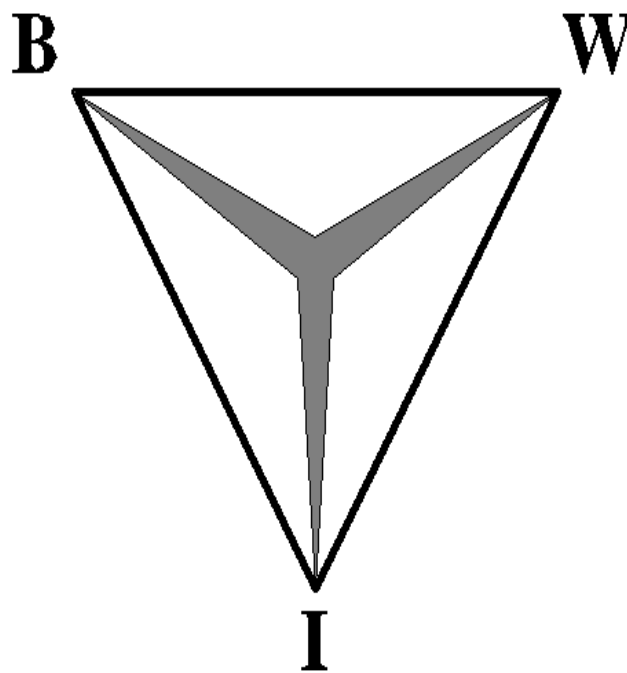


Bayesiaanse Statistiek in de Accountantscontrole

Een Bayesiaans model voor evaluatie van een postensteekproef als onderdeel van de voorraadcontrole.



A.J. Stoffelsen

Vrije Universiteit, Amsterdam

31-08-2005

Bayesiaanse Statistiek in de Accountantscontrole

Een Bayesiaans model voor evaluatie van een postensteekproef als onderdeel van de voorraadcontrole.

Annemieke Stoffelsen

Begeleiding:

Dr. Ir. G. Jongbloed, Vrije Universiteit

Dr. G.J. Franx, Vrije Universiteit

Drs. P.C. van Batenburg, Deloitte.

Drs. T.P. Degenaar, Deloitte.

Deloitte.

Enterprise Risk Services/ Data Quality and Integrity

Laan van Kronenburg 2

1183 AS Amstelveen

Vrije Universiteit

Faculteit der Exacte Wetenschappen

Divisie Wiskunde en Informatica

Studierichting Bedrijfswiskunde en Informatica

De Boelelaan 1081a

1081 HV Amsterdam

Voorwoord

Ter afsluiting van mijn Masteropleiding Bedrijfswiskunde en Informatica aan de Vrije Universiteit heb ik 7 maanden (februari –augustus 2005) stage gelopen bij Deloitte in de groep Data Quality and Integrity (DQI). Tijdens deze stage heb ik onderzoek gedaan naar nieuwe modellen voor steekproefevaluatie bij de voorraadcontrole. Hierbij is voornamelijk aandacht besteed aan een Bayesiaans model.

Naast het werken aan mijn onderzoek is mij door Deloitte de gelegenheid geboden ook andere werkzaamheden die binnen de DQI groep vallen te verrichten. Dit heb ik als erg leerzaam ervaren. Graag wil ik Deloitte en met name de DQI groep bedanken voor deze mogelijkheid en voor de hulp die mij hierbij geboden is. In het bijzonder wil ik Paul van Batenburg en Tjeerd Degenaar bedanken voor het begeleiden van mijn afstudeerproject.

Mijn begeleider vanuit de universiteit, Geurt Jongbloed, wil ik bedanken voor zijn enthousiasme en alle hulp die hij mij geboden heeft tijdens mijn onderzoek. Zijn ideeën en uitleg hebben mij enorm geholpen bij mijn onderzoek.

Annemieke Stoffelsen

Augustus 2005

Samenvatting

In deze scriptie wordt verslag gedaan van het onderzoek naar nieuwe modellen voor de evaluatie van een postensteekproef waarbij voorkennis kan worden meegenomen. Deze modellen zijn met name gericht op de voorraadcontrole, maar kunnen eventueel ook bij andere soortgelijke controlegebieden worden toegepast. Het eerste doel bij het opstellen van een nieuw model is een zo realistisch mogelijk model proberen op te stellen. Daarnaast is er ook de hoop dat er met het nieuwe model een steekproefomvang verkleining kan worden gerealiseerd.

Bij het huidige evaluatiemodel wordt gebruik gemaakt van een regressieschatter. De belangrijkste reden voor Deloitte om deze schatter te gebruiken is dat deze vergeleken met andere schatters zoals de directe schatter en de verschilschatter het kleinste betrouwbaarheidsinterval oplevert. Dat de regressieschatter ook een realistische schatter is, is echter niet erg aannemelijk.

Allereerst is onderzocht of een Bootstrap methode wellicht een realistisch model oplevert. Dit wordt niet onwaarschijnlijk geacht. Vanwege het vele werk dat nodig is om aan te tonen dat de Bootstrap werkt en vanwege het feit dat de eerste resultaten niet duiden op een steekproefomvang verkleining, is besloten dit niet verder uit te werken en eerst op zoek te gaan naar een andere methode.

Een andere soort aanpak die onderzocht is, is een Bayesiaanse aanpak gebaseerd op het huidige model. Bij een Bayesiaans model is het mogelijk voorkennis mee te nemen in de evaluatie. De voorkennis die bestaat dient te worden omschreven in a-priori parameters. Om de betrouwbaarheid van het model te waarborgen is het belangrijk dat deze parameters goed gekozen worden. Bij een correcte Bayesiaanse methode is de uitkomst in de vorm van een kans op een bepaalde hypothese. Bij Deloitte wil men echter (gelijk aan de oude methode) het criterium maximale fout gebruiken dat op basis van het 90% betrouwbaarheidsinterval van de totale fout wordt bepaald. Daarvoor moet dus het 90% betrouwbaarheidsinterval van de totale fout bepaald worden. Omdat tevens het huidige en het nieuwe model goed vergeleken kunnen worden bij het gebruik van het huidige criterium is ervoor gekozen de resultaten voor dit criterium weer te geven. Hierbij dient gezegd te worden dat dit niet leidt tot een formele Bayesiaanse aanpak. Bovendien bestaan er twijfels over de juistheid van de manier waarop het betrouwbaarheidsinterval bepaald wordt. Omdat de resultaten voor deze twee criteria in de

praktijk niet veel van elkaar verschillen voor dit probleem is besloten dat het criterium maximale fout voor dit onderzoek voldoet.

Bij een gelijke steekproefomvang is het betrouwbaarheidsinterval van de totale fout bij het nieuwe model smaller dan bij het huidige model. Dit wijst erop dat een eventuele steekproefomvang mogelijk is. Ook bij verdere analyse van dit model lijkt een verkleining van de steekproefomvang gerealiseerd te kunnen worden. Op basis van de resultaten die getoond zijn, kan gesteld worden dat het nieuwe model met kleinere steekproefomvang minimaal even betrouwbaar en precies is als het huidige. Omdat dit slechts op basis van drie datasets van elk 200 items gebaseerd is, wordt geadviseerd meer tests te doen, voordat dit model in gebruik genomen wordt.

Inhoudsopgave

Voorwoord	3
Samenvatting	5
1 Inleiding	9
1.1 Inleiding Accountantscontrole	9
1.2 Probleemstelling	9
1.3 Indeling scriptie	10
2 Deloitte	11
2.1 Deloitte Touche Tohmatsu	11
2.2 Deloitte Nederland	11
2.3 Diensten	12
2.4 Enterprise Risk Services	13
2.5 Data Quality and Integrity	13
3 Accountantscontrole	15
3.1 Organisatiegerichte Controle	16
3.2 Gegevensgerichte controle	16
4 De voorraadcontrole	22
4.1 Voorwerk	22
4.2 De steekproef	23
4.3 De huidige evaluatie	23
4.4 Probleem opgezet als formeel toetsingsprobleem	31
4.5 Kritiek op huidige model	32
5 Bootstrap	34
5.1 Inleiding Bootstrap	34
5.2 Bootstraptoets	34
5.3 Resultaten	37
6 Bayesiaanse Statistiek	39
6.1 Klassieke Statistiek	39
6.2 Bayesiaanse Statistiek	39
6.3 Klassieke versus Bayesiaanse Statistiek	40
6.4 Bayesiaanse Statistiek en Accountancy	40
7 Bestaande controlemodellen met voorkennis	42
7.1 Het Audit Assurance Model	42
7.2 Bayesiaans Risico-Analyse Model	45
8 Bayesiaans model voor de voorraadcontrole	48

8.1 Inleiding	48
8.2 Modellen	51
8.3 Resultaten	65
9 Conclusie	78
Literatuur	80
Appendix	82

1 Inleiding

1.1 Inleiding Accountantscontrole

Accountantscontrole dient bij te dragen aan de betrouwbaarheid en geloofwaardigheid van een onderneming. Dit is enerzijds noodzakelijk omdat het wettelijk verplicht is, maar anderzijds ook om het vertrouwen van aandeelhouders en klanten te behouden. Het doel van de accountantscontrole is voldoende zekerheid te verkrijgen over de waarheidsgetrouwheid van de financiële verantwoording. Om deze zekerheid te verkrijgen is het niet mogelijk om alles te controleren, daarom worden o.a. vaak steekproeven gedaan. Mede aan de hand daarvan wordt vastgesteld of er voldoende zekerheid is om goed te keuren. Natuurlijk wil men daarvoor het liefst zo min mogelijk werk doen (en dus zo min mogelijk kosten maken). Daarom wordt continu gezocht naar verbeteringen van de huidige modellen, om zo de uitvoeringskosten te verlagen zonder verhoging van het risico op onterechte goedkeuring.

1.2 Probleemstelling

Vaak is van te voren al informatie beschikbaar over wat verwacht kan worden door bijvoorbeeld kennis over het bedrijf, de historie van het bedrijf, soortgelijke bedrijven etc. maar ook door de interne controle die op het bedrijf is uitgevoerd. Hierdoor bestaat al een zeker idee over de uitkomst van de controle. Wanneer vooraf al kennis bestaat over het bedrijf, wil men dat kunnen gebruiken in de controle. Het is immers logisch dat bij de controle (waarbij de kennis wordt vergaard), de bestaande voorkennis ook gebruikt zou willen worden. Deloitte streeft naar modellen die deze voorkennis meenemen in de controle (voor elk controlegebied zoals voorraden, debiteuren, crediteuren etc. kan een verschillend model nodig zijn). De hoop bestaat dat er minder werkzaamheden nodig zijn, dan wanneer er helemaal nog geen kennis beschikbaar is. (Wanneer er al kennis beschikbaar is, hoeft er immers minder kennis te worden vergaard om tot hetzelfde kennisniveau te komen als wanneer er nog geen voorkennis beschikbaar is). Uit deze wens volgt mede het onderwerp van mijn scriptie: een model maken voor een steekproefcontrole, dat dient als onderdeel van de voorraadcontrole (of van een ander controlegebied waarbij de populatie vergelijkbare kenmerken heeft), waarbij voorkennis kan worden meegenomen en daarmee eventueel, uitgaande van de aanname van hoge kwaliteit van de data, de noodzakelijke steekproefomvang kan worden gereduceerd. De voorraad verschilt ten opzichte van de andere controlegebieden in de fouten

die in de administratie verwacht worden. In de meeste posten worden namelijk weinig fouten verwacht (zoals bijvoorbeeld openstaande debiteuren), maar in de voorraad weet men uit ervaring dat er veel fouten verwacht kunnen worden. Het bijzondere aan deze fouten is dat ze elkaar in financiële zin meestal nagenoeg opheffen. Omdat er bij de dit aspect van de controle alleen naar de financiële juistheid gekeken wordt, kan men ondanks de vele fouten in de voorraad daarom vaak toch goedkeuren.

1.3 Indeling scriptie

In deze scriptie zal allereerst het bedrijf Deloitte en de afdeling waar dit onderzoek heeft plaats gevonden behandeld worden. Daarna zal de grootste dienst die Deloitte levert, accountantscontrole, besproken worden. Vervolgens zal ingegaan worden op de huidige methode voor steekproefcontrole bij de voorraad. Daarna zal naar een oplossing gezocht worden voor het bovenstaande probleem. Hierbij zal allereerst gekeken worden naar een Bootstrap methode, vervolgens zal een Bayesiaanse aanpak nader bestudeerd worden. Als laatste zullen de tussentijdse conclusies van mijn bevindingen worden opgesomd en zal een uiteindelijke conclusie van dit onderzoek volgen.

2 Deloitte

2.1 Deloitte Touche Tohmatsu

Deloitte Nederland is een zelfstandige member firm van de wereldwijde organisatie Deloitte Touche Tohmatsu. De oorsprong van deze organisatie ligt in Londen. In 1845 opende William Welch Deloitte, daar zijn eigen accountantskantoor. Na vele overnames en fusies telt Deloitte Touche Tohmatsu inmiddels 150.000 medewerkers in 150 verschillende landen. De inkomsten op jaarbasis bedragen circa 16,4 miljard dollar. Daarmee is Deloitte een van de grootste accountant -en adviesbureaus ter wereld.

2.2 Deloitte Nederland

De geschiedenis van het huidige Deloitte Nederland is, door de vele fusies en overnames door de jaren heen, niet eenvoudig weer te geven. Er volgt nu een korte samenvatting van deze geschiedenis.

Historisch gezien is de Nederlandse Accountants Maatschap (NAM) de belangrijke rechtsvoorganger van de huidige organisatie in Nederland. Dit kantoor werd opgericht in 1955. In de zestiger en zeventiger jaren groeit de NAM in hoog tempo. Er wordt een uitgebreid landelijk netwerk van kantoren opgebouwd. In 1981 wordt ook belastingadvies onderdeel van de snel groeiende organisatie, door een samenwerking met Begheyn & Sneep. Er is dan ook al een kleine organisatieadviestak actief. Medio jaren tachtig verandert de naam in TRN Groep (Touche Ross NAM), om het belang van de internationale verbintenis met het kantoor Touche Ross te verduidelijken.

In 1988 volgt een fusie met De Tombe / Melse Groep (TMG), eveneens een multidisciplinaire organisatie, met circa 700 personeelsleden. De TRN Groep zelf heeft er op dat moment ruim 2000. De Tombe / Melse is ontstaan vanuit een oorspronkelijk in Leiden gestart accountantskantoor, reeds opgericht in 1902.

Als gevolg van een internationaal samengaan, verandert de naam in 1992 in Deloitte & Touche. De organisaties van Touche Ross International en Deloitte Haskins & Sells hebben dan een wereldwijde verbintenis gerealiseerd, met als nieuwe naam Deloitte Touche Tohmatsu International.

De grootste fusie uit de geschiedenis van de Nederlandse organisatie vindt plaats op 1 januari 1998, als Deloitte & Touche fuseert met de VB Groep, een in de overheids- en non-profitmarkt gespecialiseerde accountant- en adviesorganisatie. De VB Groep heeft dan bijna 1400 medewerkers. Door de fusie zijn bij Deloitte & Touche dan 4.800 mensen werkzaam. De jaaromzet groeit vervolgens naar circa f 780 miljoen (omgerekend is dit ongeveer 354 miljoen euro), om daarna binnen twee jaar door te groeien naar boven een miljard guldens. Een nieuwe fusie met de zelfstandige consultancyorganisatie Bakkenist draagt daar ook in belangrijke mate aan bij.

Als Andersen Nederland zich op 1 juni 2002 bij de organisatie voegt, wordt Deloitte & Touche de grootste accountant-adviesorganisatie in Nederland. Andersen brengt ongeveer 1.200 medewerkers met zich mee. Deloitte & Touche is vanaf dit moment naast marktleider in het midden- en kleinbedrijf, de public sector en in de markt voor grote en multinationale cliënten, één van de grootste accountant-adviesorganisaties.

In het najaar van 2003 wordt de merknaam ingekort tot Deloitte, maar het blijft “member of Deloitte Touche Tohmatsu”. De organisatie heeft tegenwoordig ruim 6.500 medewerkers en 50 kantoren door heel Nederland.¹

2.3 Diensten

Als wereldwijd accountants- en adviesbureau biedt Deloitte een zeer breed dienstenpakket. De kerndiensten die geleverd worden, zijn:

- Accountancy
- Belastingadvies
- Consultancy
- Financieel advies

Accountancy is hierbij nog steeds veruit de grootste, ruim de helft van de organisatie is actief in deze tak.

¹ Bron: Website Deloitte Nederland, www.Deloitte.nl

2.4 Enterprise Risk Services

Enterprise Risk Services (ERS) is een afdeling binnen de accountantsdivisie van Deloitte. In deze afdeling zijn alle diensten op het gebied van risicomanagement & control, inclusief de IT-Auditdiensten, gebundeld. De afdeling biedt naast ondersteuning aan de accountants ook diensten aan externe klanten. Vaak heeft ERS een adviserende rol. Binnen ERS kent men vijf competentiegroepen:

- Risk Consulting/Internal Audit (RI/IA)
- Control Assurance (CA)
- Data Quality and Integrity (DQI)
- Security Service Group (SSG)
- Web Services

Elk van deze afdelingen heeft zijn eigen competenties op het gebied van risicomanagement. Veelal wordt samengewerkt door mensen uit de verschillende competentiegroepen om tot een optimaal resultaat te komen.

2.5 Data Quality and Integrity

Data Quality and Integrity (DQI) is de competentiegroep waarbinnen mijn afstudeeropdracht valt. Deze groep bestaat uit econometristen, wiskundigen en statistici. DQI is in eerste instantie opgericht om accountants te ondersteunen tijdens het gegevensgerichte deel van de controlewerkzaamheden. Hierbij kan gedacht worden aan bestandsonderzoek, steekproefondersteuning en regressieanalyse. Er is gebleken dat de door DQI geleverde diensten niet beperkt moeten worden tot het aanbieden van geijkte controle gerelateerde werkzaamheden. Daarom is DQI ook diensten op de externe markt gaan aanbieden. Voorbeelden van externe klanten van DQI zijn:

- ziekenhuizen
- toezichthouders (zoals SUO en CTZ)
- ministeries (zoals VWS)
- commerciële instellingen (zoals verzekeraars)

Mede door de dynamische ontwikkelingen in de markt wordt DQI nu ook betrokken bij onder andere plausibiliteits- en benchmarking analyses.

De expertise van DQI kan opgedeeld worden in twee kerngebieden, namelijk:

1. Bestandsonderzoek:

Dit gebeurt met behulp van speciale software. Over het algemeen bestaat het resultaat van een dergelijk onderzoek uit uitzonderingsrapportages. Soms ligt het vast dat de gerapporteerde elementen fout zijn. Vaak kan voor de gerapporteerde elementen slechts een indicatie van onjuistheid worden weergegeven. Het is aan de klant of deze elementen verder onderzocht gaan worden. Met name bij onderzoeken waarbij preventief naar fraude wordt gespeurd, wordt bestandsonderzoek vaak toegepast.

2. Toepassing van statistische en econometrische technieken:

Hierbij kan gedacht worden aan het ondersteunen van een accountant bij het opzetten, uitvoeren en evalueren van een steekproefcontrole, of het uitvoeren van statistische onderzoeken of regressieanalyse om antwoord te geven op specifieke vragen van klanten.

3 Accountantscontrole

Het doel van de accountantscontrole is voldoende zekerheid verkrijgen over de getrouwheid van de financiële verantwoording van een bedrijf. Dit betekent voor accountants dat men voldoende zekerheid over de juistheid, volledigheid en rechtmatigheid van de boekhouding wil verkrijgen. Met juistheid wordt bedoeld dat alles wat op de balans staat, ook daadwerkelijk aanwezig is. Als op juistheid gecontroleerd wordt, wordt er daarom op overwaardering gecontroleerd. Met volledigheid wordt bedoeld dat alles wat er op de balans hoort te staan, er daadwerkelijk ook staat. Tijdens een volledigheidcontrole wordt op onderwaardering gecontroleerd. Volledigheid kan hierbij nog opgesplitst worden in twee categorieën:

- De individuele volledigheid: een post die op de balans staat mag geen lagere waarde hebben dan de werkelijke waarde.
- De collectieve volledigheid: alles dat waarde heeft in het bedrijf moet op de balans staan. Er mag geen voorraad in het bedrijf zijn die niet op de balans staan.

Als laatste rechtmatigheid, daarmee wordt bedoeld op de controle op naleving van regel- en wetgeving.

