Udarbejdelse af spørgeskemaer og analyser af -besvarelser

Indholdsfortegnelse

1. Baggrund og formål **2**
2. Fremgangsmåde **2**
    - 2.1 Bruttoliste 3
    - 2.2 Test af antal dimensioner 3
    - 2.3 Item Response Theory 5
    - 2.4 Revision og optimering af spørgeskema 8
    - 2.5 Integration 9
3. Referencer **9**

For notatet:
Erhvervsøkonomisk chef Klaus Kaiser
SEGES, Landbrug & Fødevarer
T +45 8740 5175
M +45 2013 5175
E kak@seges.dk

Seniorkonsulent Michael Studsgaard Sørensen
SEGES, Landbrug & Fødevarer

1. Baggrund og formål

Dette notat redegør for den fremgangsmåde og det teoretiske grundlag, der anvendes ved udarbejdelse af spørgeskemaer og til behandling af besvarelser i SEGES' Ratingmodel og Risikomodel.

Et spørgeskema anvendes typisk som værktøj, når det ønskes at måle en eller flere størrelser, der enten ikke umiddelbart kan måles, eller ikke kan måles præcist. Disse størrelser kaldes også "latente variable". I Ratingmodellen er den latente variabel **kvaliteten af management** på en bedrift, mens det i Risikomodellen er **landmandens risiko-setup** i bred forstand.

De latente variable kan hverken direkte måles eller spørges ind til, men via spørgeskemaer er det muligt at indfange information om de underliggende eller latente variable. For at estimere størrelsen af de latente variable, anvendes teorier og værktøjer som Classical Test Theory (CTT), Item Respons Theory (IRT) og Structural Equation Models (SEM). Disse anvendes blandt andet inden for psykologiske videnskaber.

Udover at estimere latente variables størrelse, kan værktøjerne tillige anvendes til at designe spørgeskemaer, således at der findes et minimalt sæt af spørgsmål, der maksimerer informationen om den eller de latente størrelser, der ønskes kvantificeret.

I forbindelse med Rating- og Risikomodellerne integreres de bearbejdede resultater fra spørgeskemabesvarelserne med en statistisk model for at forbedre prædiktionsnøjagtigheden i forhold til de bedrifter, der risikerer at gå konkurs eller på anden vis misligholde deres finansielle forpligtelser, samt i en Risikomodel. I begge tilfælde supplerer spørgeskemabesvarelsernes latente variable mere kvantitativt målbare variable, f.eks. markedsdata og regnskabsdata.

2. Fremgangsmåde

Karakteren af en latent variabel kan illustreres med følgende ligning:

$$X = T + E$$

I dette eksempel er *T* den størrelse, som man vil forsøge at finde frem til via et spørgeskema, men i realiteten observeres i stedet for *X,* der er en sum af den korrekte variable *T* samt en målefejl *E.* Målefejlen opstår, fordi respondenten ikke kan besvare spørgsmålet direkte, eller fordi besvarelserne ikke er troværdige. Eksempler på latente størrelser inkluderer selvafrapporterende mål som f.eks. betalingsvillighed, intelligens, motivation, følelser eller politiske holdninger.

For at opnå en så præcis måling for *T* som muligt, stilles en række spørgsmål, der alle har til formål at måle *T*. Besvarelserne behandles ud fra statistiske metoder med henblik på at minimere fejlsignalet *E* og estimere *T*.

For at formulere, analysere og selektere de endelige spørgsmål i en undersøgelse, der skal måle latente variable, anvendes følgende fremgangsmåde:

1. Udarbejdelse af en **bruttoliste af spørgsmål**, som ud fra en faglig betragtning skønnes at være relevante til at belyse den latente variabel. Spørgeskemaet anvendes på en relevant målgruppe.

2. Via metoden **Principal Component Analysis** (PCA) eller **Princals** testes antal dimensioner og sammenhængen mellem de enkelte spørgsmål og evt. kategorier af spørgsmål undersøges.

3. Via **Item Respons Theory** (IRT) analyseres spørgsmålene med henblik på at reducere antallet af spørgsmål samt sikre højt og ikke-redundant informationsindhold over hele respons-spekteret.

4. **Revision af spørgeskema** med reduceret, reformuleret og evt. ny-kategoriseret indhold. Det reviderede spørgeskema anvendes på ny på en relevant målgruppe. Punkt 2-3 gentages, indtil spørgeskemaet er **optimeret** i forhold til antal spørgsmål og informationsværdi.

