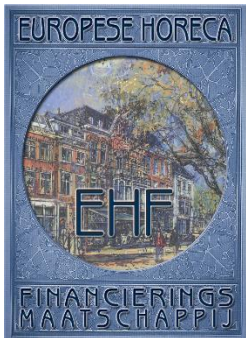


De invloed van financiële en niet-financiële variabelen op default van financieringen

Een onderzoek bij de Europese Horeca Financieringsmaatschappij



's-Gravenhage, september 2018

Inhoud

i.	Samenvatting	4
ii.	Eerdere onderzoeken bij EHF.....	4
1.	Achtergrond en aanleiding onderzoek	5
1.1.	Kredietrisico en beoordeling.....	5
1.2.	Aanleiding onderzoeksthema	5
2.	Onderzoeksopzet	6
2.1.	Doelstelling en deelvragen.....	6
2.2.	De volgende stappen worden in het onderzoek uitgevoerd:	6
2.3.	Afbakening en relevantie	6
3.	Data en methodologie	7
3.1.	Introductie	7
3.2.	Dataset	7
3.3.	Methodologie.....	7
3.4.	Model	7
4.	Resultaten van het onderzoek.....	11
4.1.	Introductie	11
4.2.	Resultaten van het onderzoek model I	11
4.3.	Resultaten van het onderzoek model II	14
4.4.	Vergelijking model I en II.....	15
4.5.	Aanbevelingen voor het ontwikkelen van een creditscore model	16
5.	Conclusies	18
6.	Discussie en toepasbaarheid	19
	Appendix A-1. Variabelen in het onderzoek	20
1.	Afhankelijke variabele.....	20
2.	Onafhankelijke variabelen model I	20
3.	Onafhankelijke variabelen model II	21
4.	Onafhankelijke variabelen model III	21
	Appendix A-2. Resultaten van het onderzoek	22
1.	Model I	22
1.1.	Testen van de assumpties.....	22
1.1.1	Lineariteit	22
1.1.2.	Onafhankelijke errors.....	22
1.1.3.	Multicollineariteit.....	22

1.2. 'Nulmodel I'	24
1.3. Resultaten model I	26
1.4. Outliers.....	27
2. Model II	28
2.1. Testen van de assumpties.....	28
2.1.1. Lineariteit	28
2.1.2. Onafhankelijke errors.....	28
2.1.3. Multicollinaeriteit.....	28
2.2. 'Nulmodel II'	30
2.3. Resulten model II	32
2.4. Outliers.....	33
3. Model III	33
3.1. Testen van de assumpties.....	33
3.2. 'Nulmodel III'	33
3.3. Resultaten model III	35

i. Samenvatting

Kredietrisico is een van de grootste risico's van banken en financiële instellingen. Het behoeft dan ook geen toelichting dat het goed inschatten van deze risico's van zeer groot belang is. Het is dan ook niet verwonderlijk dat hier in de afgelopen tientallen jaren veel onderzoek naar is gedaan. Zogenaamde 'credit score' modellen proberen in te schatten wat het risicoprofiel van een kredietnemer is. In de loop der jaren zijn er diverse soorten credit score modellen ontwikkeld die succesvol waren in het voorspellen van default (het niet kunnen terugbetalen van een krediet). Deze credit score modellen zijn over het algemeen goed bruikbaar in het geval er voldoende en betrouwbare financiële gegevens beschikbaar zijn van de kredietnemer. Echter, bij het ontbreken hiervan wordt dit een stuk lastiger. Vooral bij kleine ondernemingen speelt dit een rol. Zij hebben vaak minder uitgebreide financiële gegevens ter beschikking en omdat deze kleine ondernemingen ook geen controleplicht hebben, zijn de financiële gegevens minder betrouwbaar. Ook bij starters speelt dit probleem. Zij hebben nog geen financiële 'trackrecord' opgebouwd en de prognose die wordt afgegeven moet nog maar gehaald zien te worden. Een credit score model loslaten op een prognose brengt dan ook veel onzekerheden met zich mee. De vraag rijst dan ook of andere aanwezige variabelen iets kunnen zeggen over het voorspellen van default. Een onderzoek bij een financier van horecagelegenheden, de Europese Horeca Financieringsmaatschappij, wordt uitgevoerd om te kijken of niet-financiële variabelen default zouden kunnen voorspellen.

ii. Eerdere onderzoeken bij EHF

Bij de EHF zijn eerdere onderzoeken op gebied van default bij horecaleningen uitgevoerd; in 2008 en 2012. Dit onderzoek verschilt in vergelijking met de eerdere onderzoeken op een aantal vlakken. Ten eerste is het onderzoek uitgevoerd op een grotere dataset aan leningen. De gebruikte dataset bevat 1200 leningen, in vergelijking met respectievelijk 600 en 800 leningen in de eerdere onderzoeken. Daarnaast is een langere periode gebruikt. De dataset bevat leningen verstrekt tussen 2005 en 2016. Het grootste verschil zit hem echter in de gebruikte variabelen. Voor dit onderzoek is gebruikt gemaakt van een groter aantal variabelen, zowel financiële als niet-financiële variabelen.

1. Achtergrond en aanleiding onderzoek

1.1. Kredietrisico en beoordeling

De kredietcrisis van de afgelopen jaren heeft niet alleen de financiële wereld op zijn kop gezet, maar ook de beoordeling van kredietrisico door de financiële instellingen. In de afgelopen jaren is de beoordeling van kredietrisico bij vele ondernemingen drastisch veranderd. Door de financiële onrust werden ondernemingen genoodzaakt de praktische kant van het kredietrisico anders in te richten, omdat de traditionele fundamentele bakens voor risico en rendement door de veranderingen in de markt niet meer stand hielden. Financiële instellingen werden gedwongen om kritisch te kijken naar hoe zij risico beoordelen. Banken analyseren financiële gegevens en wat kwalitatieve factoren als onafhankelijke variabelen in statistische credit risk assessment modellen. Maar om een goed krediet- en prijsbeleid op te stellen, is het heel belangrijk om de kwaliteit van de lening portefeuille in de toekomst te voorspellen. Dit probleem kan worden aangepakt door de statistische methodes te vinden die het deel van slecht-presterende leningen in de lening portefeuille relateren aan een set van verklarende variabelen.

1.2. Aanleiding onderzoeksthema

Publicaties genoeg de afgelopen tijd over het belang van MKB-bedrijven in Nederland, die de motor zijn van de Nederlandse economie en de moeizame financiering hiervan. Grootbanken willen niet teveel tijd spenderen aan kleine kredieten voor kleine bedrijven. Immers, er wordt op deze kleine kredieten niet veel verdiend en daardoor mag het kredietbeoordelingsproces niet teveel kosten, anders is het financieren hiervan niet rendabel. Toch is een goede beoordeling van deze financieringen genoodzaakt. Een bijkomend probleem bij de beoordeling van deze kredieten is de beschikbaarheid en betrouwbaarheid van financiële gegevens. Juist bij deze groep kleinere ondernemingen speelt dit een rol. Daarnaast is bij startende ondernemingen nog geen financiële 'trackrecord' opgebouwd, wat de beoordeling op financiële gegevens bemoeilijkt. Geld lenen aan kleine ondernemingen is meer risicovol dan geld lenen aan grote ondernemingen. Dit komt doordat het voor deze ondernemingen moeilijker is om hun waarde te bepalen. Onderzoek toont aan dat de traditionele creditscore modellen die worden gebruikt om de kredietwaardigheid van kleinere ondernemingen te toetsen minder betrouwbaar zijn omdat het verzamelen van deze financiële gegevens lastiger is. Het beoordelen op basis van variabelen naast de traditionele financiële gegevens is dan ook een logische stap, maar hier is echter nog weinig onderzoek naar gedaan, met name op het terrein van kleinere ondernemingen.

2. Onderzoeksopzet

2.1. Doelstelling en deelvragen

De doelstellingen van het onderzoek luiden als volgt:

1. Trachten inzicht te verkrijgen in de invloed van zowel financiële als niet-financiële variabelen op de kans op default bij horecafinancieringen.
2. Met de mogelijke uitkomsten van bovengenoemd onderzoek trachten aanbevelingen te doen voor het ontwikkelen van een creditscore model voor horecafinancieringen.

Deze doelstellingen worden getracht verwezenlijkt te worden door de onderstaande deelvragen te beantwoorden:

- Welke variabelen hebben invloed op de kans op default bij horecafinancieringen?
- Welke aanbevelingen zijn er te doen voor het ontwikkelen van een creditscore model voor horecafinancieringen op basis van de verkregen inzichten?