Wanneer voldoende zekerheid wordt gevraagd over de getrouwheid van de jaarrekening, betekent dit dat men met een zekere betrouwbaarheid (vaak 95%) wil kunnen zeggen dat de werkelijke waarde niet meer van de balanswaarde afwijkt dan een zekere grenswaarde, deze grenswaarde wordt de materialiteit genoemd. Materialiteit is de waarde die gesteld is als een maximaal acceptabele waarde van afwijking die niet als misleidend gezien wordt. Deze wordt door de accountant bepaald.

Om voldoende zekerheid te verkrijgen, zijn twee soorten controles nodig:

- Organisatiegerichte Controle (AO/IC)
- Gegevensgerichte controle

3.1 Organisatiegerichte Controle

Tijdens de organisatiegerichte controle worden de administratieve organisatie en de interne controle (AO/IC) onderzocht. Er wordt onderzocht in hoeverre deze bestaan, of de werking ervan juist is en in welke mate men erop kan vertrouwen. Hierbij kan gedacht worden aan controles op beveiliging, procedures etc. Wanneer dit onderzocht is, kan bepaald worden in hoeverre tijdens de verdere controle gesteund kan worden op de administratieve organisatie en interne controle (oftewel AO/IC).

3.2 Gegevensgerichte controle

Na de organisatiegerichte controle wordt gegevensgericht gecontroleerd. Gegevensgericht controleren omvat onder andere het uitvoeren van integrale controles, deelwaarnemingen en steekproeven. Bij integrale controle bekijkt men de geheel te controleren populatie. Bij deelwaarnemingen en steekproeven wordt slechts naar een gedeelte van de populatie gekeken om een uitspraak te kunnen doen over het geheel. Het verschil tussen deelwaarnemingen en steekproeven is dat bij deelwaarnemingen de uitspraak voornamelijk op de ervaring van de accountant is gebaseerd terwijl bij steekproeven de uitspraak door middel van statistiek wordt onderbouwd. Dit komt doordat bij steekproeven de items willekeurig geselecteerd worden. Bij deelwaarnemingen is dit niet het geval.

3.2.1 Geldsteekproeven vs Postensteekproeven

Steekproeven worden gebruikt om statistisch onderbouwde conclusies te vormen over de getrouwheid van de financiële verantwoording. Bij het gebruik van steekproefmethodes, wordt er aan de hand van steekproefwaardes bepaald of het aannemelijk is dat de materialiteit niet wordt overschreden. Er zijn twee soorten steekproefmethodes, namelijk geldsteekproeven en postensteekproeven.

3.2.1.1 Geldsteekproeven

Vanwege winstbeschouwingen worden er bij veel controlegebieden voornamelijk overwaarderingen verwacht. Voor deze controlegebieden is het daarom voldoende om alleen op overwaardering te controleren. In feite is dit juist een conservatieve benadering, de onderwaarderingen (als die er al zijn) heffen financieel gezien de overwaarderingen gedeeltelijk op. De kans op een fout groter dan materialiteit is daarom vaak nog kleiner dan bepaald bij de controle op overwaardering alleen. Een controle op juistheid (overwaardering) kan gedaan worden door middel van een geldsteekproef. Het wordt gebruikt bij controlegebieden waarbij weinig tot geen fouten verwacht worden. Er worden

geldeenheden getrokken uit een populatie die alle geldeenheden, bijvoorbeeld euro's, bevat van de voor deze controle relevante posten die op de balans staan. Elke euro heeft dan een gelijke kans om getrokken te worden. Grote posten staan met veel euro's in de boeken en daardoor zitten er meer euro's van een grote post in de populatie dan van een kleine post. Voor elk van de getrokken geldeenheden wordt de bijbehorende post gecontroleerd. Grote posten hebben hierdoor een grotere kans om geselecteerd en daardoor gecontroleerd te worden dan kleine posten. Het kan zelfs zo zijn dat er euro's uit dezelfde post getrokken worden. In dat geval wordt die post meerdere keren meegeteld bij de evaluatie. Vanwege het trekken van geldeenheden hebben ook posten die overgewaardeerd zijn een grotere kans om in de steekproef te komen dan ondergewaardeerde posten van gelijke werkelijke waarde. Hierdoor is het niet mogelijk om zowel op juistheid (overwaardering) als volledigheid (onderwaardering) te controleren en mag er bij geldsteekproeven daarom alleen op juistheid gecontroleerd worden. Onderwaarderingen worden bij deze controle genegeerd.

Meestal maakt men bij de selectie van geldsteekproeven gebruik van 'cell sampling'. Hierbij wordt de gehele populatie euro's opgedeeld in een aantal gelijke intervallen. Uit elk interval wordt dan 1 element getrokken. Grote posten, die ten minste zo groot zijn als 2 intervallen, zullen hierdoor altijd in de steekproef terecht komen. Hierdoor worden de posten met de grootste waarden, deze worden gezamenlijk het top-stratum genoemd, in ieder geval gecontroleerd hetgeen wenselijk is, omdat deze grote risico's kunnen opleveren. Grote posten kunnen namelijk een grotere fout bevatten dan minder grootte posten, vanwege het feit dat de afstand van de boekwaarde tot nul, groter is bij grote posten.

Steekproefgrootte

De vraag is nu hoe groot de steekproef moet zijn om te kunnen goedkeuren. Dit hangt af van de eisen die gesteld worden: wat is de materialiteit (oftewel de geëiste nauwkeurigheid) en wat is de acceptabele onbetrouwbaarheid (α). Interpretatie van deze waardes is: als de materialiteit wordt overschreden, moet dat met minstens $(1 - \alpha)$ kans opgemerkt worden (In de praktijk wordt meestal met $\alpha = 0,05$ gewerkt). De formule hiervoor is:

$$\sum_{k=0}^{k_0} \binom{n}{k} (1-p)^{n-k} p^k \leq \alpha$$

Hierbij geldt:

p : de norm waaraan de fouten moeten voldoen, dit is de fractie materialiteit/totale boekwaarde.

α : onbetrouwbaarheid

n : steekproefgrootte

k_0 : aantal toegestane fouten in de steekproef

k : aantal gevonden fouten

Wanneer nul fouten toegestaan worden, dus wanneer er wordt uitgegaan van $k = 0$, zal n zo gekozen worden dat:

$$\binom{n}{0}(1-p)^{n-0}p^0 = (1-p)^n \leq \alpha$$

$$n \ln(1-p) \leq \ln \alpha$$

$$n \geq \frac{\ln \alpha}{\ln(1-p)}, \text{ voor kleine waarde van } p \text{ geldt:}$$

$$n \geq \frac{\ln \alpha}{\ln(1-p)} \approx \frac{\ln \alpha}{-p}$$

Ook wanneer de binomiale verdeling benaderd wordt met de Poisson verdelingsfunctie zien we dezelfde relatie:

$$(1-p)^n \leq \alpha \rightarrow e^{-np} \leq \alpha \rightarrow$$

$$-np \leq \ln \alpha \rightarrow np \geq -\ln \alpha \rightarrow$$

$$n \geq \frac{-\ln \alpha}{p}$$

Voor de bij de accountantscontrole veel gebruikte waardes $\alpha = 0,05$ en $p = 0,01$ wordt dit:

$$n \geq \frac{\ln \alpha}{-p} = \frac{\ln 0,05}{-0,01} = \frac{-2,995}{-0,01} \approx 300$$

Om fouten te voorkomen en accountants ingewikkelde berekeningen te besparen is in plaats van bovenstaande formules een standaard tabel geïntroduceerd om n te bepalen. Deze is met behulp van de Poisson benadering tot stand gekomen:

$$(1-p)^n \leq \alpha, \text{ voor } \alpha = 0,05 \text{ geldt als } n \uparrow \infty \text{ en } p \downarrow 0 \text{ dan } np \mapsto 3,0$$

$$e^{-np} \leq \alpha, \text{ voor } \alpha = 0,05 \rightarrow np \approx 3,0.$$

De tabel waarde voor $\alpha = 0,05$ en $k = 0$ is daarom 3,0. En de manier waarop n berekend dient te worden is:

$$n = \frac{3,0}{p}$$

Voor de n -en p -waardes die gebruikt worden bij de accountantscontrole blijkt deze benadering goed genoeg. De uitkomst (in dit geval $3,0$) wordt de R -waarde genoemd. Op vergelijkbare manier zijn ook R -waardes voor verwacht aantal fouten groter dan 0 en voor andere waardes van α bepaald. De R -tabel ziet er als volgt uit:

$\alpha \backslash p$	1%	5%	10%	14%	22%	37%	50	61%
$k=0$	4,61	3,00	2,31	2,00	1,50	1,00	0,70	0,50
$k=1$	6,64	4,75	3,89	3,51	2,84	2,14		
$k=2$	8,41	6,30	5,53	4,88	4,09	3,26		
$k=3$	10,05	7,76	6,69	6,19	5,29	4,36		
$k=4$	11,61	9,16	8,00	7,46	6,45	5,44		
$k=5$	13,11	10,52	9,28	8,70	7,60	6,51		

Tabel 1: R-tabel waarin voor verschillende combinaties van k en α de bijbehorende R -waarde wordt weergegeven.

Aan de hand van deze tabel wordt met bovenstaande formule bepaald hoe groot de steekproef moet zijn. Wanneer bijvoorbeeld 1 fout toegelaten wordt in de steekproef, en een onbetrouwbaarheid van 5% genomen wordt, dan geldt $R = 4,75$ en kan, wanneer de materialiteit vastgesteld is, n berekend worden. Als blijkt dat het aantal fouten in de steekproef kleiner dan of gelijk aan k is, kan worden goedgekeurd. Wanneer meer fouten ontdekt worden dan het aantal k dat in eerste instantie gekozen was, kan er niet worden goedgekeurd. Om dit te vermijden, wordt p vaak niet afgeleid uit de materialiteit, maar uit een percentage van de materialiteit, dit wordt controletolerantie genoemd. De p wordt dan iets lager dan de maximaal toegestane waarde waardoor de steekproefgrootte iets groter is dan nodig. Hierdoor kan vaak nog wel goedgekeurd worden wanneer het aantal gevonden fouten meer is dan het verwachte aantal fouten k .

Evaluatie

Wanneer een getrokken euro alleen als wel of niet fout kan worden gewaardeerd, zou er geen onderscheid bestaan tussen een post met boekwaarde €100 en werkelijke waarde €99 en één met werkelijke waarde 0. Hier heeft Stringer (1963) een oplossing voor bedacht. In plaats van te bepalen of een euro goed of fout is, wordt bepaald in welke mate de euro fout is. Dit gebeurt door de werkelijke waarde van de post waartoe die euro behoort te bepalen en aan de hand daarvan de

procentuele fout van de post ten opzichte van de boekwaarde te berekenen (dit wordt ook wel de ‘taint’ genoemd). Deze procentuele fout is dan ook de fout van de euro die gecontroleerd dient te worden (zoals eerder gesteld, wordt hierbij alleen de fout van overwaardering bepaald; wanneer het item ondergewaardeerd is, is de fout gelijk aan 0).

Voorbeeld:

1. Een getrokken euro behoort tot een post die in de boeken staat voor €100. De werkelijke waarde van deze post is €120. De procentuele fout is 0 met betrekking tot de steekproef, want de post is ondergewaardeerd. De te controleren euro heeft daarom met betrekking tot de steekproef ook fout 0.
2. Een getrokken euro behoort tot een post die in de boeken staat voor €100. De werkelijke waarde van deze post is €70. De procentuele fout is dan $(100-70)/100=0,30$, de fout van de te controleren euro is daarom ook 0,30.

Het evalueren van deze fout wordt nu iets lastiger. Stringer heeft het volgende bedacht:

Stel dat de controle tolerantie 1% van de totale boekwaarde (TB) is, de onbetrouwbaarheid 5% is en er wordt uitgegaan van $k = 0$. Dus de verwachte R -waarde is 3,0 (zie tabel). Bij de controle (van 300 waarnemingen) is een fout van 0,30 ontdekt, dan is de bijbehorende R -waarde:

$$R(0,30) = 0,70 \cdot R(0) + 0,30 \cdot R(1) = 0,70 \cdot 3,00 + 0,30 \cdot 4,75 = 1 \cdot 3,00 + 0,30 \cdot 1,75 = 3,525$$

De maximale fout wordt dan:

$$MF = \text{controle tolerantie} \cdot \frac{R(k_a)}{R(k_0)} = 0,01 \cdot TB \cdot \frac{3,525}{3,0} = 0,01175 \cdot TB$$

MF : maximale fout

k_a : aantal gevonden fouten

k_0 : aantal verwachte fouten

Wanneer deze maximale fout kleiner is dan de materialiteit kan er worden goedgekeurd.

3.2.1.2 Postensteekproef

Wanneer een controle nodig is op zowel juistheid als volledigheid en het controlegebied bestaat uit posten waarvan de financiële waarde van de meeste posten in een beperkt gebied liggen, kan de postensteekproef worden toegepast. Als er een aantal uitschieters in de populatie zitten kunnen deze

eerst verwijderd worden (en apart gecontroleerd worden), waardoor de overige populatie wel voldoet aan de eisen. Op deze overgebleven populatie kan dan alsnog een postensteekproef worden uitgevoerd.

Bij een populatie zoals bijvoorbeeld een voorraad worden zowel veel positieve als negatieve verschillen verwacht. Deze verschillen heffen elkaar in financiële zin gedeeltelijk (en soms zelfs volledig) op. Het aantal fouten bepalen (zoals bij geldsteekproeven gebeurd) is daarom hierbij geen goede methode. De totale waarde van de posten schatten lijkt hier een betere oplossing. Hiervoor kan de postensteekproef gebruikt worden.

Bij het trekken van een postensteekproef wordt er niet, zoals bij een geldsteekproef het geval is, gekeken naar de waarde van een post. Er wordt niet uit een populatie geldeenheden getrokken maar uit een postenpopulatie. Alle posten hebben hierbij gelijke kans om in de steekproef te komen. Bij een postensteekproef wordt niet getoetst maar juist geschat. Er wordt niet getoetst of het aantal fouten kleiner is dan een bepaalde waarde, maar er wordt een schatting gemaakt van de totale werkelijke waarde. Deze schatting kan op basis van een zelf te kiezen schatter gebeuren. De huidige postensteekproef methode die bij Deloitte gebruikt wordt zal in het volgende hoofdstuk behandeld worden.

3.2.1.3 *Verskil geldsteekproef en postensteekproef*

De twee grootste verschillen tussen geldsteekproeven en postensteekproeven zijn

1. Bij geldsteekproeven hebben de grote posten een grotere kans om gecontroleerd te worden dan kleinere posten. Bij postensteekproeven hebben alle posten gelijke kans op controle.
2. Bij geldsteekproeven kan alleen op overwaardering gecontroleerd worden, bij postensteekproeven kan zowel op overwaardering als op onderwaardering gecontroleerd worden. Mits de boeken volledig zijn. Met andere woorden er kan met een postensteekproef alleen op individuele volledigheid gecontroleerd worden en niet op collectieve volledigheid.

Geldsteekproeven zijn hierdoor geschikt voor controles waar weinig tot geen fouten te verwachten zijn. Postensteekproeven zijn geschikt voor posten waar zowel negatieve als positieve fouten te verwachten zijn (bijvoorbeeld voorraden).

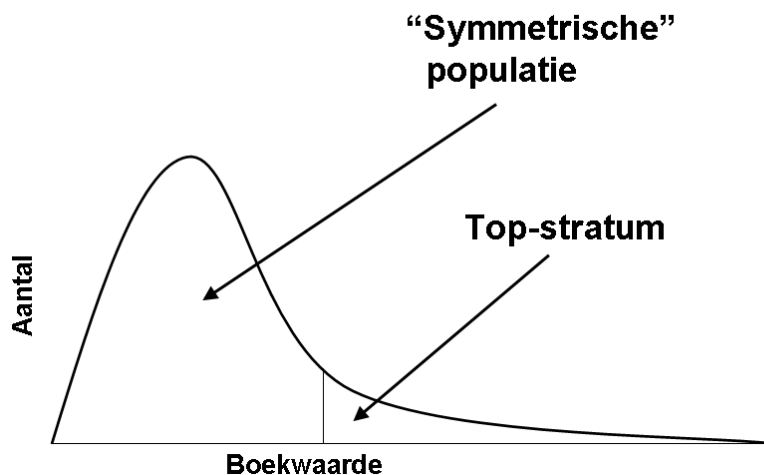
4 De voorraadcontrole

Bij de voorraadcontrole worden (zoals bij de meeste controlegebieden) de organisatiegerichte controle en de gegevensgerichte controle uitgevoerd. Een onderdeel van de gegevensgerichte controle bij voorraad is de steekproefcontrole. Met deze steekproefcontrole op de voorraad wil men voldoende zekerheid verkrijgen over de getrouwheid van de voorraadwaarde op de balans.

Nauwkeuriger geformuleerd, men wil met 95% betrouwbaarheid kunnen stellen dat het verschil tussen de voorraadwaarde in de boeken en de werkelijke voorraadwaarde kleiner is dan de vooraf door de accountant vastgestelde materialiteit. Dit wil men bereiken door willekeurig een aantal posten te controleren en aan de hand daarvan een conclusie te trekken over het wel of niet kunnen goedkeuren. Omdat men bij de voorraad zowel positieve als negatieve fouten verwacht, is het bij deze controle gebruikelijk een postensteekproef te doen.

4.1 Voorwerk

Voordat daadwerkelijk de steekproef getrokken kan worden dient eerst een aantal zaken duidelijk gesteld te worden. Zo moet duidelijk zijn wat de eenheden zijn waarop gecontroleerd dient te worden. Er is nogal een verschil wanneer bijvoorbeeld dozen mokken of losse mokken geteld worden. Om fouten bij het tellen te voorkomen, moet dit vooraf goed duidelijk zijn. Ook dient de populatie waaruit de steekproef getrokken moet worden, bepaald te zijn. Voor de huidige methode is dit de totale populatie zonder de posten met nulwaardes en negatieve waardes en zonder het top-stratum. Nulwaardes en negatieve waardes zijn een indicatie voor onjuistheden, daarom dienen deze waardes apart integraal gecontroleerd te worden. Het top-stratum, dit zijn de posten met de grootste waardes, kan grote verschillen bevatten. Door het top-stratum apart integraal te controleren, hebben deze grote verschillen geen invloed op de schatting van de werkelijke waarde van de andere posten. Het top-stratum bestaat meestal uit zo'n 30 tot 50. Het precieze top-stratum wordt subjectief bepaald. De vuistregel is dat de staart van het histogram van de totale populatie wordt verwijderd, zodat een redelijk symmetrisch histogram overblijft. De populatie die overblijft wordt de steekproefpopulatie genoemd (zie figuur 1).



Figuur 1: Top-stratum

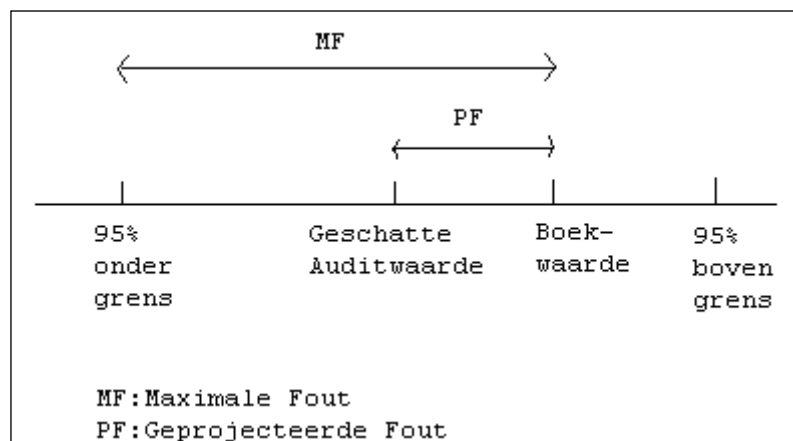
4.2 De steekproef

Nu de steekproefpopulatie bekend is, kan een postensteekproef uitgevoerd worden op deze populatie. Dit is een aselecte steekproef van grootte n (huidige aantal is normaal gesproken 200). Daarnaast dient er ook een aanvullende reserve steekproef van circa 25-50 elementen getrokken te worden. Deze kunnen in uitzonderlijke gevallen gecontroleerd worden in plaats van items uit de originele steekproef. Dit vindt plaats wanneer het originele steekproefelement onmogelijk te controleren valt. Hierbij moet dan wel gekeken worden naar de reden van het niet kunnen controleren van dat item en de bijbehorende consequenties. Voor alle posten uit de originele steekproef moet de werkelijke waarde bepaald worden, dit gebeurt door de eenheden te tellen. Wanneer voor alle posten uit de steekproef zowel de boekwaarde als de werkelijke waarde bekend is kan de evaluatie uitgevoerd worden.

