



Matematisk statistik  
Stockholms universitet

## **Flexibel konkursriskestimering med logistisk spline-regression**

Erik von Schedvin

**Examensarbete 2008:1**

**Postadress:**

Matematisk statistik  
Matematiska institutionen  
Stockholms universitet  
106 91 Stockholm  
Sverige

**Internet:**

<http://www.math.su.se/matstat>



Matematisk statistik  
Stockholms universitet  
Examensarbete 2008:1,  
<http://www.math.su.se/matstat>

# Flexibel konkursriskestimering med logistisk spline-regression

Erik von Schedvin\*

Februari 2008

## SAMMANFATTNING

Denna uppsats utvärderar huruvida en icke-lineär utveckling av den logistiska kreditriskmodellen förbättrar möjligheten att mäta ett företags konkursrisk. Det icke-lineära bidraget baseras på en kvadratisk spline utveckling av den logistiska konkursriskmodellen. Konkursriskskattningarna utvärderas utifrån en CAP-analys och dess inverkan på kreditrisken kvantifieras med hjälp av Credit Suisse kreditriskmodell CRS+. Studien utförs på ett stort datamaterial, innehållande kvartalsobservationer för samtliga svenska aktiebolag under tidsperioden 1990Q1 till 2002Q4. Det huvudsakliga resultatet som presenteras i uppsatsen är att en icke-lineär utveckling av den logistiska kreditriskmodellen avsevärt förbättrar konkursriskskattningarna. För de konkursriskmodeller som endast inbegriper bokslutsinformation så sker en förbättring för samtliga företag och för de konkursriskmodeller som även inbegriper anmärkningsinformation så sker en förbättring för de 90 % av företagen som har lägst skattad konkursrisk.

Nyckelord: Konkursriskmodellering, kreditriskmodellering, logistisk regression, spline-regression och icke-lineär logistisk regression.

---

\*E-post: [erik.von.schedvin@riksbank.se](mailto:erik.von.schedvin@riksbank.se).Handledare: Rolf Sundberg, [rolfs@math.su.se](mailto:rolfs@math.su.se)

## ABSTRACT

This thesis compares a linear logistic default risk model with a nonlinear such model, based on regression splines. The performance of the estimated default risk is evaluated by a CAP-analysis (Cumulative accuracy profile) and the impact is quantified by Credit Suisse's credit risk model CR+. The survey is based on a large panel dataset of quarterly observations, covering all Swedish limited firms (aktiebolag, AB) between 1990Q1 and 2002Q4. The main finding is that the credit scoring performance improves when accounting ratios are allowed to enter the logistic default risk model in a nonlinear fashion.

## Förord

Detta examensarbete har skrivits på Riksbankens forskningsenhet. Utgångspunkten har varit att utifrån ett pågående konkursriskprojekt utvärdera fördelarna med en icke-lineär utveckling av en logistisk konkursriskmodell. Det finns en rad personer som med sitt kunnande gjort detta examensarbete möjligt. Först bör min handledare och chef på Riksbanken, Tor Jacobson, lyftas fram. Det är utifrån hans initiativ och hjälp detta arbete har utförts. Även min handledare från Matematisk statistik vid SU, Rolf Sundberg, har genom sitt engagemang varit ett stort stöd under arbetets gång. Vidare skulle jag vilja tacka Paolo Giordani och Mattias Villani som fungerat som statistiska bollplank samt övriga kollegor på Riksbankens forskningsenhet.

# Innehållsförteckning

1	Introduktion	2
2	Syfte	4
3	Disposition	4
4	Ickelineär regression	5
5	Spline-regression	6
5.1	Den trunkerade potensfunktionen	6
5.2	Spline-modellen	7
5.3	Additiv semiparametrisk spline-regression	9
6	Logistisk regression	10
7	Semiparametrisk logistisk regression	11
8	Modellutvärdering med AIC och SC	12
9	Data	13
10	Lokala Konfidensintervall för konkursrisken	17
11	Konkursriskmodellerna	18
12	Ickelineär modellering av ett nyckeltal	20
13	Modellresultat	21
14	Utvärdering	24
15	Portföljutvärdering	32
16	Slutsatser och kommentarer	36
	Referenser	38
	Appendix A	42
	Appendix B	46
	Appendix C	51
	Appendix D	56

# 1 Introduktion

Finansiella institutioner har i allt större utsträckning börjat kvantifiera de risker som är kopplade till långivning. Denna utveckling har pådrivits av flera bakomliggande faktorer; en teknisk utveckling av kreditportföljshanteringen; minskad lönsamhet inom den traditionella utlåningsindustrin; nya möjligheter att skapa lönsamhet via kreditderivat; införandet av nya regleringar via Basel II.

Ett nyckelmått som används för att mäta kreditrisken är låntagarnas konkurssannolikhet (probability of default, PD). PDn har fått stor betydelse via de nya Basel II reglerna vilka fordrar att banker och andra finansiella institutioner utför en intern kapitalutvärdering (internal capital adequacy process, ICAAP). Regelverket är utformat så att utvärderingsprocessen baseras på respektive banks egen utlåningshistorik. Då PDn är en nyckelvariabel vid beräkningen av kreditrisken för en utlåningsportfölj så är det av yttersta betydelse att måtten återger den faktiska konkursrisken för att kreditriskbedömningen skall bli korrekt.

Konkurssannolikhet kan till stor utsträckning förklaras av företagens finansiella ställning via nyckelmått som: skuldsättningsgraden, omsättningsgraden, tillgången till likvida medel etc. Men forskning under senare år påvisar även att makrovariabler har en betydande inverkan på PDn. Fluktuationer i ett lands aggregerade konkursrisk kan, av naturliga skäl, till stor del förklaras av den allmänna konjunkturcykeln.

Ett statistiskt verktyg som är vanligt förekommande för att mäta ett företags konkursrisk är den logistiska regressionsmodellen, med finansiella nyckeltal och anmärkningshistorik som förklarande variabler. Den logistiska regressionsmodellen modellerar logit lineärt mot de förklarande variablerna. Ett problem som uppkommer då den logistiska regressionsmodellen används är att den inte fångar upp eventuellt ickelineära relationer mellan de finansiella nyckeltalen och konkursrisken. Den logistiska konkursriskmodellen har visat sig under- eller överskatta konkursrisken för olika nivåer av de finansiella nyckeltalen.

Huvuddraget i denna uppsats är att utvärdera huruvida en ickelineär utveckling av den logistiska kreditriskmodellen förbättrar möjligheten att mäta ett företags

konkursrisk. Det icke-lineära bidraget kommer att baseras på en kvadratisk spline utveckling av den logistiska konkursriskmodellen. Konkursriskskattningarna kommer att utvärderas utifrån en CAP-analys och dess inverkan på kreditrisken kommer att kvantifieras med hjälp av Credit Suisse kreditriskmodell CR<sup>+</sup>. Studien kommer att utföras på ett stort datamaterial, innehållande kvartalsobservationer för samtliga svenska aktiebolag under tidsperioden 1990Q1 till 2002Q4. Det huvudsakliga resultatet som presenteras i uppsatsen är att en icke-lineär utveckling av den logistiska kreditriskmodellen avsevärt förbättrar konkursriskskattningarna. För de konkursriskmodeller som endast inbegriper bokslutsinformation så sker en förbättring för samtliga företag och för de konkursriskmodeller som även inbegriper anmärkningsinformation så sker en förbättring för de 90 % av företagen med lägst skattad konkursrisk.

Konkursriskmodeller med finansiella nyckeltal som förklarande variabler har presenterats vid ett flertal tidigare studier, se bland annat Altman (1968), Zavgren (1985), Wilson (1997), Shumway (2001), Jacobson, Lindé & Roszbach (2005). Merparten av de tidigare studierna har baserats på börsnoterade amerikanska bolag. Altman (1968) var banbrytande då han publicerade en artikel som utvärderade sambandet mellan finansiella nyckeltal och konkursrisken i början av sextiotalet. Altman anpassade en OLS regression för att påvisa att det föreligger ett signifikant samband mellan ett företags finansiella ställning och dess konkursrisk. Zavgren (1985) tillämpade en logistisk analys av amerikanska bolag. Modellerna som Zavgren utvecklade visade sig fånga upp 99 % av samtliga konkurser, för det givna datamaterialet, under en femårsperiod. En kreditriskmodell som fångar upp relationen mellan de makroekonomiska förhållandena och konkursrisken presenterades av Wilson (1997). Wilson redogör för McKinsey's konkursriskmodell, Credit Portfolio Wiew, vilken inkorporerar en uppsättning makrovariabler i en multivariat land/marknad logistisk modell. Shumway (2001) utvecklade analysen genom att inbegripa marknadsdrivna faktorer i en konkursriskmodell. De marknadsdrivna faktorerna visade sig vara ett bra komplement till de finansiella nyckeltalen.

Ett flertal tillämpade studier rörande icke-lineära applikationer har presenterats och ett exempel av tillämpningen rörande icke-lineär modellering av konkursrisken

ges i en artikel av Muller & Härdle (2002). De visar att en uppsättning förklarande faktorer för individuella låntagare med fördel modelleras icke-lineärt mot risken att ett lån förfaller för betalning.

## 2 Syfte

Syftet med uppsatsen är att undersöka huruvida en icke-lineär utveckling av den logistiska kreditriskmodellen förbättrar skattningen av konkursrisken för svenska aktiebolag. Den icke-lineära utvecklingen av den logistiska konkursriskmodellen kommer att baseras på en kvadratisk spline-regression.

Utvärderingen av den icke-lineära utvecklingen kommer att göras genom en CAP-analys och inverkan av konkursriskskattningarna kommer att kvantifieras med Credit Suisse kreditriskmodell  $CR^+$ . Modellerna kommer dels att skattas och utvärderas med avseende på hela perioden samt utifrån två separata delperioder.

## 3 Disposition

Uppsatsens struktur följer enligt: Den icke-lineära spline-regressionen samt en icke-lineär spline utveckling av den logistiska modellen kommer att presenteras i avsnitt 4 till 8. Vidare kommer dataanalys, modellresultat, modellutvärdering och portföljberäkningar att presenteras i avsnitt 9 till 15. Slutligen kommer slutsatserna att redovisas i avsnitt 16. Kreditriskmodellen  $CR^+$  vilken ligger till grund för portföljutvärderingen i avsnitt 14 redovisas i Appendix C.



## 4 Ickelineär regression

Ett vanligt förekommande problem inom flertalet ämnesområden är att anpassa en modell som består av ett flertal förklarande variabler. Det existerar en omfattande litteratur rörande ickeparametriska och icckelineära regressionsmetoder och applikationerna har blivit allt vanligare under senare år, se bland annat P. Sprent & N.C Smeeton (2002) och C. de Boor (2001). En avgörande orsak till att tillämpningen av icckelineära metoder har ökat de senaste åren är att det har blivit allt vanligare att arbeta med stora datamaterial, där tillräcklig information finns för att estimeras en regressionsyta.

Målet är att modellera en responsvariabel  $Y$  mot en eller flera förklarande variabler  $X = (x_1, \dots, x_R)$  givet en uppsättning observationer  $\{Y_i, X_i\}_{i=1}^N$ . Det underliggande systemet som genererar de observerade värdena antas vara formulerat enligt

$$Y_i = f(X_i) + e_i,$$

över ett givet område  $X_i \in D \subset \mathbb{R}^R$  som omfattar det observerade datamaterialet. Funktionen  $f$ , med dess  $R$ -dimensionella argument, fångar den sammanslagna skattade sambandet mellan  $Y_i$  och storheten  $X_i$ . Målet med regressionsanalysen är att utifrån det observerade datamaterialet konstruera en funktion  $\hat{f}(X_i)$  över det givna området  $D$ .

Flertalet metoder kan användas för att anpassa en ickeparametrisk funktion  $f$ . En vanligt förekommande metod som kan användas för ickeparametrisk modellering är lokal regression, se Cleveland (1988, 1991). Lokal regression är en metod där en regressionensyta anpassas genom att lokalt anpassa lineära eller kvadratiske funktioner enligt en förflyttande rörelse. Andra metoder som används för att anpassa icke- och semiparametriska regressionsmodeller är thin plate metoden, se Wahba (1990), och Kernel regression, se Nadarya (1964).

Ett problem med ovan angivna regressionsmetoder är att samtliga är beräkningsintensiva. Eftersom modellerna i denna uppsats skall anpassas utifrån

ett stort datamaterial så krävs det att en metod med låg beräkningsintensitet väljs. En metod med låg beräkningsintensitet är spline-regressionen vilken presenteras nedan.

## 5 Spline-regression

Splines kan definieras som styckvis valda polynom av grad  $s$  för vilka funktionsvärdena samt de första  $s-1$  derivatorna överensstämmer för de punkter där de möts, se Smith (1979). Punkterna där polynomen möts kallas knutpunkter. En vanlig polynomutveckling kan ses som en spline-funktion utan knutpunkter. Graden på polynomen samt placeringen av knutpunkterna kan variera mellan olika tillämpningar. Placeringen av knutpunkterna, för att uppnå bästa möjliga passform, har visat sig vara en svår uppgift. För högt antal knutpunkter kan även leda till överanpassning.

### 5.1 Den trunkerade potensfunktionen

För att anpassa en spline-regression med minsta kvadratmetoden så kan den trunkerade potensfunktionen användas som basfunktion för att uttrycka de kontinuerliga/mjuka restriktioner som krävs. Om vi definierar

$$(x-t)_+ = \max\{x-t, 0\},$$

så kan vi formulera den trunkerade potensfunktionen av ordning  $s$  som

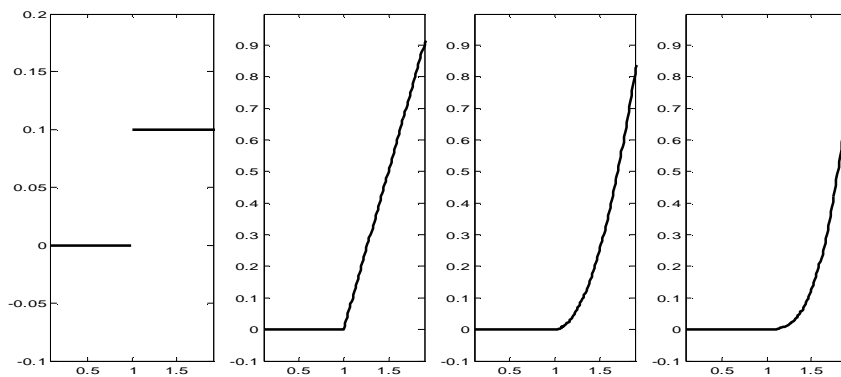
$$(x)_+^s = (x_+)^s, \quad s = 0, 1, 2, \dots$$

Funktionen  $f(x) = (x-t)_+^s$  är ett styckvis polynom, av grad  $s+1$ , med endast en aktiv knutpunkt vid  $t$ . Funktionen är kontinuerlig i  $t$  om  $s > 0$ . Men om  $s = 0$  så har  $f$  ett hopp vid knutpunkt  $t$  av storleken  $1$ . Då  $D(x-t)_+^s = s(x-t)_+^{s-1}$  kan vi se

att  $(x-t)_+^s$  har  $s-1$  kontinuerliga derivator med hopp i den  $s$ -te derivatan över  $t$ , av storleken  $s!$ . Då  $a^0 = 1$  för alla icke-negativa  $a$  så har vi att  $(x)_+^0 = 1$ . Men för  $x \leq 0$  måste vi veta värdet av  $0^0$ , vilket ej är definierat. Vi kan då göra följande definition

$$0^0 = 0,$$

vilken ger att  $(x)_+^0 = 0$ , för  $x \leq 0$ . Givet detta så är  $(x-t)_+^s$  en styckvis polynomfunktion även för  $s=0$  vilken bestäms av dess polynomdelar och knutpunkter. Figur 5.1.1 nedan illustrerar hur de trunkerade potensfunktionerna går in i spline modellen.



Figur 5.1.1: Funktionerna  $0.1(x-t)_+^0$ ,  $(x-t)_+^1$ ,  $(x-t)_+^2$  och  $(x-t)_+^3$  för  $t=1$ .