2.2. De volgende stappen worden in het onderzoek uitgevoerd:

1. Data onderzoek uitvoeren op de leningportefeuille van de Europese Horeca Financieringsmaatschappij (EHF).
2. Aanbevelingen doen voor het ontwikkelen van een creditscore model voor horecafinancieringen naar aanleiding van resultaten van het data onderzoek.

2.3. Afbakening en relevantie

Met de resultaten van het onderzoek kunnen enkel aanbevelingen worden gedaan voor het ontwikkelen van een creditscoremodel voor financieringen van kleinere ondernemingen, waar de beschikbaarheid en betrouwbaarheid van financiële gegevens veelal ontbreken. De praktische relevantie komt dan ook uit de verkregen inzichten in factoren die default bij financieringen kunnen veroorzaken. Mogelijk verkregen inzichten kunnen gebruikt worden om een creditscore model te ontwikkelen, wat een bijdrage kan leveren aan de beoordeling van financieringen. De wetenschappelijke relevantie van dit onderzoek ligt in een bijdrage aan de literatuur op het gebied van gebruik van niet-financiële informatie in creditscore modellen. Hierbij dient te worden vermeld dat de importantie van niet-financiële informatie kan verschillen tussen industrieën en branches.

3. Data en methodologie

3.1. Introductie

Zoals in paragraaf 2.2. besproken vindt het onderzoek plaats bij de Europese Horeca Financieringsmaatschappij (EHF), een financieringsmaatschappij gericht op de horeca. De markt voor horecafinanciering is van oudsher een lastige markt. Zo rond de millenniumwisseling neemt de bereidheid van reguliere banken om horecaondernemingen, zoals met name bars en cafés, te financieren opvallend snel af. Brouwerijen verstrekken van oudsher ook financiering aan klanten, bijvoorbeeld voor de overname van een exploitatievergunning, voor het uitkopen van een partner of voor het verbouwen van een etablissement. Voorwaarde voor een dergelijke financiering is uiteraard wel dat het betreffende biermerk geschonken wordt. Rond de millenniumwisseling tekent zich echter ook bij brouwerijen een kentering in de bereidheid tot het direct verstrekken van een financiering af. Brouwerijen geven in sterk toenemende mate de voorkeur aan het afgeven van een borgstelling aan financiële instellingen boven het zelfstandig verschaffen van horecafinanciering. Deze kentering komt onder andere voort uit de “hunt for cash” en uit mededingingswetgeving. Vanuit deze ontwikkelingen is de oprichting van een financieringsmaatschappij, die horecaondernemers financiert onder een gedeeltelijke borgstelling door brouwerijen, ontstaan. De gedachte was dat hierdoor een nieuwe organisatie zou ontstaan met specifieke kennis van zowel horeca als financiële dienstverlening. Eind 2004 gaat de EHF van start. In de afgelopen jaren zijn meerdere samenwerkingen aangegaan met gerenommeerde toeleveranciers van de horeca om de markt voor horecafinanciering nog beter te kunnen bedienen.

3.2. Dataset

Er is informatie van verstrekte leningen over de periode 2005-2016 verzameld. Uiteindelijk zijn er ruim 1200 leningen in het onderzoek verwerkt, die voldoende informatie boden om te worden beoordeeld. Van deze leningen zijn een aantal variabelen verzameld, welke worden gebruikt voor het onderzoek.

3.3. Methodologie

De deelvragen zoals besproken in hoofdstuk 2 zullen getracht te worden beantwoord door verschillende modellen te toetsen met een afhankelijke variabele en onafhankelijke leningspecifieke variabelen. Door middel van een logistische regressie analyse zal de invloed van de onafhankelijke variabelen op de afhankelijke variabele getoetst worden.

3.4. Model

De afhankelijke variabele in het model is het wel of niet in default geraken van een financiering. De onafhankelijke variabelen zijn de lening specifieke kenmerken. In de Appendix A-1 worden de variabelen en eventuele categorieën van de variabelen toegelicht. Omdat de afhankelijke variabele wel of geen default een categoriale variabele is met twee uitkomsten (een zogeheten dichotome variabele) kan geen gebruik worden gemaakt van normale regressie-analyse.

We willen namelijk de voorspelde kans op default berekenen (een uitkomst tussen 0 en 1) en bij lineaire regressie zou de voorspelde kans groter dan 1 en kleiner dan 0 kunnen zijn. Logistische regressie analyse is de methode die juist hiervoor gebruikt kan worden. Het logistische model gaat uit van kansverhoudingen, de zogenaamde "odds". In dit onderzoek zijn we op zoek naar de kans op default (p) gedeeld door de kans dat de lening niet in default geraakt (1-p). De "odds" is dan p/(1-p). Een "odds" kan de waarde 0 tot oneindig aannemen. Waarden onder 0 zijn niet mogelijk. Om aan de randvoorwaarde van een (lineaire) regressie te voldoen, die stelt dat de residuen normaal verdeeld moeten zijn, wordt dan ook de natuurlijke logaritme van de odds gebruikt. Deze natuurlijke logaritme van de odds wordt "logit" genoemd en kan waarden tussen min oneindig en oneindig aannemen.

De standaard vorm van logistische regressie ziet er als volgt uit:

$$\text{Logit}(P) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k$$

Als we het logistische model omzetten in een kansmodel, ziet het er als volgt uit:

$$P(Y) = \frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k)}} + \epsilon$$

waar P de kans is dat de afhankelijke variabele de waarde 1 aanneemt, α is de constante, β_k is de coëfficiënt van de onafhankelijke variabele k en X_k is de onafhankelijke variabele k en ϵ de errorterm (foutmarge) is. Met deze vergelijking worden de coëfficiënten $\beta_1 \dots \beta_k$ geschat.

Er kunnen verschillende methoden van logistische regressie worden gebruikt. In dit onderzoek is gekozen voor de "Enter" methode. Deze methode is gebaseerd op de veronderstelling dat van een vooraf aangegeven aantal variabelen wordt verwacht dat zij invloed hebben op default van een financiering. Of dit daadwerkelijk ook zo is, wordt getoetst door het model. In het eerste model worden niet-financiële variabelen als onafhankelijke variabelen meegenomen. In het tweede model zijn dit de financiële variabelen.

Model I:

$$P(Y) = \frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k)}} + \epsilon$$

In het model worden een aantal onafhankelijke leningspecifieke variabelen meegenomen. De keuze voor de onafhankelijke variabelen komt enerzijds voort uit de beschreven literatuur van de invloed van niet-financiële variabelen op default en anderzijds uit de verwachting uit de praktijk en beschikbaarheid van de variabelen bij het aangaan van de financiering. De volgende onafhankelijke variabelen worden meegenomen in model I:

Onafhankelijke variabele	Omschrijving
BKR notering	<p>het Bureau Krediet Registratie (BKR) houdt een elektronisch dossier bij van personen die in Nederland een krediet afsluiten. De kredieten en eventuele betalingsachterstanden worden geregistreerd. Het doel van het BKR is om de kredietwaardigheid van personen te kunnen beoordelen.</p> <p>Van de variabele "BKR notering" wordt verwacht dat hoe slechter de BKR notering, hoe groter de kans op default.</p>
leeftijd ondernemer	<p>van deze variabele wordt verwacht dat hoe jonger de ondernemer, hoe groter de kans op default. Er wordt verondersteld dat jongere ondernemers meer risico nemen dan de oudere ondernemers en bij het niet slagen van de onderneming 'nog een heel leven voor zich hebben'.</p>
percentage eigen inbreng	<p>van het percentage van de investering wat de ondernemer zelf inbrengt wordt verwacht dat hoe hoger de eigen inbreng, hoe lager de kans op default. De gedachte hierachter is dat de ondernemer zijn of haar eigen geld gebruikt voor de financiering en zo direct gecommitteerd is aan het positief verloop van de onderneming en het terugbetalen van de financiering.</p>
ondernemerservaring	<p>er wordt verwacht dat ondernemerservaring een negatief effect heeft op de kans op default van een lening. De ondernemer wordt met het toenemen van de ondernemerservaring in staat geacht beter te anticiperen op veranderingen in de markt en het besturen van de onderneming in het algemeen.</p>
horeca ervaring	<p>er wordt een negatief effect op de kans op default verwacht. Er wordt verondersteld dat met het toenemen van het aantal jaren horeca ervaring er beter ingespeeld kan worden op veranderingen in de markt.</p>
grootte van de onderneming	<p>er wordt een negatief effect verwacht tussen het verband van de grootte van een onderneming en de kans op default van een financiering. Bij grotere ondernemingen wordt verwacht dat zij professioneler geleid worden en hun bestaansrecht hebben bewezen. Kleinere ondernemingen daarentegen zijn gevoeliger voor veranderingen in de markt en worden geacht wat minder stabiel te zijn.</p>
locatie	<p>er wordt verwacht dat hoe beter de locatie is waar de onderneming is gevestigd, de kans op default kleiner is. Een goede locatie voor een horecaonderneming wordt gezien als een belangrijke succesfactor.</p>