4.3 De huidige evaluatie

Met behulp van de verkregen data, wordt de werkelijke waarde geschat en wordt een 90% betrouwbaarheidsinterval opgesteld rond die schatting. Aan de hand van dit betrouwbaarheidsinterval en de totale boekwaarde van de voorraad wordt een maximale fout bepaald, dit is de maximale afstand tussen de boekwaarde en de waarden die in het betrouwbaarheidsinterval vallen (zie figuur 2). De maximale fout kan geïnterpreteerd worden als met 95% eenzijdige betrouwbaarheid het grootst mogelijke verschil tussen de totale boekwaarde en de totale werkelijke waarde op basis van de steekproefbevindingen. Wanneer deze maximale fout

kleiner is dan de vooraf vastgestelde materialiteit kan er worden goedgekeurd. Hierbij dient opgemerkt te worden dat de boekwaarde hiervoor niet noodzakelijk in het interval van de werkelijke waarde hoeft te liggen, alleen de grote van de maximale fout is hierbij van belang. Pas wanneer kan worden goedgekeurd, kan er van worden uitgegaan dat de meest waarschijnlijke fout het verschil tussen de geschatte werkelijke (audit)waarde en de boekwaarde is, dit wordt de geprojecteerde fout genoemd.



Figuur 2: Aanduiding van de verschillende fouten die er bij Deloitte gehanteerd worden.

Voor de duidelijkheid volgt hier een overzicht met de definities van de verschillende fouten die bij Deloitte gehanteerd worden:

- *Bekende fout*: De totale fout die gevonden is in de steekproef; het totale verschil tussen de boekwaardes en de werkelijke waardes van de posten uit de steekproef.
- *Geprojecteerde fout*: De schatting van de totale fout; het totale verschil tussen de totale boekwaarde van de steekproefpopulatie en de geschatte totale werkelijke waarde van de steekproefpopulatie.
- *Waarschijnlijke fout*: Het verschil tussen de geprojecteerde fout en de bekende fout.
- *Maximale fout*: Het maximale verschil tussen de boekwaarde en de waardes van het betrouwbaarheidsinterval voor de werkelijke waarde. Oftewel: de 95% eenzijdige bovengrens voor de totale fout.

4.3.1 Mogelijke Schatters

De schatting van de werkelijke waarde en daarmee ook het opstellen van het 90% betrouwbaarheidsinterval rond die schatting kan aan de hand van verschillende methodes gedaan worden. De meest gebruikelijke schattingsmethodes hiervoor zijn de lineaire schatters. De belangrijkste 3 zullen daarom nu behandeld worden.

De Directe Schatter voor de gemiddelde werkelijke waarde:

$$\hat{\mu}_{x_d} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$$

Daarbij behorend:

$$(\bar{x}_n - \mu_x) \sqrt{n} \mapsto N(0, s^2)$$

$$s_d^2 = \frac{1}{T} \sum_{i=1}^T (x_i - \mu_x)^2, \quad \hat{s}_d^2 = s_{x_n}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

De Verschilsschatter voor de gemiddelde werkelijke waarde:

$$\hat{\mu}_{x_v} = \bar{x}_n + (\mu_y - \bar{y}_n)$$

Daarbij behorend:

$$(\bar{x}_n + (\mu_y - \bar{y}_n) - \mu_x) \sqrt{n} \mapsto N(0, s_v^2)$$

$$s_v^2 = \frac{1}{T} \sum_{i=1}^T [(X_i - \mu_x) - (Y_i - \mu_y)]^2 = s_x^2 + s_y^2 - 2 \text{cov}(X, Y)$$

$$\hat{s}_v^2 = s_n^2 = s_{x_n}^2 + s_{y_n}^2 - 2 \text{cov}(x, y)$$

De Regressieschatter voor de gemiddelde werkelijke waarde:

$$\hat{u}_{x_r} = \bar{x}_n + \hat{\beta} \cdot (\mu_y - \bar{y}_n)$$

Daarbij behorend:

$$(\bar{x}_n + \hat{\beta} \cdot (\mu_y - \bar{y}_n) - \mu_x) \sqrt{n} \mapsto N(0, s_r^2)$$

$$s_r^2 = \frac{1}{T} \sum_{i=1}^T [(X_i - \mu_x) - \hat{\beta}(Y_i - \mu_y)]^2 = s_x^2 + \hat{\beta}^2 s_y^2 - 2\hat{\beta} \text{cov}(X, Y)$$

$$\hat{s}_r^2 = s_n^2 = s_{x_n}^2 + \hat{\beta}^2 s_{y_n}^2 - 2\hat{\beta} \text{cov}(x, y)$$

Wanneer $\hat{\beta}$ gelijkgesteld wordt aan $\hat{\beta} = \frac{\text{cov}(x, y)}{s_{y_n}^2}$ wordt s_r^2 geminimaliseerd.

- μ_y : de gemiddelde boekwaarde van de totale steekproefpopulatie.
- μ_x : de gemiddelde werkelijke waarde van de totale steekproefpopulatie.
- \bar{x}_n : de gemiddelde werkelijke waarde van de posten in de steekproef.
- \bar{y}_n : de gemiddelde boekwaarde van de posten in de steekproef.
- T : het totaal aantal posten in de steekproefpopulatie.
- n : het aantal posten dat getrokken wordt bij de steekproef.
- $\hat{\beta}$: de geschatte regressiecoëfficiënt van de regressieschatter.
- $X = \{X_1, X_2, \dots, X_T\}$: de T individuele werkelijke waarden van alle items uit de populatie.
- $Y = \{Y_1, Y_2, \dots, Y_T\}$: de T individuele boekwaardes van alle items uit de populatie.
- $x = \{x_1, x_2, \dots, x_n\}$: de n individuele werkelijke waarden van de items uit de steekproef.
- $y = \{y_1, y_2, \dots, y_n\}$: de n individuele boekwaardes van de items uit de steekproef.

Wanneer we de varianties van die drie verschillende schatters vergelijken, kan het volgende worden opgemerkt:

- Als $s_{y_n}^2 - 2 \text{cov}(x, y) \leq 0$, dan is $\hat{s}_v^2 \leq \hat{s}_d^2$. Omdat x en y in de meeste situaties sterk gecorreleerd zijn, zal dit hier het geval zijn. x is immers de werkelijke waarde en y is de bijgehouden voorraadwaarde.
- De regressieschatter met de kleinste variantie is de regressieschatter waarvoor geldt:

$$2\hat{\beta} s_{y_n}^2 - 2 \text{cov}(x, y) = 0 \rightarrow \hat{\beta} = \frac{\text{cov}(x, y)}{s_{y_n}^2}.$$

Omdat de verschilschatter hetzelfde is als de regressieschatter met $\hat{\beta} = 1$, kunnen we stellen dat de variantie voor de regressieschatter met $\hat{\beta} = \frac{\text{cov}(x, y)}{s_{y_n}^2}$ (in bijna alle gevallen) kleiner of gelijk is aan de variantie van de verschilschatter.²

² Vanwege het verschil in vrijheidsgraden (zie eind van deze paragraaf), zal bij $\hat{\beta} \approx 1$ de variantie van de verschilschatter kleiner zijn dan die van de regressieschatter. In de praktijk zal echter de regressieschatter vrijwel altijd een kleinere variantie leveren.

We hebben aangetoond dat in de meeste situaties geldt:

$$s_r^2 \leq \hat{s}_v^2 \leq \hat{s}_d^2$$

Hierdoor is ook het betrouwbaarheidsinterval voor de regressieschatter het kleinst. Omdat volgens Deloitte alle drie de modellen te onderbouwen zijn, geeft zij de voorkeur aan de schatter met het kleinste betrouwbaarheidsinterval. Dit is de regressieschatter.

4.3.2 Regressieschatter en betrouwbaarheidsinterval

In deze paragraaf zal besproken worden hoe de schatting van de werkelijke waarde en het betrouwbaarheidsinterval rond die schatting in de huidige methode tot stand komen.

De algemene formule voor enkelvoudige lineaire regressie is:

$$X_i = \alpha + \beta \cdot y_i + \varepsilon_i$$

De totale werkelijke waarde van de voorraad kan omschreven worden als:

$$X_T = \sum_{i=1}^T X_i = T \cdot \mu_x$$

$$\mu_x = \frac{1}{T} \sum_{i=1}^T X_i = \alpha + \beta \cdot \frac{1}{T} \sum_{i=1}^T y_i + \frac{1}{T} \sum_{i=1}^T \varepsilon_i$$

De OLS (“Ordinary Least Square”) schatters voor α en β zijn:

$$\hat{\beta} = \frac{\text{cov}(x, y)_n}{S_{y_n}^2}$$

$$\hat{\alpha} = \bar{x}_n - \hat{\beta} \bar{y}_n = \bar{x}_n - \frac{\text{cov}(x, y)_n}{S_{y_n}^2} \bar{y}_n$$

De schatting voor de totale werkelijke waarde kan gemaakt worden door:

$$\hat{X}_T = T\hat{\mu}_x$$

$$\hat{\mu}_x = \hat{\alpha} + \hat{\beta} \cdot \mu_y + \bar{\hat{\epsilon}}_T = \hat{\alpha} + \hat{\beta} \cdot \mu_y \quad (\text{want } \sum_{i=1}^T \hat{\epsilon}_i = 0)$$

De schatting van de werkelijke waarde van de voorraad kan nu bepaald worden:

$$\hat{X}_T = T\hat{\mu}_x = T(\hat{\alpha} + \hat{\beta} \cdot \mu_y) = T(\bar{x}_n - \hat{\beta} \cdot \bar{y}_n + \hat{\beta} \cdot \mu_y) = T(\bar{x}_n - \hat{\beta}(\bar{y}_n - \mu_y))$$

Centrale Limietstelling:

De som en het gemiddelde van een (voldoende) groot aantal onafhankelijke en gelijk verdeelde stochastische variabelen volgt bij benadering een normale verdeling. De variabelen zelf behoeven daarvoor geen normale verdeling te hebben.

Nu moet het betrouwbaarheidsinterval nog bepaald worden. Vanwege de Centrale Limietstelling

geldt er $(X_T - \hat{X}_T) = (T(\mu_x - \bar{x}_n + \hat{\beta}(\bar{y}_n - \mu_y))) \sim N(0, S_T^2)$

(S_T^2 zal verderop verder uitgewerkt worden).

Omdat de verwachte voorraadwaarde hier de som van een voldoende groot (>30) aantal variabelen is, kan het 90% betrouwbaarheidsinterval bepaald worden met:

$$[\hat{X}_T - u_{(0,05)} \sqrt{S_T^2}, \hat{X}_T + u_{(0,05)} \sqrt{S_T^2}]$$

Bij Deloitte wordt er echter voor een meer conservatieve benadering gekozen en worden de tabelelementen uit de normale verdeling vervangen door *t-tabel* elementen. Hiermee wordt de kans op onterecht goedkeuren nog iets verkleind.

$$[\hat{X}_T - t_{(0,05;\nu)} \sqrt{S_T^2}, \hat{X}_T + t_{(0,05;\nu)} \sqrt{S_T^2}]$$

Om dit interval te kunnen berekenen is een schatting voor S_T^2 nodig, deze wordt als volgt gemaakt:

$$\begin{aligned} \hat{S}_T^2 &= T^2 \frac{S^2}{n} = T^2 \frac{1}{n(n-1)} \sum_{i=1}^n [(\mu_x - x_i + \hat{\beta}(y_i - \mu_y)) - (\mu_x - \bar{x}_n + \hat{\beta}(\bar{y}_n - \mu_y))]^2 = \\ &= \frac{T^2}{n(n-1)} \sum_{i=1}^n [\bar{x}_n - x_i - \beta(\bar{y}_n - y_i)]^2 = \frac{T^2}{n} [s_{x_n}^2 + \hat{\beta}^2 s_{y_n}^2 - 2\hat{\beta} \text{cov}(x, y)_n] = \\ &= \frac{T^2}{n} s_{x_n}^2 + \frac{T^2}{n} \frac{\text{cov}(x, y)_n^2}{s_{y_n}^2} - 2 \frac{T^2}{n} \frac{\text{cov}(x, y)_n^2}{s_{y_n}^2} = \frac{T^2}{n} s_x^2 - \frac{T^2}{n} \frac{\text{cov}(x, y)_n^2}{s_{y_n}^2} \\ \hat{S}_T &= \frac{T}{\sqrt{n}} \sqrt{\left(s_{x_n}^2 - \frac{\text{cov}(x, y)_n^2}{s_{y_n}^2} \right)} \end{aligned}$$

Uiteindelijk worden in de huidige methode nog een paar kleine aanpassingen gemaakt.

1. Er wordt een eindigheidscorrectie toegepast door de steekproefvariantie te vermenigvuldigen

met $\left(1 - \frac{n}{T}\right)$.

Hierdoor wordt het betrouwbaarheidsinterval verkleind. Dit is toegestaan omdat wanneer de steekproef een groot deel van de totale populatie omvat, de mogelijke fout kleiner wordt.

Merk op dat wanneer de steekproefgrootte een zeer klein percentage van het geheel is, deze factor 1 nadert en daardoor geen invloed heeft.³

2. Het aantal vrijheidsgraden wordt teruggebracht van $n-1$ naar $n-3$.

Normaal gesproken is het aantal vrijheidsgraden bij de steekproefvariantie $n-1$ omdat het aantal onafhankelijke deviaties $x_i - \bar{x}$ gelijk is aan 1. De laatste deviatie kan namelijk worden afgeleid uit de $n-1$ vorige. De vraag is nu wat het aantal onafhankelijke deviaties van de schatting van de variantie bij deze methode is. Vaak wordt er bij lineaire regressie gekozen voor $n-2$ vrijheidsgraden. Omdat bij deze methode nog 2 parameters (α en β) in plaats van 1 geschat dienen te worden (de regel is dat er voor elke te schatten parameter een vrijheidsgraad ingeleverd dient te worden) en om er zeker van te zijn dat de benadering conservatief genoeg is, heeft Deloitte ervoor gekozen met $n-3$ vrijheidsgraden te werken. Hierdoor wordt de *t-waarde* namelijk groter, daarmee wordt ook het betrouwbaarheidsinterval vergroot waardoor de benadering conservatiever wordt.

De variantie verandert hierdoor in:

³ Cochran W.G., 1977, *Sampling Techniques*, Wiley.

$$\hat{S}_T^2 = \frac{T^2(1-n/T)}{n(n-3)} \sum_{i=1}^n [\bar{x}_n - x_i - \beta(\bar{y}_n - y_i)]^2$$

Het *t*-tabel element wordt $t_{(0,05;n-3)}$. Het betrouwbaarheidsinterval kan nu bepaald worden door:

$$[\hat{X}_T - t_{(0,05;n-3)} \sqrt{\hat{S}_T^2}, \hat{X}_T + t_{(0,05;n-3)} \sqrt{\hat{S}_T^2}]$$

4.3.3 Conclusie

Nu het betrouwbaarheidsinterval bekend is, kan de maximale fout van de posten in de steekproefpopulatie berekend worden:

$$MF = \max \left\{ \left| Y_T - \left(\hat{X}_T - t_{(0,05;n-3)} \sqrt{\hat{S}_T^2} \right) \right|, \left| Y_T - \left(\hat{X}_T + t_{(0,05;n-3)} \sqrt{\hat{S}_T^2} \right) \right| \right\}$$

Wanneer deze maximale fout kleiner is dan de vooraf vastgestelde materialiteit, mag er worden goedgekeurd en mag de geprojecteerde fout $Y_T - \hat{X}_T$ geïnterpreteerd worden als de meest waarschijnlijke fout. Deze kan eventueel nog iets gereduceerd worden door bekende fouten te verbeteren.

De steekproefgrootte (*n*) wordt in de huidige methode rond de 200 gesteld. Dit aantal is gebaseerd op onderzoek van Kriens, Kleijnen, Timmermans en van de Wildenberg⁴. Één van de belangrijkste redenen voor deze steekproefgrootte is dat men wil voorkomen dat nul fouten worden gevonden. Wanneer nul fouten worden gevonden, bestaat het betrouwbaarheidsinterval namelijk uit maar 1 punt. Omdat er geschat wordt aan de hand van een steekproef, kan er helemaal geen 100% zekerheid bestaan over de totale fout. Dit probleem wil men daarom voorkomen.

⁴ Kleijnen J.P.C., Kriens J., Timmermans H.en Van den Wildenberg H. 1988, *Regression sampling in statistical auditing*, FEW 306, Universiteit van Tilburg.

4.4 Probleem opgezet als formeel toetsingsprobleem

In de vorige paragraaf wordt uitgelegd hoe men bij Deloitte tot het huidige model gekomen is. In deze paragraaf wordt het beslissingsprobleem uit de voorraadcontrole als een formele toets opgezet, zodat er meer inzicht in het probleem kan worden verkregen en daardoor beter naar een oplossing kan worden toegewerkt.

De vraag waar antwoord op gezocht wordt is: “Wijkt de administratieve voorraad niet te veel af van de werkelijke voorraad?” ofwel: “Is het verschil tussen de boekwaarde en auditwaarde kleiner dan de materialiteit?” Hiervoor kunnen we een toets opstellen. Het antwoord op de vraag wordt bepaald met behulp van de maximale fout, deze kunnen we daarom de toetsingsgrootheid noemen.

Probleem opgezet als toetsingsprobleem:

$$H_0 : |TF| \geq \text{Materialiteit}$$

$$H_1 : |TF| < \text{Materialiteit}$$

TF : totale boekwaarde - totale werkelijke waarde

Toetsingsgrootheid:

$$\begin{aligned} MF &= \max \left\{ |T\hat{F} - t_{\alpha, \nu} \hat{S}_T|, |T\hat{F} + t_{\alpha, \nu} \hat{S}_T| \right\} \\ &= \begin{cases} T\hat{F} + t_{\alpha, \nu} \hat{S}_T & \text{als } T\hat{F} \geq 0 \\ -T\hat{F} + t_{\alpha, \nu} \hat{S}_T & \text{als } T\hat{F} < 0 \end{cases} \\ &= |T\hat{F}| + t_{\alpha, \nu} \hat{S}_T \end{aligned}$$

MF : De maximale fout is de grootste fout waarbij de gevonden steekproef met kans groter of gelijk aan α gevonden wordt.

$T\hat{F}$: De geschatte totale fout.

\hat{S}_T : Een schatting voor de standaard deviatie van $T\hat{F}$

$t_{\alpha, \nu}$: Tabelelement voor het α kwantiel met ν vrijheidsgraden van de verdeling van $T\hat{F}$ (in de huidige methode wordt hier de *t-tabel* voor gebruikt, dit zou ook een ander soort kunnen zijn).

H_0 dient verworpen te worden als $MF < \text{Materialiteit}$.

De eis bij het verwerpen van H_0 moet zijn dat als H_0 waar is, de kans op verwerpen en dus het maken van een fout van de eerste soort, kleiner dan α is. In accountantsbewoording betekent dat, dat de kans op onterecht goedkeuren kleiner moet zijn dan α :

$$\begin{aligned}
 P_{H_0}(MF < \text{Materialiteit}) \leq \alpha &\rightarrow \sup_{|TF| \geq \text{Mat.}} P_{|TF|}(MF < \text{Materialiteit}) \leq \alpha \rightarrow \\
 P_{|TF|=\text{Mat.}}(MF < \text{Materialiteit}) \leq \alpha &\rightarrow P_{|TF|=\text{Mat.}}(|TF\hat{F}| + t_{\alpha, \nu} \hat{S}_T < \text{Materialiteit}) \leq \alpha \rightarrow \\
 P_{|TF|=\text{Mat.}}(|TF\hat{F}| - \text{Materialiteit} < -t_{\alpha, \nu} \hat{S}_T) \leq \alpha &\rightarrow P_{|TF|=\text{Mat.}}\left(\frac{|TF\hat{F}| - \text{Materialiteit}}{\hat{S}_T} < -t_{\alpha, \nu}\right) \leq \alpha
 \end{aligned}$$

Dus verwerp H_0 voor $P_{|TF|=\text{Mat.}}\left(\frac{|TF\hat{F}| - \text{Materialiteit}}{\hat{S}_T} < -t_{\alpha, \nu}\right) \leq \alpha$.