## 5.2 Spline-modellen

För att illustrera hur de trunkerade potensfunktionerna används i en spline modell betraktar vi först fallet med en styckvis kubisk funktion för  $x$  med en knutpunkt  $t$ , utan några krav på kontinuitet i knutpunkten. Funktionen formuleras då enligt följande

$$f(x) = b_{00} + b_{01}x + b_{02}x^2 + b_{03}x^3 + b_{10}(x-t)_+^0$$

$$\mathbf{b}_{11}(x-t)_+^1 + \mathbf{b}_{12}(x-t)_+^2 + \mathbf{b}_{13}(x-t)_+^3,$$

där vi definierar  $(x-t)_+^0 = 1$  om  $x-t > 0$  och noll annars. Funktionen  $f$  är definierad så att den ej behöver vara kontinuerlig för  $f$  eller  $f^{(l)}$  vid knutpunkt  $x = t$ . Om  $\beta_{10}$  tas bort från funktionen så blir  $f$  kontinuerlig, om även  $\beta_{11}$  tas bort så blir både  $f$  och  $f^{(l)}$  kontinuerliga. Genom att lägga till och ta bort termer från funktionen så kan vi införa och utvärdera restriktioner rörande kontinuitet. Utvärderingen kan göras genom att undersöka minsta kvadratskattningarna av  $\beta_{10}$  och  $\beta_{11}$ . Ett signifikanstest kan formuleras genom att testa hypotesen  $\mathbf{b}_{10} = \mathbf{b}_{11} = \mathbf{b}_{12} = \mathbf{b}_{13} = 0$  mot spline-modellen. Testet utvärderar huruvida spline-modellen ger en signifikant bättre passform än ett kubiskt polynom. Generellt så gäller det för  $J$  knutpunkter,  $t_1 < \dots < t_J$ , med  $J+1$  polynom av grad  $s$  att vi kan formulera en spline-modell utan några restriktioner rörande kontinuitet. Funktionen ges då enligt nedan

$$f(x) = \sum_{s=0}^S \mathbf{b}_{0s} x^s + \sum_{j=1}^J \sum_{s=0}^S \mathbf{b}_{js} (x-t_j)_+^s. \quad (5.2.1)$$

Närvaron av termen  $\mathbf{b}_{js} (x-t_j)_+^s$  ger att diskontinuitet tillåts vid knutpunkt  $t_j$  för den  $s$ -te ordningens derivata för  $f$  och  $f^{(s)}$ , där  $f^{(0)} = f$ , närvaron av lägre ordningens  $j$  påtvingar kontinuitet vid knutpunkt  $t_j$ . Följaktligen kan  $f$  göras kontinuerlig vid knutpunkt  $t_j$  genom att utesluta termen  $\mathbf{b}_{js} (x-t_j)_+^0$  och  $f^{(j)}$  kan göras kontinuerlig genom att utesluta termen  $\mathbf{b}_{js} (x-t_j)_+^s$ . Om vi vill att funktionen  $f$  skall vara kontinuerlig för den  $s$ -te derivatan vid en given knutpunkt  $t_j$  så vill vi även sannolikt att funktionen skall vara kontinuerlig för lägre ordningens derivator vid knutpunkt  $t_j$ . Kontinuitet införs genom att avlägsna all termer  $\mathbf{b}_{jm} (x-t_j)_+^m$  för  $m = 0, 1, 2, \dots, s$ . Olika restriktioner rörande kontinuitet kan införas genom att avlägsna termer från Ekvation 5.2.1. Eftersom avlägsnandet av en term, trunkerad potensfunktion, är det samma som att införa en kontinuitetsrestriktion så blir funktionen mjukare vid den givna knutpunkten. Den

mjukaste möjliga spline-modellen av grad  $S$  med  $J$  knutpunkter ges av Ekvation 5.2.1 enligt

$$f(x) = \sum_{s=0}^S b_{0s} x^s + \sum_{j=1}^J b_{jS} (x - t_j)_+^S,$$

där samtliga trunkerade potensfunktioner med en grad lägre än  $S$  avlägsnast. Givet att  $S \geq 2$  så är funktionen  $f$  samt första ordningens derivator kontinuerliga vid knutpunkt  $t_j$ .

### 5.3 Additiv semiparametrisk spline-regression

Definiera liksom tidigare  $X_i$  som en  $R$ -dimensionell vektor bestående av förklarande variabler samt  $Z_i$  som en  $P$ -dimensionell vektor av förklarande variabler och  $Y_i$  som en observation av en responsvariabel. Antag att relationen mellan  $Z_i$  och  $Y_i$  är linjär och att relationen mellan  $X_i$  och  $Y_i$  är okänd. En semiparametrisk modell kan då definieras enligt

$$Y_i = f(X_i) + Z_i \mathbf{b} + e_i,$$

där  $f$  är en ospecificerad funktion och  $Z_i \mathbf{b}$  är den lineära delen av regressionsmodellen. Vektor  $\mathbf{b}$  kan formuleras enligt  $\mathbf{b} = (\alpha, \beta_{01}, \beta_{02}, \dots, \beta_{0(P-1)})^T$  där  $\alpha$  är interceptet,  $\beta_{0p}$  är lutningskoefficienten för  $Z_i$  och  $e_i$  är feltermen vilken definieras som en oberoende och likafördelad  $N(0, \sigma^2)$  variabler. Den additiva semiparametriska modellen är ett specialfall av den ovan formulerade semiparametriska modellen. Ett additivt tillägg kan göras genom att utveckla  $f$  och  $X_i$  med  $R$ -dimensioner vilket ger en modell som kan formuleras enligt

$$Y_i = \sum_{r=1}^R f_r(X_{ir}) + Z_i \mathbf{b} + e_i, \quad (5.3.1)$$

där funktionen  $f_r$  är formulerad enligt en modifierad spline-funktion

$$f_r(X_{ir}) = \sum_{j=1}^{J_r} \mathbf{b}_{jr} (x_{rh} - t_{rj})_+^S. \quad (5.3.2)$$

Knutpunkten  $t_{rj}$  är den  $j$ -te knutpunkten för den  $h$ -te förklarande variabeln. En modell bestående av ett högt antal förklarande variabler kräver ett omfattande datamaterial för att ge goda koefficientskattningar. Fördelen med att använda sig av en spline-regression vid ickelineär modellering är att den kan skattas utifrån minsta kvadratmetoden vilken är en metod med låg beräkningsintensitet.

## 6 Logistisk regression

Den logistiska regressionen definieras av att  $n_i$ ;  $i = 1, 2, \dots, k$ , är oberoende stokastiska variabler och

$$n_i \sim \text{Bin}(N_i, p(X_i)), \quad i = 1, 2, \dots, k,$$

där

$$\text{logit}(p(X_i)) = \log\left(\frac{p(X_i)}{1 - p(X_i)}\right) = X_i \mathbf{b}, \quad i = 1, 2, \dots, k, \quad (6.1)$$

där  $X_i$  är en vektor bestående av förklarande variabler,  $(1, x_{1i}, \dots, x_{Ri})$ , och  $\mathbf{b}$  är en vektor bestående  $R$  parametrar. Den logistiska funktionen ligger i intervallet  $[0, 1]$  då  $X_i \mathbf{b}$  går från  $-\infty$  och  $+\infty$ . Den logistiska funktionen medför att det ej krävs några bivillkor på vektor  $\mathbf{b}$  för att uppfylla  $0 < p(X_i) < 1$ .

ML-skattningarna som används för att anpassa den multipla logistiska regressionsmodellen redovisas i Appendix D.

## 7 Semiparametrisk logistisk regression

Den logistiska regressionen är ett specialfall av den generaliserade lineära modellen vilken definieras enligt

$$E[Y_i | X_i] = F(a + X_i b),$$

där  $F(\bullet)$  är en länkfunktion. Då  $Y_i$  är binär gäller att

$$E[Y_i | X_i] = P(Y_i = 1 | X_i).$$

Vilket ger att den logistiska regressionsmodellen är en generaliserad lineär modell med en fördelningsfunktion  $F(\bullet)$  som länkfunktion. Detta ger att de utvidgningar som kan genomföras för en generaliserad lineär modell även kan göras för en logistisk regressionsmodell, se McCullagh & Nelder (1989).

Den semiparametriska modifikationen som avses i denna uppsats generaliserar det lineära argumentet till ett partiellt lineärt argument. Antag liksom tidigare två vektorer bestående av förklarande variabler,  $X_i$  och  $Z_i$ , där  $Z_i$  har en lineär relation till  $Y_i$  och  $X_i$  har en okänd relation till  $Y_i$ . Den generaliserade partiellt lineära modellen kan då formuleras enligt

$$E[Y_i | X_i, Z_i] = F(f(X_i) + Z_i b), \quad (7.1)$$

där vi liksom tidigare antar att  $F(\bullet)$  är den logistiska länkfunktionen. Det icke-lineära bidraget  $f$  och vektor  $\beta$  skattas då enligt samma metod som parametrarna för den vanliga logistiska regressionsmodellen skattas, enligt ML-metoden.

## 8 Modellutvärdering med AIC och SC

För att utvärdera modeller så kan *Akaike information criterion* (AIC) och/eller *Schwartz criterion* (SC) användas, se Akaike (1974) och Schwartz (1978). SC kan även benämnas som *Bayesian information criterion* (BIC). De två kriterierna ställer förbättringen av den logaritmerade *likelihoodfunktionen* mot kostnad för införandet av variabler i en modell. Det som skiljer de två kriterierna åt är strafftermen. Om vi låter  $r$  vara antalet förklarande variabler och  $s$  vara antalet responsnivåer minus ett så kan de två kriterierna definieras enligt

$$AIC = -2 \log L + 2(r + s),$$

och

$$SC = -2 \log L + (r + s) \log(k),$$

där  $\log L$  är den logaritmerade *likelihoodfunktionen* och  $k$  är antalet observationer. Den logaritmerade *likelihoodfunktionen* följer enligt

$$\log L = \sum_{i=1}^k \left\{ y_i \log \left( \frac{1}{1 + \exp(-X_i \mathbf{b})} \right) + (1 - y_i) \log \left( \frac{1}{1 + \exp(X_i \mathbf{b})} \right) \right\},$$

där  $y_i$  är den binära responsvariabeln. De två kriterierna användas för att utvärdera modeller som anpassats på samma datamaterial. Strafftermen för införandet av ytterligare variabler är större för SC jämfört med AIC vilket resulterar i val av modeller med färre antal förklarande variabler. För en icke-lineära spline-modell så kan AIC och SC användas för att utvärdera införandet av, samt antalet, trunkerade potensfunktioner. Det optimala antalet knutpunkter fås genom att minimera AIC och/eller SC.



## 9 Data

Datamaterialet som legat till grund för studien är en panel bestående av ca 10 miljoner kvartalsobservationer av svenska aktiebolag (AB). Panelen sträcker sig över tretton år och omfattar samtliga aktiebolag som levererat ett årsbokslut under perioden 1 januari 1990 till 31 december 2002. Totalt så inbegriper materialet 78 213 konkurser över hela perioden.

Det som karakteriserar ett aktiebolag är att det fordras 100 000 Kr för nyregistrering vid Patent- och registreringsverket (PRV). Aktiva bolag är även skyldiga att årligen inkomma med ett bokslut till PRV. Syftet med att endast inkludera aktiebolag i studien är att de utgör huvuddelen av den totala långivningen hos svenska banker. Under den finansiella krisen i början av 90-talet så stod aktiebolagen för 75 % av den totala kreditförlusten inom banksystemet.

Den företagsspecifika informationen har inhämtats från Upplysningscentralen AB (UC) vilket är det största kreditupplysningsföretaget i Sverige. Datamaterialet som inhämtats från UC innehåller huvudsakligen två källor till information: Först, företagsspecifik bokslutsinformation vilket UC inhämtat från de årsbokslut som årligen inlämnats till PRV. Bokslutsinformationen täcker perioden 1 januari 1989 till 31 december 2003. Dessutom innehåller materialet information rörande företagsspecifik anmärkningshistorik samt betalningshistorik. Den företagsspecifika informationen presenteras årsvis och anmärkningsinformationen presenteras enligt de datum anmärkningen uppkom. En mer detaljerad beskrivning av bokslutsinformationen följer nedan.

Datamaterialet rörande årsboksluten innehåller standardinformation från aktiebolagens balansrapporter (del av årsbokslutet) samt variabler rörande omsättning och vinst. Som komplement till bokslutsmaterialet återfinns även information rörande företagets betalningsbeteende och anmärkningshistorik. Betalningsanmärkningarna är baserade på 61 olika slags utestående kredit- och skattebetalningar. Lagringen och hanteringen av anmärkningsinformationen regleras av kreditupplysningslagen och datalagen samt övervakas av Datainspektionen. Exempel på händelser som registrerats är: försenad skattebetalning, utestående eller försenad kreditbetalning, konfiskering av

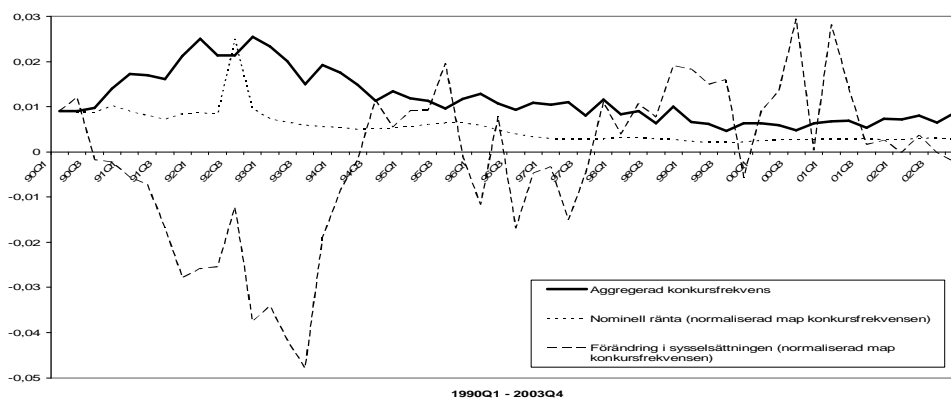
egendom, rekonstruktion av lån och företagskonkurs. En betalningsanmärkning medför problem för ett företag som ansöker om lån eller krediter. Populationen i undersökningen definieras av antalet företag som existerar vid tidpunkten  $t$  och som har inkommit med ett bokslut som täcker det aktuella kvartalet samt är klassificerade som aktiva. Ett företag klassificeras som aktivt om det har tillgångar samt en omsättning som överstiger 1000 kr. Studien exkluderar även företag som har blivit satta i konkurs under ett tidigare kvartal. De finansiella nyckeltalen som kommer att användas vid anpassningen av konkursriskmodellerna presenteras i Tabell 9.1 nedan. Givet värdena på förklaringsvariablerna så betraktas de successiva kvartalsobservationerna som oberoende.

<b>Nyckeltal</b>	<b>Kvot</b>
<b>EBITDA/TA</b>	<i>Vinst / Totala tillgångar</i>
<b>TL/TA</b>	<i>Total skuld / Totala tillgångar</i>
<b>I/TS</b>	<i>Inventarier / Omsättning</i>
<b>LA/TL</b>	<i>Likvida medel / Total skuld</i>
<b>IP/EBITDA</b>	<i>Räntebetalningar / Vinst</i>

Tabell 9.1: De finansiella nyckeltal

Utöver de finansiella nyckeltalen så tillkommer även en uppsättning indikatorvariabler:  $UTDE$ ,  $ANM1$ ,  $ANM2$  och  $TTBS$ .  $UTDE$  är en indikatorvariabel med  $UTDE = 1$  om företagets aktieägare erhåller utdelning under det aktuella kvartalet.  $ANM1$  indikatorvariabel med  $ANM1 = 1$  om någon av följande anmärkningar registreras under det aktuella kvartalet: utebliven betalning, inkommen konkursansökan eller utmätning av egendom.  $ANM2$  är en variabel med  $ANM2 = 1$  om företaget uteblir med skatteinbetalningar och  $TTBS$  är en variabel med  $TTBS = 1$  om företaget ej inkommer med årsbokslutet i tid. Utöver utdelningsvariabeln och anmärkningsvariablerna så tillkommer även en säsongvariabel som sätts till 1 för det tredje kvartalet varje år. Orsaken till att en indikatorvariabel inkluderas för det tredje kvartalet är att antalet konkursansökningar av naturliga skäl avtar under sommarmånaderna.