Model II:

$$P(Y) = \frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k)}} + \varepsilon$$

In dit model worden wederom een aantal onafhankelijke variabelen meegenomen. De keuze voor de variabelen komt voort uit de literatuur en de praktijk. Daarnaast speelt ook de beschikbaarheid van de gegevens hier een rol, aangezien het onderzoek uit veelal kleinere ondernemingen bestaat, waar de beperkte aanwezigheid van financiële gegevens een rol speelt. Bij het beoordelen van kredieten wordt in het algemeen gekeken naar de solvabiliteitspositie van de onderneming, de liquiditeitspositie van de onderneming, de winstgevendheid van de onderneming en de kasstromen van de onderneming. Onderstaande variabelen zijn hierop gebaseerd. De onafhankelijke variabelen in dit model zijn:

Onafhankelijke variabele	Omschrijving
vaste kosten als percentage van de omzet	de verwachting is dat hoe hoger het percentage vaste kosten ten opzichte van de omzet is, hoe hoger de kans op default. De gedachte hierachter is, dat hoe hoger het percentage vaste kosten, hoe minder flexibel de onderneming zich kan aanpassen aan onder andere eventuele veranderingen in de markt.
solvabiliteitsratio	de verwachting is dat een hoger percentage van solvabiliteit een lagere kans op default tot gevolg heeft. Er wordt verwacht dat een met een hogere solvabiliteitsratio eventuele tegenvallers makkelijker opgevangen kunnen worden.
liquiditeitsratio	er wordt verondersteld dat hoe hoger de liquiditeitsratio, hoe lager de kans op default wordt. De verwachting is dat met een hogere liquiditeitsratio de onderneming beter in staat is om aan de verplichtingen te voldoen.
percentage brutomarge	er wordt verondersteld dat een hoger percentage brutomarge een lagere kans op default heeft. Van een hogere brutomarge wordt verondersteld dat deze de onderneming winstgevender maakt.
free cashflow ten opzichte van de verschuldigde rente en aflossing	de verwachting is dat deze ratio een negatieve invloed heeft op de kans op default van een financiering. Hoe sterker de free cashflow ten opzichte van de verschuldigde rente en aflossing, hoe kleiner de kans op default.

4. Resultaten van het onderzoek

4.1. Introductie

In de Appendix A-2 worden het testen van de assumpties, de modellen, de logistische regressie analyse en de resultaten uitgebreid toegelicht. Dit hoofdstuk zal zich beperken tot de kern van de resultaten van de modellen en zullen er een aantal onderdelen van de logistische regressie analyse worden besproken.

4.2. Resultaten van het onderzoek model I

Tabel 1 laat zien dat er van de 1204 cases 993 cases zijn gebruikt in het model. Dit komt doordat er bij 211 cases in de dataset één of meerdere variabelen missen. Hierdoor kunnen deze cases niet worden meegenomen in de logistische regressie analyse.

Tabel 1. Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	993	82,5
	Missing Cases	211	17,5
	Total	1204	100,0
Unselected Cases		0	,0
Total		1204	100,0

a. If weight is in effect, see classification table for the total number of cases.

Omdat er een paar variabelen in het model zitten die verschillende categorieën bevatten (zie tabel 2), worden de uitkomsten van de verschillende categorieën vergeleken met één van de categorieën. Bij BKR wordt elke BKR notering vergeleken met categorie “geen bijzonderheden”. Bij locatie wordt elke locatie vergeleken met categorie “A”.

In tabel 3 is zichtbaar hoe goed een voorspelling kan worden gedaan met model I, waarin de onafhankelijke variabelen zijn meegenomen. Indien er geen variabelen in het model zouden zijn meegenomen (het zogeheten ‘nulmodel’; met alleen maar een constante), kan er ‘gegokt’ worden dat in 67,6% van de gevallen de lening in default geraakt, omdat in 67,6% van alle cases dat het geval is (zie voor een toelichting Appendix A-2 1.2). In het model waar de onafhankelijke variabelen zijn meegenomen is dit 71,3%. Het model geeft dus een betere voorspelling dan het ‘nulmodel’. Dit betekent dat met behulp van de variabelen in het model een betere voorspelling kan worden gegeven op de kans op default dan zonder deze variabelen.

Tabel 2. Categorical Variables Codings

		Frequency	Parameter coding			
			(1)	(2)	(3)	(4)
BKR notering	Geen bijzonderheden	823	,000	,000	,000	,000
	AH	42	1,000	,000	,000	,000
	A1	45	,000	1,000	,000	,000
	A2	52	,000	,000	1,000	,000
	A3	31	,000	,000	,000	1,000
Locatie	C	35	1,000	,000		
	B	420	,000	1,000		
	A	538	,000	,000		

Tabel 3. Classification Table^a

		Observed		Predicted		Percentage Correct
				Default	Percentage	
		Nee	Ja	Nee	Ja	
Step 1	Default	Nee	638	33	95,1	
		Ja	252	70	21,7	
Overall Percentage					71,3	

a. The cut value is ,500

In tabel 4 zijn de variabelen te zien die in model I zijn meegenomen en welke significant zijn. BKR(1) betekent de vergelijking tussen de notering "AH" en de notering "geen bijzonderheden". Dit blijkt niet significant te zijn, dus dit betekent dat er geen verschil is tussen de notering "AH" en de notering "geen bijzonderheden" op de kans op default. De BKR noteringen "A1", "A2" en "A3" vergeleken met "geen bijzonderheden" blijkt wel van invloed te zijn. In de kolommen "B" en "Exp(B)" is te zien wat er gebeurt. Bij een BKR notering "A3" ten opzichte van "geen bijzonderheden" is de kans op default versus de kans op niet in default geraken 4.992 maal groter. In de kolommen "95% C.I. for Exp(B)" (het 95% betrouwbaarheidsinterval) zijn de onder- en bovenwaarde te zien. Ligt het getal 1 tussen deze waarden, dan heeft de onafhankelijke variabele geen invloed op de kans op default en is het dus ook niet significant. Dit geldt voor "eigen inbreng", "leeftijd", "ondernemerservaring" en "locatie". Er is wel te zeggen dat "eigen inbreng", "leeftijd" en "ondernemerservaring" binnen de steekproef zich bewegen naar hoe meer, hoe minder de kans op default. Dus dat de 'richting' van deze variabelen overeenkomt met de verwachting die is gesteld. Dit is te zien in kolom "B" (de negatieve getallen), maar dit is niet significant. Locatie is

vergeleken met “locatie A”. Er is te zien dat “locatie C” en “locatie B” voor een kleinere kans op default zorgen, maar wederom dus niet significant is. De onafhankelijke variabele “balanstotaal” is wel significant, wat inhoudt dat een balanstotaal van minder dan EUR 150.000 de kans op default versus niet in default geraken 1.381 maal groter is ten opzichte van een balanstotaal van meer dan EUR 150.000. Als laatste blijkt de onafhankelijke variabele “horeca ervaring” significant, waarbij de richting negatief is. Dit betekent hoe minder horeca ervaring, hoe groter de kans op default. Elk jaar meer horeca ervaring zorgt voor een “oddsratio” van .966. Dit betekent dat de kans op default versus de kans om niet in default te geraken .966 maal kleiner wordt. Uit de resultaten blijkt dat niet-financiële variabelen invloed hebben op de kans op default te voorspellen.