Het probleem dat tot nu toe geformuleerd is, is in algemene vorm, en kan met verschillende schattingsmethodes opgelost worden. De huidige methode die gebruikt wordt is de OLS-regressieschatter, zoals in de vorige paragraaf behandeld is.

$$\begin{aligned}
 -TF\hat{F} &= T(\mu_y - \hat{\mu}_x) = T(\mu_y - (\bar{x}_n - \hat{\beta}(\bar{y}_n + \mu_y))) \\
 -\hat{S}_T^2 &= \frac{T^2(1 - n/T)}{n(n-3)} \sum_{i=1}^n [\bar{x}_n - x_i - \hat{\beta}(\bar{y}_n - y_i)]^2 \\
 -t_{\alpha, \nu} &\text{ is een element uit de } t\text{-tabel.}
 \end{aligned}$$

Of met deze methode aan de eis dat de fout van de eerste soort kleiner dan α moet zijn voldaan wordt, is niet duidelijk. Daarvoor dient de verdeling van $\frac{TF\hat{F} - \text{Mat}}{\hat{S}_T}$ onder de conditie $|TF| = \text{Mat}$ bekend te zijn.

4.5 Kritiek op huidige model

4.5.1 Betrouwbaarheid

Zoals aan het eind van de vorige paragraaf is aangegeven is het niet duidelijk of de kans op een fout van de eerste soort, dit is de kans op het onterecht goedkeuren van de massa, kleiner is dan α . Dit is wel hetgeen wat men zou moeten kunnen aantonen, om te kunnen goedkeuren.

4.5.2 Steekproefgrootte

Over de hoeveelheid te controleren posten bestaan verschillende meningen. De steekproefgrootte van het huidige model is gebaseerd op het onderzoek van Kriens en Kleijnen⁴. Daarin wordt met een simulatiestudie bewezen dat de steekproefgrootte minimaal 200 items moet zijn om een voldoende aantal fouten te hebben gevonden. Accountants vinden deze hoeveelheid nogal groot. Dit komt met name omdat onder bepaalde omstandigheden bij een geldsteekproef met veel minder items (zie hoofdstuk 7) kan worden voldaan.

4.5.3 Normaliteits

In dit model wordt aangenomen dat de afwijkingen (ε_i) normaal verdeeld zijn. Dit is bij de voorraden niet het geval. De reden hiervoor is dat veel afwijkingen rond de waarde nul liggen. Als argument om de OLS-schatter toch te kunnen gebruiken wordt aangedragen dat er bij de werkelijk verdeling van de afwijkingen minder extreme waardes voorkomen dan bij de normale verdeling en dat de OLS-schatter daarom een conservatieve benadering geeft (zie Geertman⁵). Daarnaast is er door T. Lew⁶ onderzoek gedaan naar de voorwaarde waaraan een populatie dient te voldoen om de toepasbaarheid van de regressieschatter aannemelijke te kunnen maken:

1. De steekproef dient minimaal uit 180 – 200 elementen te bestaan.
2. Er moeten fouten aanwezig zijn in de steekproef. Statistisch gezien maakt het niet uit als het alléén positieve of alléén negatieve fouten zijn.

Hieraan voldoen de meeste steekproeven in de voorraadcontrole wel.

⁵ Geertman F. 1994, *Over de toepasbaarheid van regressieschatter en verschilschatter in de accountantscontrole*, Intern Document Deloitte.

⁶ Lew T. 1998, *Regressieschatter, literatuurstudie naar de praktijkaanname voor toepassen van normaliteit*, Vrije Universiteit.

5 Bootstrap

5.1 Inleiding Bootstrap

Bootstrap is een methode die gebruikt kan worden om een onbekende grootte te benaderen (in dit geval de totale fout). In de huidige aanpak wordt dit op basis van normaliteitsaanname bepaald. Omdat er nogal wat kritiek bestaat over de aanname dat de afwijkingen (ε_i) normaal verdeeld zijn, is het goed om Bootstrap nader te bekijken. Bij de Bootstrap is namelijk geen verdelingsaanname nodig. Het enige dat bij Bootstrap dient te worden aangenomen is dat de steekproef representatief is voor de gehele populatie. Wanneer dit een goede methode lijkt kunnen we wellicht een Bootstrap methode met voorkennis implementeren. Zonder moderne computers zou de Bootstrap praktisch niet bruikbaar zijn. In plaats van het theoretisch afleiden van een kansverdeling wordt bij de Bootstrap namelijk gebruik gemaakt van computersimulaties. Het algemene idee bij Bootstrap is om aan de hand van de getrokken steekproef, steekproeven te simuleren. Omdat de steekproef waaruit getrokken wordt representatief is voor de gehele populatie (volgens bovenstaande aanname), kunnen we aan de hand van de B gesimuleerde steekproeven een uitspraak doen over de gehele populatie. Voordat een Bootstrap methode gebruikt kan worden moet nog wel aangetoond worden dat de Bootstrap werkt. Helaas is dit vaak niet zo eenvoudig.

5.2 Bootstrap toets

Aan de hand van het in paragraaf 4.4 geformuleerde toetsingsprobleem wordt nu een Bootstrap toets opgesteld. Hierbij wordt gebruik gemaakt van een toetsingsgrootte en wordt met behulp van simulatie een overschrijdingskans bepaald.

Zoals in hoofdstuk 5 is laten zien kunnen we het probleem als volgt opzetten:

$$H_0 : |TF| \geq \text{Materialiteit}$$

$$H_1 : |TF| < \text{Materialiteit}$$

Verwerp H_0 voor $P_{|TF|=Mat.} (T_n < -t) \leq \alpha$, waarbij $T_n = \frac{|T\hat{F}| - \text{Materialiteit}}{\hat{S}_T}$.

$-t$ kan worden afgeleid uit de verdeling van de toetsingsgrootte T_n .

De individuele fouten kunnen gemodelleerd worden als $\varepsilon_i = Y_i - X_i$ en deze zijn onafhankelijke identiek verdeeld (iid):

$$\varepsilon_i \stackrel{iid}{\sim} G.$$

De verdeling van T_n is $L_G(T_n)$. Als G bekend zou zijn, dan $L_G(T_n)$ te benaderen met Monte Carlo simulatie. Dit is niet het geval, daarom moet eerst een schatter voor G bepaald worden.

$$H_0 : \sum_1^T \varepsilon_i = Mat \text{ of } \sum_1^T \varepsilon_i = -Mat, \rightarrow E_G(\varepsilon_i) = \frac{Mat}{T} \text{ (1) of } E_G(\varepsilon_i) = -\frac{Mat}{T} \text{ (2)}$$

Voor beide randsituaties uit H_0 moet aangetoond worden dat ze verworpen kunnen worden. Pas dan kan de totale nulhypothese verworpen worden.

Situatie 1, $\sum_1^T \varepsilon_i = Mat$:

De verdeling G is onbekend en zal daarom geschat moet worden. Dit doen we met een empirische verdelingsfunctie van $\varepsilon_1 \dots \varepsilon_n$

De schatting van G onder de nulhypothese wordt bepaald door een verschuiving van de oorspronkelijke schatting, namelijk:

Schatting van G onder H_0 is de empirische verdelingsfunctie van $\tilde{\varepsilon}_1 \dots \tilde{\varepsilon}_n$ waarbij $\frac{1}{n} \sum_{i=1}^n \tilde{\varepsilon}_i = \frac{mat}{T}$ en

$\{\#i : \tilde{\varepsilon}_i = 0\} = \{\#i : \varepsilon_i = 0\}$. Dit wordt bereikt door:

$$\tilde{\varepsilon}_i = \begin{cases} 0 & \forall \varepsilon_i = 0 \\ \varepsilon_i - \bar{\varepsilon}_n^+ + \frac{mat}{T} \cdot \frac{n}{\#\{i : \varepsilon_i \neq 0\}} & \forall \varepsilon_i \neq 0 \end{cases}$$

Waarbij $\bar{\varepsilon}_n^+$: het gemiddelde van de afwijkingen die ongelijk aan 0 zijn.

Te zien is nu dat:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \tilde{\varepsilon}_i &= \frac{1}{n} \sum_{i:\varepsilon_i=0} \tilde{\varepsilon}_i + \frac{1}{n} \sum_{i:\varepsilon_i \neq 0} \tilde{\varepsilon}_i = 0 + \frac{1}{n} \sum_{i:\varepsilon_i \neq 0} \tilde{\varepsilon}_i = \\ &= \frac{1}{n} \sum_{i:\varepsilon_i \neq 0} \varepsilon_i - \frac{1}{n} \sum_{i:\varepsilon_i \neq 0} \bar{\varepsilon}_n^+ + \frac{1}{n} \sum_{i:\varepsilon_i \neq 0} \frac{mat}{T} \cdot \frac{n}{\#\{i : \varepsilon_i \neq 0\}} = \frac{mat}{T} \end{aligned}$$

Dat is precies de aanname die gemaakt wordt in de nulhypothese.

We hebben nu $\tilde{\varepsilon}_1 \dots \tilde{\varepsilon}_n$. Hiermee hebben we ook een schatting voor G onder H_0 , namelijk de empirische verdelingsfunctie van $\tilde{\varepsilon}_1 \dots \tilde{\varepsilon}_n$.

Nu er een schatter voor G onder H_0 is, kan $L_{G|H_0}(T_n)$ benaderd worden door $L_{\hat{G}}(T_n)$ met behulp van Monte Carlo simulatie:

1. Trek $\varepsilon_1^* \dots \varepsilon_n^*$ uit \hat{G} met teruglegging (\hat{G} is een empirische verdelingsfunctie van $\tilde{\varepsilon}_1 \dots \tilde{\varepsilon}_n$)
2. Bereken $T\hat{F}$ en \hat{S}_T :

$$T\hat{F} = \frac{T}{n} \sum_{i=1}^n \varepsilon_i$$

$$\hat{S}_T = \sqrt{\frac{T^2}{n^2} \cdot n \cdot \text{var}(\varepsilon_i)} \approx \sqrt{\frac{T^2}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}$$

3. Bepaal T_n
4. Herhaal stap 1 t/m 3 B keer. De empirische verdeling van $T_n \hat{L}_{\hat{G}}(T_n)$ is een benadering voor $L_{\hat{G}}(T_n)$ en dus voor $L_G(T_n)$.

De kritieke waarde $-t$ kan nu benaderd worden door het α -kwantiel van $\hat{L}_{\hat{G}}(T_n)$ te bepalen. Nu kan de toetsingsgrootheid T_n uit de oorspronkelijke steekproefelementen $\varepsilon_1 \dots \varepsilon_n$ afgeleid worden:

$$T\hat{F} = \frac{T}{n} \sum_{i=1}^n \varepsilon_i$$

$$\hat{S}_T = \sqrt{\frac{T^2}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}$$

Als T_n kleiner is dan de $-t$ waarde, kan H_0 verworpen worden, omdat de kans op een fout van de eerste soort (dit is de kans op onterecht goedkeuren) dan kleiner α is.

Voor situatie 2, ($\sum_1^T \varepsilon_i = -Mat$) kan nu op soortgelijke manier bepaald worden of er verworpen kan worden. Er moet gekeken worden naar het $1-\alpha$ van $\hat{L}_{\hat{G}}(T_n)$ om de $-t$ te bepalen. Wanneer T_n groter is dan deze waarde kan verworpen worden.

Als situatie 1 verworpen is houdt dat in dat er met voldoende zekerheid kan worden gesteld dat er geen materiele overwaardering is. Als situatie 2 verworpen kan worden, is er voldoende zekerheid over het niet voorkomen van materiele onderwaardering. Als beide situaties verworpen kunnen worden kan er daarom worden goedgekeurd.

5.3 Resultaten

Voor verschillende datasets (uit oorspronkelijke data), is de Bootstrap toets toegepast. Hieronder worden de resultaten hiervan vergeleken met de resultaten van de huidige methode. Met de grensmaterialiteit wordt bedoeld dat wanneer de materialiteit boven deze grens is vastgesteld, kan er worden goedgekeurd. Wanneer de materialiteit kleiner of gelijk aan de grenswaarde is, kan er niet worden goedgekeurd. Vanwege het feit dat de resultaten van de Bootstrap toets simulatie resultaten zijn, kunnen deze resultaten per berekening wat verschillen, vandaar dat een precieze bepaling van de materialiteitsgrens hiervoor niet mogelijk is⁷:

	<i>Grens mat. Huidige methode</i>	<i>Grens mat. Bootstrap toets</i>
<i>Dataset 1</i>	<i>136.671</i>	<i>≈165.000</i>
<i>Dataset 2</i>	<i>1.171.576</i>	<i>≈2.100.000</i>
<i>Dataset 3</i>	<i>102.445</i>	<i>≈171.000</i>

Tabel 2: Materialiteitsgrensen waar vanaf kan worden goedgekeurd.

Zoals te zien is, ligt de grens van materialiteit waarbij mag worden goedgekeurd voor deze Bootstrap toets een stuk hoger dan voor de huidige methode, dit is niet waarop gehoopt werd, gezien de probleemstelling van dit onderzoek. De methode lijkt echter tot zover wel een realistische methode, er zijn namelijk geen aannames gemaakt die niet aannemelijk te maken zijn. De belangrijkste eis voor een bruikbaar model is dat het een realistisch model moet zijn. Dit is het wellicht toch een reden om de Bootstrap methode verder uit te werken. De Bootstrap methode uitwerken is echter niet eenvoudig. Hierbij moet onder andere aangetoond worden dat de Bootstrap werkt, hetgeen geen

⁷ Voor meer uitgebreide resultaten zie appendix.

gemakkelijke opgave is. Vanwege de beperkte tijd die voor dit onderzoek beschikbaar is, is besloten dit niet te doen.

6 Bayesiaanse Statistiek

Statistiek kan opgedeeld worden in twee hoofdrichtingen; klassieke statistiek en Bayesiaanse statistiek. Om een duidelijk beeld te geven van wat Bayesiaanse statistiek inhoudt, zullen beide richtingen en de verschillen er tussen besproken worden.

6.1 Klassieke Statistiek

De klassieke statistiek is ontwikkeld tussen 1930 en 1960, met als grote voorman Ronald Fisher. Deze berust op de frequentistische definitie van een kans. Bij deze vorm van statistiek wordt bepaald of een gevonden uitkomst statistisch wel of niet significant is, oftewel of er statistisch gezien voldoende bewijs is om iets te concluderen aan de hand van de verkregen data. Dit gebeurt aan de hand van een toets. In die toets wordt een hypothese geformuleerd en aan de hand van de verkregen data wordt met een toetsingsgrootte bepaald of er bewijs genoeg is om de hypothese te verwwerpen. Op deze manier hoopt men tot een juiste conclusie te komen. Deze theorie vult vrijwel alle leerboeken en ligt ten grondslag aan 80% van de wetenschappelijke publicaties over statistiek.⁸

6.2 Bayesiaanse Statistiek

De oorsprong van deze tak van statistiek ligt in de 18^{de} eeuw. Dominee Thomas Bayes (1702-1761) was de eerste die volgens deze lijn redeneerde. Zijn artikel hierover werd postuum in 1763 gepubliceerd. Veertien jaar later zette de beroemde wiskundige Laplace dezelfde stappen opnieuw. De basis voor de huidige vorm van de Bayesiaanse Statistiek ligt net als die van de klassieke, in de dertiger jaren van de vorige eeuw, maar pas na 1960 kreeg deze wat meer aandacht. Belangrijke personen voor de ontwikkeling van deze statistiek zijn B. de Finetti, H. Jeffreys, L.J. Savage en D. Lindley.

Bij Bayesiaanse statistiek wordt uitgegaan van een bepaalde geloofswaarde over de hypothese. Er bestaat al een verwachting over het wel of niet waar zijn van een hypothese voordat er ook maar enige data verkregen is. Wanneer de data verkregen is, wordt bepaald hoe dit de geloofswaarde over de hypothese beïnvloedt om vervolgens een uitspraak te doen over de waarschijnlijkheid van een

⁸ Zie Dr. de Vos A.F., *De godsdienstoorlog der statistici*

hypothese. De geloofswaarden over de hypothese voordat er data is verkregen, worden in de Bayesiaanse statistiek de a-priori kansen genoemd. Een belangrijke opmerking die hierbij geplaatst dient te worden is dat de a-priori kansen, die invloed hebben op de uiteindelijke conclusie, niet altijd objectief zijn. Vaak zijn de geloofswaarden over hypothesen niet voor iedereen gelijk. Er dient gewaakt te worden voor het gebruik van onjuiste a-priori kansen, deze kunnen namelijk leiden tot onjuiste a-posteriori kansen, hetgeen onjuiste conclusies kan opleveren. Dit zorgt voor nogal wat kritiek op deze methode door de aanhangers van de klassieke statistiek. De gecombineerde geloofswaarde van zowel de a-priori kansen als de data noemt men de a-posteriori kansen. De regel van Bayes voor de berekening van de a-posteriori kansen is:

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

De a-priori kans is hier $P(A)$ en de a-posteriori kans is $P(A|B)$.

6.3 Klassieke versus Bayesiaanse Statistiek

Het grootste verschil tussen deze twee soorten statistiek is dat de klassieke aanpak een eenduidig verband tussen een model en waarnemingen probeert te vinden terwijl de Bayesiaanse aanpak helemaal nog niet zeker is over wat het juiste model is. Bij de klassieke aanpak wordt de overschrijdingskans van een bepaald onderzoeksresultaat berekend onder voorwaarde dat de gestelde hypothese juist is. Als daartoe reden wordt gezien, wordt de nulhypothese verworpen. Aan de hypothesen zelf worden geen kansen toegekend. In de Bayesiaanse statistiek daarentegen, richt men zich op de waarschijnlijkheid van de hypothese, samen met de verkregen data leiden deze tot de uiteindelijke a-posteriori waarschijnlijkheid. Aart de Vos: *“De klassieke kant gaat uit van het bestaan van een vaste, zij het onbekende, waarheid die zich manifesteert in waarnemingen met een toevalskarakter. De Bayesiaan hangt het kansbegrip op aan het feit dat de mogelijke waarheden onzeker zijn. En hij gebruikt de waarschijnlijkheidsrekening om over deze kansen te leren.”*⁷

6.4 Bayesiaanse Statistiek en Accountancy

In de accountancy wil men een goede massa zo snel mogelijk kunnen goedkeuren, dit betekent met zo min mogelijk werkzaamheden. Natuurlijk moet deze goedkeuring wel te verantwoorden zijn. Wanneer gunstige voorkennis in het controlemodel kan worden meegenomen, zal men hoogstwaarschijnlijk met minder werkzaamheden al kunnen goedkeuren. Bayesiaanse statistiek kan goed gebruikt worden voor het meenemen van voorkennis in de kansverdeling en kan een goede

methode zijn om dit te verwezenlijken. Bij de controle met behulp van een geldsteekproef is al een dergelijk model ontwikkeld (zie hoofdstuk 8), een soortgelijk model zou ook voor de controle met een postensteekproef (deze wordt voornamelijk gebruikt bij voorraden) wenselijk zijn.

7 Bestaande controlemodellen met voorkennis

Deloitte is al een aantal jaren bezig geweest met het ontwikkelen van modellen voor de accountantscontrole waarin voorkennis kan worden meegenomen. Tot nu toe is voornamelijk aandacht besteed aan het ontwikkelen van modellen voor controles waarin weinig fouten verwacht worden en dan met name de controles waarbij gebruik wordt gemaakt van een geldsteekproef. De ontwikkelde modellen zijn daarom ook gebaseerd op een geldsteekproef. In dit hoofdstuk wordt de werking van deze modellen uitgelegd.

7.1 Het Audit Assurance Model

Het Audit Assurance Model of Audit Risk Model (ARM) wordt gehanteerd als uitgangspunt voor het doen van een uitspraak over de foutenfractie in een te onderzoeken populatie. Het maakt gebruik van een geldsteekproef en is bedoeld voor een controle waarbij weinig fouten in de populatie verwacht worden. In dit model wordt de informatie vanuit de AO/IC gecombineerd met de gegevensgerichte controle. Het ARM is gebaseerd op het kwantificeren van informatie. Voor de verschillende controlegebieden wordt aan de hand van bestaande tabellen een risicoprocentage voor het overschrijden van de foutfractie bepaald. Het model stelt dat de totale (on)zekerheid gelijk is aan het product van de afzonderlijke (on)zekerheden.

$$ARM: OR = IR \times ICR \times SR$$

- *OR*: “Overall Risk”, risico op het ten onrechte goedkeuren van een populatie.
- *IR*: “Inherent Risk”, risico dat de populatie voorafgaande aan de AO/IC een materiële fout bevat.
- *ICR*: “Internal Control Risk”, risico dat de aanwezige materiële fout niet ontdekt wordt tijdens de AO/IC.
- *SR*: Steekproefrisico α , risico dat de aanwezige materiële fout met de steekproefcontrole niet ontdekt wordt.