5. I forhold til dette notats to modeller (Rating- og Risiko-), **integreres** besvarelserne fra de optimerede spørgeskemaer vedrørende latente variable med observérbare variable via Structural Equation Models til estimering af henholdsvis en ”rating-score” og en ”risiko-score”.

På baggrund af denne fremgangsmåde opnås en model, hvor vi kan anvende besvarelserne fra spørgeskemaet til at få et mål for den underliggende latente variabel. De følgende afsnit indeholder en uddybning af hvert af disse punkter.

2.1 Bruttoliste

Indledningsvist opstilles et antal spørgsmål, som empirisk eller fagligt intuitivt vurderes at indeholde information om den latente variabel. Om muligt bør der i den indledende spørgerunde ikke være antalsmæssige eller indholdsmæssige begrænsninger udover rent faglige betragtninger. Den efterfølgende behandling af besvarelserne gør det muligt at udvælge de spørgsmål, der indeholder mest information.

Spørgsmålene kan evt. opdeles i kategorier, der alle er relateret til genstanden for analysen. I Ratingmodellen er det landmandens evne til management. Her blev spørgsmålene indledningsvist opdelt i følgende kategorier: Strategi/forretningsmodel/innovation, ejerfamilie, samarbejdspartnere og virksomheden.

I Risikomodellen er det alle væsentlige forhold, som udgør en økonomisk risiko i bred forstand for en landbrugsbedrift, og her blev spørgsmålene opdelt i følgende kategorier: Markedsrisiko, finansiel risiko, menneskelig risiko, institutionel risiko og produktionsrisiko.

2.2 Test af antal dimensioner

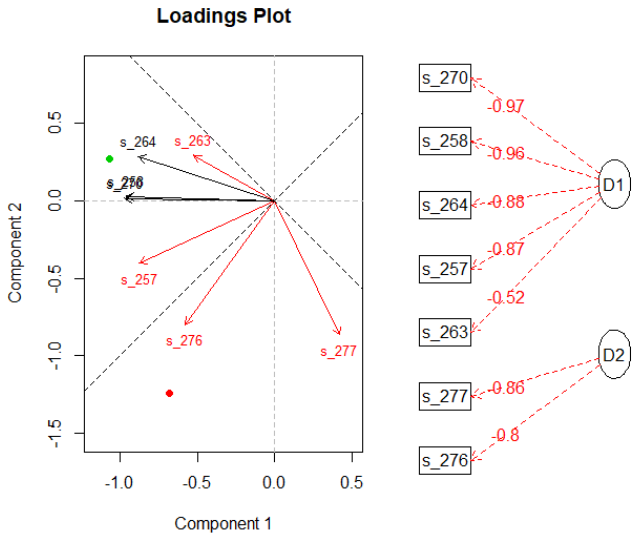
Analysen af spørgeskemabesvarelserne indledes med en undersøgelse af sammenhængen mellem de enkelte spørgsmål både samlet set og for hver kategori ved hjælp af en metode til dimensions-reducering. Det er vigtigt, at hver kategori er éndimensionel, fordi det sikrer, at alle spørgsmål fra en given kategori måler den latente værdi, som vi er interesseret i. I det følgende fokuseres der på to af de mest anvendte, nemlig Principal Components Analysis (PCA) og Princals.

PCA er en kendt teknik, der opdeler datamatricen i delelementer kaldet principalkomponenter under de begrænsninger, at de enkelte principalkomponenter skal være lineært uafhængige, samt at den første principalkomponent skal forklare den største andel af variansen i datasættet. PCA som metode begrænses af, at den kun kan finde lineære sammenhænge, og at spørgeskemabesvarelserne behandles som numeriske data (fx en Lickert skala fra 1-5). Det betyder bl.a., at forskellen mellem ”1” og ”2” er den samme som forskellen mellem besvarelserne ”4” og ”5”.

Som et alternativt til PCA er Princals (Mair, 2018, kap. 8) meget anvendt i behandling af spørgeskemaer, da denne metode kan håndtere forskellige former for nonlinearitet, og spørgeskemabesvarelserne behandles som kategoriske data. Det betyder, at der i vurderingen og tolkningen af værdien af resultaterne ikke kun kan differentieres mellem de forskellige spørgsmål, men også mellem de enkelte svarmuligheder. F.eks. kan svarmulighederne ”2” og ”3” bidrage i samme grad til information om den latente variable, mens svarmulighed ”4” kan bidrage med mere information til estimatet for landmandens evne til management. Se eksempel på test af dimensionalitet i Boks 1.

