hypothesis (H1) be no insolvency. From Table 4, H0 is equal to A+C and H1 is equal to B+D. A type I error is rejecting H0 while H0 is true, which is a missing signal. The proportion of missing signals is equal to C/(A+C). Similarly, a type II error is not rejecting H0 when H0 is false, which is a false signal. The proportion of false signals is equal to B/(B+D).

Table 4. Performance evaluation matrix

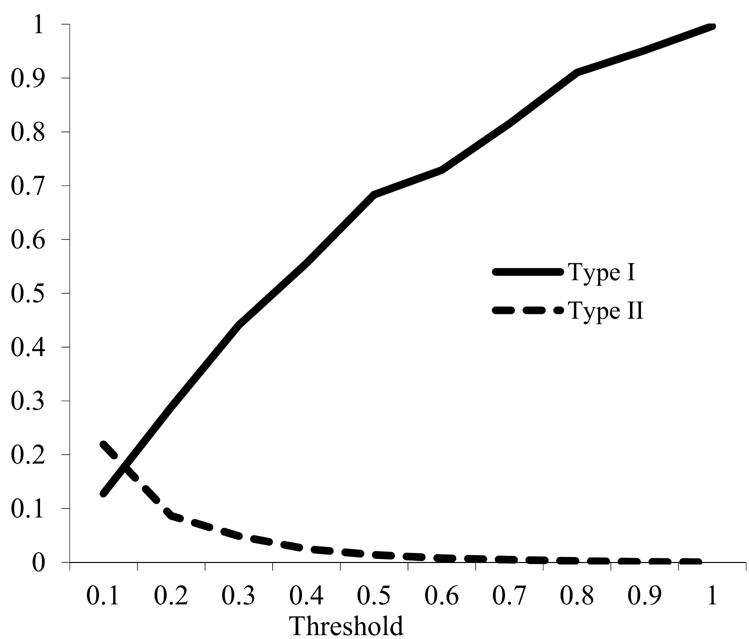
	Insolvency next year	Solvency next year
Signal issued	A	B
No signal issued	C	D

When using a probit model for predictions, one has to choose a threshold for the probability of an event. The model will give a signal if this threshold is crossed. A high (low) value for the threshold decreases (increases) the number of signals issued. Consequently, the choice of the threshold will affect the relative frequencies of type I and II errors. The latter are drawn in Figure 3 for

[7] See, e.g. Kaminsky and Reinhart (1999), Edison (2003), Alessi and Detken (2011).

our estimated model for different thresholds. The figure shows, firstly, that the higher the threshold, the higher is the proportion of missing signals (type I error) and the lower the proportion of false signals (type II error). Secondly, it shows that this trade-off is asymmetric. The type II curve begins at 0.219 for a threshold of 0.1 and quickly bends off towards zero for thresholds higher than 0.5, while the type I curve nearly linearly rises from 0.128 for a threshold of 0.1 to 0.996 for a threshold of 1.0.

Figure 3. Relative frequencies of Type I and Type II errors, by threshold



Note: Following Table 4, Type I = C/(A+C), Type II = B/(B+D).

At face value, one might be inclined to choose a threshold between 0.1 and 0.2 where the two lines intersect at a proportion of 0.185 for both types of errors. However, in that case one assumes that both types of errors are equally important. Alessi and Detken (2011) argue that, when choosing the threshold value, one has to incorporate the relative risk aversion of the model user with

Table 5. Losses for different probability thresholds, given θ

θ =	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Threshold									
0.1	0.210	0.201	0.192	0.182	**0.173**	**0.164**	**0.155**	**0.146**	**0.137**
0.2	0.107	0.127	**0.147**	**0.167**	0.187	0.207	0.228	0.248	0.268
0.3	0.088	**0.127**	0.166	0.205	0.244	0.284	0.323	0.362	0.401
0.4	**0.078**	0.131	0.184	0.237	0.290	0.343	0.396	0.449	0.502
0.5	0.081	0.148	0.215	0.282	0.349	0.415	0.482	0.549	0.616
0.6	0.080	0.152	0.224	0.296	0.368	0.440	0.512	0.584	0.656
0.7	0.086	0.167	0.248	0.329	0.410	0.419	0.572	0.653	0.734
0.8	0.093	0.184	0.274	0.365	0.456	0.547	0.637	0.728	0.819
0.9	0.095	0.190	0.285	0.381	0.476	0.571	0.666	0.761	0.856
1.0	0.100	0.199	0.299	0.398	0.498	0.598	0.697	0.797	0.896

Note: θ is the relative risk aversion between missing signals and false alarms. Minimum losses per θ have been printed in bold.

respect to type I and type II errors. Therefore, they define the following loss function:

$$L = \theta \frac{C}{A+C} + (1-\theta)\frac{B}{B+D} \qquad (2)$$

where $0 \le \theta \le 1$ is the parameter revealing the relative risk aversion between type I and type II errors. The loss can be interpreted as the weighted sum of both types of errors. A θ smaller than 0.5 reveals less aversion towards missing a signal than towards receiving false alarms, a θ greater than 0.5 reveals the opposite risk attitude.

Alessi and Detken (2011) apply (2) to the case of a central banker who uses an indicator of costly asset boom/bust cycles to determine whether or not to ‘lean against the wind’ with monetary policy. They argue that a θ smaller than 0.5 is a realistic description of central bankers’ loss functions, as the economic cost of reacting to a false signal is higher than missing a signal. In our case, where we take the perspective of a supervisor, matters are different. The supervisor uses the model to earmark insurers one year ahead as running a high risk of getting insolvent. A signal from the model may lead to additional supervisory scrutiny of the insurer in question. Hence, the cost of a false alarm is the cost of the additional supervisory effort. The cost of a missing signal, on the other hand, is relatively large, as insolvency may lead to a failure. Hence, we argue that a θ larger than 0.5 is a realistic description of a supervisor's loss function.