De twee meest belangrijke aannames die in dit model gemaakt worden zijn:

1. De informatie over de verschillende controlegebieden is uit te drukken in een betrouwbaarheidspercentage.
2. De verschillende zekerheden zijn onderling onafhankelijk zodat de betrouwbaarheidspercentages vermenigvuldigd kunnen worden.

Ook in dit model wordt gebruik gemaakt van de R-waardes (zie paragraaf 3.2.1). Wanneer het risico 5% mag zijn, betekent dit dat het totale risico 5% mag zijn. De formule van het ARM wordt dan:

$$0,05 = IR \cdot ICR \cdot SR \rightarrow$$

$$3,0 = \ln(IR) + \ln(ICR) + R(k, \alpha) \rightarrow R(k, \alpha) = 3,0 - \ln(IR) - \ln(ICR)$$

Wanneer dus een kleiner risico bij *IR* en *ICR* kan worden aangetoond, mag het risico van de steekproefuitkomst wat groter zijn en hoeft daarbij minder werk verricht te worden. Omdat het nogal subjectief is om percentages te geven aan bepaalde risico's, zijn er tabellen ontwikkeld waarin aangegeven staat hoeveel nog gecontroleerd dient te worden. Hierin wordt ook het wel of niet aanwezig zijn van het specifieke risico meegenomen. Dit staat voor het risico dat niet onder de bovengenoemde risico's valt, maar dat door onderzoek toch naar boven is gekomen. De cijfers die in de tabel staan zijn de $-\ln(x)$ -waarden, waarbij x het risicosoort is. Bij elkaar opgeteld moeten deze altijd 3,0 zijn, omdat bij accountantscontrole altijd met onbetrouwbaarheid van 5% gewerkt wordt.

-The Audit Assurance Model [Revised March 31, 1998]					
Sources of Assurance	SPECIFIC RISK IDENTIFIED		SPECIFIC RISK NOT IDENTIFIED		
	Do Not Rely on Controls Plan and Perform <i>Focused</i> Substantive Tests	Adopt a Control Reliance Strategy Plan and Perform Tests of Control Activities that Mitigate Specific Identified Risk—Plan and Perform a <i>Basic</i> Level of Substantive Tests	Adopt a Control Reliance Strategy Plan and Perform Tests of Control Activities That Address Relevant Potential Errors—Plan and Perform a <i>Low</i> Level of Substantive Tests	Include Tests of Control Activities in our Rotation Plan, but do not Test Them in Current Year—Plan and Perform a <i>Basic</i> Level of Substantive Tests	Do Not Rely on Controls Plan and Perform an <i>Intermediate</i> Level of Substantive Tests
Inherent Assurance	0.0 <i>none</i>	0.0 <i>none</i>	1.0 <i>maximum</i>	1.0 <i>maximum</i>	1.0 <i>maximum</i>
Control Assurance	0.0 <i>none</i>	2.3 <i>maximum</i>	2.0 <i>moderate</i>	1.3 <i>basic</i>	0.0 <i>none</i>
Substantive Assurance	3.0 <i>focused</i>	0.7 <i>basic</i>	— <i>low</i>	0.7 <i>basic</i>	2.0 <i>intermediate</i>
Overall Assurance	3.0	3.0	3.0	3.0	3.0

Tabel 3: Het AAM-model, hierin staat wanneer en waarmee er bepaalde R -waardes te behalen zijn. Uiteindelijk zal $R = 3,0$ behaald moeten worden om te kunnen goedkeuren.

We zien dat er 4 niveaus van controle zijn bij de steekproefcontrole. Namelijk uitgebreide controle ($R=3,0$), gematigde controle ($R=2,0$), beperkte controle ($R=0,7$) en geen steekproefcontrole ($R \approx 0,0$).

De omschrijving en interpretatie die tot dusver gegeven is van het ARM, is zoals deze bij Deloitte bekend is. Om ook de leken op het gebied van accountantscontrole inzicht te kunnen geven in hoe dit model werkt, volgt nu een meer statistisch onderbouwde uitleg van het model:

- OR : $P(\text{de fout is groter dan materialiteit en er wordt goedgekeurd})$:

$$P(\text{fout} > \text{mat} \cap \text{goedkeuring})$$

- IR : $P(\text{de fout is groter dan materialiteit, voordat verdere gegevens bekend zijn})$:

$$P(\text{fout} > \text{mat})$$

- ICR : $P(\text{goedkeuring na AO/IC controle, als de fout groter dan materialiteit is})$:

$$P(\text{goedkeuring AO/IC} \mid \text{fout} > \text{mat})$$

- SR : $P(\text{goedkeuring na steekproefcontrole, als de fout groter dan materialiteit is})$: α

$$P(\text{goedkeuring steekproef} \mid \text{fout} > \text{mat} \cap \text{goedkeuring AO/IC})$$

Dan $OR = IR \times ICR \times SR$ is:

$$\begin{aligned} P(\text{fout} > \text{mat} \cap \text{goedkeuring}) &= \\ &= P(\text{fout} > \text{mat} \cap \text{goedkeuring AO/IC} \cap \text{goedkeuring steekproef}) = \\ &= P(\text{goedkeuring AO/IC} \mid \text{fout} > \text{mat}) * \\ &P(\text{goedkeuring steekproef} \mid \text{fout} > \text{mat} \cap \text{goedkeuring AO/IC}) * P(\text{fout} > \text{mat}) \end{aligned}$$

Daar waar bij Deloitte over risico gesproken wordt, wordt in de statistiek liever het woord kans of waarschijnlijkheid gebruikt. Risico wordt in de statistiek gebruikt als het gaat om bijvoorbeeld eenheden geld die op het spel staan, hier gaat het om de waarschijnlijkheid dat iets zich voordoet en niet om wat er op het spel staat. Daarom hoort hier over een kans gesproken te worden.

Over de juistheid van het ARM zijn nogal wat discussies ontstaan. De belangrijkste punten van kritiek richten zich op eerder genoemde aannames. Ten eerste zijn de risicopercentages niet achteraf te controleren op juistheid, hierdoor blijft de conclusie twijfelachtig. Ten tweede wordt de aanname

dat de risicopercentages van de verschillende controlegebieden onafhankelijk zijn niet waarschijnlijk geacht. Daarnaast worden vraagtekens gezet bij de bruikbaarheid van de formule. Wanneer de *IR* en *ICR* zeer laag zijn, bijvoorbeeld 30%, dan geeft de formule:

$$R(k, \alpha) = 3,0 - \ln(0,3) - \ln(0,3) \approx 0,60$$

Dit zou een betrouwbaarheid geven van $< 50\%$. Dit betekent dat de onbetrouwbaarheid van de controle groter is dan de betrouwbaarheid, dat lijkt geen juiste controle meer.

Vanwege deze kritieken is door R.H. Veenstra en P.C. Van Batenburg⁹ een alternatief op het ARM bedacht, het Bayesian Risk Assurance Model (BRAM).

7.2 Bayesiaans Risico-Analyse Model

Bij het Bayesiaans Risico-Analyse Model (BRAM) wordt net als bij ARM uitgegaan van weinig fouten in de massa. Er wordt gebruik gemaakt van een ‘Bayesiaanse methode’. Het idee is dat aan de hand van de voorkennis al een maximale fout (*MP**) kan worden vastgesteld. Dit is de maximale fout waarvan de accountant 100% zeker is dat die niet overschreden wordt, deze wordt geïnterpreteerd als de 95% bovengrens van de maximale fout. Hierdoor is de maximale fout op voorhand niet meer zo groot als de totale massa (dit is de grootst mogelijke fout, er kan namelijk nooit een grotere fout zijn dan dat de gehele populatie fout is) en is er al een stukje zekerheid over de fout verworven. Het BRAM stelt aan de hand van deze *MP** een a-priori verdeling op voor de maximale fout. In het originele model wordt uitgegaan van de formule:

$$n = \frac{3}{p_k}$$

In het BRAM wordt het aantal te controleren items, het aantal te controleren items dat uit het originele model volgt, min een aantal omdat je al een stukje zekerheid hebt. Dit wordt als volgt weergegeven:

$$n = \frac{3}{p_m} - \frac{3}{p^*}$$

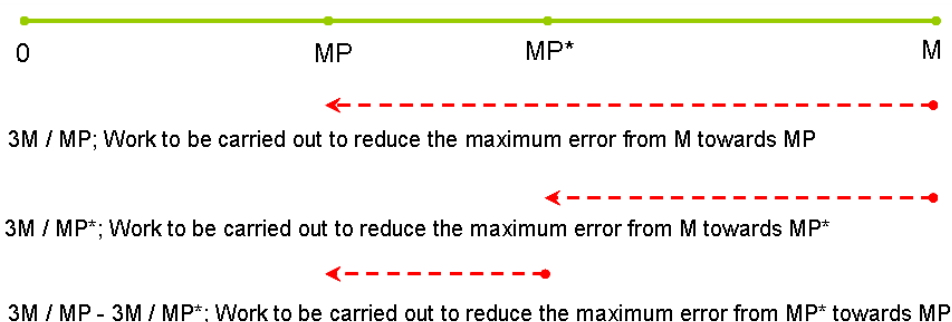
⁹ Veenstra, R.H., Van Batenburg, P.C., 1989, *Een doorbraak in steekproeftoepassingen door Bayesiaanse statistiek*, De Accountant, nr.11, p.561-564.

Waarbij $p^* = \frac{MP^*}{M}$ met M : totale waarde van de te controleren massa.

Wanneer de maximale foutfractie (p^*), die we vooraf vaststellen, kleiner is zal daarom minder gecontroleerd hoeven te worden. Belangrijk om hierbij te vermelden is dat het een benadering blijft van de originele formule. Dit kunnen we zien wanneer we vaststellen dat er geen voorkennis is, dan

is p^* gelijk aan 1, maar dan $n = \frac{3}{p_m} - \frac{3}{1}$ in plaats van $n = \frac{3}{p_m}$. Deze benadering blijkt voor de

praktijk goed genoeg te zijn.



Figuur 3: Het BRAM-model

Wanneer we nu het BRAM koppelen aan het ARM en we stellen dat er 4 niveaus van controle zijn dan kunnen we deze niveaus in BRAM als volgt schrijven:

- Uitgebreide niveau: $R = 3,0 \rightarrow n = 3,0 \cdot M / MP \approx 3,0 \cdot M / MP - 3,0 \cdot M / MP^*$ als $MP^* = M$
- Gematigd niveau: $R = 2,0 \rightarrow n = 2,0 \cdot M / MP \approx 3,0 \cdot M / MP - 3,0 \cdot M / MP^*$ als $MP^* = 3,0 \cdot MP$
- Beperkt niveau: $R = 0,7 \rightarrow n = 0,7 \cdot M / MP \approx 3,0 \cdot M / MP - 3,0 \cdot M / MP^*$ als $MP^* = 1,3 \cdot MP$
- Geen controle: $R \approx 0,0 \rightarrow n = 0,0 \approx 3,0 \cdot M / MP - 3,0 \cdot M / MP^*$ als $MP^* = MP$

Deze MP^* waardes zijn wel achteraf op juistheid te controleren. Dit is het grote voordeel van BRAM ten opzichte van ARM. Doordat deze MP^* waardes op juistheid te controleren zijn, zijn nu ook de *IR* en *ICR* uit het ARM achteraf op juistheid te controleren. Een ander belangrijk aspect van BRAM is dat het “professional judgement” bespreekbaar heeft gemaakt. In de accountantswereld wordt er vaak een schatting op basis van ervaring gemaakt voor het aantal te controleren items. Deze kan nu onderbouwd worden:

Stel dat de accountant 150 items wil controleren bij 1% onnauwkeurigheid en 95% betrouwbaarheid, dan:

$$N = 3 \cdot \frac{M}{MP} - 3 \cdot \frac{M}{MP^*} \rightarrow 150 = 300 - 3 \frac{M}{MP^*} \rightarrow MP^* = 3 \cdot \frac{M}{150} = \frac{M}{50}$$

Dit betekent dat wanneer je kunt aantonen dat $p^* = 2\%$, deze “professional judgement” onderbouwd is.

8 Bayesiaans model voor de voorraadcontrole

8.1 Inleiding

Het doel van een Bayesiaans model in de voorraadcontrole is, door voorkennis mee te kunnen nemen een realistisch model voor de voorraad te creëren en daarmee ook de variantie van de verdeling van de individuele fouten te verkleinen. Wanneer deze namelijk verkleind kan worden, is de onnauwkeurigheid bij bepaalde onzekerheid over de fouten minder groot en kan vaker worden goedgekeurd. Wanneer we bijvoorbeeld naar het criterium maximale fout kijken, zien we dat deze direct afhangt van de variantie van de fouten:

$$MF = TF\hat{F} + t\sqrt{\hat{\sigma}^2}$$

Wanneer de schatting van de variantie verkleind kan worden kan daardoor bij een kleinere materialiteit worden goedgekeurd. Het idee is dat de schatting $\hat{\sigma}^2$ een combinatie wordt van de oorspronkelijke schatter en de schatter van de variantie op basis van de voorkennis (dit wordt de a-priori schatting genoemd). Ook voor andere criteria geldt een soortgelijke redenering.

Met een Bayesiaans model zou een verkleining van de variantie bereikt kunnen worden door a-priori verdelingen te zetten op de parameters van het model. Deze a-priori verdelingen zijn dan gebaseerd op gunstige voorkennis, welke op allerlei verschillende manieren verkregen kan zijn. Voorbeelden van informatiebronnen hiervoor zijn:

- Expertmening, in dit geval zijn dat vaak accountants.
- Historische data.
- Gegevens van vergelijkbare bedrijven.

8.1.1 Algemene Baysiaanse Formule

De algemene Bayesiaanse formule is:

$$f_{post}(\theta | x) = \frac{f(x | \theta) \cdot f_{prior}(\theta)}{f(x)}$$

Met:

$f(x | \theta)$: De dichtheidsfunctie van x onder aanname van de parameter θ .

$f_{prior}(\theta)$: De a-priori dichtheidfunctie op de modelparameter(s) θ .

$f(x)$: De overall dichtheidfunctie van de data x .

$f_{post}(\theta | x)$: De a-posteriori dichtheidfunctie op de modelparameter(s) θ gegeven de data x .

Omdat de $f(x)$ constant is in de variabele θ , kunnen we stellen:

$$f_{post}(\theta | x) \propto f(x | \theta) \cdot f_{prior}(\theta)$$

Dit houdt in dat $f_{post}(\theta | x)$ zich als functie van θ hetzelfde gedraagt als $f(x | \theta) \cdot f_{prior}(\theta)$. Ze verschillen alleen in een positieve multiplicatieve constante.

8.1.2 Geconjugeerde a-priori verdelingen

Er zijn uiteraard vele mogelijkheden om a-priori verdelingen op te stellen. Om voor in de praktijk te ingewikkelde berekeningen te voorkomen, gebruiken we alleen geconjugeerde a-priori verdelingen. Dit zijn verdelingen die, volgens bovenstaande formule, gecombineerd met de likelihoodfunctie, een a-posteriori verdeling opleveren die tot dezelfde familie van verdelingen behoort als de a-priori verdeling. Hierdoor is de a-posteriori verdeling volledig gekarakteriseerd door de a-posteriori parameters en daarmee is de a-posteriori verdeling relatief eenvoudig te bepalen. Tevens is een vergelijking tussen deze twee verdelingen en daarmee ook tussen kansen onder de a-priori en a-posteriori verdeling gemakkelijk te maken. Voor de meeste likelihoodfuncties bestaan er a-priori verdelingen die conjugeren met deze likelihoodfunctie.

8.1.3 Beslissingscriteria

Wanneer een a-posteriori verdeling tot stand gekomen is, moet mede aan de hand van die verdeling besloten worden of wel of niet goedgekeurd kan worden. Hier kunnen verschillende criteria voor gebruikt worden. De meest voor de hand liggende bij dit probleem zijn:

1. Het huidige criterium dat bij Deloitte gehanteerd wordt: de maximale fout zoals in hoofdstuk 5 beschreven. Als de maximale fout kleiner is dan de materialiteit, kan worden goedgekeurd. De maximale fout zou aan de hand van de a-posteriori parameters berekend moeten worden:

$$MF = TF + t\sqrt{\hat{S}_T}$$

De t -waarde is hier een t -tabel element. De schattingen TF en \hat{S}_T worden gedaan met een Bayes schatting. De algemene regel voor de Bayes schatter voor $g(\theta)$ is:

$$BS(g(\bar{\Theta})) = E_{post}(g(\bar{\Theta})) = \int g(\theta) p_{post}(\theta) d\theta$$

Dit criterium behoort echter niet tot een formeel Bayesiaans model. Bij een Bayesiaans model hoort namelijk naar de kans op een bepaalde hypothese gekeken te worden. Daarnaast is het vanwege de a-priori verdelingen ook niet aannemelijk dat het t -tabel element gebruikt dient te worden. Omdat men bij Deloitte het criterium maximale fout graag wil blijven hanteren, is dit criterium toch meegenomen in de evaluatie.

Een nadeel van dit criterium is dat de a-priori parameters zo kunnen worden gekozen dat de waarnemingen weinig tot geen invloed meer hebben op de a-posteriori parameters. Dit gebeurt bijvoorbeeld wanneer de a-priori kennis de waarde toegekend krijgt van 10000 waarnemingen. Het aantal waarnemingen van de steekproef en daarmee ook de invloed van deze waarnemingen op de a-posteriori parameters vallen hierbij in het niet.

2. De a-posteriori kans op de nulhypothese. Wanneer de a-posteriori kans op de nulhypothese kleiner dan $0,05$ is, kan de nulhypothese verworpen worden en kan worden goedgekeurd. Een voordeel van deze manier is dat wanneer er a-priori een kleine kans op de nulhypothese is, snel kan worden goedgekeurd. Een nadeel van deze methode is dat de a-priori parameters zo gekozen kunnen worden, dat de waarnemingen geen invloed meer hebben op de a-posteriori parameters en dat daardoor ongeacht de waarnemingen de nulhypothese verworpen mag worden. De a-posteriori kans kan als volgt worden bepaald:

$$H_0 : |TF| \geq \text{Materialiteit} \rightarrow$$

$$P_{post}(H_0) = \int_{\lambda: |TF| \geq \text{Mat}} f_{post}(\lambda) d\lambda$$

3. De Bayes factor. Met deze factor worden a-priori kansen vergeleken met a-posteriori kansen. Wanneer de a-posteriori kans op de nulhypothese (veel) kleiner is dan de a-priori kans, kan de nulhypothese worden verworpen en kan worden goedgekeurd.

$$\text{Bayes Factor} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P_{\text{post}}(H_0) / P_{\text{post}}(H_1)}{P_{\text{prior}}(H_0) / P_{\text{prior}}(H_1)}$$

De kansen worden op dezelfde manier berekend als de a-posteriori kans op de nulhypothese. Het voordeel van deze methode is dat de data ook moet wijzen op het verwerpen van H_0 , hierdoor kan er minder makkelijk door een verkeerde a-priori gedachte een onjuiste beslissing gemaakt worden. Een nadeel hiervan is, dat wanneer de ‘prior odds’ al erg klein zijn de Bayes Factor vaak niet klein genoeg is om te kunnen verwerpen ondanks dat zowel de a-priori als a-posteriori kansen daar wel reden toe geven.

8.2 Modellen

Zoals in hoofdstuk 4 te zien is, wordt bij Deloitte gebruik gemaakt van de regressieschatter. De regressieschatter ziet er voor het voorraadprobleem als volgt uit:

$$\hat{E}_i = \bar{Y}_T - \hat{X}_i = \bar{Y}_T - \hat{\alpha} - \beta Y_i = \bar{Y}_T - X_i - \beta(\bar{Y}_T - Y_i)$$

Een correct Bayesiaans model zou een a-priori verdeling moeten hebben op alle parameters van het model, dus ook op de parameter β . Omdat de a-priori verwachting hiervan altijd rond 1 ligt, is dit onlogisch om te doen, in feite creëren we daarmee de verschilschatter. Een Bayesiaans model op basis van de regressieschatter is hierom geen natuurlijke keuze. De verschilschatter daarentegen lijkt wel een goede keuze:

$$\hat{E}_i = \bar{Y}_T - \hat{X}_i = \bar{Y}_T - X_i - (\bar{Y}_T - Y_i) = Y_i - X_i$$

Daarom gaan we Bayesiaanse modellen bouwen op basis van de verschilschatter.