Tabel 4. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
			41,947	4	,000			
BKRnotering								
BKRnotering(1)	-,003	,352	,000	1	,992	,997	,500	1,988
BKRnotering(2)	1,207	,321	14,151	1	,000	3,345	1,783	6,275
BKRnotering(3)	1,232	,301	16,769	1	,000	3,429	1,901	6,184
BKRnotering(4)	1,608	,415	15,019	1	,000	4,992	2,214	11,258
Eigeninbreng	-,001	,003	,100	1	,752	,999	,993	1,005
Leeftijd	-,003	,009	,142	1	,706	,997	,980	1,014
Step 1 ^a Ondernemersserva ring	-,014	,013	1,131	1	,288	,986	,961	1,012
Horecaervaring	-,035	,012	8,577	1	,003	,966	,943	,989
Balanstot_CAT	,323	,152	4,507	1	,034	1,381	1,025	1,862
Locatie			2,682	2	,262			
Locatie(1)	,401	,386	1,080	1	,299	1,494	,701	3,185
Locatie(2)	,222	,153	2,124	1	,145	1,249	,926	1,684
Constant	-,817	,422	3,756	1	,053	,442		

a. Variable(s) entered on step 1: BKRnotering, Eigeninbreng, Leeftijd, Ondernemerservaring, Horecaervaring, Balanstot_CAT, Locatie.

4.3. Resultaten van het onderzoek model II

Tabel 5 laat zien dat er in model II 955 van de 1204 cases zijn gebruikt.

Tabel 5. Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	955	79,3
	Missing Cases	249	20,7
	Total	1204	100,0
Unselected Cases		0	,0
Total		1204	100,0

a. If weight is in effect, see classification table for the total number of cases.

In tabel 6 is te zien dat er geen verandering in de voorspelling op is getreden ten opzichte van het 'nulmodel' bij model II (zie Appendix A-2, 2.2, tabel 14).

Tabel 6. Classification Table^a

	Observed	Predicted		
		Default	Percentage	
		Nee	Ja	Correct
Step 1	Default	659	0	100,0
	Nee	296	0	,0
Overall Percentage				69,0

a. The cut value is ,500

In tabel 7 zijn de onafhankelijke variabelen te zien die in model II zijn meegenomen en welke significant zijn. De enige variabele die significant is, is "free cashflow /(rente + aflossing)". Wat te zien is bij 1 punt (0.01) verschil, wordt de kansverhouding op default .947 groter. Dit houdt in dat hoe hoger de free cashflow, hoe kleiner de kans op default is. De andere variabelen zijn niet significant. Er zijn wel verschillen in de steekproef (de gebruikte cases), maar deze verschillen zijn niet te bewijzen voor de gehele populatie.

Tabel 7. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a								
Solvabiliteitsratio	,000	,001	,004	1	,950	1,000	,998	1,002
Liquiditeitsratio	-,067	,038	3,037	1	,081	,936	,868	1,008
Brutomarge	,011	,015	,508	1	,476	1,011	,981	1,042
Nettomarge	,001	,009	,010	1	,919	1,001	,983	1,019
Vastekosten	-,017	,009	3,333	1	,068	,983	,966	1,001
Freecashflowrent eaflossing	-,054	,018	8,635	1	,003	,947	,914	,982
Constant	-,690	1,289	,286	1	,593	,502		

a. Variable(s) entered on step 1: Solvabiliteitsratio, Liquiditeitsratio, Brutomarge, Nettomarge, Vastekosten, Freecashflowrenteaflossing.

4.4. Vergelijking model I en II

De resultaten van model I en II worden met elkaar vergeleken. De veranderingen van de “Chi²-waarden” van model I en model II worden naast elkaar gezet. De verandering van de Chi² houdt in hoeveel de invloed is van de verschillende variabelen op het model (de “-2 Log likelihood”, zie Appendix A-2, 1.3, tabel 7) ten opzichte van het ‘nulmodel’. Bij model I was de Chi²-verandering nog 95,479 (zie Appendix A-2, 1.3, tabel 8) terwijl bij model II de Chi² 20,989 (zie Appendix A-2, 2.3, tabel 18) was. Dit geeft aan dat model I met de niet-financiële variabelen voor een grotere verklaring van de kans op default zorgt dan model II met de financiële variabelen. Daarnaast worden de “Cox & Snell R²” en de “Nagelkerke R²” uit de verschillende modellen met elkaar vergeleken. Deze ‘pseudo’ R²-maten (want logistische regressie analyse geeft geen proportie verklaarde variantie (R²) zoals bij een lineair model) zijn vergelijkbaar met de R² uit de lineaire regressie analyse. Zowel de “Cox & Snell R²” als de “Nagelkerke R²” (Appendix A-2, 1.3, tabel 9) zijn hoger dan de “Cox & Snell R²” en de “Nagelkerke R²” van model II (Appendix A-2, 2.3, tabel 19). Dit is wederom een bewijs dat model I een betere voorspelling op de kans op default geeft dan model II. Hieruit blijkt dus dat het model met de niet-financiële variabelen een betere voorspelling op de kans op default geeft dan het model met de financiële variabelen.

4.5. Aanbevelingen voor het ontwikkelen van een creditscore model

Met de resultaten van model I en model II kan een model III gebouwd worden, wat op basis van de significante onafhankelijke variabelen de kans op default zou kunnen voorspellen. Deze variabelen zijn dus: BKR-notering, horeca ervaring, balanstotaal en free cashflow / (rente + aflossing). Hieronder worden kort de resultaten weergegeven van dit ‘combinatiemodel’. Tabel 8 laat zien dat er in model II 1014 van de 1204 cases zijn gebruikt.

Tabel 8. Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	1014	84,2
	Missing Cases	190	15,8
	Total	1204	100,0
Unselected Cases		0	,0
Total		1204	100,0

a. If weight is in effect, see classification table for the total number of cases.

In tabel 9 is zichtbaar hoe goed een voorspelling kan worden gedaan met model III, waarin de onafhankelijke variabelen zijn meegenomen. Bij het ‘nulmodel’ (zie Appendix A-2, 3.2, tabel 22) werd er 68,4% goed voorspeld. Nu is dit 72,7%. Het model geeft dus een betere voorspelling dan het ‘nulmodel’. Dit betekent dat met behulp van de variabelen in het model een betere voorspelling kan worden gegeven op de kans op default dan zonder deze variabelen.

Tabel 9. Classification Table^a

	Observed	Predicted			
		Default	Percentage		
		Nee	Ja	Correct	
Step 1	Default	Nee	666	28	96,0
		Ja	249	71	22,2
Overall Percentage				72,7	

a. The cut value is ,500

Tabel 10 laat zien dat alle onafhankelijke variabelen significant zijn behalve “BKRnotering 1” (dit betekent geen verschil tussen de BKR noteringen “geen bijzonderheden” ten opzichte van “AH”). Dit houdt in dat een BKR notering “A1” ten opzichte van “geen bijzonderheden” voor een kansverhouding op default versus niet in default geraken zorgt, die 3.320 maal groter is. Bij een BKR notering “A2” is dit 3.728 maal groter en bij een BKR notering “A3” zelfs 6.369 maal. Qua horeca ervaring geldt per extra jaar ervaring een kansverhouding van .952, dus hoe meer ervaring, hoe kleiner de kans op default. Bij de variabele

balanstotaal geldt dat een balanstototaal van minder dan EUR 150.000 een grotere kansverhouding (1.528 maal) op default geeft. Bij de onafhankelijke variabele free cashflow geldt: hoe groter de free cashflow, hoe kleiner de kans op default.

Tabel 10. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
			47,249	4	,000			
Step 1 ^a	BKRnotering							
	BKRnotering(1)	,077	,351	,047	1	,828	1,080	,542 2,150
	BKRnotering(2)	1,200	,314	14,613	1	,000	3,320	1,794 6,142
	BKRnotering(3)	1,316	,296	19,773	1	,000	3,728	2,087 6,659
	BKRnotering(4)	1,851	,432	18,405	1	,000	6,369	2,733 14,839
	Horecaervaring	-,049	,009	28,315	1	,000	,952	,935 ,969
	Balanstot_CAT	,424	,145	8,581	1	,003	1,528	1,151 2,030
	Freecashflowrent eaflossing	-,049	,017	8,654	1	,003	,952	,921 ,984
	Constant	-,897	,269	11,124	1	,001	,408	

a. Variable(s) entered on step 1: BKRnotering, Horecaervaring, Balanstot_CAT, Freecashflowrenteaflossing.

Uit dit model kan een formule worden gebouwd om de kans op default te kunnen berekenen. Deze formule ziet er als volgt uit:

$$P(\text{Default}) = e^{(-0.897) + (.077 * \text{BKR1}) + (1.2 * \text{BKR2}) + (1.316 * \text{BKR3}) + (1.851 * \text{BKR4}) + (-0,49 * x2i) + (0,424 * x3i) + (-0,49 * x4i)} / 1 + e^{(-0.897) + (.077 * \text{BKR1}) + (1.2 * \text{BKR2}) + (1.316 * \text{BKR3}) + (1.851 * \text{BKR4}) + (-0,49 * x2i) + (0,424 * x3i) + (-0,49 * x4i)}$$

Dit model kan gebruikt worden bij de beoordeling van de kredietaanvraag. Na invulling van deze factoren wordt dan dus de kans dat een lening in default geraakt weergegeven.