Boks 1: Test af dimensionalitet

Figuren nedenfor, der består af to diagrammer, er en illustration af, hvordan éndimensionaliteten for hver kategori kan testes. Pilene i figuren til venstre er farvet afhængigt af kategori (sorte pile er f.eks. spørgsmål, der går på anvendelsen af rådgivere, mens røde pile kan være relateret til medarbejdere på bedriften). Da alle besvarelser er i samme enhed (en Lickert-skala fra 1 til 5), angiver længden på pilene variationen i besvarelserne på det enkelte spørgsmål. Denne variation kan måles via egenværdierne i matricen med spørgsmål.



Alle kvadratiske matricer – som f.eks. kovariansmatricen for den tabel med besvarelser af spørgeskemaet, hvor hver kolonne er et spørgsmål og hver række er besvarelse – har både egenværdier λ og egenvektorer *x*, der kan findes ved at løse følgende ligning:

$$Ax = \lambda x$$

hvor *A* er kovariansmatricen, *x* er egenvektorer og λ er egenværdierne. En egenvektor til *A* ændrer ikke retning, når *A* ganges på, men ændrer kun længde afhængig af λ. I figuren svarer hver pil til en egenvektor, der er navngivet efter det spørgsmål, egenvektoren relaterer sig til. Den egenværdi med højest værdi knytter sig til spørgsmålet med mest variation. Derfor kan længden på pilene i figuren siges at være en illustration af, hvor stor en del af variationen i *A* det pågældende spørgsmål kan forklare.

Boks 1 (fortsat)
For at sikre, at hver underkategori i spørgeskemaet er éndimensionelt, er det vigtigt, at pilene peger i samme retning (inden for de stiplede linjer). To pile, der peger i modsat retning af hinanden, måler det samme, men negeret. Hvis det sker, bør besvarelserne transformeres, så en besvarelse på "1" bliver til "5", en besvarelse på "2" bliver til "4". Det sikrer, at høje besvarelser er lig med en høj evne til management.

Figuren til højre viser, hvorvidt hver enkelt spørgsmål korrelerer mest med den første principalkomponent (D1) eller den anden principalkomponent (D2).

2.3 Item Response Theory

Efter at have sikret, at hver kategori er éndimensionel, anvendes IRT (Item Response Theory) til at analysere spørgsmålene inden for hver kategori. IRT er en samlebetegnelse for en række forskellige teorier for, hvordan man kan analysere kategoriske data som f.eks. spørgeskemabesvarelser for at opnå et estimat for en latent værdi. Ved brug af IRT-teknikker er det muligt yderligere at optimere indhold, formuleringer og antallet af spørgsmål i spørgeskemaet ved blandt andet at fjerne spørgsmål uden informationsindhold samt redundante spørgsmål.

Boks 2: Eksempler på IRT-modeller

Den simpleste metode inden for IRT til at analysere éndimensionelle spørgeskemaer med mulighed for polytome besvarelser er Rating Scale Model (RSM). Her modelleres sandsynligheden for, at en respondent *v* svarer *h* på spørgsmål *i*

$$P(X_{vi} = h) = \frac{\exp(h(\theta_v - \beta_i) + \omega_h)}{\sum_{l=0}^{k} \exp(l(\theta_v - \beta_i) + \omega_l)}$$

Tager vi udgangspunkt i Rating-undersøgelsen, angiver θ_v den enkelte respondent *v*'s latente management-evne, β_i angiver spørgsmål *i*'s sværhedsgrad og ω_h er en konstant, der er ens for alle svarmuligheder $h = \{1,2,3,4,5\}$ på tværs af spørgsmål. Nævneren garanterer, at sandsynligheden for alle svarmuligheder *h* summer til 1.

RSM kan ikke anvendes til formålet i nærværende undersøgelser, da det forudsætter, at alle spørgsmål har samme antal svarmuligheder *h*, og at alle skal være anvendt. Derudover behandles alle spørgsmål ens i RSM, dvs. der tages ikke højde for eventuelle forskelle i sværhedsgrad.

En anden metode er den generaliserede Partial Credit Model (PCM), der ikke på samme måde er begrænset af, at alle spørgsmål skal have samme antal svarmuligheder. Dermed kan nogle spørgsmål være af "ja/nej"-typen, mens andre anvender en Lickert-skala.

