

Opmerkingen op het conceptrapport van de AFM

"Toetsingsuitkomsten bij niet-OOB-accountantsorganisaties"

De opmerkingen op het rapport '*Toetsingsuitkomsten bij niet-OOB-accountantsorganisaties*' kunnen worden onderverdeeld in twee gedeelten, namelijk 'ontbrekende informatie' en 'methodologische en statistische tekortkomingen'.

1. Ontbrekende informatie

In het rapport ontbreekt belangrijke informatie om een goed beeld te krijgen van de situatie met betrekking tot niet-OOB-accountantsorganisaties en de uitkomst van de toetsing. Zeven punten kunnen worden onderscheiden.

1.1. Informatie over de tijd

In de analyse zijn de jaren 2014-2020 samengenomen en is er geen verdere informatie gegeven over verschillen tussen jaren of periodes. Omdat de markt in die periode behoorlijk in beweging is geweest waarbij het aantal Wta-vergunningen sterk is verminderd, zou hier op zijn minst informatie over moeten worden gegeven. Uiteindelijk gaat het erom met deze analyse te leren over met name organisaties die nu een wta-vergunning hebben.

1.2. Informatie over hertoetsingen

In de analyse zijn zowel eerste toetsingen als hertoetsingen meegenomen. Informatie over het aantal accountantsorganisaties met zowel een eerste toetsing als een hertoetsing ontbreekt. Het aantal hertoetsingen is van belang voor de interpretatie van de uitkomsten (zie ook par. 2.1).

1.3. Informatie over steekproef en populatie

Omdat de controles op accountsorganisaties met een Wta-vergunning in een cyclus van 6 jaren plaatsvinden, zijn over de periode 2014-2020 (7 jaren) in principe alle accountantsorganisaties aan de beurt geweest, of zijn er organisaties reeds tweemaal gecontroleerd. Informatie over het totaal aantal organisaties en informatie over het aantal organisaties dat geen Wta-vergunning meer heeft, ontbreekt.

1.4. Informatie over andere kenmerken van accountantsorganisaties

Het doel in dit onderzoek is om mogelijke verbanden te onderzoeken tussen kenmerken van accountantsorganisaties en de uitkomst van toetsingen. Als ongeveer 50 kenmerken beschikbaar zijn, is het te beperkt om over alleen drie van deze kenmerken te rapporteren. Het is vrij eenvoudig is om van alle, of in ieder geval de belangrijkste kenmerken aan te geven in welke mate samenhang bestaat met de toetsingsuitkomst. Dit geeft relevante informatie en sluit beter aan bij het doel van dit onderzoek.

1.5. Informatie van verschillende databronnen

De toetsingsuitkomsten zijn afkomstig van twee verschillende bronnen, namelijk SRA en NBA. Dat betekent dat er verschillen kunnen zijn in het percentage negatieve beoordelingen en het type accountantsorganisaties dat is gecontroleerd. Dergelijke informatie ontbreekt en is vanuit statistisch oogpunt belangrijk omdat deze factor (SRA/NBA) verstrend zou kunnen werken bij het onderzoeken van de relaties tussen toetsingsuitkomst en kenmerken.

1.6. Informatie bij de chi-kwadraattoetsen

De feitelijke tabellen waarop de chi-kwadraattoetsen en de conclusies zijn gebaseerd, ontbreken.

1.7. Informatie over de decision tree

De belangrijkste analyse, de decision tree, is niet volledig weergegeven. In de tak '< 16 K' is het onderscheid naar omzet weggelaten en in de tak '> 16 K' is het onderscheid naar aantal wettelijke controles weggelaten. De reden hiervan is niet toegelicht.

2. Statistische en methodologische tekortkomingen

Gegeven de vraagstelling, de soort data, de afhankelijke variabele, de afhankelijkheid in cases en het aantal en type potentiële factoren, hadden andere keuzes gemaakt moeten worden bij het analyseren van deze data. Vier punten kunnen hierbij genoemd worden.