De toepasbaarheid van het model is geschikt voor het moment van de kredietaanvraag. Het is dus een momentopname en kan niet met ongewijzigde input gebruikt worden om de kredietwaardigheid van een kredietnemer te toetsen gedurende de looptijd van de financiering. Immers, als de financiering een aantal jaren loopt zijn er idealiter (betrouwbare) financiële gegevens beschikbaar die getoetst kunnen worden. Indien dit niet het geval is, kan worden teruggegrepen naar het bovengenoemde model, dat met geüpdate gegevens de kans op default kan voorspellen. Een belangrijk aspect is de beschikbaarheid en betrouwbaarheid van de variabelen. De variabelen zullen indien mogelijk getoetst moeten worden op betrouwbaarheid. Dit is een vereiste voor het goed functioneren van het model. Het aantal jaren horeca ervaring wat de ondernemer opgeeft bijvoorbeeld is makkelijk te manipuleren en kan door middel van opgegeven referenties worden gecontroleerd.

5. Conclusies

Kredietrisico is een van de grootste risico's van een bank. Kredietrisico is het risico dat de leningnemer niet in staat is het volledige bedrag van hoofdsom en rente van het geleende bedrag terug te betalen. Banken dienen de kredietwaardigheid van hun klanten grondig te analyseren om het niveau van kredietwaardigheid te beoordelen. In het geval niet aan de betalingsverplichtingen van de lening wordt voldaan, spreekt men van default. Een default van een lening kan uiteindelijk leiden tot kredietverliezen. Het voorspellen van de kans op default (probability of default) is daarom een essentieel onderdeel voor banken en andere financieringsinstellingen om hun prijsbeleid voor financieringen en voorzieningen voor oninbare vorderingen vast te stellen. Het niet goed inschatten hiervan kan grote gevolgen hebben voor de kosten van kapitaal van deze financieringsinstellingen. Om de kans op default in te schatten, wordt er veelal gebruik gemaakt van creditscore modellen. Het principe van creditscore modellen is om factoren te bepalen die de kans op default kunnen beïnvloeden. Aan deze factoren kan een bepaalde weging worden meegegeven en deze kunnen gecombineerd worden in een relevante creditscore.

In dit onderzoek is de rol van niet-financiële informatie naast financiële informatie op het voorspellen van de kans op default onderzocht. Literatuur uit het verleden heeft gewezen op een potentiële rol van niet-financiële informatie in het voorspellen van de kans op default, maar heeft niet de omvang waar welke niet-financiële informatie toegevoegde voorspellende waarde heeft boven financiële informatie benadrukt. In dit onderzoek is geprobeerd hier een bijdrage aan te leveren. Door gebruik te maken van een model met niet-financiële variabelen is onderzocht of deze variabelen van invloed zijn op het voorspellen van de kans op default.

De eerste deelvraag luidde: "Welke variabelen hebben invloed op de kans op default bij horecafinancieringen? De resultaten zoals beschreven in paragraaf 4.2 laten zien dat bij het eerste model de variabelen BKR-notering, balanstotaal en horeca ervaring van significante invloed zijn op het voorspellen van de kans op default van een financiering. Bij het tweede model, waarvan de resultaten zijn beschreven in paragraaf 4.3, komt naar voren dat de variabele free cashflow van significante invloed is op het voorspellen van de kans op default van een financiering. Een vergelijking van beide modellen, zoals in paragraaf 4.4. beschreven, laat zelfs zien dat het model met de niet-financiële variabelen een grotere voorspellende waarde op de kans op default heeft dan het model met de financiële variabelen.

De tweede deelvraag luidde: "Welke aanbevelingen zijn er te doen voor het ontwikkelen van een credit score model voor horecafinancieringen op basis van de verkregen inzichten?". In het tweede deel van het onderzoek zijn de significante variabelen van de beide modellen samengevoegd in een nieuw model, wat een nog grotere voorspellende waarde heeft dan het model met enkel de niet-financiële variabelen.

Met deze verkregen inzichten vanuit het onderzoek zijn aanbevelingen gedaan voor het ontwikkelen van een creditscore model voor financieringen, zoals beschreven in paragraaf 4.5. Door de onafhankelijke variabelen in het model in te vullen wordt de kans op default weergegeven. De toepasbaarheid van het model en gerelateerde zaken waar met name op moet worden gelet zijn eveneens beschreven in deze paragraaf.

6. Discussie en toepasbaarheid

Literatuur uit het verleden heeft gewezen op de potentiële rol van niet-financiële informatie op het voorspellen van de kans op default, maar heeft niet de omvang waar welke niet-financiële informatie toegevoegde voorspellende waarde heeft boven financiële informatie benadrukt. In dit onderzoek is aangetoond dat niet-financiële informatie een voorspellende waarde op de kans op default heeft en uit de resultaten van het onderzoek blijkt zelfs dat deze voorspellende waarde groter is dan bij financiële informatie. Met deze resultaten en het voorgestelde model is het mogelijk dat ondernemingen die financieringen verstrekken dit model, wat de kans op default voorspelt, gebruiken voor het toekennen van een creditscore aan een kredietnemer. Zoals reeds besproken in paragraaf 4.5 is de toepasbaarheid van het model geschikt voor het moment van de kredietaanvraag. Het model is namelijk een momentopname en kan niet met ongewijzigde input gebruikt worden om de kredietwaardigheid van een kredietnemer te toetsen gedurende de looptijd van de financiering. Een belangrijk aspect is de beschikbaarheid en betrouwbaarheid van de variabelen. De gebruikte variabelen zullen indien mogelijk getoetst moeten worden op betrouwbaarheid. Een ander belangrijk aspect is dat het onderzoek is uitgevoerd op een leningportefeuille in de horecabranche, waar de ondernemer in de onderneming als geheel een zeer belangrijke rol vervult. Een aantal van de niet-financiële variabelen (eigen inbreng, BKR notering, leeftijd, ondernemerservaring en horeca ervaring) zijn dan ook direct gerelateerd aan de ondernemer. Dit brengt ons tot het punt van de toepasbaarheid in andere branches en industrieën. De toepasbaarheid van het model in andere branches en industrieën hangt af van aspecten van het type onderneming die in belangrijke mate overeen dienen te komen met de ondernemingen in de horecabranche, waarin het onderzoek is uitgevoerd. Belangrijke aspecten hierin zijn de veelal kleinere ondernemingen, waar de beperkte aanwezigheid van betrouwbare financiële gegevens een rol speelt, de relatief hoge investeringen, maar de belangrijkste reden is de rol van de ondernemer zelf.

Appendix A-1. Variabelen in het onderzoek

1. Afhankelijke variabele

De afhankelijke variabele “default” is als volgt gedefinieerd. Indien een financiering vanwege betalingsachterstanden is opgeëist, wordt deze financiering bestempeld als “default”. Daarnaast worden financieringen die meer dan zes maanden achterstand hebben ook gekwalificeerd als default. Ervaring wijst uit dat de kans dat deze financieringen alsnog gaan betalen namelijk zeer gering is.

2. Onafhankelijke variabelen model I

De onafhankelijke variabelen die in model I worden meegenomen zijn:

- BKR (1 = Geen bijzonderheden, 2 = AH, 3 = A1, 4 = A2, 5 = A3)
- Leeftijd (in jaren)
- Eigen inbreng (in percentage van totale investering)
- Ondernemerservaring (in jaren)
- Horeca ervaring (in jaren)
- Balanstotaal (1 = > 150.000 euro, 2 = < 150.000 euro)
- Locatie (1 = A, 2 = B, 3 = C)

-de variabele BKR notering wordt ingedeeld in vijf categorieën;

1. geen bijzonderheden
2. AH codering
3. A codering
4. A2 codering
5. A3 codering

De categorie “geen bijzonderheden” houdt in dat personen geen registratie van kredieten hebben bij het BKR of dat er op de geregistreerde kredieten geen achterstandsmeldingen hebben plaatsgevonden.

De categorie “AH” betekent dat er in het verleden een achterstandsmelding heeft plaatsgevonden, maar dat die achterstand door de persoon hersteld is.

Bij de categorie “A codering” is er een achterstandsmelding geregistreerd die nog niet hersteld is.

De categorie “A2 codering” betekent dat de nog openstaande lening opeisbaar is gesteld.

De categorie “A3 codering” houdt in dat er een bedrag op de lening is afgeboekt.