Given the relative risk preference θ, one may find the threshold that minimizes the loss using loss function (2). Table 5 gives all possible losses for different thresholds and θ. The optimal threshold is 0.4 for $\theta = 0.1$, 0.3 for $\theta = 0.2$, 0.2 for $\theta = 0.3$ and $\theta = 0.4$, and 0.1 for $\theta \geq 0.5$. Hence, assuming that a supervisor's θ is larger than 0.5, a threshold of 0.1 should be chosen.

7 - Conclusion

We estimate a probit model for insolvency risk among Dutch insurance companies, using a much larger dataset covering a longer sample period than previous studies for the Netherlands. The model relates the probability of high insolvency risk for an insurance company to several of its characteristics. The results suggest that surplus capital, company size, profitability, long-tailed business and being a mutual insurer reduce the risk of insolvency. The model can be used to predict one year ahead which insurance companies are running high insolvency risk. It is shown that the choice of the threshold above which an insurer is classified close to insolvency, is an important determinant of the relative occurrence of type I and type II prediction errors. We use a loss function to find the optimal threshold given the supervisor's relative preferences with respect to missing insolvencies and false alarms.

References

Alessi, L. and C. Detken, 2011. Quasi real time early warning indicators for costly asset price boom/bust cycles: a role for global liquidity. *European Journal of Political Economy*, 27(3), 520-533.

Altman, E., 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, 23, 589-609.

BarNiv, R. and R.A. Hershbarger, 1990. Classifying financial distress in the life insurance industry. *Journal of Risk and Insurance*, 57(1), 110-136.

BarNiv, R. and J.B. McDonald, 1992. Identifying financial distress in the insurance industry: a synthesis of methodological and empirical issues. *Journal of Risk and Insurance*, 59(4), 543-574.

Bliss, C.I., 1935. The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22, 134-167.

Cummins, J.D., S.E. Harrington and R. Klein, 1995. Insolvency experience, risk-based capital, and prompt corrective action in property-liability insurance. *Journal of Banking & Finance*, 19, 511-527.

Cummins, J.D., M.F. Grace and R.D. Phillips, 1999. Regulatory solvency prediction in property-liability insurance: risk-based capital, audit ratios, and cash flow simulation. *Journal of Risk and Insurance*, 66(3), 417-458.

Cummins, J.D. and G.P. Nini, 2002. Optimal capital utilization by financial firms: evidence from the property-liability insurance industry. *Journal of Financial Services Research*, 21(1/2), 15-53.

De Haan, L. and J. Kakes, 2010. Are non-risk based capital requirements for insurance companies binding? *Journal of Banking and Finance*, 34(7), 1618-1627.

Edison, H.J., 2003. Do indicators of financial crises work?: an evaluation of an early warning system. *International Journal of Finance and Economics*, 8, 11-53.

Grace, M.F., S.E. Harrington and R.W. Klein, 1998. Risk-based capital and solvency screening in property-liability insurance: hypotheses and empirical tests. *Journal of Risk and Insurance*, 65(2), 213-243.

Guo, D. and R.A. Winter, 1997. The capital structure of insurers: theory and evidence. Working Paper, Sauder School of Business.

Harrington, S. and G. Niehaus, 2002. Capital structure decisions in the insurance industry: Stocks versus mutuals. *Journal of Financial Services Research*, 21(1/2), 145-163.

Hocking, R.R., 1976. The analysis and selection of variables in linear regression. *Biometrics*, 32(1), 1-49.

Jarrow, R.A. and S. Turnbull, 1995. Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50(1), 53-85.

Jensen, M.C. and W.H. Meckling, 1976. Theory of the firm: managerial behaviour, agency costs, and ownership structure. *Journal of Financial Economics*, 3(4), 305-360.

Kaminski, G.L. and C.M. Reinhart, 1999. The twin crises: the causes of banking and balance-of-payments problems. *American Economic Review*, 89(3), 473-500.

Klein, R.W., R.D. Phillips and W. Shiu, 2002. The capital structure of firms subject to price regulation: evidence from the insurance industry. *Journal of Financial Services Research*, 21(1/2), 79-100.

Kramer, B., 1996. An ordered logit model for the evaluation of Dutch non-life insurance companies. *De Economist*, 144(1), 79-91.

Kramer, B., 1997. N.E.W.S.: A model for the evaluation of non-life insurance companies. *European Journal of Operational Research*, 98, 419-430.

Lamm-Tennant, J. and L.T. Starks, 1993. Stock versus mutual ownership structures: the risk implications. *Journal of Business*, 66(1), 29-46.

Lennox, C., 1999. Identifying failing companies, a re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business*, 51, 347-364.

Merton, R.C., 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29, 449-470.

Meyers, G.G., 1989. Analysis of the capital structure of an insurance company. *Proceedings of the Casualty Actuarial Society*, 76, 147-171.

Myers, S.C. and N. Majluf, 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13, 187-221.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18, 109-131.

Pottier, S.W. and D.W. Sommer, 2000. Capital ratios and property-liability insurer insolvencies. Mimeo.

Schmeiser, H., 2004. New risk-based capital standards in the European Union: a proposal based on empirical data. *Risk Management and Insurance Review*, 7(1), 41-52.