PCM anvender en ordinal skala, dvs. at de enkelte svarmuligheder i et spørgsmål kan rangeres. Derudover behandler PCM hvert spørgsmål asymmetrisk ved at estimere en parameter for hver spørgsmål, α_i, der måler sværhedsgraden af hvert spørgsmål. Derfor vil nogle spørgsmål kunne målrettes, så de kan anvendes til at adskille respondenter i den gode (alternativt lave) ende af skalaen for den latente management-evne θ_v.

$$P(X_{vi} = h) = \frac{\exp(\alpha_i(h\theta_v - \beta_{ih}))}{\sum_{l=0}^{k} \exp\big(\alpha_i(l\theta_v - \beta_{ih})\big)}$$

IRT-modeller kan opdeles i tre forskellige grupper afhængig af antallet af mulige besvarelser og antal af underliggende dimensioner, jf. ovenstående afsnit:

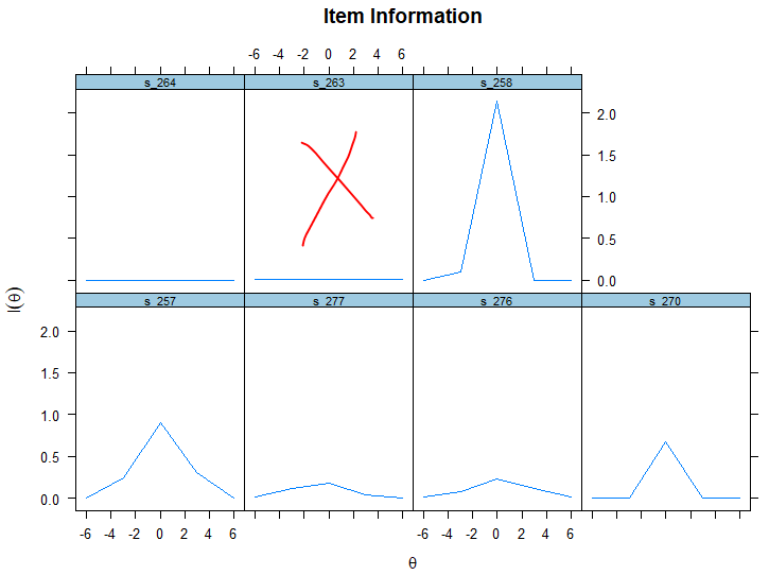
- Éndimensionelle, dikotomiske modeller, dvs. hvor respondenten kun kan svare 0/1 eller ja/nej
- Éndimensionelle, polytome modeller, hvor der er mulighed for at svare på en skala fra f.eks. 1:5
- Flerdimensionelle modeller

PCM-modellen, beskrevet i Boks 2, opfylder ovennævnte kriterier, og denne modeltype er anvendt i nærværende sammenhæng.

I nedenstående figur vises informationsindholdet af hvert enkelt spørgsmål i en kategori på baggrund af en PCM. Y-aksen angiver informationsindholdet af det enkelte spørgsmål. X-aksen angiver den latente evne til management.

Spørgsmål med højtliggende punkter/kurver bidrager med information om management. Helt eller tilnærmelsesvist flade kurver, der ligger lavt, indikerer spørgsmål, der ikke bidrager med information om management.

Hvis toppunktet på en kurve er til venstre (højre) i diagrammet betyder det, at spørgsmålet er godt til at skelne mellem bedrifter med lav (høj) evne til management. Set henover alle spørgsmål skal så bredt en del af spektret som muligt helst dækkes.