- de variabele “percentage eigen inbreng ten opzichte van de investering” houdt in het deel van de investering die de ondernemer zelf inbrengt.

- de variabele “leeftijd” wordt aangeduid in jaren.

- de variabele “ondernemerservaring” wordt gemeten in jaren.

- de variabele “horecaervaring” wordt gemeten in jaren.

- de variabele “grootte van de onderneming” wordt gemeten naar balanstotaal van de onderneming. Er wordt onderscheid gemaakt tussen de categorieën minder dan EUR 150.000 balanstotaal en meer dan EUR 150.000 balanstotaal.

- de variabele “locatie” wordt gemeten in drie categorieën; een A-locatie, een B-locatie en een C-locatie. Deze categorieën zijn bepaald door de kredietanalist en zijn daardoor voor een deel subjectief te

noemen. De keuze voor een A-, B-, of C-locatie hangt van meerdere factoren af zoals omgeving, zichtbaarheid en bereikbaarheid.

3. Onafhankelijke variabelen model II

De onafhankelijke variabelen die in model II worden meegenomen zijn:

- Solvabiliteitsratio (gestandaardiseerd met 1000 als gemiddelde, in hele getallen)
- Liquiditeitsratio (in twee decimalen)
- Brutomarge (in percentages)
- Nettomarge (in percentages)
- Vastekosten (in percentages)
- Free Cashflow/(rente+aflossingen)

- vaste kosten als percentage van de omzet; de verwachting is dat hoe hoger het percentage vaste kosten ten opzichte van de omzet is, hoe hoger de kans op default is.
- solvabiliteitsratio; de verwachting is dat een hoger percentage van solvabiliteit (gemeten als eigen vermogen ten opzichte van het totale vermogen van de onderneming) een negatieve invloed heeft op de kans op default van een lening.
- liquiditeitsratio; er wordt verondersteld dat hoe hoger de liquiditeitsratio, gemeten als kortlopende vorderingen minus kortlopende schulden, hoe kleiner de kans op default.
- percentage bruto marge; er wordt verondersteld dat dat een hoger percentage brutomarge een negatief effect heeft op de kans op default van een lening. De brutomarge wordt gemeten door de omzet minus de inkoopwaarde van de omzet te delen door de omzet.
- percentage netto marge; de verwachting is dat het nettowinstpercentage (gemeten als netto winst gedeeld door de omzet) een negatieve invloed heeft op de kans op default.
- free cashflow ten opzichte van de verschuldigde rente en aflossing; de verwachting is dat deze ratio een negatieve invloed heeft op de kans op default van een lening. Hoe sterker de free cash flow ten opzichte van de verschuldigde rente en aflossing, hoe kleiner de kans dat de lening nemer de verschuldigde bedragen niet kan betalen.

4. Onafhankelijke variabelen model III

De onafhankelijke variabelen die in model III worden meegenomen zijn:

- BKR (1 = Geen bijzonderheden, 2 = AH, 3 = A1, 4 = A2, 5 = A3)
- Horeca-ervaring (in jaren)
- Balanstotaal (1 = > 150.000 euro, 2 = < 150.000 euro)
- Freecashflowrenteaflissing (in decimalen)

Appendix A-2. Resultaten van het onderzoek

1. Model I

1.1. Testen van de assumpties

Er dient te worden voldaan aan een aantal assumpties voordat de logistische regressie kan worden uitgevoerd. Deze betreffen:

- lineariteit: er dient een lineaire relatie te zijn tussen de intervalpredictoren en de logit van de afhankelijke variabele. Het blijkt dat de interactietermen niet significant zijn. Dit betekent dat er sprake is van een lineaire relatie tussen de intervalpredictoren en de logit van de afhankelijke variabele. Er wordt aan de assumptie voldaan.
- onafhankelijke errors: geen van de dossiers elkaar hebben beïnvloed qua predictoren. Er wordt aan de assumptie voldaan.
- multicollineariteit: De predictoren mogen onderling niet te hoog correleren. Ook aan deze assumptie wordt voldaan.

1.1.1 Lineariteit

Zoals met groen gemarkeerd in tabel 1, blijkt dat de interactietermen niet significant zijn. Dit betekent dat er sprake is van een lineaire relatie tussen de intervalpredictoren en de logit van de afhankelijke variabele. Er wordt aan de assumptie voldaan.

1.1.2. Onafhankelijke errors

Geen van de dossiers hebben elkaar beïnvloed qua predictoren.

1.1.3. Multicollineariteit

Uit tabel 2 is op te maken dat er bij kolom "Condition Index" geen hele grote verschillen zitten. Daarnaast lijken de varianties van de regressie coëfficiënten niet afhankelijk van elkaar te zijn aangezien de 'kleine' "eigenvalues" (vanaf model stap 6) niet hoog scoren op verschillende predictoren. Ook dit is een bewijs voor het kunnen aannemen van de assumptie.

Tabel1. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
BKRnotering			14,512	4	,006			
BKRnotering(1)	,227	,579	,154	1	,695	1,255	,404	3,905
BKRnotering(2)	,559	,571	,958	1	,328	1,749	,571	5,361
BKRnotering(3)	1,666	,568	8,607	1	,003	5,293	1,739	16,114
BKRnotering(4)	2,017	,841	5,757	1	,016	7,519	1,447	39,076
Eigeninbreng	-,071	,086	,679	1	,410	,931	,786	1,103
Leeftijd	-,570	,491	1,348	1	,246	,566	,216	1,480
Ondernemerservaring	-,104	,200	,270	1	,603	,901	,609	1,333
Horecaervaring	-,260	,201	1,668	1	,197	,771	,520	1,144
Balanstot_CAT	,788	,281	7,887	1	,005	2,199	1,269	3,812
Locatie			,661	2	,719			
Locatie(1)	-,853	1,127	,573	1	,449	,426	,047	3,880
Locatie(2)	-,115	,278	,170	1	,680	,891	,517	1,538
Eigeninbreng by LN_Eigeninbreng	,015	,019	,646	1	,421	1,015	,979	1,053
LN_Leeftijd by Leeftijd	,122	,103	1,414	1	,234	1,130	,924	1,382
LN_ondernemerservaring by Ondernemerservaring	,038	,057	,442	1	,506	1,039	,929	1,162
Horecaervaring by LN_horecaervaring	,048	,056	,709	1	,400	1,049	,939	1,171
Constant	4,914	4,451	1,219	1	,270	136,159		

a. Variable(s) entered on step 1: BKRnotering, Eigeninbreng, Leeftijd, Ondernemerservaring, Horecaervaring, Balanstot_CAT, Locatie, Eigeninbreng * LN_Eigeninbreng , LN_Leeftijd * Leeftijd , LN_ondernemerservaring * Ondernemerservaring , Horecaervaring * LN_horecaervaring .

Tabel 2. Collinearity Diagnostics^a

Model	Dimensie	Eigenvalue	Condition Index	Variance Proportions							
				(Constant)	BKR notering	Percentage eigen inbreng	Leeftijd in jaren	Ondernemservaring in jaren	Horeca ervaring in jaren	Balans tot CAT	Locatie
1	1	6,310	1,000	,00	,01	,01	,00	,00	,00	,00	,00
	2	,716	2,968	,00	,01	,57	,00	,05	,03	,00	,00
	3	,469	3,667	,00	,22	,33	,00	,10	,03	,01	,00
	4	,258	4,941	,00	,74	,04	,00	,07	,00	,05	,01
	5	,112	7,518	,00	,00	,00	,00	,57	,93	,00	,00
	6	,086	8,558	,00	,00	,01	,00	,06	,00	,55	,21
	7	,038	12,857	,00	,00	,00	,56	,11	,00	,18	,27
	8	,010	24,740	,99	,03	,03	,43	,03	,00	,20	,50

a. Dependent Variable: Default

1.2. 'Nulmodel I'

In tabel 3 is zichtbaar hoe goed het model bij de data past. Onderstaande tabel geeft dit weer voor het model waar de onafhankelijke variabelen niet zijn meegenomen. Alleen de constante is in dit model opgenomen. Dit noemen we het 'nulmodel'.

Het "-2 Log likelihood" geeft aan hoe goed het model bij de data past. Hoe hoger dit getal, hoe slechter. Dit kan straks vergeleken worden met het model waarin de onafhankelijke variabelen zijn opgenomen.