Konkursriskmodellerna kommer även att inbegripa den nominella räntan som en makrovariabel, benämnd *RÄNT*. Syftet med att inkludera den nominella räntan i modellerna är att den medför information rörande det aktuella konjunkturläget. Den aggregerade konkursfrekvensen tycks till stor utsträckning följa det allmänna konjunkturläget, se Figur 9.1 vilken illustrerar konkursandelen, den nominella räntan och en konjunkturindikator i form av förändringen i sysselsättningen. Relationerna antyder att ytterligare makrovariabler utöver den nominella räntan med fördel kan inkluderas i modellerna men då syftet i första hand är att undersöka om icke-lineär modellering av de finansiella nyckeltalen förbättrar de logistiska konkursriskmodellerna så har studien avgränsats till att endast inbegripa en makrovariabel.



Figur 9.1: Konkursandel per kvartal med avseende på samtliga svenska aktiebolag samt den procentuella förändringen i sysselsättningen och den nominella räntan (normaliserade med avseende på konkursfrekvensen, med avseende att förenkla den grafiska framställningen).

Medelvärden för de finansiella nyckeltalen samt indikatorvariablerna, med avseende på perioden 1990Q1 till 2002Q4, presenteras i Tabell A.1 Appendix A. Tabellen särskiljer företag som gått i konkurs och företag som ej gått i konkurs. Medelvärdet av en indikatorvariabel kan ses som ett procentuellt mått på hur stor andel av populationen som har egenskapen. Utifrån presentationen så är det uppenbart att företag som gått i konkurs generellt har en lägre vinst, högre skuldsättningsgrad samt lägre tillgång till likvida medel. Anmärkningsvariablerna tycks också indikera en förhöjd sannolikhet för att ett företag skall gå i konkurs: företag som gått i konkurs tenderar att ha en betydligt högre andel

betalningsanmärkningar samt utestående årsbokslut. Ett mått som exkluderar inverkan av eventuella extremobservationer är medianen vilken presenteras i Tabell A.3 Appendix A. Medianen återger liksom medelvärdet att det föreligger en skillnad mellan företag som går i konkurs och företag som ej går i konkurs. Skillnaden mellan de två grupperna är dock mindre då medianen beaktas vilket tyder på att datamaterialet rörande de finansiella nyckeltalen har en skev fördelning, eventuellt innehållande extremobservationer.

Relationen mellan konkursrisken och var och en av nyckeltalen redovisas i Figur A.1 Appendix A. Graferna i Figur A.1 illustrerar att det föreligger ett mer eller mindre icke-lineärt förhållande mellan samtliga nyckeltal och konkursrisken. Det är mer eller mindre uppenbart i det univariata fallet att den logistiska regressionsmodellen får det svårt att återge relationen mellan nyckeltalen och konkursrisken. En viktig slutsats som kan dras utifrån figuren är att nivåerna för de finansiella nyckeltalen har en inverkan på konkursrisken vilket antyder att de kommer att bidra med information i de empiriska modellerna. Utifrån denna slutsats så blir det tydligt att Figur A.1 har en koppling till Tabell A.3 genom att de båda illustrerar att de finansiella nyckeltalen ger en signal rörande företagets konkursrisk.

Ytterligare en egenskap som är intressant att studera är relationen mellan två nyckeltal och konkursrisken, se Figur A.2 Appendix A. Figur A.2 illustrerar de tio möjliga relationerna för de fem nyckeltalen. En intressant observation är att de icke-lineära relationerna mellan nyckeltalen och konkursrisken blir mindre tydliga när relationen mellan två nyckeltal och konkursrisken beaktas. Detta blir extra tydligt då det finansiella nyckeltalet likvida medel genom totala lån ( $LA/TL$ ) beaktas. Då ett företag har tillgång till likvida medel tycks olika nivåer av de övriga variablerna ha en låg inverkan på konkursrisken. Ickelineära relationer som framkommer i det univariata fallet tycks ej vara lika framträdande då relationen mellan två nyckeltal och konkursrisken beaktas.

Utöver sambanden mellan två nyckeltal så kan det vara intressant att beakta hur relationerna mellan de finansiella nyckeltalen och konkursrisken har förändrats över de femtiotvå kvartalen, se Figur A.3 Appendix A. Om vi beaktar den finansiella krisen i början av 1990 talet så ser vi att andelen konkurser med

avseende på det allmänna konjunkturläget främst påverkar nyckeltalens svansar. Figuren antyder att en neråtgående trend i det allmänna konjunkturläget främst leder till en förhöjd marginell konkursrisk med avseende på de finansiella nyckeltalen. Sambandet mellan konjunkturläget och konkursfrekvensen föranleder inbegripandet av makrovariabler i konkursriskmodellerna.

## 10 Lokala konfidensintervall för konkursrisken

Låt  $n_L$  beteckna antalet konkurser i ett givet intervall  $L$  med avseende på en förklarande variabel. En medelvärdesbildning fås då genom att låna information från angränsande intervall enligt  $\hat{p} = \frac{1}{2J+1} \sum_{i=L-J}^{L+J} \frac{n_i}{N_i}$ , där  $N_i$  är antalet observationer i intervall  $i$  och  $J$  är antalet intervall som används för att skatta  $p_i$  i ett område runt  $L$ . Vi förutsätter oberoende mellan olika intervall i antal konkurser, att  $\hat{p}$  är lokalt linjär med god approximation och att små  $p$  ger att  $p \approx p(1-p)$ . Antag att antalet konkurser i intervall  $i$  är Binomialfördelad,  $n_i \sim Bin(N_i, p_i)$ . Då följer väntevärdet och variansen enligt (1) respektive (2) nedan.

$$\begin{aligned} (1) \quad E[\hat{p}_i] &= E\left[\frac{1}{2J+1} \sum_{i=L-J}^{L+J} \frac{n_i}{N_i}\right] = \frac{1}{2J+1} \sum_{i=L-J}^{L+J} \frac{1}{N_i} E[n_i] \\ &= \frac{1}{2J+1} \sum_{i=L-J}^{L+J} \frac{1}{N_i} N_i p_i = \frac{1}{2J+1} \sum_{i=L-J}^{L+J} p_i. \end{aligned}$$

$$(2) \quad Var[\hat{p}_i] = Var\left[\frac{1}{2J+1} \sum_{i=L-J}^{L+J} \frac{n_i}{N_i}\right] = [Oberoende]$$

$$\begin{aligned} &= \frac{1}{(2J+1)^2} \sum_{i=L-J}^{L+J} \frac{1}{N_i^2} Var[n_i] = \frac{1}{(2J+1)^2} \sum_{i=L-J}^{L+J} \frac{1}{N_i^2} N_i p_i (1-p_i) \\ &= \frac{1}{(2J+1)^2} \sum_{i=L-J}^{L+J} \frac{p_i(1-p_i)}{N_i} \approx \frac{1}{(2J+1)^2} \sum_{i=L-J}^{L+J} \frac{p_i}{N_i}. \end{aligned}$$

Normalapproximation ger att

$$\hat{p}_i \sim N\left(\frac{1}{2J+1} \sum_{i=L-J}^{L+J} p_i, \frac{1}{(2J+1)^2} \sum_{i=L-J}^{L+J} \frac{p_i}{N_i}\right).$$

Med hjälp av normalapproximationen kan ett konfidensintervall för  $p_i$ , på en  $\alpha$ -nivå, formuleras enligt

$$CI_a = \frac{1}{2J+1} \sum_{i=L-J}^{L+J} p_i \pm I_{1-a} \sqrt{\frac{1}{(2J+1)^2} \sum_{i=L-J}^{L+J} \frac{p_i}{N_i}}.$$

(9.1)

Normalapproximerade konfidensintervall har beräknats för samtliga bokslutsvariabler och redovisas i Figur A.1 Appendix A, där  $J$  har satts till 2. Den grafiska framställningen illustrerar även antalet observationer som befinner sig inom det 95 procentiga konfidensintervallet. Under den förenklade förutsättningen att alla fem  $p_i$  är lika och att skattningarna är normalfördelade så skall ca 60 % av observationerna ligga inom konfidensintervallen vilket visar sig stämma för samtliga nyckeltal.

## 11 Konkursriskmodellerna

Konkursriskmodellerna som används i arbetet baseras på den multipla logistiska regressionsmodellen och en utvecklade multipla logistisk spline-regression. Den multipla logistiska regressionsmodellen som används följer enligt Ekvation 6.1 där samtliga finansiella nyckeltal och indikatorvariabler kommer att modelleras lineärt mot logit enligt

$$\log\left(\frac{p(Z_i)}{1-p(Z_i)}\right) = Z_i b ,$$

där  $Z_i$  är en radvektor med den första kolumnen i form av ettor följt av kolonner bestående av de finansiella nyckeltalen, indikatorvariablerna och makrovariabeln. Kolumnvektorn  $\mathbf{b} = (\mathbf{a}, \mathbf{b}_{10}, \mathbf{b}_{20}, \dots, \mathbf{b}_{p0})^T$  består av ett interceptet vid första positionen följt av lutningskoefficienter för de förklarande variablerna. Den multipla logistiska spline modellen som kommer att anpassas definieras enligt Ekvation 5.3.1 och 7.1. Modellen är en utvidgning av den multipla logistiska modellen och följer enligt

$$\log\left(\frac{p(X_i, Z_i)}{1 - p(X_i, Z_i)}\right) = \sum_{r=1}^R f_r(X_{ir}) + Z_i \mathbf{b}_r,$$

där  $Z_i$  och  $\mathbf{b}$  definieras liksom ovan och funktionen  $f_r(X_{ir})$  följer enligt Ekvation 5.3.2 med grad  $S = 2$ :

$$f_r(X_{ir}) = \sum_{j=1}^{J_r} \mathbf{b}_{jr} (x_{ir} - t_{rj})_+^2, \quad (11.1)$$

där  $x_{ir}$  är observation  $i$  av förklarande variabel  $r = 1, \dots, R$ . I uppsatsen så kommer vi att ha  $R = 5$  förklarande variabler och  $J_r = 3$  knutpunkter, placerade på 25e, 50e och 75e percentilen för vardera av de fem nyckeltalen. Syftet med att placera knutpunkterna enligt dessa percentiler är att uppnå bästa möjliga passform för de intervall där flertalet av observationerna återfinns. Koefficienterna ML estimeras enligt Newtons metod, skattningarna görs med hjälp av det statistiska paketet SAS.

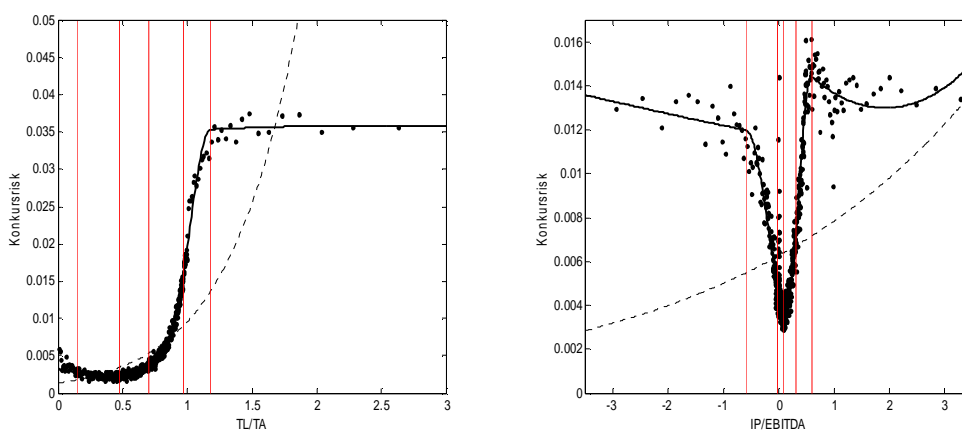
## 12 Ickelineär modellering av nyckeltalens inverkan på konkursrisken

För att illustrera hur en logistisk spline-regression kan användas för att fånga upp ickelineära förhållanden mellan de finansiella nyckeltalen och konkursrisken så presenteras två fall nedan. De två relationerna som kommer att presenteras är konkursriskens beroende av dels skuldsättningsgraden ( $TL/TA$ ), dels kvoten räntebetalningar/vinst ( $IP/EBITDA$ ). De två variablerna uppvisar en tydligt ickelineär relation till konkursrisken. Den logistiska spline modellen som kommer att anpassas följer enligt

$$\log\left(\frac{p(x_i)}{1-p(x_i)}\right) = a + b_{10}x_i + \sum_{j=1}^J b_{j1}(x_i - t_j)_+^2,$$

där fem knutpunkter,  $t_j$ , är placerade på 10e, 25e, 50e, 75e och 90e percentilen för vardera av de två finansiella nyckeltalen. Vi jämför också med en anpassning av den enkla logistiska modellen, som svarar mot  $\forall j: b_{j1} = 0$ . Syftet med att placera knutpunkterna enligt de angivna percentilerna är att uppnå bästa möjliga passform i de intervall där merparten av observationerna återfinns. Resultaten presenteras i Figur 11.1 nedan. De två graferna illustrerar den observerade konkursrisken för givna nivåer av de två finansiella nyckeltalen vilka är markerade med punkter, se även Figur A.1 Appendix A. Den heldragna linjen illustrerar den logistiska spline modellen och den streckade linjen illustrerar den vanliga logistiska modellen. De fem vertikala linjerna återger placeringen av de fem knutpunkterna. De två grafiska framställningarna visar att den logistiska spline modellen fångar upp de ickelineära sambanden mellan nyckeltalen och konkursrisken på ett bra sätt. Den vanliga logistiska modellen tenderar däremot att under- eller överskatta beroendet av  $TL/TA$  samt ge ett totalt set helt missvisande återgivelse av relationen mellan konkursrisken och  $IP/EBITDA$ .





**Figur 12.1:** Den logistiska modellen samt logistiska spline modellen anpassade för de finansiella nyckeltalen  $TL/TA$  och  $IP/EBITDA$ . Punkterna avser observerad konkursrisk, heldragen linje avser logistisk spline modell, streckad linjen avser enkel logistisk och de lodräta linjerna avser knutpunkternas placering.

Vidare så kommer de multivariata spline modellerna som presenteras i uppsatsen att bestå av tre knutpunkter för vardera av de fem finansiella nyckeltalen. Knutpunkterna kommer att placeras enligt den 24e, 50e och 75e percentilen för vardera av de fem nyckeltalen. Tre knutpunkter per finansiellt nyckeltal visar sig vara tillräckligt för att avsevärt förbättra konkursriskmodellernas egenskaper.

## 13 Modellresultat

I detta avsnitt redovisas resultat rörande de konkursriskmodeller som anpassats på hela panelen, 1990Q1 till 2002Q4, samt modeller som anpassats med avseende på de två panelerna där åren 1994 respektive 2002 utelämnats. Syftet med att utelämna ett år för vardera av de två panelerna är att få en fristående period för utvärdering, en period som ej legat till grund för modellanpassningen. Koefficientskattningarna som fås då en logistisk kreditriskmodell anpassas med avseende på hela panelen presenteras nedan, i Tabell 13.1.

Parameter	<i>Logistisk – ursprungspanel</i>			<i>Logistisk – trunkerad panel</i>		
	Estimat	s.e.	Pr> $\chi^2$	Estimat	s.e.	Pr> $\chi^2$
<b>Intercept</b>	-6,143600	0,008900	<,0001	-6,49390	0,01240	<,0001
<b>EBITDA/TA</b>	-0,000004	0,000011	0,6891	-0,57620	0,01230	<,0001
<b>TL/TA</b>	0,000010	0,000002	<,0001	0,56090	0,00900	<,0001
<b>I/TS</b>	0,000000	0,000000	0,3444	0,20090	0,01030	<,0001
<b>LA/TL</b>	-0,000040	0,000018	0,0163	-0,29510	0,00692	<,0001
<b>IP/EBITDA</b>	-1,928900	0,034500	<,0001	-1,45810	0,03320	<,0001
<b>UTDELNING</b>	0,000000	0,000000	0,8875	0,08120	0,00504	<,0001
<b>ANM1</b>	1,784600	0,016000	<,0001	1,63600	0,01500	<,0001
<b>ANM 2</b>	2,701100	0,010200	<,0001	2,57240	0,00986	<,0001
<b>TTBS</b>	4,549100	0,010400	<,0001	3,92240	0,00950	<,0001
<b>Säs – indikator</b>	-0,453600	0,011100	<,0001	-0,41930	0,01060	<,0001
<b>RÄNT</b>	0,045500	0,000672	<,0001	0,04180	0,00065	<,0001
<b>Pseudo R<sup>2</sup></b>		<b>0,337</b>			<b>0,354</b>	

Tabell 13.1: Koefficientskattningar samt förklaringsgraden<sup>1</sup> baserade på den ursprungliga panelen samt den trunkerade panelen.