8.2.1 Model naar aanleiding van eerst aanzet Deloitte

De eerste stap naar een Bayesiaans model voor de voorraad controle wordt bij Deloitte gemaakt in het onderzoek van Ferry Geertman¹⁰:

¹⁰ Ferry Geertman, “Over de toepasbaarheid van regressieschatter en verschilschatter in de accountantscontrole”, 1994

Toepassing van een Bayesiaans model is slechts aantrekkelijk als de posterior-spreiding kleiner is dan de spreiding van de klassieke aanpak (de likelihood in het Bayesiaanse model). Uit Zellner(1971) blijkt dat bij een normaal verdeelde prior en een normaal verdeelde likelihood de posterior normaal verdeeld is met een spreiding die altijd kleiner is dan de spreiding van de likelihood. Hieruit volgt dat de ontwikkeling van een Bayesiaans model zin heeft.

Dit idee zullen we daarom allereerst proberen uit te werken. Door Zellner wordt gesteld:

- De dichtheidsfunctie van de geobserveerde data:

$$f(x_i | \mu, \sigma^2) \sim N(\mu, \sigma^2)$$

De likelihoodfunctie van μ is de dichtheidsfunctie geëvalueerd in de geobserveerde data:

$$\begin{aligned} L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) &= \prod_{i=1}^n L(\mu, \sigma^2 | x_i) = \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ \frac{-1}{2\sigma^2} \left[(n-1)s_n^2 + n(\bar{x}_n - \mu)^2 \right] \right\} \end{aligned}$$

- a-priori kennis:

$$- f_{prior}(\mu) \sim N(\mu_0, \sigma_0^2)$$

- σ^2 : σ_c^2 is bekend

- a-posteriori verdeling:

$$f_{post}(\mu) \sim N(\mu_n, \sigma_n^2)$$

Met:

$$\mu_n = \frac{\bar{\epsilon}_n (\sigma^2 / n)^{-1} + \mu_0 (\sigma_0^2)^{-1}}{(\sigma^2 / n)^{-1} + (\sigma_0^2)^{-1}} \text{ en } \sigma_n^2 = \frac{1}{(\sigma^2 / n)^{-1} + (\sigma_0^2)^{-1}}$$

(zie ook Zellner pag. 15)¹¹

¹¹ Zellner, A., 1971, *An introduction to Bayesian inference in Econometrics*, John Wiley & Sons, Inc.

Om dit concept te kunnen toepassen op het voorraadprobleem zullen we daarom dezelfde aannames moeten kunnen maken over de parameters en hun verdelingen als bovenstaande.

Allereerst zal de aanname dat de individuele waarnemingen normaal verdeeld zijn gemaakt moeten kunnen worden.

Aanname: de individuele verschillen zijn normaal verdeeld:

$$E_i = Y_i - X_i \sim N(\mu, \sigma^2)$$

Deze aanname wordt ook in het huidige model gemaakt. Ten tweede kijken we naar de voorkennis. Hierbij zien we dat in het model van Zellner de variantie van de individuele verschillen als bekend verondersteld wordt. Omdat deze aanname niet realistisch is bij de voorraad, kunnen we dit model niet toepassen op de voorraad. Een schatting voor σ^2 gebruiken in plaats van een bekende σ^2 ($\sigma^2 \rightarrow \hat{\sigma}^2$), is in dit geval onjuist. De onzekerheid over de schatting van de variantie wordt dan namelijk volledig buiten beschouwing gelaten. Het model van Zellner kan hierdoor niet toegepast worden op het voorraad probleem. Het verkleint weliswaar de variantie van de verdeling op μ , maar de voorraad voldoet niet aan de voorwaarden die gesteld worden voor het gebruik van dit model, namelijk dat de variantie van de individuele verschillen bekend is.

8.2.2 Model met a-priori verdeling op zowel μ als σ^2

Laten we de aanname dat de verschillen normaal verdeeld zijn (zoals ook in het bovenstaande model wordt gedaan), behouden:

$$E_i = Y_i - X_i \sim N(\mu, \sigma^2)$$

De algemene Bayesiaanse formule hierbij is van de volgende vorm:

$$f_{post}(\mu, \sigma^2 | \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \propto f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n | \mu, \sigma^2) \cdot f_{prior}(\mu, \sigma^2)$$

Waarbij:

- $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ de n verschillen uit de steekproef zijn.

- $f_{post}(\mu, \sigma^2 | \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$: de a-posteriori verdeling van μ en σ^2 , deze kansverdeling is gebaseerd op zowel de waarnemingen als op de voorkennis.
- $f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n | \mu, \sigma^2)$: de dichtheidsfunctie van de waargenomen verschillen onder μ en σ^2 .
- $f_{prior}(\mu, \sigma^2)$: de a-priori verdeling van μ en σ^2 , dit is de kansverdeling die gebaseerd is op voorkennis.

8.2.2.1 Likelihoodfunctie

De likelihoodfunctie van μ is de dichtheidsfunctie geëvalueerd in de geobserveerde verschillen. De dichtheidsfunctie van de individuele verschillen is zoals eerder gesteld:

$$E_i \sim N(\mu, \sigma^2) \quad \rightarrow \quad f_{prior}(\varepsilon_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon_i - \mu)^2\right\}$$

De likelihoodfunctie van μ bij de waargenomen verschillen is:

$$\begin{aligned} L(\mu, \sigma^2 | \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) &= \prod_{i=1}^n L(\mu, \sigma^2 | \varepsilon_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\varepsilon_i - \mu)^2\right\} = \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{\frac{-1}{2\sigma^2} \left[\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon}_n)^2 + n(\bar{\varepsilon}_n - \mu)^2\right]\right\} = \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{\frac{-1}{2\sigma^2} \left[(n-1)s_n^2 + n(\bar{\varepsilon}_n - \mu)^2\right]\right\} \end{aligned}$$

Waarbij:

$$\bar{\varepsilon}_n : \text{Het gemiddelde van de individuele verschillen, } \frac{1}{n} \sum_{i=1}^n \varepsilon_i .$$

$$s_n^2 : \text{De steekproefvariantie van de individuele verschillen, } \frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon}_n)^2 .$$

8.2.2.2 Meest aannemelijke schatter (Maximum likelihood estimator)

Om de meest aannemelijke schatters te bepalen voor μ en σ^2 , wordt eerst de log aannemelijkheidsfunctie bepaald:

$$\begin{aligned} & \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{\frac{-1}{2\sigma^2}[(n-1)s_n^2 + n(\bar{\varepsilon}_n - \mu)^2]\right\} = \\ & = -\frac{1}{2}n \log 2\pi - \frac{1}{2}n \log \sigma^2 - \frac{1}{2} \frac{1}{\sigma^2} \sum_{i=1}^n (E_i - \mu)^2 \end{aligned}$$

De meest aannemelijke schatter, is de waarde van de te schatten parameter die deze functie maximaliseert.

Maximaliseren van de aannemelijkheidsfunctie naar μ is gelijk aan het minimaliseren van

$$\sum_{i=1}^n (E_i - \mu)^2 \text{ naar } \mu:$$

$$\begin{aligned} \frac{\delta y}{\delta \mu} \sum_{i=1}^n (E_i - \mu)^2 &= -2 \sum_{i=1}^n (E_i - \mu) \\ -2 \sum_{i=1}^n (E_i - \mu) &= 0 \rightarrow \hat{\mu} = \bar{E}_n \end{aligned}$$

Maximaliseren van de aannemelijkheidsfunctie naar σ^2 met de schatter van μ gesubstitueerd is

gelijk aan minimaliseren van $\frac{1}{2}n \log \sigma^2 + \frac{1}{2} \frac{1}{\sigma^2} \sum_{i=1}^n (E_i - \bar{E}_n)^2$ naar σ^2 :

$$\begin{aligned} \frac{\delta y}{\delta \sigma^2} \left(\frac{1}{2}n \log \sigma^2 + \frac{1}{2} \frac{1}{\sigma^2} \sum_{i=1}^n (E_i - \bar{E}_n)^2 \right) &= \frac{1}{2} \frac{n}{\sigma^2} + \frac{1}{2} \frac{1}{\sigma^4} \sum_{i=1}^n (E_i - \bar{E}_n)^2 \rightarrow 0 \\ \rightarrow \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (E_i - \bar{E}_n)^2 \end{aligned}$$

De meest aannemelijke schatters zijn $\hat{\mu} = \bar{E}_n$ en $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (E_i - \bar{E}_n)^2$. Dit is ook wat de schatters

aan de hand van de verschilschatter ongeveer opleveren. (De schatting van de variantie met de verschilschatter is alleen net iets groter vanwege het kleinere aantal vrijheidsgraden wat daarbij gebruikt wordt).

8.2.2.3 A-priori verdelingen

Om een model op te stellen dat bruikbaar is voor de voorraad, mag de variantie van de individuele verschillen niet als bekend worden verondersteld (zoals in paragraaf 8.2.1 wordt gedaan). Daarom zal naast een a-priori verdeling op het gemiddelde van de verschillen (μ), ook een a-priori verdeling

moeten worden opgesteld op de variantie van de individuele verschillen (σ^2). Een geconjugeerde a-priori verdeling over zowel de μ als de σ^2 hierbij is:

$$\mu | \sigma^2 \sim N(\mu_0, \frac{\sigma^2}{k_0}) \quad \rightarrow f_{prior}(\mu | \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2/k_0}} \exp\left\{-\frac{1}{2\sigma^2/k_0}(\mu_0 - \mu)^2\right\}$$

$$\sigma^2 \sim Inv - Gamma\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right) \quad \rightarrow f_{prior}(\sigma^2) = \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{\frac{-\nu_0\sigma_0^2}{2\sigma^2}\right\}$$

De a-priori verdeling wordt nu:

$$\begin{aligned} f_{prior}(\mu, \sigma^2) &= f_{prior}(\mu | \sigma^2) \cdot f_{prior}(\sigma^2) = \\ &= \frac{1}{\sqrt{2\pi\sigma^2/k_0}} \exp\left\{-\frac{1}{2\sigma^2/k_0}(\mu_0 - \mu)^2\right\} \cdot \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{\frac{-\nu_0\sigma_0^2}{2\sigma^2}\right\} = \\ &= \left[\frac{1}{\sqrt{2\pi/k_0}} \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} \sigma_0^{\nu_0} \right] \cdot \frac{1}{\sigma} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + k_0(\mu_0 - \mu)^2]\right\} \end{aligned}$$

Om een juiste waarde te bepalen voor de a-priori parameters is een goede interpretatie noodzakelijk.

De interpretaties die bij de parameters van dit model horen:

- μ_0 : a-priori verwachting van het gemiddelde van de verschillen.
- k_0 : Aantal waarnemingen dat de a-priori verwachting voor μ (dit is μ_0) waard geacht wordt.
- ν_0 : Aantal vrijheidsgraden voor de a-priori variantie van de individuele verschillen.
- σ_0^2 : a-priori verwachte variantie van de individuele verschillen.

8.2.2.4 A-posteriori verdeling

Vanwege het feit dat we gebruik maken van een geconjugeerde a-priori verdeling, zal de a-posteriori verdeling tot dezelfde familie van verdelingen behoren als de a-priori verdeling. De a-posteriori verdeling bezit daarom ook dezelfde soort parameters als de a-priori verdeling. Voor het vergemakkelijken van de notatie stellen we:

$$\lambda_0 = \{\mu_0, k_0, \nu_0, \sigma_0^2\} \text{ en } \lambda_n = \{\mu_n, k_n, \nu_n, \sigma_n^2\}$$

Het bewijs dat de a-posteriori verdeling tot dezelfde familie van verdelingen behoort als de a-priori verdeling en daarmee ook het bewijs dat de a-priori verdeling een geconjugeerde verdeling is voor de eerder geformuleerde likelihood:

$$\begin{aligned}
 f_{post}(\mu, \sigma^2 | \lambda_n) &= f_{post}(\mu, \sigma^2 | \lambda_0, \varepsilon) = f_{prior}(\mu, \sigma^2 | \lambda_0) \cdot f(\varepsilon | \mu, \sigma^2) = \\
 &\left[\frac{1}{\sqrt{2\pi/k_0}} \frac{(v_0 \sigma_0^2 / 2)^{v_0/2}}{\Gamma(v_0 / 2)} \sigma_0^{v_0} \right] \cdot \frac{1}{\sigma} (\sigma^2)^{-(\frac{v_0}{2}+1)} \exp\left\{-\frac{1}{2\sigma^2} [v_0 \sigma_0^2 + k_0 (\mu_0 - \mu)^2]\right\} \cdot \\
 &\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left\{\frac{-1}{2\sigma^2} [(n-1)s_n^2 + n(\bar{\varepsilon}_n - \mu)^2]\right\} \propto \\
 &\propto \frac{1}{\sigma} \cdot (\sigma^2)^{-(\frac{v_n}{2}+1)} \exp\left\{\frac{-1}{2\sigma^2} [v_n \sigma_n^2 + k_n (\mu_n - \mu)^2]\right\}
 \end{aligned}$$

Hieruit volgt dat de a-posteriori parameters uitgedrukt in de a-priori parameters en de data worden gegeven door:

- $\mu_n = \frac{k_0}{k_0 + n} \mu_0 + \frac{n}{k_0 + n} \cdot \bar{\varepsilon}_n$
- $k_n = k_0 + n$
- $v_n = v_0 + n$
- $\sigma_n^2 = \frac{v_0 \sigma_0^2 + (n-1)S_n^2 + \frac{k_0 n}{k_0 + n} (\bar{\varepsilon}_n - \mu_0)^2}{v_0 + n}$

Met deze parameters is de a-posteriori verdeling $f_{post}(\mu, \sigma^2 | \lambda_n)$ namelijk van gelijke soort als de a-priori verdeling (alleen de waarden van de parameters zijn verschillend).

8.2.2.5 Bayes schatters

De Bayes schatters voor respectievelijk μ en σ^2 aan de hand van dit model zijn:

- $T\hat{F} = E_{post}(TF) = E_{post}(T \cdot \mu) = T \cdot E_{post}(\mu) = T \cdot \mu_n$
- $\hat{S}_T^2 = E_{post}(S_T^2) = E_{post}(T^2 \cdot \bar{\sigma}^2) = T^2 \cdot E_{post}(\bar{\sigma}^2) = T^2 \cdot \frac{\sigma_n^2}{k_n}$

8.2.2.6 Beslissingscriteria

De 3 genoemde beslissingscriteria kunnen nu bepaald worden:

1) Maximale fout:

$$MF = |TF| + t\sqrt{\hat{S}_T^2} = |T \cdot \mu_n| + t\sqrt{T^2 \frac{\sigma_n^2}{k_n}}$$

Er mag worden goedgekeurd als deze maximale fout kleiner is dan de vooraf vastgestelde materialiteit.

2) De a-posteriori kans op de nulhypothese:

$$\begin{aligned} P_{post}(H_0) &= P(\mu, \sigma^2 \text{ waarbij } T | \mu | \geq mat) = \\ &= 1 - \int_{\sigma^2=0}^{\infty} \int_{\mu=-\frac{Mat}{T}}^{\frac{Mat}{T}} f_{post}(\mu, \sigma^2 | \lambda_n) \delta\mu \delta\sigma \end{aligned}$$

Er mag worden goedgekeurd als deze a-posteriori kans op de nulhypothese kleiner dan 0,05 is.

3) De Bayesfactor:

$$Bayes\ Factor = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\frac{P_{post}(H_0)}{P_{post}(H_1)}}{\frac{P_{prior}(H_0)}{P_{prior}(H_1)}} = \frac{P_{post}(H_0)}{1 - P_{post}(H_0)} \cdot \frac{1 - P_{prior}(H_0)}{P_{prior}(H_0)}$$

Waarbij:

$$P_{prior}(H_0) = 1 - \int_{\sigma^2=0}^{\infty} \int_{\mu=-\frac{Mat}{T}}^{\frac{Mat}{T}} f_{prior}(\mu, \sigma^2 | \lambda_0) \delta\mu \delta\sigma$$

$$P_{post}(H_0) = 1 - \int_{\sigma^2=0}^{\infty} \int_{\mu=-\frac{Mat}{T}}^{\frac{Mat}{T}} f_{post}(\mu, \sigma^2 | \lambda_n) \delta\mu \delta\sigma$$

Er mag worden goedgekeurd als deze Bayes factor klein genoeg is. Hoe klein deze precies moet zijn is afhankelijk van de data en de vooraf gestelde eisen en kan daarom niet zomaar worden vastgesteld.

8.2.3 Model met a-priori verdeling op de parameters μ , σ^2 en p

Tot nu toe is ervan uitgegaan dat alle individuele verschillen normaal verdeeld zijn. Wanneer we deze verschillen nader bekijken, zien we echter dat een groot aantal van de verschillen gelijk aan nul is. Hierdoor blijkt de aanname dat de verschillen normaal verdeeld zijn niet realistisch. Daarom is besloten de verschillen op een andere manier te modelleren namelijk:

$$E_i = Y_i - X_i \sim pN(\mu, \sigma^2) + (1-p)\delta_0 \quad \text{waarbij } p \in [0,1] \text{ en } E_i | E_i \neq 0 \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

Interpretatie hiervan is met kans p komt het verschil uit een normale verdeling en met kans van $1-p$ is het verschil nul. Het model is een mengsel van een normale verdeling en de gedegeneerde (*Dirac*) verdeling op $\{0\}$.

8.2.3.1 Likelihoodfunctie

Nu de individuele fouten op een andere manier gemodelleerd zijn, verandert ook de likelihoodfunctie ten opzichte van de voorgaande modellen. Deze functie is nu niet meer zo eenvoudig weer te geven. Vanwege het feit dat het wel of niet aanwezig zijn van een verschil ongelijk aan nul een discrete verdeling (binomiaal) heeft en de grootte van een fout een continue verdeling (normaal), kunnen we dit niet zomaar combineren. Hiervoor zullen we eerst een nieuwe dominerende maat moeten introduceren:

$$l = Leb + \delta_0 \quad (\text{Lebesgue-maat} + \text{Dirac op } 0)$$

De dichtheid van 1 observatie ten opzichte van l kan worden gedefinieerd als:

$$f(\varepsilon_i | \mu, \sigma^2, p) = p\varphi_{\mu, \sigma^2}(\varepsilon_i)1_{\{\varepsilon_i \neq 0\}} + (1-p)1_{\{\varepsilon_i = 0\}}, \text{ want voor ieder interval } [l, \mu) \text{ geldt:}$$

$$P(E_i \in [l, \mu)) = \int_{[l, \mu)} f(\varepsilon | \mu, \sigma^2, p) dl(\varepsilon)$$

Voor het gemak wordt de volgende notatie ingevoerd:

$$k = \#\{i : \varepsilon_i \neq 0\}$$

zodat: $n - k = \#\{i : \varepsilon_i = 0\}$

De totale likelihoodfunctie van μ, σ^2 en p bij de geobserveerde n waarnemingen ten opzichte van l is dan:

$$\begin{aligned}
 f(\mu, \sigma^2, p | \varepsilon) &= \prod_{i=1}^n f(\mu, \sigma^2, p | \varepsilon_i) = \prod_{\{i:\varepsilon_i=0\}} (1-p) \cdot \prod_{\{i:\varepsilon_i \neq 0\}} p \varphi_{\mu, \sigma^2}(\varepsilon_i) = \\
 &= (1-p)^{n-k} p^k \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^k \exp^{-\frac{1}{2\sigma^2}((k-1)s_{\varepsilon_i \neq 0}^2 + k(\bar{\varepsilon}_{\varepsilon_i \neq 0} - \mu)^2)}
 \end{aligned}$$

Waarbij $s_{\varepsilon_i \neq 0}^2$ en $\bar{\varepsilon}_{\varepsilon_i \neq 0}$ respectievelijk de steekproefvariantie en het gemiddelde van de verschillen ongelijk aan nul zijn.