Tabel 3. Iteration History^{a,b,c}

Iteration	-2 likelihood	Log Coefficients Constant
1	1251,486	-,703
Step 0 2	1251,272	-,734
3	1251,272	-,734

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 1251,272

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

In het zogeheten 'nulmodel' zou men kunnen 'gokken' dat een lening niet in default geraakt, omdat in 67,6% van alle cases dat het geval is. Dit betekent dat men zonder predictoren het in 67,6% van de gevallen goed zou hebben. Dit is in tabel 4 te zien.

Tabel 4. Classification Table^{a,b}

	Observed	Predicted			Percentage Correct
		Default		Percentage	
		Nee	Ja		
Step 0	Nee	671	0	100,0	
	Default	Ja	322	0	,0
Overall Percentage				67,6	

a. Constant is included in the model.

b. The cut value is ,500

Tabel 5. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-,734	,068	117,295	1	,000	,480

In tabel 5 is te zien dat alleen de constante in het 'nulmodel' is opgenomen.

De onafhankelijke variabelen zijn dus niet opgenomen in het model. De X^2 geeft de verwachte verandering weer als die variabele wel zou worden opgenomen in het model. Zoals in tabel 6 te zien, lijkt BKR, leeftijd, ondernemerservaring, horeca ervaring, balanstotaal en locatie significante invloed te hebben op het model. Zoals verderop zal blijken, is dit niet het geval als alle variabelen samen in het model worden meegenomen.

Tabel 6. Variables not in the Equation

		Score	df	Sig.	
Step 0	Variables	BKRnotering	53,929	4	,000
		BKRnotering(1)	,044	1	,835
		BKRnotering(2)	13,825	1	,000
		BKRnotering(3)	13,645	1	,000
		BKRnotering(4)	21,692	1	,000
		Eigeninbreng	,152	1	,697
		Leeftijd	9,527	1	,002
		Ondernemersserva ring	27,373	1	,000
		Horecaervaring	33,753	1	,000
		Balanstot_CAT	15,232	1	,000
		Locatie	12,402	2	,002
		Locatie(1)	1,801	1	,180
		Locatie(2)	8,955	1	,003
Overall Statistics		94,963	11	,000	

1.3. Resultaten model I

In tabel 7 is de invloed te zien van de verschillende variabelen op het model (“-2Log likelihood”). Zoals aangegeven had het ‘nulmodel’ een “-2Log Likelihood” van 1251,272 terwijl dit nu 1155,793 bij stap 4 is. Deze verschillende stappen houdt in dat SPSS de informatie zo probeert te rangschikken dat hier het beste model uitkomt. SPSS gaat niet verder dan stap 4 omdat er een verschil van minder dan .001 is ten opzichte van stap 3.

Tabel 7. Iteration History^{a,b,c,d}

Iteration	-2 likelihood	Log ant	Coefficients										
			Const	BKRnoteri ng(1)	BKRnoteri ng(2)	BKRnoteri ng(3)	BKRnoteri ng(4)	Eigeninbr eng	Leeft ijd	Onderne merserva ring	Horecaer varing	Balanstot _CAT	Locatie (1)
1	1159,593	-,788	-,010	1,099	1,099	1,471	-,001	-,003	-,009	-,027	,276	,357	,180
Step 2	1155,810	-,820	-,004	1,201	1,224	1,602	-,001	-,003	-,013	-,034	,322	,401	,220
1 3	1155,793	-,817	-,003	1,207	1,232	1,608	-,001	-,003	-,014	-,035	,323	,401	,222
4	1155,793	-,817	-,003	1,207	1,232	1,608	-,001	-,003	-,014	-,035	,323	,401	,222

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 1251,272

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

De “Chi²” (zie tabel 8) geeft aan hoeveel verandering er is opgetreden ten opzichte van het ‘nulmodel’ (1251,272 – 95,479 = 1155,793). Deze “Chi²” wordt bij “Step”, “Block” en “Model” aangegeven omdat hier de “Entermethode” gebruikt wordt.

Tabel 8. Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
	Step	95,479	11	,000
Step 1	Block	95,479	11	,000
	Model	95,479	11	,000

Tabel 9. Model Summary

Step	-2 likelihood	Log Cox & Snell R Square	Nagelkerke R Square
1	1155,793 ^a	,092	,128

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

In tabel 9 is nogmaals te zien dat het “-2Log likelihood” beter is dan het ‘nulmodel’. Daarnaast staan de “Cox & Snell R²” en de “Nagelkerke R²” die iets zeggen over de proportie verklaarde variantie (hoeveel variantie betreft default kan worden verklaard door de predictoren in het model). Oftewel hoe goed past het model op de data. Beide geven een indicatie over de proportie verklaarde variantie. Bij Cox is dit dus 9,2% en bij Nagelkerke 12,8%.

1.4. Outliers

In tabel 10 zijn de residuen (outliers) te zien.

Tabel 10. Casewise List^b

Case	Selected Status ^a	Observed Default	Predicted	Predicted Group	Temporary Variable Resid	ZResid
412	S	J**	,065	N	,935	3,779
467	S	J**	,129	N	,871	2,601
490	S	J**	,114	N	,886	2,784
932	S	J**	,110	N	,890	2,840
960	S	J**	,131	N	,869	2,571
1000	S	J**	,112	N	,888	2,811

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2,000 are listed.

2. Model II

2.1. Testen van de assumpties

Er dient te worden voldaan aan een aantal assumpties voordat de logistische regressie kan worden uitgevoerd. Deze betreffen:

- lineariteit: er dient een lineaire relatie te zijn tussen de intervalpredictoren en de logit van de afhankelijke variabele. Het blijkt dat de interactietermen niet significant zijn. Dit betekent dat er sprake is van een lineaire relatie tussen de intervalpredictoren en de logit van de afhankelijke variabele. Er wordt aan de assumptie voldaan.
- onafhankelijke errors: geen van de dossiers elkaar hebben beïnvloed qua predictoren. Er wordt aan de assumptie voldaan.
- multicollineariteit: De predictoren mogen onderling niet te hoog correleren. Ook aan deze assumptie wordt voldaan.

2.1.1. Lineariteit

Zoals met groen gemarkeerd in tabel 11, blijkt dat de interactietermen niet significant zijn. Dit betekent dat er sprake is van een lineaire relatie tussen de intervalpredictoren en de logit van de afhankelijke variabele. Er wordt aan de assumptie voldaan.

2.1.2. Onafhankelijke errors

Geen van de dossiers hebben elkaar beïnvloed qua predictoren.

2.1.3. Multicollinaeriteit

De verschillen tussen "Condition Index" zijn groter dan bij model I, maar nog steeds geen indicatie voor een slechte fit tussen model en data vanwege het niet hoog laden van meerdere regressie coëfficiënten op dezelfde "Eigenvalue". Dit betekent dat aan de assumptie is voldaan. Dit is te zien in tabel 12.

Tabel 11. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Solvabiliteitsratio	,053	,051	1,060	1	,303	1,054
Liquiditeitsratio	,192	,195	,968	1	,325	1,211
Brutomarge	-,485	,473	1,054	1	,305	,615
Nettomarge	-,015	,067	,048	1	,827	,985
Vastekosten	,076	,167	,206	1	,650	1,079
Freecashflowrent eaflossing	-,050	,019	7,041	1	,008	,952
LN_Solvabiliteit by	-,007	,007	1,093	1	,296	,993
Step 1 ^a Solvabiliteitsratio						
LN_liquiditeit by Liquiditeitsratio	-,102	,082	1,564	1	,211	,903
Brutomarge by LN_Brutomarge	,098	,091	1,153	1	,283	1,102
LN_Nettomarge by Nettomarge	,003	,017	,023	1	,879	1,003
LN_Vastekosten by Vastekosten	-,021	,036	,349	1	,555	,979
Constant	-,422	8,039	,003	1	,958	,656

a. Variable(s) entered on step 1: Solvabiliteitsratio, Liquiditeitsratio, Brutomarge, Nettomarge, Vastekosten, Freecashflowrenteaflossing, LN_Solvabiliteit * Solvabiliteitsratio, LN_liquiditeit * Liquiditeitsratio, Brutomarge * LN_Brutomarge, LN_Nettomarge * Nettomarge, LN_Vastekosten * Vastekosten.