Resultaten under rubriken ”Logistisk – ursprungspanel” i illustrerar att det av de fem finansiella nyckeltalen endast är skuldsättningsgraden samt räntebetalningar genom vinst som har en signifikant inverkan på konkursrisken. Samtliga indikatorvariabler samt den nominella räntan har en signifikant inverkan på konkursrisken. Det är tydligt då man betraktar koefficientskattningarna för den logistiska kreditriskmodellen att samtliga av de fem finansiella nyckeltalen har en relativt liten inverkan på konkursrisken. En orsak till detta kan vara att de observerade nyckeltalen innehåller extremobservationer vilka bidrar till de små koefficientskattningarna. En metod som använts vid tidigare studier rörande logistisk kreditriskmodellering är trunkering, se Shumway (2001). Genom att tilldela värdena för de finansiella nyckeltalen som är mindre än den första percentilen värdet av den första percentilen samt de som är över den nittionionde percentielen värdet av den nittionionde percentilen fås en trunkerad panel där eventuella extremobservationer normaliserats. Resultaten som fås då den logistiska kreditriskmodellen anpassats med avseende på den trunkerade panelen redovisas under rubriken ”Logistisk – trunkerad panel”, i Tabell 13.1.

Då den logistiska kreditriskmodellen anpassas på den trunkerade panelen så har samtliga av de fem finansiella nyckeltalen en signifikant inverkan på konkursrisken. Man kan utläsas ur tabellen att koefficientskattningarna för

<sup>1</sup> McFadden´s Pseudo R<sup>2</sup>, se appendix D

samtliga nyckeltalet, bortsett från *IP/EBITDA*, går i linje med graferna i Figur A.1. Förklaringsgraden samt koefficientskattningarna för de två modellerna pekar på att den främsta källan till information fås från indikatorvariablerna rörande anmärkningar och försenat årsbokslut. En modell som exkluderar indikatorvariablerna redovisas i Tabell B.1 Appendix B. Förklaringsgraden som fås för modellen som ej inbegriper indikatorvariablerna är 0,0571 vilket är avsevärt lägre än för den modell som inbegriper indikatorvariablerna. Då modellerna ovan visar att det främst är de variabler rörande anmärkningar samt uteblivet/försenat årsbokslut som förklarar konkursrisken så kommer även en logistisk spline modell utan dessa variabler att utvärderas.

Utöver de modeller som anpassats med avseende på hela perioden, 1990Q1-2002Q4 så har tre konkursriskmodeller anpassats för vardera av de två delperioderna: en logistisk modell baserad på den ursprungliga panelen, en logistisk modell baserad på en trunkerad panel och en logistisk spline modell baserad på den trunkerade panelen. Modellerna som utvärderas för de två delperioderna kommer att inkludera samtliga variabler. Koefficientskattningarna för de tre konkursriskmodellerna för vardera av de två delperioderna redovisas i Tabell B.2, B.3 och B.5 Appendix B. Placeringen av knutpunkterna för samtliga logistiska spline modeller redovisas i Tabell B.6 Appendix B och analysen av modellernas egenskaper utförs enligt en CAP-analys vilken redovisas i efterföljande avsnitt.

Skattningarna som redovisas i Appendix B gör gällande att de logistiska konkursriskmodellerna för de två delperioderna har koefficientskattningar samt en förklaringsgrad som är i linje med de konkursriskmodeller som anpassats med avseende på hela perioden. Resultaten rörande den logistiska spline-modellen gör gällande att flertalet av de trunkerade potensfunktionerna har en signifikant inverkan på konkursriskmodellerna. Förklaringsgraden för de modeller som inkluderar indikatorvariablerna får endast en mindre förbättring då de finansiella nyckeltalen modelleras icke-lineärt. En klar förbättring uppkommer dock för de modeller som ej inkluderar indikatorvariablerna, här går förklaringsgraden från 0.050 till 0.094 då de finansiella nyckeltalen modelleras icke-lineärt.

Två mått som kan användas för att utvärdera införandet av de trunkerade potensfunktionerna i modellen är AIC och SC vilka redovisas i avsnitt 8. Nedan så följer en tabell över värden för de modeller som anpassats med avseende på hela perioden, resultatet avser med och utan indikatorvariabler.

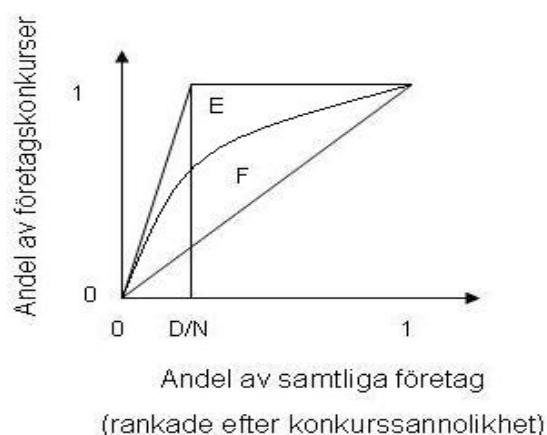
	<i>Total, inklusive indikatorvariabel</i>		<i>Total, exklusive indikatorvariabel</i>	
	<b>Logistisk</b>	<b>Logistisk spline</b>	<b>Logistisk</b>	<b>Logistisk spline</b>
<b>Log likelihood</b>	-298366	-290385	-435557	-418526
<b>AIC</b>	596758	580826	871130	837101
<b>SC</b>	596942	581223	871243	837441
<b>Pseudo R<sup>2</sup></b>	<b>0,354</b>	<b>0,371</b>	<b>0,057</b>	<b>0,093</b>

**Tabell 13.2:** Log likelihood, AIC och SC för den logistiska och den logistiska splinmodellen, med och utan indikatorvariabler, anpassade med avseende på hela perioden, 1990Q1 till 2002Q4. Måtten för samtliga modeller redovisas i Appendix B.

Både AIC och SC är högre för den logistiska modellen jämfört med den logistiska spline modellen. Skillnaden är störst mellan de modeller som ej inkluderar indikatorvariablerna vilket antyder att den ickelineära utvecklingen får en större inverkan då endast informationen från nyckeltalen beaktas. AIC och SC indikerar att införandet av de trunkerade potensfunktionerna ger en förbättrad anpassning.

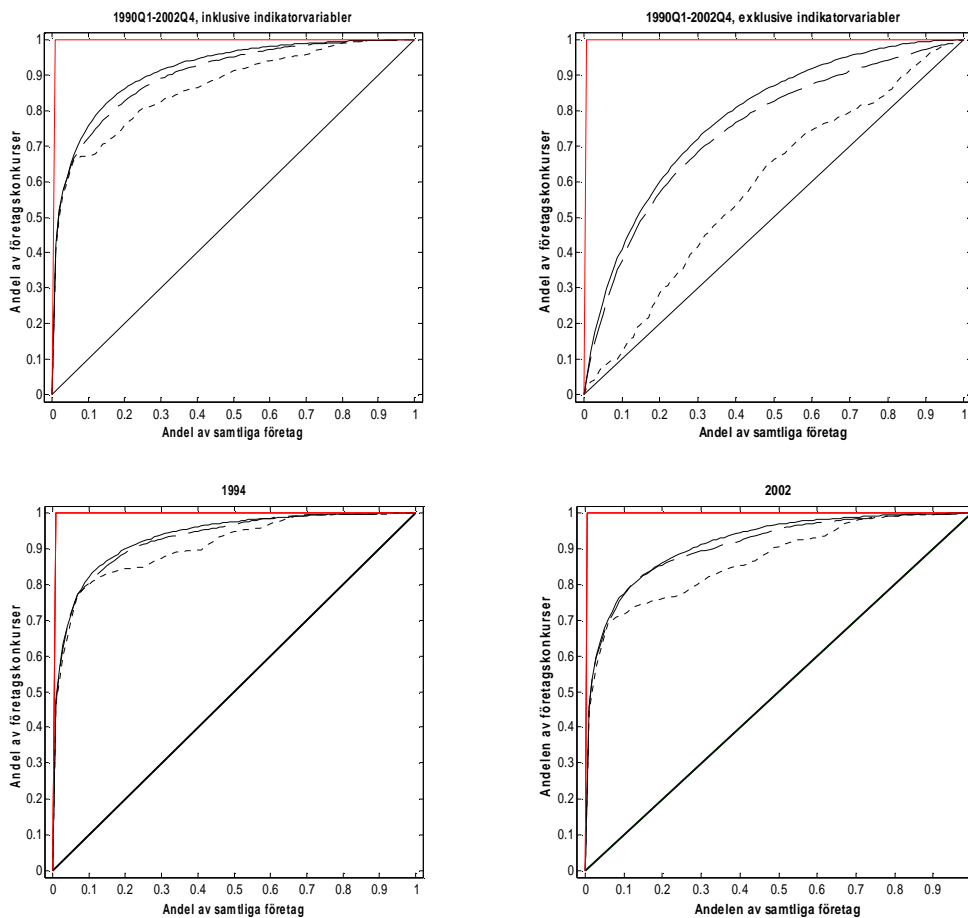
## 14 Utvärdering

Utvärderingen av konkursriskmodellerna kommer att göras med avseende på hela perioden, 1990Q1 till 2002Q4, samt på de två delperioderna 1994 och 2002. Modellerna som kommer att utvärderas utifrån de två delperioderna har baserats på två paneler där vardera av de två delperioderna har exkluderats. Syftet med att inkludera denna övning i analysen är att utvärdera hur väl konkursriskmodellerna förutsäger konkurser ur ett material som ej legat till grund för modellenanpassningen. Valet av de två åren har gjorts baserat på att 1994 och 2002 skiljer sig åt i form av att det första av de två åren karakteriseras av en relativt hög konkursfrekvens och det andra av en relativt låg konkursfrekvens.



Figur 14.1: CAP-kurvan.

En vanligt förekommande metod som används för att utvärdera konkursriskmodellens prestanda är CAP-analysen (Cumulative accuracy profil), se De Servigny & Renault (2004). CAP-analysen utvärderar konkursriskskattningarna genom att rangordna den skattade konkursrisken mot den faktiska konkursrisken, se Figur 14.1 ovan. Detta görs genom att först rangordna företagen i en avtagande ordning, baserat på konkursriskskattningarna som fås via modellerna, vilka redovisas på den vertikala axeln. Den lodräta axeln avser andelen av samtliga företagskonkurserna. Den perfekta modellen skulle tilldela de  $D$  högsta konkursriskskattningarna till de  $D$  utav  $N$  företag som går i konkurs. Den perfekta modellen skulle således vara en rät linje mellan  $(0,0)$  och  $(D/N,1)$ . Motsatsen är en modell som ger en slumpvist tilldelad konkursriskskattning mellan företag med hög och låg risk. CAP-kurvan blir då en rät linje mellan  $(0,0)$  och  $(1,1)$ . Ett kvantitativt mått som kan användas för att jämföra olika modeller är AR-kvoten (Accuracy ratio), även kallad Gini-kvoten, vilket mäter en modells prestanda genom kvoten  $G=F/(E+F)$ , se Figur 14.1 ovan. Värdet på AR-kvoten påverkas positivt av konkursriskmodellernas förmåga att förutsäga konkurser.



Figur 14.2: CAP-kurvor för samtliga modeller med avseende på den totala panelen samt de två utvärderingsperioderna 1994 och 2002. Den heldragna linjen avser den logistiska spline modellen, den streckade linjen under den heldragna linjen avser den logistiska modellen baserad på den trunkerade panelen och den fint streckade linjen avser den logistiska modellen baserad på den ursprungliga panelen. En kompletterande illustration av CAP-kurvan återfinns i Appendix B Figur B.1, där en grafisk illustration av intervallet  $[0, D/N)$  återges för delperioderna 1994 och 2002 .

Det framgår utifrån Figur 14.2 att samtliga modeller som inkluderar indikatorvariablerna uppvisar samma prestanda för de företag med högst skattad konkursrisk. För hela perioden samt för 1994 så uppkommer avvikelser mellan modellerna efter den åttonde percentilen och för 2002 så uppkommer en avvikelse efter den femte percentilen. Den motsvarande förklaringen motsvarar mellan 70 och 78 % av samtliga konkurser. För de modeller som ej inkluderar indikatorvariablerna så framgår det att modellernas prestanda klart förbättras då de anpassas med avseende på den trunkerade panelen. CAP-kurvan för den

logistiska spline-modellen ligger över kurvan för den vanliga logistiska modellen för samtliga företag. CAP-kurvorna indikerar liksom modellernas förklaringsgrad att den främsta källan till information rörande ett företags konkursrisk fås via anmärkningsinformationen och försenade/uteblivna årsbokslut. AR-kvoterna som fås i Figur 14.2 ges i Tabell 14.1 nedan.

<i>Total, inklusive indikator variabler</i>			<i>Total, exklusive indikatorvariabler</i>		
<b>Logit</b>	<b>Logistisk – trunk</b>	<b>Logistisk spline</b>	<b>Logit</b>	<b>Logistisk – trunk</b>	<b>Logistisk spline</b>
0,747	0,817	0,845	0,186	0,507	0,577
<i>1994</i>			<i>2002</i>		
<b>Logit</b>	<b>Logistisk – trunk</b>	<b>Logistisk spline</b>	<b>Logit</b>	<b>Logistisk – trunk</b>	<b>Logistisk spline</b>
0,832	0,872	0,884	0,751	0,832	0,848

Tabell 14.1: AR-kvoter för samtliga konkursriskmodeller med avseende på den totala perioden samt de två utvärderingsperioderna 1994 respektive 2002. Modellerna som anpassats med avseende på de två utvärderingsperioderna inkluderar samtliga variabler.

AR-kvoterna som fås utifrån den logistiska spline-modellen är högre både för de modeller som anpassats på hela perioden samt för de två utvärderingsperioderna. AR-kvoten är 3 respektive 14 % högre mellan den logistiska och den logistiska spline-modellen för de modeller som inkluderar respektive exkluderar indikatorvariablerna. Skillnaden för de två utvärderingsperioderna ligger på 1 respektive 2 %.

En intressant iakttagelse är att AR-kvoten för samtliga konkursriskmodeller är högre med avseende på den första utvärderingsperioden. Orsaken till detta tycks vara att indikatorvariablerna rörande betalningsanmärkningar och försenat årsbokslut har en högre frekvens under år som karakteriseras av en hög konkursandel. Det tydligaste exemplet rör frekvensen avseende försenat årsbokslut, TTBS, vilken går mellan 0,0237 och 0,0025 för 1994 respektive 2002 vilket indikerar en betydande skillnad mellan de två åren, se Tabell A.2. Även anmärkningsvariabeln ANM1 förändras från 0,0072 till 0,0046 mellan 1994 och 2002. Enligt de skattade parametrarna så framgår det att betalningsanmärkningar samt försenat årsbokslut har en avsevärd inverkan på ett företags konkursrisk. Då andelen betalningsanmärkningar och uteblivna årsbokslut ökar, för företag som

tenderar att gå i konkurs, under perioder karakteriserade av hög konkursfrekvens så ökar modellernas prestanda.

En kompletterande metod som är vanligt förekommande vid utvärdering av konkursriskmodeller är en metod som benämns decilranking. Decilrankingen görs liksom CAP-analysen genom att rangordna samtliga konkursskattningar i en avtagande ordning samt dela in dem i deciler. För varje decil beräknas sedan andelen av samtliga konkurser vilket redovisas i Tabell 14.2 nedan.