8.2.3.2 Meest aannemelijke schatters (maximum likelihood schatters)

De log aannemelijkheidsfunctie:

$$(n-k) \log(1-p) + k \log p - \frac{1}{2} n \log 2\pi - \frac{1}{2} n \log \sigma^2 - \frac{1}{2} \frac{1}{\sigma^2} \sum_{i=1:\{\varepsilon_i \neq 0\}}^n (\varepsilon_i - \mu)^2$$

De meest aannemelijke schatters voor de parameters μ , σ^2 en p zijn dan:

- Maximaliseer aannemelijkheidsfunctie naar $\mu \rightarrow$ minimaliseer $\sum_{i=1:\{\varepsilon_i \neq 0\}}^n (\varepsilon_i - \mu)^2$ naar μ :

$$\frac{\delta y}{\delta \mu} \sum_{i=1:\{\varepsilon_i \neq 0\}}^n (\varepsilon_i - \mu)^2 = -2 \sum_{i=1:\{\varepsilon_i \neq 0\}}^n (\varepsilon_i - \mu) \rightarrow \hat{\mu} = \bar{\varepsilon}_{n:\{\varepsilon_i \neq 0\}}$$

- Maximaliseer aannemelijkheidsfunctie naar σ^2 met de schatter van μ gesubstitueerd \rightarrow

minimaliseer $\frac{1}{2} n \log \sigma^2 + \frac{1}{2} \frac{1}{\sigma^2} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon}_n)^2$ naar σ^2 :

$$\begin{aligned}
 \frac{\delta y}{\delta \sigma^2} \left(\frac{1}{2} n \log \sigma^2 + \frac{1}{2} \frac{1}{\sigma^2} \sum_{i=1:\{\varepsilon_i \neq 0\}}^n (\varepsilon_i - \bar{\varepsilon}_{n:\{\varepsilon_i \neq 0\}})^2 \right) &= \frac{1}{2} \frac{n}{\sigma^2} + \frac{1}{2} \frac{1}{\sigma^4} \sum_{i=1:\{\varepsilon_i \neq 0\}}^n (\varepsilon_i - \bar{\varepsilon}_{n:\{\varepsilon_i \neq 0\}})^2 \\
 \rightarrow \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1:\{\varepsilon_i \neq 0\}}^n (\varepsilon_i - \bar{\varepsilon}_{n:\{\varepsilon_i \neq 0\}})^2
 \end{aligned}$$

- Maximaliseer aannemelijkheidsfunctie naar $p \rightarrow$ maximaliseer $(n-k) \log(1-p) + k \log p$ naar p :

$$\frac{d}{dp} (n-k) \log(1-p) + k \log p = -\frac{n-k}{1-p} + \frac{k}{p} \rightarrow \hat{p} = \frac{k}{n}$$

8.2.3.3 A-priori verdelingen

Er zijn nu drie parameters waarop een a-priori verdeling gezet kan worden. We kunnen nu voorkennis over de fractie verschillen gelijk aan nul, voorkennis over het gemiddelde van de verschillen ongelijk aan nul en voorkennis over de variantie van de individuele verschillen ongelijk aan nul in de a-priori verdeling stoppen. Omdat we willen werken met een geconjugeerde a-priori verdeling, is de a-priori verdeling op de verschillende parameters de volgende:

$$\begin{aligned}
 - \quad \mu \mid \sigma^2 &\sim N\left(\mu_0, \frac{\sigma^2}{k_0}\right) &\rightarrow f_{\text{prior}}(\mu \mid \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2/k_0}} \exp\left\{-\frac{1}{2\sigma^2/k_0}(\mu_0 - \mu)^2\right\} \\
 - \quad \sigma^2 &\sim \text{InvGamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right) &\rightarrow f_{\text{prior}}(\sigma^2) &= \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\} \\
 - \quad p &\sim \text{Beta}(\alpha_0, \beta_0) &\rightarrow f_{\text{prior}}(p) &= \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\alpha_0-1} (1-p)^{\beta_0-1}
 \end{aligned}$$

Voor de notatie voeren we in:

$$\lambda_0 = \{\mu_0, \sigma_0^2, k_0, \nu_0, \alpha_0, \beta_0\}$$

Aanname: De kans op een fout is onafhankelijk van de grootte van een fout (in geldeenheden). Met andere woorden: p is onafhankelijk van (μ, σ^2)

Nu kunnen we de totale a-priori verdeling bepalen:

$$\begin{aligned}
 f_{\text{prior}}(\mu, \sigma^2, p) &= f_{\text{prior}}(\mu \mid \sigma^2) \cdot f_{\text{prior}}(\sigma^2) \cdot f_{\text{prior}}(p) = \\
 &= \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\alpha_0-1} (1-p)^{\beta_0-1} \cdot \frac{1}{\sqrt{2\pi\sigma^2/k_0}} \exp\left\{-\frac{1}{2\sigma^2/k_0}(\mu_0 - \mu)^2\right\} \cdot \\
 &\quad \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\} \propto \\
 &\propto p^{\alpha_0-1} (1-p)^{\beta_0-1} \cdot \frac{1}{\sigma} \cdot (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + k_0(\mu_0 - \mu)^2]\right\}
 \end{aligned}$$

Om juiste waarden voor de a-priori parameters in te kunnen vullen, is een goede interpretatie van deze parameters noodzakelijk. De parameters van dit model hebben de volgende interpretatie:

- μ_0 : a-priori verwachting van het gemiddelde van de verschillen ongelijk aan nul.
- k_0 : aantal waarnemingen dat de a-priori verwachting van μ (μ_0) waard geacht wordt.
- ν_0 : aantal vrijheidsgraden voor de a-priori variantie van de individuele verschillen ongelijk aan nul.
- σ_0^2 : a-priori verwachte variantie van de individuele verschillen ongelijk aan nul.
- α_0 : a-priori verwachting van de kans op een verschil ongelijk aan nul vermenigvuldigd met het totaal aantal waarnemingen dat deze a-priori informatie waard geacht wordt.
- β_0 : a-priori verwachting van de kans op een verschil gelijk aan nul vermenigvuldigd met het totaal aantal waarnemingen dat deze a-priori informatie waard geacht wordt.

8.2.3.4 a-posteriori verdeling

Omdat de a-priori verdeling die we gekozen hebben een geconjugeerde verdeling is bij de bovenstaande likelihood, is de a-posteriori van dezelfde vorm als de a-priori verdeling, alleen de parameters verschillen. Het bewijs hiervan gaat op dezelfde manier als in paragraaf 8.2.2.4 en is te vinden in de appendix. De a-posteriori parameters zijn:

$$\lambda_n = \{\mu_n, \sigma_n^2, k_n, \nu_n, \alpha_n, \beta_n\}$$

Waarbij:

- $$\mu_n = \frac{k_0}{k_0 + k} \mu_0 + \frac{k}{k_0 + k} \cdot \bar{\varepsilon}_{\varepsilon_i \neq 0}$$
- $$\sigma_n^2 = \frac{\nu_0 \sigma_0^2 + (k-1) S_{\varepsilon_i \neq 0}^2 + \frac{k_0 k}{k_0 + k} (\bar{\varepsilon}_{\varepsilon_i \neq 0} - \mu_0)^2}{\nu_0 + k}$$
- $k_n = k_0 + k$
- $\nu_n = \nu_0 + k$
- $\alpha_n = \alpha_0 + k$
- $\beta_n = \beta_0 + n - k$

8.2.3.5 Bayes schatters

De Bayes schatters voor respectievelijk μ en σ^2 en p aan de hand van dit model zijn:

$$\begin{aligned}
 - \quad T\hat{F} &= E_{post}(TF) = E_{post}(T \cdot p \cdot \mu) = T \cdot E_{post}(\mu) \cdot E_{post}(p) = T \cdot \mu_n \cdot \frac{\alpha_n}{\alpha_n + \beta_n} \\
 - \quad \hat{S}_T^2 &= E_{post}(S_T^2) = E_{post}((T \cdot p)^2 \cdot \bar{\sigma}^2) = T^2 \cdot E_{post}(p^2 \bar{\sigma}^2) = T^2 \cdot \left(\frac{\alpha_n}{\alpha_n + \beta_n} \right)^2 \frac{\sigma_n^2}{k_n}
 \end{aligned}$$

8.2.3.6 Beslissingscriteria

De 3 genoemde beslissingscriteria kunnen nu bepaald worden:

1) Maximale fout:

$$MF = |TF| + t\sqrt{\hat{S}_T^2} = \left| T \cdot \mu_n \frac{\alpha_n}{\alpha_n + \beta_n} \right| + t \sqrt{T^2 \left(\frac{\alpha_n}{\alpha_n + \beta_n} \right)^2 \frac{\sigma_n^2}{k_n}} = t \cdot T \cdot \frac{\alpha_n}{\alpha_n + \beta_n} \cdot \sqrt{\frac{\sigma_n^2}{k_n}}$$

Er kan worden goedgekeurd wanneer deze maximale fout kleiner is dan de vooraf vastgestelde materialiteit.

2) De a-posteriori kans op de nulhypothese:

$$\begin{aligned}
 P_{post}(H_0) &= P(\mu, \sigma^2, p \text{ waarbij } Tp \mid \mu \geq mat) = \\
 &= 1 - \int_{p=0}^1 \int_{\sigma^2=0}^{\infty} \int_{\mu=-\frac{Mat}{Tp}}^{\frac{Mat}{Tp}} f_{post}(\mu, \sigma^2, p \mid \lambda_n) \delta\mu \delta\sigma^2 \delta p
 \end{aligned}$$

Er kan worden goedgekeurd wanneer deze a-posteriori kans op H_0 kleiner is dan 0,05

3) De Bayesfactor:

$$\begin{aligned}
 \text{Bayes Factor} &= \frac{\text{posterior odds}}{\text{prior odds}} = \\
 &= \frac{P_{post}(H_0) / P_{post}(H_1)}{P_{prior}(H_0) / P_{prior}(H_1)} = \frac{P_{post}(H_0)}{1 - P_{post}(H_0)} \cdot \frac{1 - P_{prior}(H_0)}{P_{prior}(H_0)}
 \end{aligned}$$

Waarbij:

$$P_{prior}(H_0) = 1 - \int_{p=0}^1 \int_{\sigma^2=0}^{\infty} \int_{\mu=-\frac{Mat}{Tp}}^{\frac{Mat}{Tp}} f_{prior}(\mu, \sigma^2, p | \lambda_0) \delta\mu \delta\sigma^2 \delta p$$

$$P_{post}(H_0) = 1 - \int_{p=0}^1 \int_{\sigma^2=0}^{\infty} \int_{\mu=-\frac{Mat}{Tp}}^{\frac{Mat}{Tp}} f_{post}(\mu, \sigma^2, p | \lambda_n) \delta\mu \delta\sigma^2 \delta p$$

Er kan worden goedgekeurd wanneer deze Bayes factor kleiner is dan een vooraf vastgestelde waarde.

Dit model blijkt meer realistisch te zijn. Dit komt doordat de verschillen op een meer realistische manier gemodelleerd zijn. Door de splitsing van de verschillen gelijk en ongelijk aan nul, is ook de eventuele toevoeging die de regressieschatter biedt ten opzichte van de verschilschatter verdwenen. De regressieschatter is namelijk juist in gebruik genomen vanwege het relatief grote aantal verschillen gelijk aan nul. Aan de hand van dit model zullen daarom de resultaten berekend worden.

8.3 Resultaten

Om te bepalen of het nieuwe model (zie paragraaf 8.2.3) een toegevoegde waarde heeft voor de voorraadcontrole, zullen in dit hoofdstuk resultaten worden weergegeven voor een drietal datasets van reeds uitgevoerde voorraadcontroles. Deze datasets bestaan ieder uit 200 boekwaardes en bijbehorende werkelijke waardes van de voor de huidige methode geselecteerde items. Aan de hand van de resultaten zal een vergelijking gemaakt worden tussen het huidige en het nieuwe model. Natuurlijk geven deze resultaten geen garantie voor resultaten van andere voorraden. Het idee is dat het wel een indruk kan geven over de bruikbaarheid van het model. De resultaten zijn berekend aan de hand van de a-posteriori parameters. Deze zijn berekend volgens de in de vorige paragraaf weergegeven formules. Verder is er bij sommige berekeningen gebruik gemaakt van simulatie omdat er willekeurig elementen geselecteerd dienden te worden.

Het doel van het Bayesiaanse model is voorkennis implementeren in een model zodat deze kan worden meegenomen bij de controle en het model daardoor meer overeenkomt met de werkelijkheid. Het model is daardoor op meer kennis gebaseerd. Doordat het model op meer kennis gebaseerd is, is het misschien mogelijk om met een kleinere steekproefomvang dan het huidige een even betrouwbaar model met een (ongeveer) even breed 90% betrouwbaarheidsinterval voor de totale fout te verkrijgen als het huidige model. Om vast te kunnen stellen of dit daadwerkelijk het geval is, gaan we als volgt te werk:

1. Allereerst controleren we of het nieuwe model met a-priori parameters die (vrijwel) geen voorkennis bevatten, ongeveer dezelfde uitkomsten geeft als het huidige model.
2. Als blijkt dat de modellen vergelijkbaar zijn, wordt bekeken wat de uitkomsten met het nieuwe model met a-priori verdelingen met gunstige voorkennis zijn. Geeft het nieuwe model met voorkennis betere resultaten dan het huidige?
3. Als blijkt dat het nieuwe model met voorkennis voor eenzelfde steekproef een minstens zo goed model is, kan worden bepaald of het nieuwe model met een kleinere steekproef een even betrouwbaar model oplevert als het huidige model met steekproefomvang 200.

8.3.1 Vergelijking huidige model en nieuwe model (1)

Allereerst dient het nieuwe model gevalideerd te worden. Omdat er geen werkelijke resultaten beschikbaar zijn, wordt dit gedaan door het nieuwe model te vergelijken met het huidige model. Omdat in het huidige model geen voorkennis kan worden meegenomen, worden de resultaten van het

huidige model vergeleken met die van het nieuwe model, waarbij er gebruik gemaakt wordt van een a-priori verdeling zonder voorkennis.

8.3.1.1 a-priori verdeling opstellen zonder voorkennis

Omdat voor het ontwikkelde model a-priori parameters noodzakelijk zijn, moeten ook wanneer er geen voorkennis aanwezig is, toch a-priori parameters opgegeven worden. Omdat de a-priori verdeling zo weinig mogelijk invloed mag hebben op de resultaten (immers er is vooraf niets bekend), dienen de a-priori parameters zo gekozen te worden dat de invloed ervan op de a-posteriori parameters en dus op de a-posteriori verdeling minimaal is. De uiteindelijke resultaten (die op de a-posteriori verdeling gebaseerd zijn), zullen hierdoor bijna geheel bepaald worden door de resultaten van de waarnemingen.

Voor het nieuwe model worden de volgende a-priori parameters gebruikt wanneer er geen voorkennis beschikbaar is:

$$\mu_0 = 0, k_0 = 1, \nu_0 = 1, \sigma_0^2 = 1000, \alpha_0 = 1 \text{ en } \beta_0 = 1$$

Met het gebruik van deze parameters zullen de a-posteriori parameters zijn:

$$\mu_n = \frac{k_0}{k_0 + k} \mu_0 + \frac{k}{k_0 + k} \cdot \bar{\varepsilon}_{\varepsilon_i \neq 0} \quad \rightarrow \quad \mu_n \approx \bar{\varepsilon}_{\varepsilon_i \neq 0}$$

$$k_n = k_0 + k \quad \rightarrow \quad k_n \approx k$$

$$\nu_n = \nu_0 + k \quad \rightarrow \quad \nu_n \approx k$$

$$\sigma_n^2 = \frac{\nu_0 \sigma_0^2 + (k-1) S_{\varepsilon_i \neq 0}^2 + \frac{k_0 k}{k_0 + k} (\bar{\varepsilon}_{\varepsilon_i \neq 0} - \mu_0)^2}{\nu_0 + k} \quad \rightarrow \quad \sigma_n^2 \approx S_{\varepsilon_i \neq 0}^2$$

$$\alpha_n = \alpha_0 + k \quad \rightarrow \quad \alpha_n \approx k$$

$$\beta_n = \beta_0 + n - k \quad \rightarrow \quad \beta_n \approx n - k$$

Te zien is dat deze a-posteriori parameters nu vrijwel alleen bestaan uit op data gebaseerde parameters. Hierdoor is het mogelijk om het nieuwe model te vergelijken met het huidige. Op deze manier is er immers in het nieuwe model net als in het huidige model geen voorkennis meegenomen.

In de volgende tabel volgt een vergelijking. Het criterium $P(H_0) < 0,05$ kan met de huidige methode niet bepaald worden. De Bayes factor heeft bij gebruik van een a-priori verdeling zonder voorkennis geen toegevoegde waarde en zal daarom in deze tabel niet worden weergegeven.

<i>a-priori verdeling zonder voorkennis</i>	<i>90% bbhi. van de totale fout</i>	<i>Breedte bbhi.</i>	<i>Maximale fout</i>	<i>Mat: $P(H_0 \lambda_{post}) \approx 0.05$</i>
<i>Dataset 1</i>				
<i>Huidige model</i>	<i>[-149.029, 1.171.576]</i>	<i>1.320.605</i>	<i>1.171.576</i>	<i>NA.</i>
<i>Nieuwe model</i>	<i>[-286.018, 1.206.609]</i>	<i>1.492.627</i>	<i>1.206.609</i>	<i>$\approx 1.230.000$</i>
<i>Dataset 2</i>				
<i>Huidige model</i>	<i>[93.060, 1.662.804]</i>	<i>1.569.744</i>	<i>1.662.804</i>	<i>NA.</i>
<i>Nieuwe model</i>	<i>[124.906, 1.649.691]</i>	<i>1.524.785</i>	<i>1.649.691</i>	<i>$\approx 1.720.000$</i>
<i>Dataset 3</i>				
<i>Huidige model</i>	<i>[-439.270, 586.821]</i>	<i>1.026.091</i>	<i>586.821</i>	<i>NA.</i>
<i>Nieuwe model</i>	<i>[-535.307, 500.513]</i>	<i>1.035.820</i>	<i>535.307</i>	<i>≈ 620.000</i>

Tabel 4: Resultaten van het huidige model en het nieuwe model zonder voorkennis.

We zien dat de resultaten van beide modellen redelijk overeen komen. Daarnaast is te zien dat het criterium maximale fout ongeveer dezelfde materialiteitsgrens geeft als het criterium $P(H_0 | \lambda_{post})$. Een vergelijk maken tussen het nieuwe model en het huidige model op basis van de maximale fout en het 90% betrouwbaarheidsinterval blijkt daardoor mogelijk.

8.3.2 Resultaten nieuwe model met voorkennis (2)

Nu we weten dat de modellen vergelijkbaar zijn, wordt bekeken of met behulp van het Bayesiaanse model met voorkennis bij eenzelfde steekproefomvang de breedte van het betrouwbaarheidsinterval verkleind kan worden. Natuurlijk zijn de resultaten erg afhankelijk van de voorkennis en daarmee van de a-priori parameters. Alleen wanneer de voorkennis gunstig is, heeft het zin dit model te gebruiken.

8.3.2.1 Invulling a-priori parameters

Er zijn vele a-priori verdelingen mogelijk, die op basis van verschillende informatiebronnen bepaald kunnen worden. Natuurlijk is een juiste invulling van de a-priori parameters van groot belang voor de resultaten. Daarom volgt hier nogmaals de bij de a-priori parameters behorende interpretatie:

- μ_0 : a-priori verwachting van het gemiddelde van de verschillen ongelijk aan nul.

- k_0 : aantal waarnemingen dat de a-priori verwachting van μ (μ_0) waard geacht wordt.
- ν_0 : aantal vrijheidsgraden voor de a-priori variantie van de individuele verschillen ongelijk aan nul.
- σ_0^2 : a-priori verwachte variantie van de individuele verschillen ongelijk aan nul.
- α_0 : a-priori verwachting van de kans op een verschil ongelijk aan nul vermenigvuldigd met het totaal aantal waarnemingen dat deze a-priori informatie waard geacht wordt.
- β_0 : a-priori verwachting van de kans op een verschil gelijk aan nul vermenigvuldigd met het totaal aantal waarnemingen dat deze a-priori informatie waard geacht wordt.

Om de a-priori verdeling zo objectief mogelijk te houden, is ervoor gekozen om bij dit onderzoek de a-priori parameters zo veel mogelijk op historische data te baseren. Hierbij wordt uitgegaan van gunstige historische data. Er is immers gesteld dat alleen positieve voorkennis wordt gebruikt. Het aantal waarnemingen dat de a-priori informatie waard geacht wordt kan niet uit de historische data gehaald worden. Daarom zal deze aan de hand van een expertmening moeten worden geschat. Dit aantal waarnemingen wordt n_0 genoemd. Vanwege het feit dat in de loop der tijd een bedrijf behoorlijk kan veranderen is ervoor gekozen de parameters te baseren op de resultaten voor het voorgaande jaar. Dit is als volgt gedaan:

- μ_0 : het gemiddelde van de verschillen ongelijk aan nul die in de steekproef van het voorgaande jaar gevonden zijn.
- σ_0^2 : de variantie van de verschillen ongelijk aan nul die in de steekproef van het voorgaande jaar gevonden zijn.
- k_0 : fractie gevonden verschillen ongelijk aan nul in de steekproef van vorig jaar vermenigvuldigd met n_0 .
- ν_0 : $k_0 - 2$.
- α_0 : fractie gevonden verschillen ongelijk aan nul in de steekproef van vorig jaar vermenigvuldigd met n_0 .
- β_0 : fractie gevonden verschillen gelijk aan nul in de steekproef van vorig jaar vermenigvuldigd met n_0 .