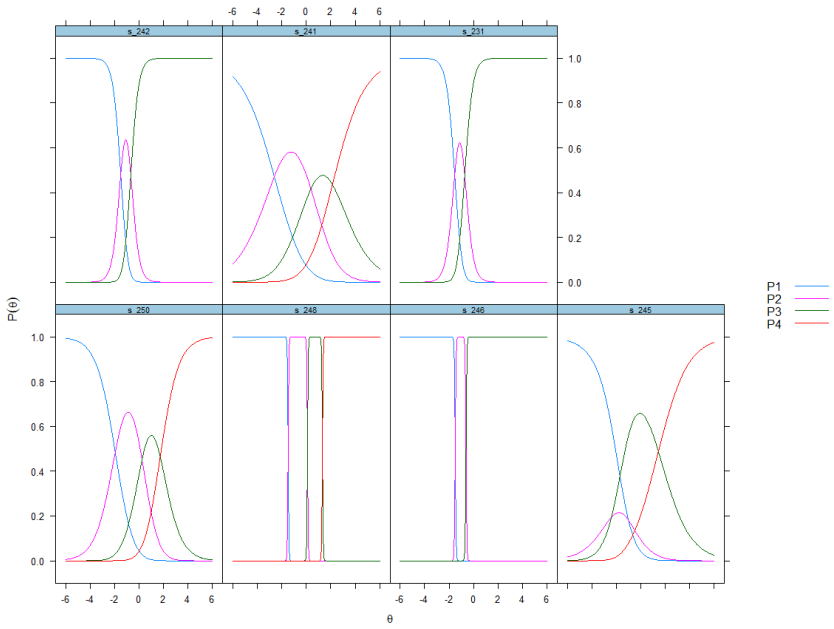


I ovenstående figur er den kurve, der repræsenterer informationsindholdet for spørgsmål 263 (markeret med et rødt kryds) helt flad, og dermed kan dette spørgsmål fjernes. Hver gang ét spørgsmål fjernes, gentages analysen med de resterende spørgsmål, da det kan påvirke informationsindholdet i andre spørgsmål. Til at sammenligne to modeller sammenlignes relevante metrikker[1] for den samlede modelperformance for at tjekke, at modelperformance ikke falder signifikant, når et spørgsmål fjernes.

[1] Det kan eksempelvis være Root Mean Square Error of Approximation (RMSEA), Akaike's Information Criteria (AIC) eller Schwartz Information Criteria (SIC – også kendt som Bayesian Information Criteria (BIC))

Det er også relevant at betragte de såkaldte *item-category characteristic curves* eller *item response curves*, der viser sandsynligheden (målt på y-aksen) for, hvad en respondent svarer, i forhold til den latente evne θ_v (x-aksen). Der fremkommer en kurve for hver anvendt svarmulighed *h*. Dermed fremgår det, at alle fem svarmuligheder ikke er blevet anvendt i nogle af de syv spørgsmål i figuren nedenfor.

I figuren øverst til venstre er der en meget høj sandsynlighed for, at en respondent har en latent evne $\theta_v < -2$ ved en besvarelse af spørgsmål 1. Omvendt er der en ligeså stor sandsynlighed for, at respondenten har en $\theta_v > 0{,}5$ ved en besvarelse af spørgsmål 3. Ved en besvarelse på 2 er der højest sandsynlighed for en latent evne omkring 0. Dermed er dette ene spørgsmål bedst til at skelne mellem respondenter med lave hhv. høje latente evner.



I figuren nederst til højre er der anvendt fire svarmuligheder. Den lilla kurve har aldrig højst sandsynlighed for en given værdi af θ_v. Det betyder, at I det tilfælde vil man kunne nøjes med færre svarmuligheder, da P2 ikke giver ekstra information til IRT-modellen.

Af figuren fremgår det også, at de to spørgsmål nederst til venstre hhv. i midten øverst ligner hinanden meget. Det betyder, at der er et stort sammenfald mellem respondenternes besvarelser på disse spørgsmål, og derfor er det ene spørgsmål tilnærmelsesvis redundant. Under en optimering af spørgeskemaet kan det derfor overvejes at fjerne spørgsmålet med lavest informationsindhold.

Boks 3: Test af IRT-modellen

To eller flere forskellige IRT-modeller – det kunne f.eks. være en model med alle syv spørgsmål fra figuren ovenover, og en uden spørgsmålet nederst til venstre – kan sammenlignes via log-likelihood værdien *L*. Det er et udtryk for, hvor godt den statistiske model passer på de observerede dataværdier.

En svaghed ved denne metrik er, at den ikke tager højde for modellens kompleksitet. To ofte anvendte alternativer hertil er AIC og BIC, der hver især tilføjer en "straf" for modelkompleksitet målt ved, hvor mange parametre, der svarer til $2k$ hhv. $\log(n) \cdot k$, hvor *k* er antallet af parametre og *n* er antallet af observationer:

$$AIC = -2\log L + 2 \cdot k$$
$$BIC = -2\log L + \log n \cdot k$$

Generelt straffer BIC modelkompleksitet højere end AIC, da $\log n > 2$ for alle heltal $n \geq 8$. For alle tre mål gælder det, at jo lavere værdi desto bedre model.

Normalt udtrykkes ønske om at "maksimere likelihood funktionen", da observerede værdier, der passer godt til sandsynlighedstæthedsfunktionen, vil give en høj værdi af L. I praksis vil man i stedet "*minimere* log-likelihood funktionen", da dette giver bedre numerisk stabilitet. Det er også denne værdi, der rapporteres i mange statistikværktøjer som f.eks. R.