<i>Total, inklusive indikatorvariabler</i>			<i>Total, exklusive indikatorvariabler</i>			
<b>Decile</b>	<b>Logistisk - ursprunglig panel</b>	<b>Logistisk - trunkerad panel</b>	<b>Logistisk spline</b>	<b>Logistisk - ursprunglig panel</b>	<b>Logistisk - trunkerad panel</b>	<b>Logistisk spline</b>
<b>1</b>	0,673	0,724	0,757	0,120	0,378	0,410
<b>2</b>	0,083	0,104	0,105	0,165	0,191	0,189
<b>3</b>	0,069	0,063	0,052	0,130	0,114	0,123
<b>4</b>	0,039	0,037	0,032	0,117	0,083	0,086
<b>5</b>	0,047	0,024	0,021	0,130	0,062	0,063
<b>6 to 10</b>	0,089	0,048	0,033	0,338	0,173	0,129

<i>1994</i>			<i>2002</i>			
<b>Decil</b>	<b>Logistisk – ursprungs panel</b>	<b>Logistisk – trunkerad panel</b>	<b>Logistisk spline</b>	<b>Logistisk – ursprungs panel</b>	<b>Logistisk – trunkerad panel</b>	<b>Logistisk spline</b>
<b>1</b>	0,801	0,803	0,822	0,718	0,771	0,776
<b>2</b>	0,043	0,084	0,078	0,043	0,083	0,085
<b>3</b>	0,030	0,040	0,039	0,049	0,039	0,050
<b>4</b>	0,023	0,022	0,023	0,044	0,029	0,034
<b>5</b>	0,051	0,020	0,013	0,053	0,033	0,023
<b>6 till 10</b>	0,053	0,031	0,024	0,094	0,045	0,032

Tabell 14.2: Andel av samtliga konkurser per decil med avseende på den skattade konkursrisken, företagen rangordnade efter avtagande konkursrisk. Decilrankingen med avseende på de två utvärderingsperioderna baseras på modeller som inkluderar samtliga variabler.

Decilrankingen visar liksom AR-kvoterna att indikatorvariablerna har en klart positiv inverkan på modellernas rangordningsegenskaper. En skillnad är dock att spline-modellen enligt decilrankingen även tycks förbättra möjligheten att urskilja de företag med den högsta faktiska konkursrisken vilket resulterar i att andelen konkurser ökar i den övre decilen. Med utgångspunkt från att CAP-analysens AR-kvoter illustrerar att spline-modellen främst förbättrar konkursriskskattningarna



för 2002 så ger decilrankingen ett intressant resultat. Resultaten för de modeller som inkluderar indikatorvariablerna tyder på att den icke-lineära modelleringen av de finansiella nyckeltalen främst bidrar till en förbättrad möjlighet att skatta konkursrisken för de 90 % av företagen med lägst skattad konkursrisk. Förbättringen för de modeller som endast baseras på de finansiella nyckeltalen uppvisar en förbättring för samtliga företag.

En svaghet med analysen ovan är att varken CAP-analysen eller decilrankingen återger sambandet mellan konkursriskskattningarna och den faktiska konkursrisken. De två analysmetoderna utvärderar endast rangordningsegenskaperna för modellerna och inte hur den skattade konkursrisken förhåller sig till den skattade konkursrisken. Utifrån resultaten ovan kan det vara av betydelse att utvärdera hur väl den skattade konkursrisken återger den faktiska konkursrisken. För de modeller som inkluderar indikatorvariablerna så kommer utvärderingen att ske för de 90 % av företagen med lägst skattad konkursrisk och för de modeller som ej inkluderar indikatorvariablerna så kommer utvärderingen att ske för samtliga företag.

För att återge hur väl modellerna återger den faktiska konkursrisken så har företagen med avseende på den totala perioden samt för de två delperioderna, 1994 och 2002, liksom tidigare rangordnats efter modellernas skattade konkursrisk. För varje percentil beräknades sedan medelvärdet av den skattade konkursrisken samt medelvärdet av konkursvariabeln. I Figur 14.3 illustreras utfallen med avseende på de tre konkursriskmodellerna för den totala perioden, där den skattade konkursrisken ställs mot den faktiska. I Figur B.2 Appendix B så återges utfallet för de två delperioderna.

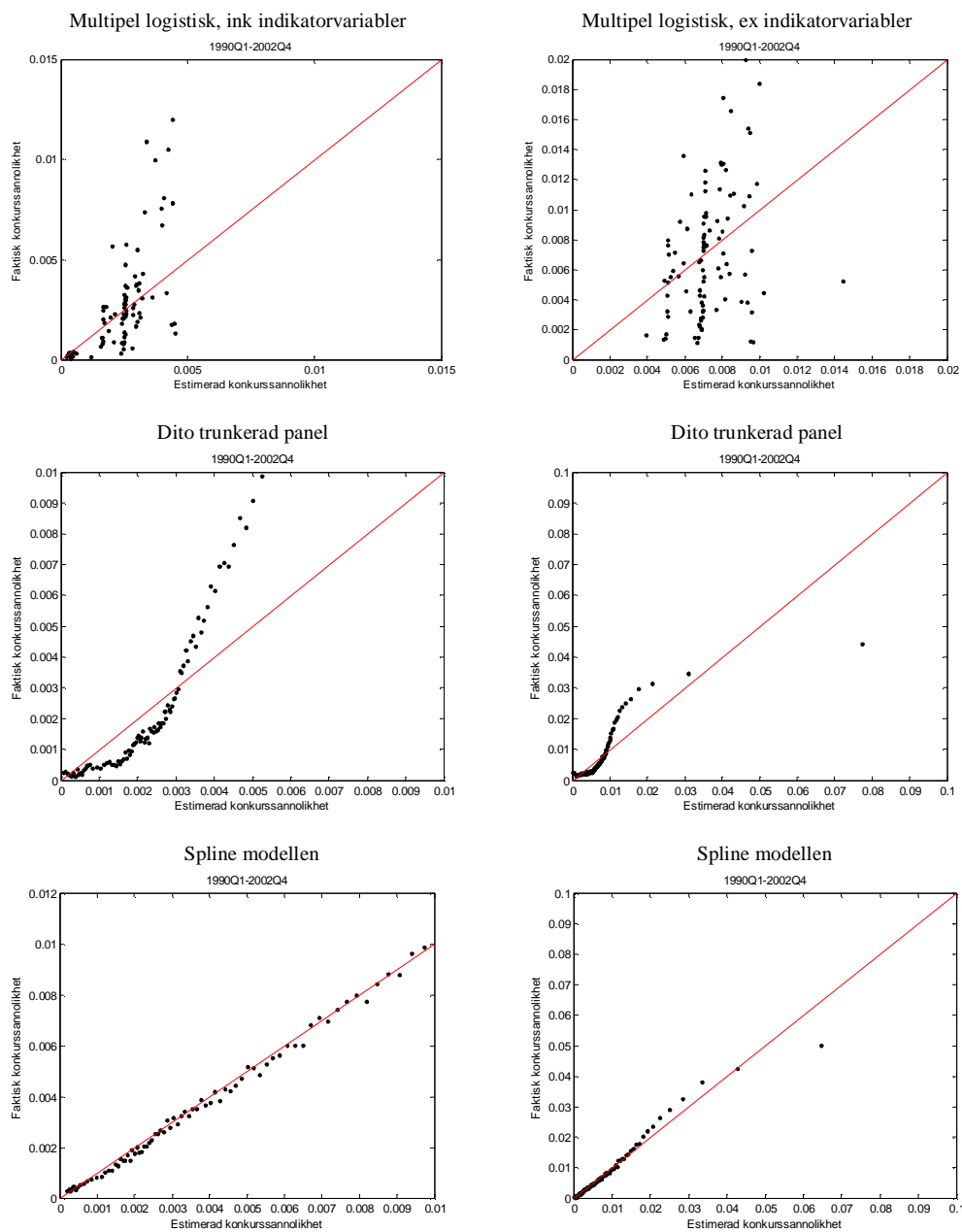
Den perfekta modellen ska följa ett 1 till 1 förhållande mellan genomsnittlig skattad konkursrisken och genomsnittlig faktisk konkursrisk vilket grafiskt framställs av 45-graderslinjen. De övre två graferna i Figur 14.3 och Figur B.2 avser den logistiska modellen som anpassats med avseende på den otrunkerade panelen. Den grafiska framställningen illustrerar att det föreligger en stor varians mellan den skattade konkursrisken och det faktiska utfallet. Om man endast beaktar interceptet så hamnar modellen som inkluderar indikatorvariablerna på en

skattade konkursrisken om 0.2 % och för modellen som ej inkluderar indikatorvariablerna så ligger respektive skattning på 0.6 %. Ur graferna som baserats på hela perioden så går det att urskilja en ansamling kring 0.2 och 0.6 % för de två modellerna. Nivåskillnaden kan förklaras med att indikatorvariabeln bidrar till att differentiera konkursriskskattningarna mellan företag med hög och låg konkursrisk. För de två delperioderna, se Figur B.1, så uppstår en tydlig bandbildning. Nivåskillnader för banden kan härledas till de tre indikatorvariablerna rörande anmärkningar och försenat/uteblivet årsbokslut. Mindre fluktuationer i banden beror dock på ingående bokslutsvariabler.

De två mellersta graferna i Figur 14.3 och B.2 avser den logistiska kreditriskmodellen baserad på den trunkerade panelen. Här framgår det tydligt att trunkeringen medför att informationen i bokslutsvariablerna bidrar till en differentiering av konkursriskskattningarna. Modellerna tenderar dock att överskatta konkursrisken för de lägre nivåerna samt underskatta konkursrisken för de högre nivåerna. Sambanden mellan skattad och faktisk konkursrisk följer samma mönster för samtliga perioder men skiljer sig något mellan modellen med och utan indikatorvariabler. Då indikatorvariablerna inkluderas så ökar spridningen mellan konkursriskskattningarna vilket följer av att indikatorvariablerna ökar möjligheten att särskilja företag med olika risknivåer. Detta är extra tydligt för år 1994 vilket kan förklaras av att frekvensen för indikatorvariablerna rörande betalningsanmärkningar och försenat årsbokslut var hög under detta år vilket bidrar till att en högre andel företag får en förhöjd skattning av konkursrisken.

De sista två graferna i Figur 14.3 och B.2 avser de skattningar som den logistiska spline-modellen ger. Det framgår tydligt att det ickelineära bidragen leder till konkursriskskattningar får en klart förbättrad återgivning av den faktiska konkursrisken. Sambandet mellan den skattade konkursrisken och den faktiska konkursrisken pressas mot 45-graderslinjen och den tydliga över och underskattningen uteblir. Den logistiska spline-modell som ej inkluderar indikatorvariablerna uppvisar dock en mindre avvikelse för de högre risknivåerna, avvikelserna är dock avsevärt mindre i jämförelse med den vanliga logistiska modellen. En viktig iakttagelse är även att differentieringen tenderar att öka,

spline modellen tycks förbättra möjligheten att särskilja företag som går i konkurs. För de två utvärderingsperioderna, se Figur B.2, så ökar spridningen vilket kan förklaras med att perioderna inbegriper ett lägre antal konkursobservationer. Mönstret för de två perioderna följer samma mönster som för den totala perioden, den logistiska spline-modellen visar sig ge en förbättrad återgivning av den faktiska konkursrisken.



Figur 14.3: Faktisk mot estimerad konkurssannolikhet med avseende på de tre konkursriskmodellerna, för den totala perioden.

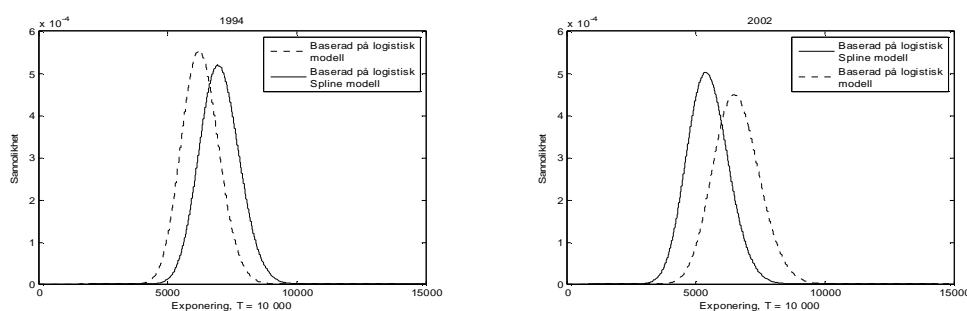
Den huvudsakliga förklaringen till att den logistiska spline-modellen ger avsevärt bättre konkursriskskattningar är att de trunkerade potensfunktionerna i spline-modellen bidrar till att fånga upp information från de finansiella nyckeltalen. Resultatet blir en högre differentiering av konkursriskskattningarna samt bättre återgivning av den faktiska konkursrisken. Det är dock tydligt utifrån graferna att även den logistiska spline-modellen tenderar att överskatta konkursrisken för företag med låg risk under år som karakteriserade av hög andel konkurser. Men överskattningen är inte lika omfattande och systematisk som för den vanliga logistiska konkursriskmodellen.

Resultaten som presenteras ovan bör ses som ett bra komplement till CAP-analysen och decilrankingen. Fördelarna med att modellera de finansiella nyckeltalen icke-lineärt är inte enbart att rangordningsegenskaperna förbättras utan även att den skattade konkursrisken ger en bättre återgivning av den faktiska konkursrisken. Rangordningen av företags risknivåer är viktigt ur ett kreditbedömningshänseende då de faktiska värdena ej är av betydelse utan måttet kan betraktas som kategoriskt. Men då konkursriskskattningarna skall ligga till grund för kreditriskbedömningen av en banks utlåningsportfölj så är det även av yttersta vikt att konkursriskskattningen ger en korrekt mått av den faktiska konkursrisken. Skillnaden mellan att beräkna kreditrisken för en bankportfölj baserat på den logistiska modellens och den logistiska spline-modellens kreditriskskattningar kommer att kvantifieras i efterföljande kapitel. Syftet är att visa att ett förbättrat konkursriskmått får en betydande inverkan på bankers kreditriskbedömning.

## 15 Portföljutvärdering

Den huvudsakliga orsaken till att ett korrekt konkursriskmått, PD, har fått en så stor betydelse enligt de nya Basel II reglerna är att det utgör en viktig variabel vid beräkningen av en banks kreditrisk. Om konkursriskmålet förbättras så blir effekten att ett bättre kreditriskmått erhålls. För att kvantifiera inverkan av de förbättrade konkursriskskattningarna som fås via den logistiska spline-modellen

så har kreditrisken för två exempelportföljer beräknats enligt Credit Suisse kreditriskmodell CR<sup>+</sup>, se Appendix C. Kreditriskberäkningarna har baserats på 20 000 slumpvist valda företag från de två åren 1994 och 2002. För att få två homogena portföljer så innehåller stickproven endast företag som har en omsättning som ligger inom den tionde och nittionde percentilen för vardera av de två åren. Exponeringen givet konkurs har satts till 40 % och lånen har beräknats enligt en basenhet på 10 000 kr. Figur 15.1 nedan illustrerar kreditriskfördelningarna<sup>1</sup> med avseende på de två exempelportföljerna. Det framgår tydligt av den grafiska framställningen att kreditriskfördelningen baserad på konkursriskskattningarna från den logistiska spline-modellen ger en högre förväntad kreditförlust för 1994 och en lägre förväntad kreditförlust för 2002 jämfört med den logistiska modellen.



Figur 15.1: Fördelning för kreditförlusten med avseende på de två åren 1994 och 2002.

Ett mått som kan användas för att utvärdera kreditriskfördelningarna är VaR<sub>α</sub> (Value at Risk). För en konfidensnivå, även kallad solvensnivå eller solvenssannolikhet,  $1 - \alpha \in (0,1)$ . Definieras VaR<sub>α</sub> för den stokastiska variabeln  $L$  enligt

$$\text{VaR}_\alpha [L] = \inf \{c \in R : P(L \leq c) \geq \alpha\}.$$

Då förlustfördelningen beräknas diskret så ges VaR<sub>α</sub> enligt

<sup>1</sup> Beräkningen av förlustfördelningen görs enligt Ekvation C.3 Appendix C.

$$\text{VaR}_a [L] = \bar{r}_a (L) = \inf \{n \in N : P(L \leq n) \geq a\}.$$

Tabell 15.1 nedan presenterar en uppsättning VAR-nivåer baserat på de två modellernas konkursriskskattningar för de två perioderna.

Model	a -nivåer					Förväntad förlust
	0,95	0,975	0,99	0,995	0,999	
<b>1994: Logit</b>	0,94%	0,98%	1,01%	1,04%	1,09%	0,79%
<b>1994: Spline</b>	1,05%	1,08%	1,12%	1,15%	1,20%	0,88%
<b>2002: Logit</b>	0,71%	0,74%	0,78%	0,80%	0,85%	0,58%
<b>2002: Spline</b>	0,60%	0,63%	0,66%	0,68%	0,72%	0,48%

Tabell 14.1: Value at risk – procentuell förlust av exponering med avseende på en uppsättning solvensnivåer.