2.1. Afhankelijkheid in de data

Een belangrijke veronderstelling bij de uitgevoerde analyses is dat de 269 cases onafhankelijk van elkaar zijn. Dat is duidelijk niet het geval, omdat er zowel eerste toetsingen en hertoetsingen van dezelfde organisaties in de 269 cases zijn opgenomen. Omdat de periode 7 jaren beslaat, zouden ook organisaties reeds tweemaal kunnen zijn gecontroleerd.

Er is bij deze analyses dus geen rekening gehouden met afhankelijkheid in de data. Dit is een ernstige en bekende statistische fout. De in het rapport opgenomen argumentatie dat de cases als onafhankelijk kunnen worden beschouwd (p. 23) is onjuist. Zowel de toetsingsuitkomst (accountantsorganisatie reageert op aanvankelijk negatieve beoordeling) als de kenmerken (het is nog steeds dezelfde organisatie) zijn juist afhankelijk van elkaar. De impact van het aantal 'dubbele' cases is niet in te schatten, omdat geen informatie is gegeven over het aantal hertoetsingen en over het aantal organisaties dat in die periode al tweemaal is gecontroleerd, hetgeen in par. 1.2 en 1.3 is benoemd.

2.2. De decision tree als belangrijkste analyse

Het gebruik van de decision tree is om verschillende redenen geen goede keuze:

– Intuïtief is dit reeds te begrijpen omdat het in dit onderzoek niet gaat om het maken van beslissingen op basis van kenmerken, maar zoals de onderzoekers aangeven, om het onderzoeken van mogelijke verbanden. De decision tree is meer

gericht op het maken van voorspellingen, ongeacht hoe die voorspellingen tot stand zijn gekomen.

– In de decision tree wordt de oorspronkelijke kwantitatieve informatie over het aantal wettelijke controles, de omzet van de accountantsorganisatie en de vergoeding, in de minimale variant van twee klassen geforceerd (dichotomiseren). Hoewel er voor presentatiedoeleinden niets op tegen is om te classificeren, is het simpelweg het beste om zoveel mogelijk oorspronkelijk gemeten variabelen in de analyse op te nemen om relaties te kunnen blootleggen.

2.3. De factor 'tijd'

Hoewel dit punt reeds in par. 1.1 aan de orde is gekomen, is het in deze paragraaf van belang aan te geven dat het eveneens als een statistische tekortkoming moet worden beschouwd, dat de factor tijd nauwelijks bij de analyse is betrokken. Als data uit verschillende jaren komt en er is beweging in de markt over de jaren heen, dan dient te worden gekeken hoe deze factor bij de analyse kan worden betrokken. Het alleen aangeven dat ook naar de periode 2017-2020 is gekeken is dan onvoldoende.

2.4. Chi-kwadraattoetsen

Het uitvoeren van chi-kwadraattoetsen op de drie geselecteerde kenmerken is geen goede keuze en overbodig omdat er betere manieren zijn om de samenhang tussen een kwantitatief kenmerk en de toetsingsuitkomst te bepalen.

Daarbij komt dat in de decision tree de grenzen die bij de drie genoemde kenmerken zijn gekozen, niet vooraf of inhoudelijk zijn bepaald, maar het gevolg zijn van een zoektocht naar de optimale grens, een grens met maximale samenhang. Het daarna toetsen van die maximale samenhang is dan niet correct. Een niet-significante samenhang van de oorspronkelijke variabele zou op deze manier als significant uit de chi-kwadraattoets naar voren kunnen komen. Wel is het zo dat die kans kleiner is naarmate de p-waarde kleiner is.

Iets anders is nog dat significantie niet de betekenis heeft die er in het rapport aan wordt toegekend. De data betreft namelijk de gehele populatie accountantsorganisaties met een Wta-vergunning en er is geen sprake van een steekproef.

Tot slot: omdat belangrijke informatie ontbreekt over o.a. het aantal dubbeltellingen en de samenhangen van andere kenmerken met de toetsingsuitkomst, en omdat de data niet ter beschikking staat, is het lastig om de invloed van o.a.

- de afhankelijkheid in de data,
 - het meenemen van organisaties die geen vergunning meer hebben en
 - de keuze voor de decision tree
- goed in te schatten.