Det er også muligt at anvende disse testværdier til at undersøge antagelsen om éndimensionalitet ved at sammenligne en IRT-model, der er estimeret som en éndimensionel model, med en IRT-model, der er estimeret under antagelse om to underliggende dimensioner.

2.4 Revision og optimering af spørgeskema

Efter at have analyseret spørgeskemabesvarelserne ved hjælp af statistiske metoder som beskrevet ovenfor, kan processen gentages efter behov med det reviderede spørgeskema. Det skyldes, at analysen formentligt vil afsløre, at nogle af spørgsmålene fra det oprindelige spørgeskema måske kan udelades, mens andre med fordel kan omformuleres eller tilføjes, hvis det synliggøres, at der mangler spørgsmål i en eller flere kategorier, der kan anvendes til at skelne mellem respondenter med enten en lav eller høj θ_v.

En anden mulig årsag til, at det kan være nødvendigt at indsamle et nyt sæt besvarelser på et spørgeskema er, hvis der er spørgsmål, hvor besvarelserne ikke indeholder variation, og som derfor ikke kan bidrage til et estimat for θ_v. Det kan ske i de tilfælde, hvor alle respondenter stort set har svaret det samme.

Processen beskrevet i afsnit 2.2 og 2.3 er iterativ, så når besvarelserne på det reviderede spørgeskema er modtaget, gentages analysen, indtil spørgeskemaet er tilstrækkelig optimeret i forhold til indhold, formulering, informationsværdi og antal spørgsmål.

I forhold til de spørgeskemaer, der er udviklet i forbindelse med Rating- og Risikomodellerne, er der endvidere foretaget "back test", hvor der er testet på en gruppe af virksomheder, som har misligholdt deres økonomiske forpligtelser i forhold til en kontrolgruppe af økonomisk velkørende virksomheder.

Denne test er blevet anvendt i forhold til den endelige finjustering af spørgeskemaerne samt vægtning af de enkelte spørgsmål til brug i den endelige fastlæggelse af de enkelte spørgsmåls relative betydning for hhv. risiko og rating.

2.5 Integration

I ovenstående afsnit er det beskrevet, hvordan der anvendes en IRT-analyse til at opnå estimater for hver af de underliggende spørgsmål samt for hver kategori, der er defineret i spørgeskemaet, og som kan have selvstændig interesse. Som et resultat af IRT-analysen fremkommer et estimat for eksempelvis respondentens management-evne (den latente værdi) på hvert af spørgsmålene og for hver af de kategorier, som spørgeskemaet er delt op i.

Undertiden kan det være hensigtsmæssigt at samle sådanne estimater for flere kategorier i én samlet score eller værdi. Scoren/værdien kan i princippet antage værdier i intervallet fra minus uendelig til (plus) uendelig. Da den latente værdi pr. definition ikke er målbar, kan de velkendte regressionsmodeller ikke anvendes til at finde de vægte, der kan anvendes til at kombinere de latente variable fra IRT-analysen til én samlet score.

I vægtningen af de enkelte spørgsmål er der derfor taget højde for IRT-analysens resultater i forhold til spørgsmålenes informationsværdi.

Som ovenfor nævnt er der dog også anvendt resultater fra ”back tests” i vægtningen af de enkelte spørgsmål, i forhold til spørgsmålenes evne til at differentiere mellem misligholdende og ikke-misligholdende bedrifter. og endelig er der suppleret med fagligt baserede skøn, for eksempel i tilfælde, hvor der til spørgsmålet ikke forud kan kobles empirisk baserede ”korrekte” svar.

For at kunne integrere resultaterne fra spørgeskemaerne, er der i svarmulighederne anvendt en skala, der scorer svarene efter, i hvor høj grad de pågældende besvarelser er gældende. Scoren kalibreres til en fælles skala for de moduler, der indgår i de samlede Risiko- og Ratingmodeller, jf. notater om Rating- og Risikomodellerne. Resultaterne fra spørgeskemaerne indgår med en fastsat vægtning i forhold til de øvrige elementer i de samlede modeller. På denne måde foregår der en kvantificering af kvalitative data, så der kan foretages en direkte integration i Risiko- og Ratingmodellerne med sammenlignelige enheder.

3. Referencer

Patrick Mair, 2018, *Modern Psychometrics with R*, Springer International Publishing AG, ISBN 9783319931753

Yves Rossel, 2019, *The lavaan tutorial*, accessed via: http://lavaan.ugent.be/tutorial/tutorial.pdf (20.08.2019)