Resultaten i Tabell 15.1 återger att den logistiska spline-modellen genererar konkursriskskattningar som resulterar i en högre förväntad kreditrisk under den första perioden och lägre förväntad kreditrisk under den andra perioden. Skillnaden mellan de två perioderna blir extra tydlig då svansarna för fördelningen beaktas. En orsak till detta tycks vara att 1994 karakteriseras av en relativt högre konkursrisk vilket bidrar till att den logistiska modellen bidrar till en underskattning av kreditrisken. Det framgår tydligt av Figur 14.3 att den skattade konkursrisken för den logistiska kreditriskmodellen har en lägre differentiering mellan låg- och högriskföretag samt att skattningarna ej ger en god återspeglning av risken för de företagen med hög konkursrisk. Den logistiska spline-modellen ger konkursriskskattningar med högre spridning mellan låg- och högriskföretag vilka även ger en avsevärt bättre återspeglning av den faktiska konkursrisken. Resultaten som då uppkommer vid en kreditriskvärdering är att den vanliga logistiska konkursriskmodellen tenderar att leda till en underskattning av kreditrisken. Motsatt förhållande ges då portföljen baserad på 2002 beräknas. Resultatet blir att kreditrisken överskattas med avseende på den logistiska konkursriskmodellen. Även detta kan utläsas ur Figur 14.3 där det tydligt framgår att merparten av observationerna för den logistiska spline-modellen ligger längst 45-graderslinjen nära origo till skillnad från skattningarna med avseende på den logistiska modellen. Resultatet blir att den logistiska konkursriskmodellen ger

konkursriskskattningar som leder till en överskattning av exempelportföljens kreditrisk.

En slutsats som kan dras utifrån Figur 15.1 och Tabell 15.1 är att kreditriskfördelningen baserad på den logistiska spline-modellen har en högre varians för den första perioden. En högre varians påverkar längden på svansarna samt kreditriskmålet för de övre percentilerna. Den andra perioden illustrerar det motsatta. Konkursriskskattningarna för den logistiska spline-modellen resulterar i en lägre förväntad kreditrisk samt lägre varians. Även detta är en konsekvens av de skattade konkursriskerna som fås från de två modellerna. Det framgår tydligt av Figur 15.1 att spridningen som uppkommer för den logistiska konkursriskmodellen är högre under 1994 och lägre under 2002 vilket resulterar i att variansen för kreditrisken för de två perioderna påverkas.

Beräkningarna visar att konkursriskskattningarna har en tydlig inverkan på beräkningen av kreditrisken. Givet tidigare resultat så är det tydligt att kreditrisken som beräknas med avseende på den logistiska spline-modellen leder till en korrektare bedömning av den beräknade kreditriskfördelningen. Ett förbättrat konkursriskmått visar sig således ge en klar inverkan på de reservavsättningar som banker måste hålla för att täcka eventuella kreditförluster.

## 16 Slutsatser

Syftet med uppsatsen var att utvärdera om en icke-lineär utveckling av den logistiska konkursriskmodellen förbättrar möjligheten att skatta konkursrisk för svenska aktiebolag. Teorin rörande icke-lineära spline-regressionen och icke-lineär logistisk spline-regression presenterades i avsnitt 2 till 8. Vidare så följde en utvärdering av den logistiska och den icke-lineära logistiska konkursriskmodellen i efterföljande avsnitt.

Resultaten som presenteras i uppsatsen gör gällande att den huvudsakliga källan till information rörande ett företags konkursrisk fås från anmärkningshistorik och försenade/uteblivna årsbokslut. Förklaringsgraden gick från 0.057 till 0.363 då indikatorvariabler rörande anmärkningar och försenade uteblivna årsbokslut inkluderas i modellerna. Dessa indikatorvariabler visade sig enligt CAP-analysen stå för den huvudsakliga förklaringen, 70 – 80 %, av konkurserna med avseende på hela panelen samt de två utvärderingsperioderna. För de två utvärderingsperioderna så varierade konkursriskmodellernas prestanda mellan år med relativt hög och låg konkursrisk vilket kan förklaras av att frekvensen för anmärkningsvariablerna är högre för år karakteriserade av en hög konkursandel.

Den icke-lineära utvecklingen av den logistiska konkursriskmodellen gav modeller med högre förklaringsgrad samt konkursriskskattningar med förbättrade rangordningsegenskaper. Förbättringen påvisades dels genom CAP-analysens AR-kvoter samt resultaten med avseende decilrankingen. Den icke-lineära utvecklingen fick störst inverkan på de modeller som ej inkluderade indikatorvariabler rörande anmärkningsinformation och försenat/uteblivet årsbokslut. Förklaringsgraden för denna modell uppvisade en fördubbling då de finansiella nyckeltalen modellerades icke-lineärt.

För de modeller som inkluderade indikatorvariabler så framgick det att den icke-lineära utvecklingen främst förbättrar rangordningsegenskapen med avseende på de 90 % av företagen med lägst skattad konkursrisk. Vidare så framgick det att konkursriskskattningarna för de 90 % av företagen med lägst skattad konkursrisk avsevärt förbättrades då de finansiella nyckeltalen modelleras



ickelineärt. Den grafiska framställningen av de skattade konkursriskerna mot de faktiska konkursriskerna visade att den logistiska spline modellen gav konkursriskskattningar med högre differentiering mellan låg- och högriskföretag samt en klart förbättrad återgivning av den faktiska konkursrisken.

Avslutningsvis så visade utvärderingen av de två exempelportföljerna att de förbättrade konkursriskskattningarna får en betydande inverkan på kreditriskfördelningen. Differentieringen av de förbättrade konkursriskskattningarna påverkade dels den förväntade kreditförlusten samt variansen för den totala fördelningen. Ett förbättrat konkursriskmått visar sig således få en betydande inverkan på beräkningen av bankers reservavsättningar.

Utifrån resultaten i uppsatsen så har följande fyra förslag för vidare arbete formulerats:

- i. För att belysa fördelen med icke-lineär modellering så kan vidare studier med fördel avgränsas till modeller som endast inbegriper de finansiella nyckeltalen.
- ii. Expandera antalet knutpunkter i modellerna. Det optimala antalet knutpunkter kan väljas genom att minimera AIC och/eller SC. Vidare så kan antalet knutpunkter även utvärderas utifrån att beräkna en förlustfunktion baserat på en utvärderingsperiod. Ett förslag på förlustfunktion är förklaringsgraden, pseudo  $R^2$ .
- iii. Undersöka eventuella för och/eller nackdelar med alternativa basfunktioner. Förslag på alternativ basfunktion är ”Natural Cubic splines” vilken har fördelen att den endast ger ett konstant bidrag bortom de två ändknutpunkten.
- iv. Utvärdera alternativa metoder för placeringen av knutpunkterna. Ett alternativ till att placera knutpunkterna enligt angivna percentiler är att basera placeringen utifrån en klusterindelning med funktionen *K-means*.

## Referenser

Altman, E. I., 1968, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy" *Journal of Finance*, Vol. XXIII No. 4, pp. 589-611.

Akaike, H., 1974, "A new look at the statistical model identification". *IEEE Transactions on Automatic Control* 19 (6): 716–723.

Avesani R. G. & Liu K., 2006, "Review and Implementation of Credit Risk Models of the Financial Sector Assessment Program" IMF Working Paper, WP/06/134.

Cleveland W. S., Devlin S. J & Grosse E., 1988, "Regression by Local Fitting – Methods, Properties, and Computational Algorithms" *Journal of Econometrics* 37, 87-114.

Cleveland W. S., 1991, "Computational methods for local regression" *Statistics and Computing* 1, 47-62.

CSFP, 1997, "Credit Risk+: A Credit Risk Management Framework," Credit Suisse Firs Boston.

De Boor C., 2001, "A Practical Guide to Splines", Springer-Verlag, New York.

De Servigny A. och Renault O., 2004, "Measuring and Managing Credit Risk", McGraw-Hill, New York.

Smith P. L., 1979, "Splines As a Useful and Convenient Statistical Tool" *The American Statistician*, Vol. 33, No. 2. (Maj, 1979), pp. 57-62.

Shumway T., 2001, "Forecasting bankruptcy more accurately: a simple hazard model" *Journal of Business*, Vol 74, No 1, pp. 101-124.

Schwarz, G., 1978, "Estimating the dimension of a model". *Annals of Statistics* 6(2):461-464.

Jacobson T., Lindé J. & Roszbach K., 2005, "Credit Risk Versus Capital Requirements under Basel II: Are SME Loans and Retail Credit Really Different?" *Journal of Financial Services Research*, 28:1/2/3, 43-75.

McCullagh, P. & Nelder, J. A., 1989. *Generalized Linear Models*, Vol. 37 of *Monographs on Statistics and Applied Probability*, 2 edn, Chapman and Hall, London.

McFadden, D., 1974, "The Measure of Urban Travel Demand" *Journal of Public Economics*, 3, pp. 303-328.

Muller, M. & Härdle, W., 2002, "Exploring Credit Data" tillgänglig från: <http://edoc.hu-berlin.de/series/sfb-373-papers/2002-79/PDF/79.pdf>.

Ohlsson, E., 2000, "Log-linjära modeller och Logistisk regression", *Kompendium*, Stockholms universitet.

Zavgren, C. V., 1985, "Assesing the Vulnurability to Failure of American Industrial Firms: A Logistic Analysis" *Journal of Business Finance & Accounting*, 12(1).

Wahba, G., 1990, "Spline Models for Observational Data" Philadelphia: Society for Industrial and Applied Mathematics.

Wilson T., 1997, *Portfolio credit risk (I)*, Risk, pp. Xx-xx+5.

## Appendix A

Variabel	Total	No defaults	Defaults
<b>EBITDA/TA</b>	-0,1316	-0,0941	-4,9679
<b>TL/TA</b>	3,9803	2,6936	169,8050
<b>I/TS</b>	3,8756	3,7257	23,9826
<b>LA/TL</b>	1,1942	1,1984	0,6391
<b>IP/EBITDA</b>	-1.3512E10	-1,36E+10	0,2933
<b>UTDE</b>	0,1523	0,1539	0,0089
<b>ANM1</b>	0,0054	0,0035	0,1665
<b>ANM2</b>	0,0445	0,0399	0,4452
<b>TTBS</b>	0,0163	0,0120	0,3890

Tabell A.1: Medelvärde av samtliga förklarande variabler som ingår i den totala panelen 1990Q1 till 2002Q4.

Variabel	<i>1990Q1-1992Q4 &amp; 1995Q1-2002Q4</i>	<i>1990Q1- 2001Q4</i>	<i>1994Q1-1994Q4</i>	<i>2002Q1- 2002Q4</i>
<b>EBITDA/TA</b>	-0,1120	-0,1483	0,3091	0,0667
<b>TL/TA</b>	3,8308	4,2501	5,5034	0,8565
<b>I/TS</b>	3,6783	4,1844	6,1890	0,2780
<b>LA/TL</b>	1,1879	1,1723	1,2961	1,3815
<b>IP/EBITDA</b>	-1,47E+10	-1,15E+10	0,4385	2,84E+11
<b>UTDE</b>	0,1432	0,1480	0,2607	0,2063
<b>ANM1</b>	0,0051	0,0054	0,0072	0,0046
<b>ANM2</b>	0,0436	0,0434	0,0534	0,0572
<b>TTBS</b>	0,0154	0,0174	0,0237	0,0025

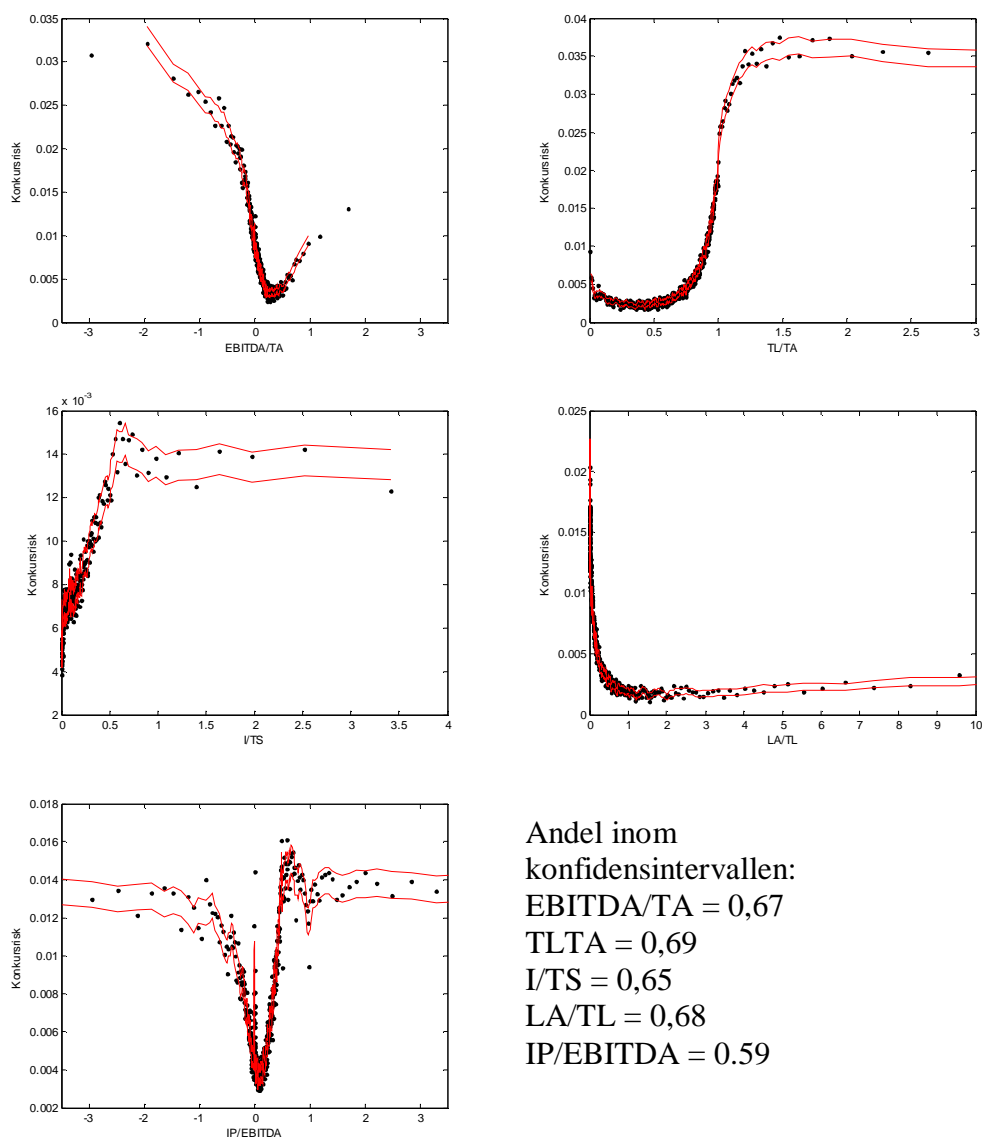
Tabell A.2: Medelvärde av samtliga förklarande variabler som ligger till grund för modellanpassning och utvärdering.