Tabel 12. Collinearity Diagnostics^a

Model	Dimensie	Eigenvalue	Condition Index	Variance (Constant)	Proportions Gestandaardiseerd naar duizend, hele getallen	Liquiditeitsratio in twee decimale n	Bruto marge in percenta ges	Netto marge in percenta ges	Free cash flow (rente + aflossing)	Vaste kosten in percenta ges
1	4,993	1,000	,00	,00	,01	,00	,01	,01	,00	
2	,789	2,516	,00	,00	,43	,00	,02	,46	,00	
3	,730	2,616	,00	,00	,56	,00	,00	,30	,00	
1 4	,458	3,303	,00	,00	,00	,00	,34	,22	,01	
5	,024	14,571	,02	,03	,00	,01	,58	,00	,87	
6	,005	33,001	,01	,55	,00	,49	,03	,00	,07	
7	,002	51,108	,98	,41	,00	,51	,02	,00	,04	

a. Dependent Variable: Default

2.2. 'Nulmodel II'

Net zoals bij model I te zien was, geeft “-2 Log likelihood” aan hoe goed het model bij de data past. Hoe hoger deze waarde, hoe slechter. Dit is te zien in tabel 13.

Tabel 13. Iteration History^{a,b,c}

Iteration	-2 likelihood	Log Coefficients Constant
1	1182,733	-,760
Step 0 2	1182,402	-,800
3	1182,402	-,800

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 1182,402

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Tabel 14. Classification Table^{a,b}

	Observed	Predicted			
		Default	Percentage		
		Nee	Ja	Correct	
Step 0	Default	Nee	659	0	100,0
		Ja	296	0	,0
Overall Percentage				69,0	

a. Constant is included in the model.

b. The cut value is ,500

Bij het 'nulmodel' (zonder predictoren) zou men kunnen 'gokken' dat een lening niet in default geraakt, omdat dit in 69% van alle cases het geval is. Dit betekent dat men zonder predictoren het in 69% van de gevallen goed zou hebben. Tabel 14 laat dit zien.

Tabel 15. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	
Step 0	Constant	-,800	,070	130,843	1	,000	,449

In tabel 15 is te zien dat alleen de constante in het nulmodel is opgenomen.

De onafhankelijke variabelen zijn dus niet opgenomen in het model. De X^2 geeft de verwachte verandering weer als die variabele wel zou worden opgenomen in het model. Zoals in onderstaande tabel is te zien, lijken vaste kosten en free cashflow / (rente+aflossing) significante invloed te hebben op het model. Zoals verderop zal blijken, is alleen free cashflow / (rente+aflossing) significant als alle variabelen in het model worden meegenomen. Dit is te zien in tabel 16.

Tabel 16. Variables not in the Equation

	Score	df	Sig.		
Step 0	Solvabiliteitsratio	,363	1	,547	
	Liquiditeitsratio	3,448	1	,063	
	Brutomarge	,067	1	,796	
	Variables	Nettomarge	1,522	1	,217
	Vastekosten	4,456	1	,035	
	Freecashflowrent eaflossing	7,287	1	,007	
	Overall Statistics	17,083	6	,009	

2.3. Resulten model II

In tabel 17 is de invloed te zien van de verschillende variabelen op het model (“-2 Log likelihood”). Zoals aangegeven had het ‘nulmodel’ een “-2Log likelihood” van 1181,402 terwijl dit nu 1164,414 bij stap 5 is. Deze verschillende stappen houdt in dat SPSS de informatie zo probeert te rangschikken dat hier het beste model uitkomt. SPSS gaat niet verder dan stap 5 omdat er een verschil van minder dan .001 is t.o.v. stap 4.

Tabel 17. Iteration History^{a,b,c,d}

Iteration	-2 likelihood	Log Coefficients							
		Consta nt	Solvabilitei tsratio	Liquiditeits ratio	Brutomar ge	Nettomar ge	Vastekost en	Freecashflo wreteafl o ssing	
	1	1164,831	-,581	,000	-,033	,008	,001	-,012	-,031
Step 1	2	1161,534	-,669	,000	-,057	,011	,001	-,016	-,050
	3	1161,414	-,688	,000	-,066	,011	,001	-,017	-,054
	4	1161,414	-,690	,000	-,067	,011	,001	-,017	-,054
	5	1161,414	-,690	,000	-,067	,011	,001	-,017	-,054

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 1182,402

d. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

De “Chi²” geeft aan hoeveel verandering er is opgetreden ten opzichte van het ‘nulmodel’ (1181,402 – 20,989 = 1164,414). Deze “Chi-square” wordt bij “Step”, “Block” en “Model” aangegeven omdat hier de “Enter methode” gebruikt wordt. Dit is te zien in tabel 18.

Tabel 18. Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step	20,989	6	,002
Step 1 Block	20,989	6	,002
Model	20,989	6	,002

Tabel 19. Model Summary

Step	-2 likelihood	Log Cox & Snell R Square	Nagelkerke Square
1	1161,414 ^a	,022	,031

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

In tabel 19 is nogmaals te zien dat het “-2Log likelihood” beter is dan het ‘nulmodel’.

2.4. Outliers

In tabel 20 zijn de residuen (outliers) te zien.

Tabel 20. Casewise List^b

Case	Selected Status ^a	Observed Default	Predicted	Predicted Group	Temporary Variable Resid	ZResid
64	S	J**	,137	N	,863	2,509
78	S	J**	,026	N	,974	6,069

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2,000 are listed.

3. Model III

3.1. Testen van de assumpties

Omdat de assumpties reeds zijn getest bij model I en model II is dit hier niet meer nodig.

3.2. 'Nulmodel III'

Net zoals bij model I en model II te zien was, geeft “-2 Log likelihood” in tabel 21 aan hoe goed het model bij de data past. Hoe hoger deze waarde, hoe slechter.

Tabel 21. Iteration History^{a,b,c}

Iteration	-2 likelihood	Log Coefficients Constant
1	1264,739	-,738
Step 0 2	1264,446	-,774
3	1264,446	-,774

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 1264,446

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

In het zogeheten ‘nulmodel’ zou men kunnen ‘gokken’ dat een lening niet in default geraakt, omdat in 68,4% van alle cases dat het geval is. Dit betekent dat men zonder predictoren het in 68,4% van de gevallen goed zou hebben. Dit is te zien in tabel 22.

Tabel 22. Classification Table^{a,b}

	Observed	Predicted			Percentage Correct
		Default		Percentage	
		Nee	Ja		
Step 0	Default	Nee	694	0	100,0
		Ja	320	0	,0
Overall Percentage					68,4

a. Constant is included in the model.

b. The cut value is ,500

Tabel 23. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-,774	,068	131,257	1	,000	,461

In tabel 23 is te zien dat alleen de constante in het nulmodel is opgenomen.

In tabel 24 is te zien dat wordt voorspeld dat alleen BKRnotering contrast 1 (“geen bijzonderheden” ten opzichte van “AH”) niet significant is. Dit blijkt later ook zo te zijn.

Tabel 24. Variables not in the Equation

	Score	df	Sig.	
Step 0 Variables	BKRnotering	56,724	4	,000
	BKRnotering(1)	,000	1	,983
	BKRnotering(2)	12,883	1	,000
	BKRnotering(3)	15,543	1	,000
	BKRnotering(4)	22,994	1	,000
	Horecaervaring	40,472	1	,000
	Balanstot_CAT	15,925	1	,000
	Freecashflowrent eaflossing	7,307	1	,007
Overall Statistics	106,379	7	,000	

3.3. Resultaten model III

In tabel 25 is de invloed te zien van de verschillende variabelen op het model (“-2 Log likelihood”). Zoals aangegeven had het ‘nulmodel’ een “-2Log likelihood” van 1264,446 terwijl dit nu 1154,553 bij stap 4 is.

Tabel 25. Iteration History^{a,b,c,d}

Iteration	-2 likelihood	Log Coefficients								
		Const ant	BKRnoter ing(1)	BKRnoter ing(2)	BKRnoter ing(3)	BKRnoter ing(4)	Horecaer varing	Balanstot _CAT	Freecashf lowrente aflossing	
1	1161,313	-,863	,045	1,062	1,137	1,609	-,036	,340	-,028	
Step 2	1154,640	-,901	,072	1,187	1,299	1,829	-,048	,418	-,046	
1 3	1154,553	-,897	,076	1,200	1,316	1,851	-,049	,424	-,049	
4	1154,553	-,897	,077	1,200	1,316	1,851	-,049	,424	-,049	

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 1264,446

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

De “Chi-square” in tabel 26 geeft aan hoeveel verandering er is opgetreden ten opzichte van het ‘nulmodel’ (1264,446 – 109,893 = 1154,553). Deze “Chi-square” wordt bij “Step”, “Block” en “Model” aangegeven omdat hier de “Entermethode” gebruikt wordt.

Tabel 26. Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
	Step	109,893	7	,000
Step 1	Block	109,893	7	,000
	Model	109,893	7	,000