Variabel	Total	No defaults	Defaults
<b>EBITDA/TA</b>	0,1093	0,1099	0,0242
<b>TL/TA</b>	0,7020	0,7002	0,9296
<b>I/TS</b>	0,0088	0,0087	0,0278
<b>LA/TL</b>	0,1520	0,1543	0,0178
<b>IP/EBITDA</b>	0,0836	0,0833	0,1798

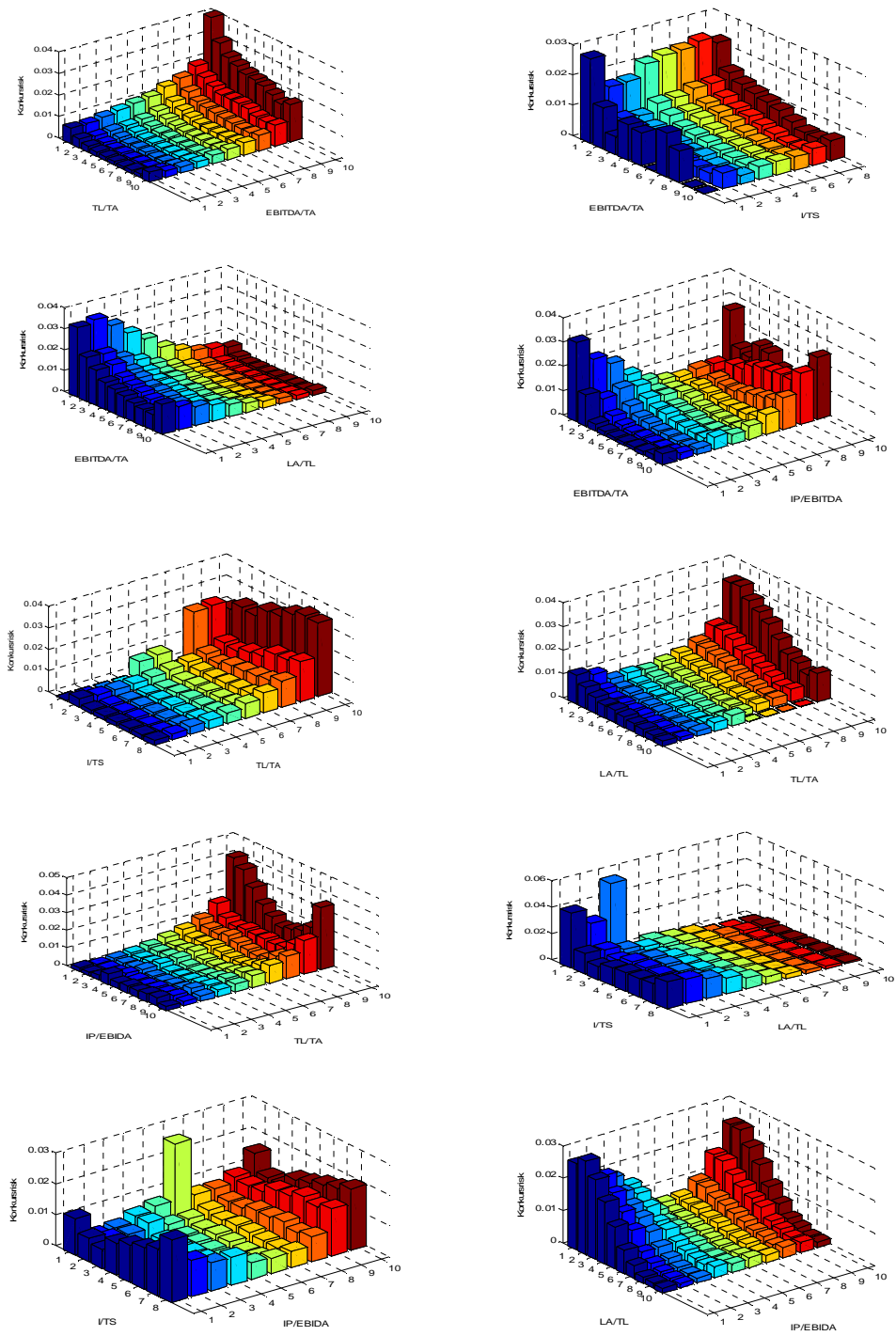
Tabell A.3: Medianen för samtliga finansiella nyckeltal som ingår i den totala panelen 1990Q1 till 2002Q4.

Variabel	1990Q1-1992Q4 & 1995Q1-2002Q4	1990Q1- 2001Q4	1994Q1- 1994Q4	2002Q1-2002Q4
<b>EBITDA/TA</b>	0,1089	0,1105	0,1152	0,0990
<b>TL/TA</b>	0,6989	0,7086	0,7405	0,6206
<b>I/TS</b>	0,0086	0,0096	0,0114	0,0012
<b>LA/TL</b>	0,1551	0,1460	0,1191	0,2407
<b>IP/EBITDA</b>	0,0813	0,0862	0,1181	0,0612

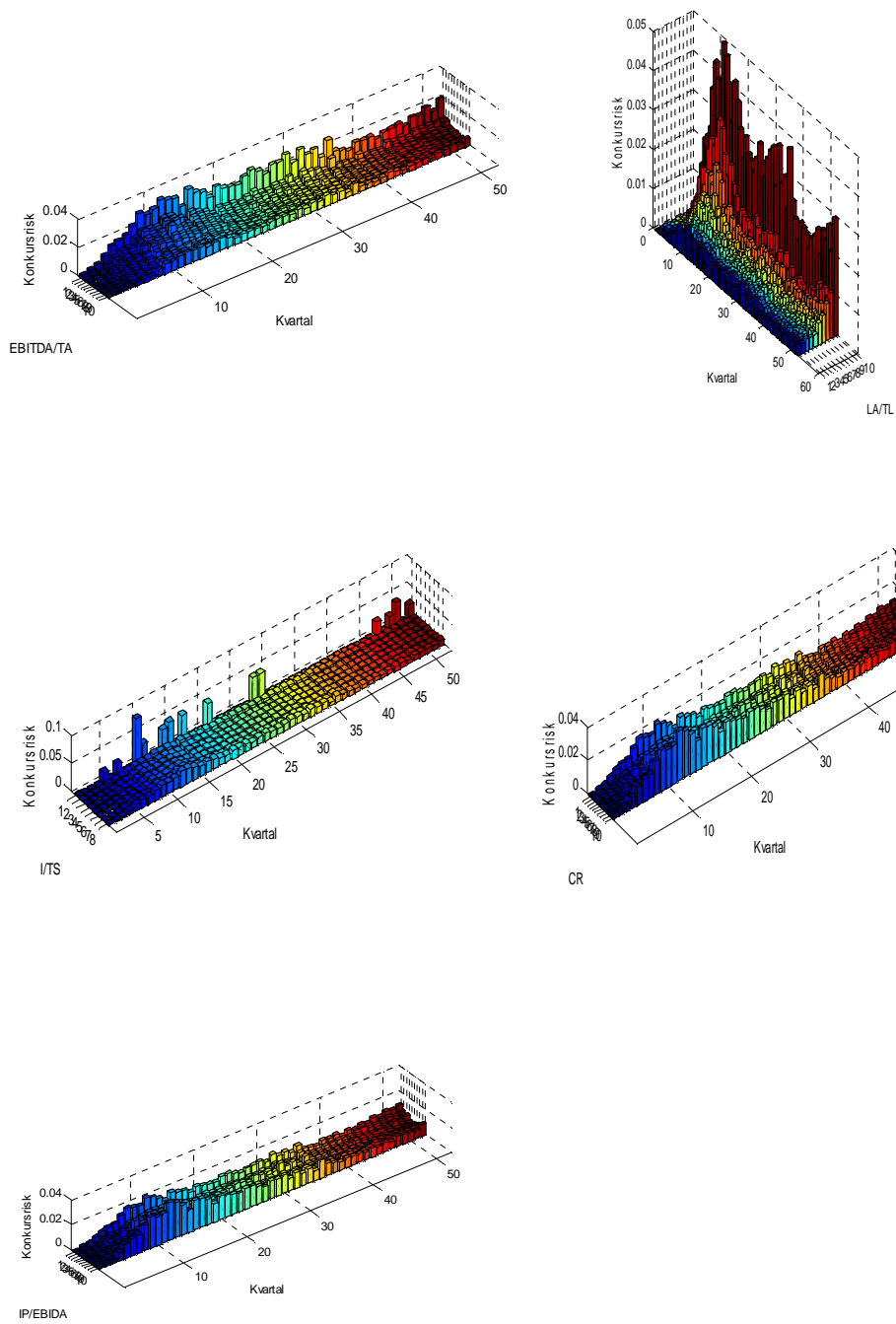
Tabell A.4: Medianen för samtliga finansiella nyckeltal som ligger till grund för modellanpassning och utvärdering.



Figur A.1: Observerad konkursrisk samt 95 procentigt konfidensintervall per bokslutsvariabel (konfidensintervallet beräknat enligt Ekvation 9.1).



Figur A.2: Konkursrisk per decil; finansiella nyckeltal mot finansiella nyckeltal.



Figur A.3: Konkursrisken per decil; förändring under den observerade perioden 1990Q1-2002Q4.

## Appendix B

Parameter	Logistisk			Logistisk - trunkerad panel		
	Estimat	s.e.	Pr> $\chi^2$	Estimat	s.e.	Pr> $\chi^2$
Intercept	-5,0884	0,00582	<,0001	-5,8430	0,0106	<,0001
EBITDA/TA	0,000025	0,000011	0,0298	-0,7705	0,0113	<,0001
TL/TA	2,707E-6	4,925E-7	<,0001	1,0911	0,00807	<,0001
I/TS	1,748E-7	1,992E-7	0,3803	0,2862	0,00884	<,0001
LA/TL	-0,0162	0,000851	<,0001	-0,3179	0,00814	<,0001
IP/EBITDA	-6,41E-15	5,61E-14	0,9090	0,0947	0,00459	<,0001
Säs-indikator	-0,3245	0,00916	<,0001	-0,3008	0,00913	<,0001
RÄNT	0,0339	0,000543	<,0001	0,0273	0,000563	<,0001
<b>Pseudo R<sup>2</sup></b>		<b>0,005</b>			<b>0,057</b>	

Tabell B,1: Parameterskattningar för den logistiska kreditriskmodellen exklusive indikatorvariabler, skattade med avseende på perioden 1990Q1-2002Q4,

Parameter	Logistisk			Logistisk - trunkerad panel		
	Estimat	s,e,	Pr> $\chi^2$	Estimat	s,e,	Pr> $\chi^2$
Intercept	-6,165000	0,008910	<,0001	-6,514000	0,013100	<,0001
EBITDA/TA	0,000003	0,000013	0,7993	-0,595100	0,013000	<,0001
TL/TA	0,000003	0,000001	0,0007	0,594000	0,009660	<,0001
I/TS	0,000000	0,000000	0,3635	0,213600	0,011000	<,0001
LA/TL	-0,000050	0,000021	0,0241	-0,295400	0,007360	<,0001
IP/EBITDA	0,000000	0,000000	0,8944	0,080700	0,005350	<,0001
UTDE	-1,776300	0,034800	<,0001	-1,374500	0,034600	<,0001
ANM1	1,825200	0,015800	<,0001	1,661200	0,015900	<,0001
ANM2	2,803100	0,010200	<,0001	2,561500	0,010500	<,0001
TTBS	3,859000	0,009900	<,0001	3,944800	0,010100	<,0001
Säs-indikator	-0,425200	0,011100	<,0001	-0,423800	0,011200	<,0001
RÄNT	0,044300	0,000642	<,0001	0,041800	0,000654	<,0001
<b>Pseudo R<sup>2</sup></b>		<b>0,327</b>			<b>0,346</b>	

Tabell B,2: Parameterskattningar för den logistiska kreditriskmodellen skattade med avseende på perioden 1990Q1-1993Q4 & 1995Q1-2002Q4,



Parameter	Logistisk			Logistisk - trunkerad panel		
	Estimat	s,e,	Pr> $\chi^2$	Estimat	s,e,	Pr> $\chi^2$
<b>Intercept</b>	-6,163400	0,008940	<,0001	-6,502200	0,013000	<,0001
<b>EBITDA/TA</b>	0,000010	0,000012	0,3687	-0,554200	0,012800	<,0001
<b>TL/TA</b>	0,000003	0,000001	0,0001	0,576500	0,009310	<,0001
<b>I/TS</b>	0,000000	0,000000	0,3672	0,201200	0,010500	<,0001
<b>LA/TL</b>	-0,000050	0,000019	0,0081	-0,319200	0,007520	<,0001
<b>IP/EBITDA</b>	0,000000	0,000000	0,9839	0,081500	0,005110	<,0001
<b>UTDE</b>	-1,859100	0,035300	<,0001	-1,470200	0,035100	<,0001
<b>ANM1</b>	1,785400	0,015600	<,0001	1,631600	0,015600	<,0001
<b>ANM2</b>	2,814500	0,010000	<,0001	2,572800	0,010200	<,0001
<b>TTBS</b>	3,794600	0,009520	<,0001	3,888200	0,009710	<,0001
<b>Säs-indikator</b>	-0,434900	0,010900	<,0001	-0,433700	0,011000	<,0001
<b>RÄNT</b>	0,045100	0,000643	<,0001	0,042700	0,000656	<,0001
<b>Pseudo R<sup>2</sup></b>		<b>0,334</b>			<b>0,352</b>	

Tabell B,3: Parameterskattningar för den logistiska kreditriskmodellen skattade med avseende på perioden 1990Q1-2001Q4,

Parameter	1990Q1-1992Q4 & 1995Q1-2002Q4			1990Q1-2001Q4		
	Estimat	s,e,	Pr> $\chi^2$	Estimat	s,e,	Pr> $\chi^2$
<b>Intercept</b>	-6,226	0,032	<,0001	-6,300	0,032	<,0001
<b>EBITDA/TA</b>	-0,711	0,022	<,0001	-0,649	0,022	<,0001
<b>TL/TA</b>	-0,347	0,067	<,0001	-0,245	0,065	0,000
<b>I/TS</b>	-1,999	4,371	0,648	1,707	3,851	0,658
<b>LA/TL</b>	-5,721	0,189	<,0001	-5,996	0,197	<,0001
<b>IP/EBITDA</b>	-0,111	0,009	<,0001	-0,117	0,008	<,0001
<b>UTDE</b>	-1,038	0,035	<,0001	-1,133	0,035	<,0001
<b>ANM1</b>	1,641	0,016	<,0001	1,612	0,016	<,0001
<b>ANM2</b>	2,416	0,011	<,0001	2,427	0,010	<,0001
<b>TTBS</b>	4,018	0,010	<,0001	3,973	0,010	<,0001
<b>Säs-indikator</b>	-0,414	0,011	<,0001	-0,425	0,011	<,0001
<b>RÄNT</b>	0,038	0,001	<,0001	0,039	0,001	<,0001
<b>TP<sub>11</sub>: EBITDA/TA</b>	-18,695	1,108	<,0001	-18,715	1,069	<,0001
<b>TP<sub>12</sub>: EBITDA/TA</b>	40,411	2,392	<,0001	40,429	2,307	<,0001
<b>TP<sub>13</sub>: EBITDA/TA</b>	-21,509	1,530	<,0001	-21,770	1,475	<,0001
<b>TP<sub>21</sub>: TL/TA</b>	7,804	0,340	<,0001	7,972	0,337	<,0001
<b>TP<sub>22</sub>: TL/TA</b>	-11,914	0,692	<,0001	-12,959	0,689	<,0001
<b>TP<sub>23</sub>: TL/TA</b>	2,750	0,389	<,0001	3,754	0,389	<,0001
<b>TP<sub>31</sub>: I/TS</b>	379,600	273,900	0,166	135,200	217,800	0,535
<b>TP<sub>32</sub>: I/TS</b>	-402,500	275,800	0,144	-156,700	219,600	0,476
<b>TP<sub>33</sub>: I/TS</b>	22,841	2,262	<,0001	21,468	2,137	<,0001
<b>TP<sub>41</sub>: LA/TL</b>	16,713	0,916	<,0001	18,717	1,008	<,0001
<b>TP<sub>42</sub>: LA/TL</b>	-15,632	0,993	<,0001	-17,578	1,090	<,0001
<b>TP<sub>43</sub>: LA/TL</b>	-1,074	0,097	<,0001	-1,131	0,103	<,0001
<b>TP<sub>51</sub>: CR</b>	-1,367	0,495	0,006	-0,571	0,442	0,196
<b>TP<sub>52</sub>: CR</b>	3,307	0,736	<,0001	2,159	0,665	0,001
<b>TP<sub>53</sub>: CR</b>	-2,075	0,251	<,0001	-1,722	0,233	<,0001
<b>Pseudo R<sup>2</sup></b>		<b>0,364</b>			<b>0,373</b>	

Tabell B,4: Parametrar för logistiska spline modellen skattade med avseende på perioden 1990Q1-1992Q4 & 1995Q1-2002Q4 och 1990Q1-2001Q4, Parameter  $K_{nj}$  avser de trunkerade potensfunktionerna med avseende på knutpunkter  $j = 1,2,3$  för variabel  $r = 1,2,\dots,5$ , Knutpunkternas placering redovisas i Tabell 12,2 och trunkerade potensfunktionerna formuleras enligt Ekvation 10,1,

Total					
Quartile	EBITDATA	TLTA	ITS	LATL	CR
25% Q1	0,0202	0,4685	0	0,0178	-0,0173
50% Median	0,1093	0,702	0,0088	0,152	0,0836
75% Q3	0,2199	0,8779	0,1083	0,6267	0,3152

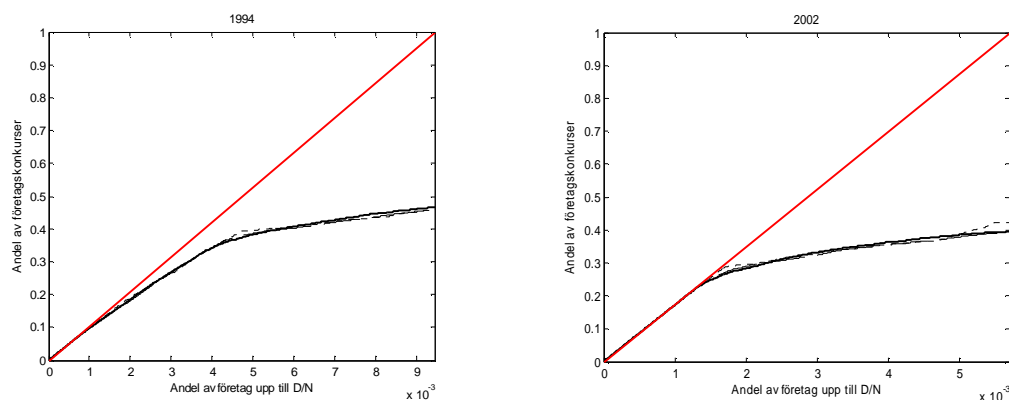
  

Delperiod 1					
Quartile	EBITDATA	TLTA	ITS	LATL	CR
25% Q1	0,0202	0,4655	0	0,0183	-0,0172
50% Median	0,1089	0,6989	0,0086	0,1551	0,0813
75% Q3	0,2193	0,8755	0,1075	0,6341	0,3105

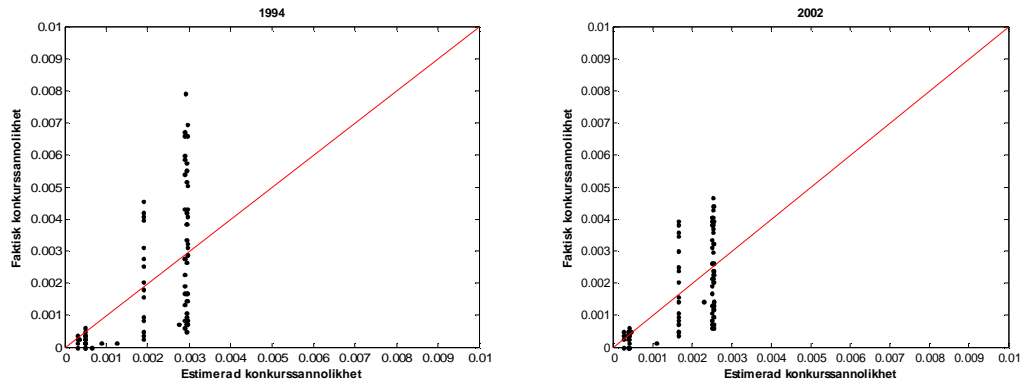
Delperiod 2					
Quartile	EBITDATA	TLTA	ITS	LATL	CR
25% Q1	0,021	0,4777	0	0,0171	-0,0185
50% Median	0,1105	0,7086	0,0096	0,146	0,0862
75% Q3	0,2215	0,8816	0,11	0,607	0,3203

Tabell B,2: Placering av knutpunkter med avseende på de modeller som anpassats på hela perioden samt de två delperioderna, Placeringen har gjorts efter den 25e, 50e och 75e percentilen för vardera av de fem nyckeltalen

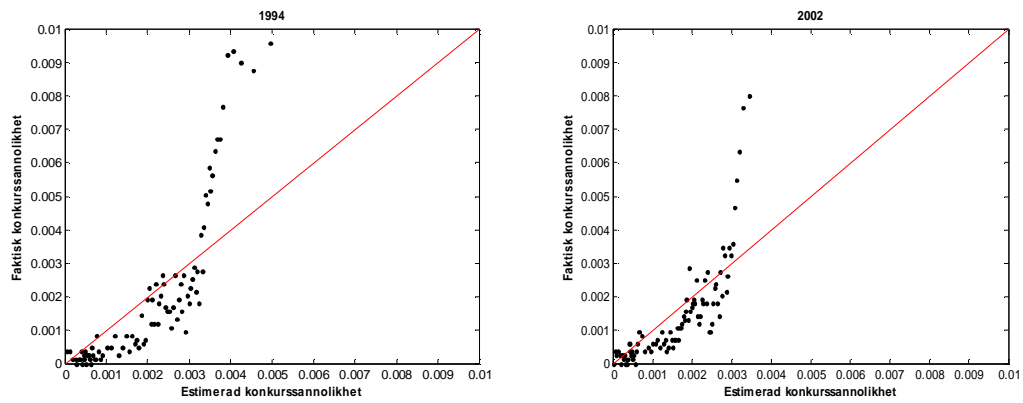


Figur B,1: CAP-kurvor för de två delperioderna 1994 samt 2002 med avseende på intervallet  $(0, D/N]$  den raka linjen illustrerar den ”perfekta” modellen

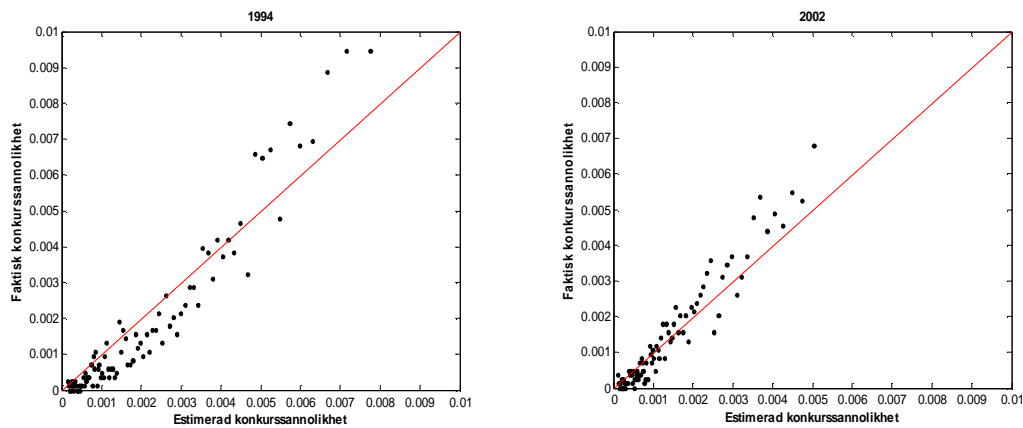
### Multipel logistisk



### Dito trunkerad panel



### Spline modellen



Figur B,2: Faktisk mot estimerad konkurssannolikhet med avseende på de tre konkursriskmodellerna och de två utvärderingsperioderna, 1994 och 2002,

## Appendix C

### Kreditriskmodellen CR<sup>+</sup>

CR<sup>+</sup> är en modell som används vid beräkningen av kreditförlustfördelningen för en utlåningsportfölj, se Avesani & Kexue (2006), Beakta en portfölj bestående av  $N_A$  låntagare, indexerade enligt  $A = 1, \dots, N_A$ , Låt låntagare  $A$  stå för en exponering  $E_A$ , Förlusten för låntagare  $A$  representeras då av den stokastiska variabeln  $L_A = D_A E_A$ , där  $D_A$  är bernoullifördelad där sannolikheten för konkurs ges av  $P(D_A = 1) = p_A$  och sannolikheten för att ej gå i konkurs ges av  $P(D_A = 0) = 1 - p_A$ , Sannolikhetsfördelningen för  $L_A$  ges då av  $P(L_A = E_A) = p_A$  och  $P(L_A = 0) = 1 - p_A$ , Den totala förlusten för en given portfölj kan då formuleras enligt

$$L = \sum_{A=1}^{N_A} L_A = \sum_{A=1}^{N_A} D_A \times E_A ,$$

Målet är att beräkna sannolikhetsfördelningen för  $L$  vilken avser den totala förlusten för en utlåningsportfölj,

#### **Konkurshändelser**

Om vi antar att varje exponering, som sträcker sig över ett givet kvartal, har en känd konkurssannolikhet,

$$p_A = \text{Konkursrisken per kvartal för låntagare } A,$$

Låt den sannolikhetsgenererande funktionen vara definierad enligt

$$P(z) = \sum_{n=0}^{\infty} P(n)z^n,$$

För en individuell låntagare så är den sannolikhetsgenererande funktionen given enligt

$$P(z) = 1 - p_A + p_A z = 1 + p_A(z - 1),$$

Som en konsekvens av den individuella konkursrisken mellan låntagarna antas vara oberoende så blir den sammanslagna sannolikhetsgenererande funktionen en produkt av de individuella:

$$P(z) = \prod_A P_A(z) = \prod_A (1 + p_a(z - 1)),$$

vilket kan skrivas enligt

$$\log P(z) = \sum_A \log(1 + p_a(z - 1)), \quad (\text{C},1)$$

Om vi antar att låntagarnas konkurssannolikhet är uniformt små så kan logaritmen ersättas med uttrycket

$$\log(1 + p_a(z - 1)) \approx p_a(z - 1),$$

Ekvation C,1 kan då uttryckas enligt

$$P(z) = \exp\left(\sum_A p_a(z - 1)\right) = \exp(\mathbf{m}(z - 1)),$$

där  $m = \sum_A p_A$  är det förväntade antalet förluster för en given portfölj, över ett kvartal, För att identifiera fördelningen som är kopplad till den sannolikhetsgenererande funktionen så kan  $P(z)$  Taylorutvecklas:

$$P(z) = \exp(m(z-1)) = \exp(-m)\exp(mz) = \sum_{n=0}^{\infty} \frac{\exp(-m)m^n}{n!} z^n, \quad (C,2)$$

Om konkurssannolikheterna är små, så fås från ekvation 5, att sannolikheten för  $n$  konkurser ges av

$$P(n) = \frac{\exp(-m)m^n}{n!},$$

vilket är en poissonfördelning med väntevärde  $m$  och varians  $m$ ,

### Notation

Låt liksom ovan  $E_A$  vara exponeringen och  $L_A$  den förväntade förlusten givet att låntagare  $A$  går i konkurs, För att beräkningarna av den totala förlustfördelningen skall förenklas så väljs en enhet  $T$  som basenhet, För varje låntagare  $A$  fås då enheterna  $d_A$  och  $n_A$  enligt

$$E_A = T \times n_A \text{ och } L_A = T \times d_A,$$

där  $d_A$  och  $n_A$  är exponeringen och den förväntade förlusten uttryckt i basenheten  $T$ ,

### Normaliserade exponeringar

Genom att normalisera och avrunda så begränsas de möjliga utfallen för  $L$  samt reducerar tiden som krävs för att beräkna sannolikhetsfördelningen för  $L$ , Då

normaliseringsfaktorn är liten i förhållande till den totala exponeringen så blir avrundningsfelen små, Vi kan formulera  $n_A$  enligt

$$n_A = \text{ceil}(E_A / T),$$

där funktionen  $\text{ceil}(\bullet)$  avrundar den normaliserade exponeringen till närmaste övre heltal, Portföljen kan delas upp i  $J$  exponeringsband, Då de individuella exponeringarna har normaliserats och avrundats så kan exponeringsbanden formuleras enligt: det totala antalet exponeringsband  $J$  är givet av den största exponeringen  $J = \max\{n_n\}_{n=1}^N$ , Om  $j$  noterar indexeringen för exponeringsbanden så gäller det att  $j$  är givet enligt  $n_j = j$ , Fördelningen av låntagarna mellan exponeringsbanden är gjord så att varje låntagare  $A$  tillhör ett exponeringsband  $j$  om  $n_a = n_j = j$ , Det förväntade antalet konkurser i exponeringsband  $j$ ,  $m_j$ , ges av

$$m_j = \sum_{n.n_A=j} p_A = \frac{d_j}{n_j},$$

Där den totala mängden konkurser för en given portfölj,  $m$ , ges enligt

$$m = \sum_{j=1}^J m_j = \sum_{j=1}^J \frac{d_j}{n_j},$$

### Förlustfördelningen

Låt  $C_n$  vara sannolikheten att förlora  $n \times B$  eller  $n$  enheter av en utlåningsportfölj

$$P(\text{förlust} : n \times B) = \frac{1}{n!} \left. \frac{d^n G(z)}{dz^n} \right|_{z=0},$$

där  $G(z)$  definieras enligt Ekvation C,2



$$\begin{aligned} \left. \frac{1}{n!} \frac{d^n G(z)}{dz^n} \right|_{z=0} &= \frac{1}{n!} \frac{d^{n-1}}{dz^{n-1}} \left( G(z) \frac{d}{dz} \sum_{j=1}^J m_j z^{n_j} \right) \Big|_{z=0} \\ &= \frac{1}{n!} \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{d^{n-k-1}}{dz^{n-k-1}} G(z) \frac{d^{k+1}}{dz^{k+1}} \left( \sum_{j=1}^J m_j z^{n_j} \right) \Big|_{z=0}, \end{aligned}$$

Då gäller att

$$\left. \frac{d^{k+1}}{dz^{k+1}} \left( \sum_{j=1}^J m_j z^{n_j} \right) \right|_{z=0} = \begin{cases} m_j (k-1)! & \text{om } k = n_j, \\ 0 & \text{annars} \end{cases},$$

och av definition

$$\left. \frac{d^{n-k-1}}{dz^{n-k-1}} G(z) \right|_{z=0} = (n-k-1)! = C_{n-k-1},$$

följer det att

$$C_n = \sum_{k \leq n-1} \frac{1}{n!} \binom{n-1}{k} (k+1)! (n-k-1)! m_j C_{n-k-j} = \sum_{j: n_j \leq n} \frac{m_j n_j}{n} C_{n-n_j},$$

genom relationen  $d_j = n_j \times m_j$ , vilket ger

$$C_n = \sum_{k \leq n-1} \frac{d_j}{n} C_{n-n_j}, \quad (\text{C,3})$$

där

$$C_0 = G(0) = F(P(0)) = \exp \left( - \sum_{j=1}^J \frac{d_j}{n_j} \right) = \exp(-\mathbf{m}),$$

Ekvation C,3 illustrerar den rekursionsformel som används för beräkning av den totala kreditförlustfördelningen för  $L$ ,

## Appendix D

### McFadden's $R^2$

McFadden's förklaringsgrad formuleras enligt

$$R^2 = 1 - \frac{\log L_{(1)}}{\log L_{(0)}},$$

där  $\log L_{(0)}$  och  $\log L_{(1)}$  avser den logaritmerade likelihoodfunktionen för grundmodellen respektive hypotesmodellen, se McFadden (1974), Den logaritmerade likelihoodfunktionen definieras i kapitel 8.

### ML-skattningar

ML-skattningarna av  $b$  fås enligt följande sats vilken ger "likelihood-ekvationerna", se Ohlsson (2000),

**Sats** Antag att design-matrisen  $X$  har maximal rang  $r + 1$ , Antag vidare att det går att finna en lösning  $\hat{b}$  till systemet

$$X^T n = X^T m, \tag{D,1}$$

där väntevärdet  $m$  skall uppfylla

$$m_i = (1 - y_i) p(X_i) = (1 - y_i) \frac{1}{1 + \exp(-X_i b)}, \quad i = 1, 2, \dots, k,$$

Då är  $\hat{b}$  en unik ML-skattning,

Newton-Raphsons metod utgår från att finna en lösning till ekvationssystemet (B,1), vilken kan skrivas  $X^T n = X^T m = 0$ , Iterationsstegen i Newton-Raphsons metod blir:

$$\mathbf{b}^{[t+1]} = \mathbf{b}^{[t]} + [\mathbf{I}(\mathbf{b}^{[t]})]^{-1} (X^T n - X^T \mathbf{m}^{[t]})$$

där  $\mathbf{I}(\mathbf{b}^{[t]}) = X^T D X$  med  $D$  i form av en diagonalmatris med elementen  $d_{ii} = (1 - y_i)p(X_i)(1 - p(X_i))$  för  $i = 1, 2, \dots, k$ , Som startvärde kan  $\mathbf{b}_j = 0$  väljas,