Zoals al eerder vermeld is dit niet de enige manier waarop de parameters tot stand kunnen komen. Voor meer informatie over hoe expertmeningen om te zetten in a-priori parameters wordt u verwezen naar de Appendix.

8.3.2.2 Vergelijking huidige model en nieuwe model met voorkennis bij steekproefomvang 200

Vanwege het feit dat n_0 niet bepaald kan worden met behulp van de historische data en dat het daarom op basis van een erg subjectieve informatiebron wordt ingevuld, zijn resultaten berekend voor verschillende waarden van n_0 ($n_0 = 100$, $n_0 = 200$). Omdat de parameters k_0 , ν_0 , α_0 en β_0 afhankelijk zijn van n_0 betekent dit ook dat de resultaten voor verschillende waarden van deze parameters zijn berekend.

$n = 200$	90% bbhi. van de totale fout	Breedte bbhi.	Maximale fout	$P(H_0 \lambda_{post}) \approx 0.05$	Bayes factor
<i>Dataset 1</i>					
<i>Nieuwe model, $n_0=200$</i>	<i>[-56.870, 778.363]</i>	<i>835.233</i>	<i>778.363</i>	<i>≈ 800.000</i>	<i>≈ 4</i>
<i>Nieuwe model, $n_0=100$</i>	<i>[-139.333, 928.832]</i>	<i>1.068.165</i>	<i>928.832</i>	<i>≈ 960.000</i>	<i>≈ 2</i>
<i>Huidige model</i>	<i>[-149.029, 1.171.576]</i>	<i>1.320.605</i>	<i>1.171.576</i>	<i>NA.</i>	<i>NA.</i>
<i>Dataset 2</i>					
<i>Nieuwe model, $n_0=200$</i>	<i>[60.278, 1.143.655]</i>	<i>1.083.377</i>	<i>1.143.655</i>	<i>$\approx 1.160.000$</i>	<i>$\approx 1,3$</i>
<i>Nieuwe model, $n_0=100$</i>	<i>[73.505, 1.323.605]</i>	<i>1.250.100</i>	<i>1.323.605</i>	<i>1.350.000</i>	<i>$\approx 0,6$</i>
<i>Huidige model</i>	<i>[93.060, 1.662.804]</i>	<i>1.569.744</i>	<i>1.662.804</i>	<i>NA.</i>	<i>NA</i>
<i>Dataset 3</i>					
<i>Nieuwe model, $n_0=200$</i>	<i>[-423.992, 476.935]</i>	<i>900.927</i>	<i>476.935</i>	<i>550.000</i>	<i>$\approx 0,17$</i>
<i>Nieuwe model, $n_0=100$</i>	<i>[-477.195, 500.786]</i>	<i>977.981</i>	<i>500.786</i>	<i>600.000</i>	<i>$\approx 0,1$</i>
<i>Huidige model</i>	<i>[-439.270, 586.821]</i>	<i>1.026.091</i>	<i>586.821</i>	<i>NA.</i>	<i>NA</i>

Tabel 5: Vergelijking huidige model en nieuwe model met voorkennis (waarde n_0) en steekproefomvang = 200.

We zien dat de 90% betrouwbaarheidsintervallen voor alle drie de datasets versmallen. Zowel wanneer de a-priori informatie de waarde van 200 waarnemingen krijgt als voor een waarde van 100 waarnemingen. Wat een juiste waardering voor de a-priori informatie is, is voor iedere situatie weer anders; hoe hoger de waardering des te groter de invloed. Ook de maximale fout en de materialiteit waarbij $P(H_0) < 0,05$ zijn gedaald ten opzichte van de resultaten zonder voorkennis. Bij de Bayes

factor zien we dat hoe meer de a-priori kennis meetelt, hoe hoger deze factor. Dit duidt erop dat de a-priori kennis positiever is dan de kennis die uit de waarnemingen gehaald kan worden.

8.3.2.3 *Vergelijking huidige model en nieuwe model met voorkennis bij steekproefomvang < 200*

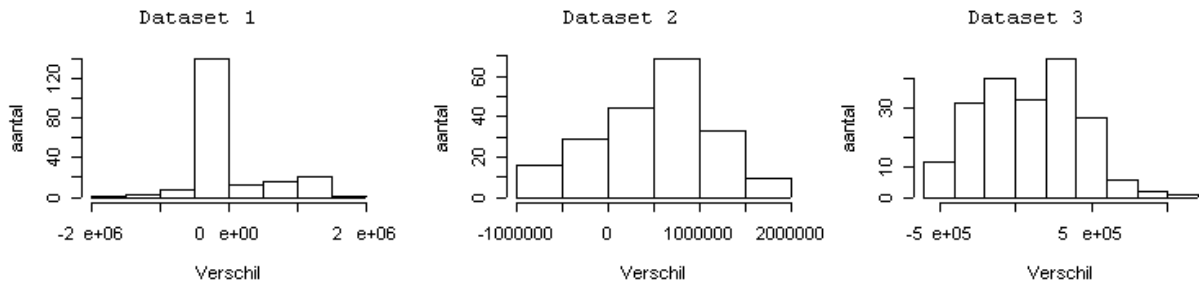
Om nog een wat beter inzicht te krijgen in de invloed van de a-priori verdeling, worden ook het huidige en het nieuwe model vergeleken bij een kleinere steekproefomvang. Het voordeel hiervan is dat we hiermee verschillende trekkingen kunnen simuleren en daarmee beter inzicht kunnen krijgen in de invloed van de a-priori verdeling.

De waarde van de voorkennis en grootte van de steekproefomvang waarvan de resultaten tijdens dit onderzoek berekend zijn, zijn bepaald aan de hand van de bij Deloitte bekende R-waardes. Daarbij is ervan uitgegaan dat het totaal aantal waarnemingen (waarde van de voorkennis + werkelijke aantal waarnemingen) gelijk moet zijn aan het huidige aantal waarnemingen (200). Dit is op de volgende manier gedaan:

- $R = 3,0$. Alle kennis moet uit de steekproef gehaald worden. $n_0 = 0$ en aantal waarnemingen is volledig, daarom $n = 200$. (Deze situatie is al in paragraaf 8.3.1 behandeld en zal daarom hier niet behandeld worden).
- $R = 2,0$. Er is enige voorkennis waarop men kan steunen. $(3,0-2,0)/3,0 * 200 = 67$ waarnemingen is de waarde van de voorkennis waarop gesteund mag worden, $n_0 = 67$. $2,0/3,0 * 200 = 133$ waarnemingen moeten nog gedaan worden om tot het kennisniveau van 200 waarnemingen te komen, $n = 133$.
- $R = 0,7$. Er kan bijna volledig op de voorkennis gesteund worden. $(3,0-0,7)/3,0 * 200 = 153$ waarnemingen is de waarde van de voorkennis waarop gesteund mag worden, $n_0 = 153$. $0,7/3,0 * 200 = 47$ waarnemingen moeten nog gedaan worden om tot het kennisniveau van 200 waarnemingen te komen, $n = 47$

Daarnaast is ook de situatie $n_0 = 100$, $n = 100$ bekeken. Bij de simulatie is gebruikt gemaakt van gepaarde simulatie. Om een vergelijk te kunnen maken tussen het huidige model en het nieuwe model, vergelijken we resultaten van hetzelfde criterium voor de beide methodes. Bij het huidige model worden alleen het betrouwbaarheidsinterval en de maximale fout bepaald. De breedte van het betrouwbaarheidsinterval is een indicatie voor de precisie van het model, daarom wordt dit criterium gebruikt om de modellen te vergelijken. Per gesimuleerde trekking is voor elk van de methodes de breedte van het betrouwbaarheidsinterval bepaald en is het verschil tussen de twee bepaald.

Hieronder zijn de resultaten te vinden. Voor de situatie $n_0 = 100$, $n = 100$ zijn histogrammen van de verschillen in intervalbreedtes weergegeven. Voor de overige situaties zijn de resultaten in tabellen weergegeven.



Figuur 4: Histogram per dataset voor de 200 gesimuleerde verschillen tussen de intervalbreedte van het huidige model ($n = 100$) en het nieuwe model ($n_0=100$, $n=100$).

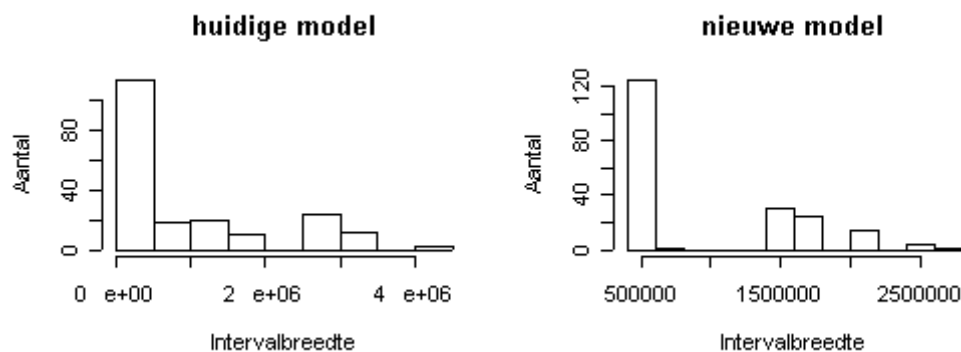
n_0 : voorkennis waarde, n : # waarnemingen	90% bbhi. van verschil in intervalbreedte	gem. verschil intervalbreedte	Fractie verschillen > 0
Dataset 1			
$n_0 = 100, n = 100$	[-772329, 1043359]	-56759	0.24
$n_0 = 67, n = 133$	[-846155, 555609]	-128165	0.23
$n_0 = 153, n = 47$	[-557945, 960999]	-25624	0.39
Dataset 2			
$n_0 = 100, n = 100$	[-572970, 1389815]	492272	0.82
$n_0 = 67, n = 133$	[-386152, 778112]	274239	0.81
$n_0 = 153, n = 47$	[-1075896, 3344240]	720963	0.59
Dataset 3			
$n_0 = 100, n = 100$	[-488197, 665468]	68057	0,54
$n_0 = 67, n = 133$	[-299049, 350787]	3189	0,49
$n_0 = 153, n = 47$	[-787007, 2216601]	416504	0,61

Tabel 6: Resultaten van het verschil in intervalbreedte (huidig – nieuw) van 200 simulaties bij eenzelfde steekproefomvang.

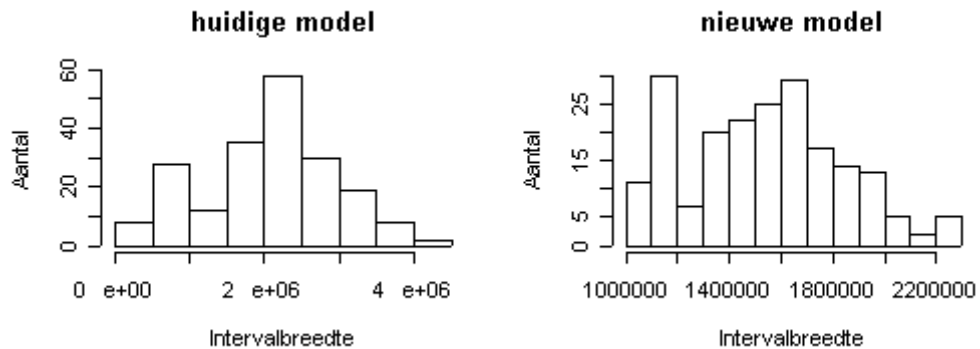
Te zien is dat de intervalbreedtes berekend met de verschillende methodes behoorlijk van elkaar verschillen. Bij de ene dataset is het gemiddelde van het huidige model kleiner, bij de andere dataset juist het gemiddelde van het nieuwe model. Ook per dataset zijn er zowel trekkingen waarbij de

intervalbreedte van het huidige model kleiner dan die van het nieuwe model als omgekeerd. Het blijkt dus dat er behoorlijk wat variatie in één van de of beide modellen zit. Een reden waarom de intervalbreedte van het nieuwe model soms juist groter is dan die van het oude model is dat wanneer de waarnemingen gunstigere resultaten hebben dan de voorkennis, dat de voorkennis het interval kan vergroten. De voorkennis wordt echter wel bestempeld als gunstig, want er is afgesproken dat alleen gunstige voorkennis gebruikt wordt. Dat de waarnemingen meer gunstige resultaten leveren dan de voorkennis en dat de voorkennis in dit geval dus het interval vergroot is geen enkel probleem. De intervalbreedte zal dan immers altijd kleiner worden dan de breedte die aan de hand van de voorkennis kan worden bepaald. Omdat aan de hand van de voorkennis kan worden goedgekeurd, zal er na het doen van de waarnemingen daarom ook kunnen worden goedgekeurd.

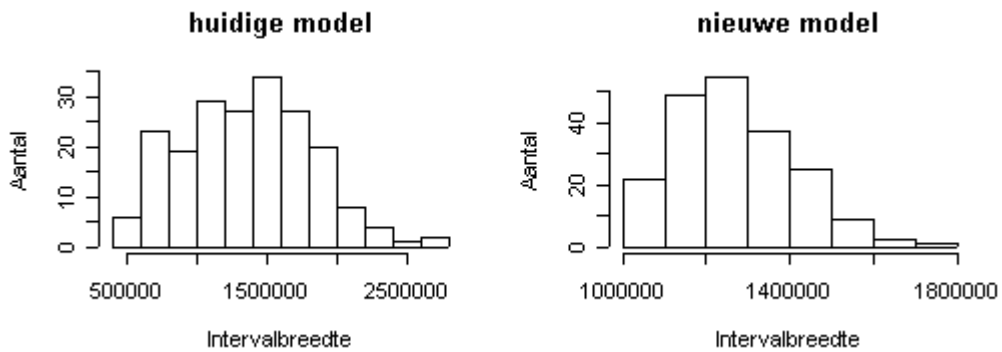
Er is gezien dat de intervalbreedtes van de twee methodes nogal van elkaar kunnen verschillen. Om te bepalen of een model betrouwbaar is, is het ook belangrijk te bepalen wat de spreiding is van de verschillende resultaten. De verschillen worden veroorzaakt doordat er telkens een andere trekking gedaan wordt. Een vergelijk tussen de spreiding van de intervalbreedtes van de huidige en de nieuwe methode zal daarom nu gemaakt worden. Bij deze vergelijking is weer gebruik gemaakt van gepaarde simulatie.



Figuur 5: Histogrammen voor de 200 gepaard gesimuleerde intervalbreedtes voor dataset 1. De eerste aan de hand van het huidige model ($n = 100$), de tweede aan de hand van het nieuwe model ($n_0 = 100, n = 100$).



Figuur 6: Histogrammen voor de 200 gepaard gesimuleerde intervalbreedtes voor dataset 2. De eerste aan de hand van het huidige model ($n = 100$), de tweede aan de hand van het nieuwe model ($n_0 = 100, n = 100$).



Figuur 7: Histogrammen voor de 200 gepaard gesimuleerde intervalbreedtes voor dataset 3. De eerste aan de hand van het huidige model ($n = 100$), de tweede aan de hand van het nieuwe model ($n_0 = 100, n = 100$).

Te zien is dat de spreiding van de intervalbreedtes bij het nieuwe model veel kleiner is dan de spreiding van het huidige model bij steekproefomvang $n = 100$. Om meer informatie hierover te verkrijgen is de standaard deviatie berekend van de intervalbreedtes voor verschillende combinaties van n_0 en n .

n_0 : voorkenniswaarde (voor nieuwe model) n : # waarnemingen	Standaard deviatie intervalbreedte huidige model	Standaard deviatie intervalbreedte nieuwe model
<i>Dataset 1</i>		
$n_0 = 100, n = 100$	1.044.031	636.396
$n_0 = 67, n = 133$	789.088	701.429
$n_0 = 153, n = 47$	1.001.615	449.404

<i>Dataset 2</i>		
$n_0 = 100, n = 100$	900.555	298.831
$n_0 = 67, n = 133$	625.300	327.109
$n_0 = 153, n = 47$	1.630.230	188.801
<i>Dataset 3</i>		
$n_0 = 100, n = 100$	463.681	136.382
$n_0 = 67, n = 133$	369.994	176.803
$n_0 = 153, n = 47$	1.005.991	92.151

Tabel 7: Vergelijking van de standaarddeviaties van de intervalbreedtes tussen het nieuwe model met voorkennis en het huidige model.

Ook deze resultaten geven aan dat het nieuwe model een veel kleinere spreiding oplevert dan het huidige model bij een gelijk aantal waarnemingen. Een verklaring hiervoor is dat de grote verschillen tussen de verschillende trekkingen in het nieuwe model iets afgevlakt worden doordat de voorkennis gelijk blijft. In het huidige model is dit niet het geval en daardoor hebben de verschillen meer invloed.

Inmiddels is aangetoond dat gemiddeld gezien de nieuwe methode een minstens zo smal betrouwbaarheidsinterval oplevert als het huidige model. Bovendien is de spreiding van de mogelijke resultaten een stuk kleiner bij het nieuwe model dan bij het huidige model. Hieruit kan geconcludeerd worden dat voor gelijke steekproefomvang het nieuwe model geprefereerd wordt boven het huidige model (Mits er goede a-priori kennis beschikbaar is).

8.3.3 Vergelijking van het huidige model ($n = 200$) met het nieuwe model ($n < 200$)

Om te kunnen beoordelen of het nieuwe model met minder aantal waarnemingen in gebruik kan worden genomen, dient aangetoond te worden dat dit minstens even goed is als het huidige model met steekproefomvang 200. Er kan een vergelijk gemaakt worden tussen de breedte van het betrouwbaarheidsinterval aan de hand van het huidige model en de gemiddelde breedte van het betrouwbaarheidsinterval voor de nieuwe methode:

<i>N₀: voorkennis waarde, n: # waarnemingen</i>	<i>Gem. breedte 90% bbhi. van de totale fout</i>
<i>n₀ = 100, n = 100</i>	<i>975.883</i>
<i>n₀ = 67, n = 133</i>	<i>1.064.147</i>
<i>n₀ = 153, n = 47</i>	<i>879.470</i>
<i>Huidige model n = 200</i>	<i>1.320.605</i>

Tabel 8: Resultaten van het verschil in intervalbreedte (huidig – nieuw) van 200 simulaties voor Dataset 1

<i>N₀: voorkennis waarde, n: # waarnemingen</i>	<i>Gem. breedte 90% bbhi. van de totale fout</i>
<i>n₀ = 100, n = 100</i>	<i>1.517.933</i>
<i>n₀ = 67, n = 133</i>	<i>1.477.735</i>
<i>n₀ = 153, n = 47</i>	<i>1.504.672</i>
<i>Huidige model n = 200</i>	<i>1.569.744</i>

Tabel 9: Resultaten van het verschil in intervalbreedte (huidig – nieuw) van 200 simulaties voor Dataset 2

<i>N₀: voorkennis waarde, n: # waarnemingen</i>	<i>Gem. breedte 90% bbhi. van de totale fout</i>
<i>n₀ = 100, n = 100</i>	<i>1.286.565</i>
<i>n₀ = 67, n = 133</i>	<i>1.183.070</i>
<i>n₀ = 153, n = 47</i>	<i>1.390.228</i>
<i>Huidige model n=200</i>	<i>1.026.091</i>

Tabel 10: Resultaten van het verschil in intervalbreedte (huidig – nieuw) van 200 simulaties voor Dataset 3

Hieruit blijkt dat deze redelijk overeenkomen. Om te kunnen stellen dat het nieuwe model met voorkennis en kleinere steekproefgrootte minstens even goed werkt als het huidige model met steekproefomvang $n = 200$, zal echter ook naar de spreiding van deze resultaten gekeken moeten worden voor verschillende trekkingen. Daarom zullen trekkingen gesimuleerd worden door te trekking uit de 200 items met teruglegging. Er kan echter geen gepaarde vergelijking gemaakt worden vanwege het feit dat het nieuwe model minder waarnemingen gebruikt dan het huidige.

<i>Dataset 1</i>	
<i>n₀: voorkenniswaarde, n: # waarnemingen</i>	<i>Standaard deviatie interval- breedte (200 simulaties)</i>
<i>Nieuw n₀ = 100, n = 100</i>	<i>636.396</i>
<i>Nieuw n₀ = 67, n = 133</i>	<i>701.429</i>
<i>Nieuw n₀ = 153, n = 47</i>	<i>449.404</i>
<i>Huidige model n = 200</i>	<i>568.389</i>

Tabel 11: Standaard deviatie van de intervalbreedte voor 200 simulaties bij verschillend aantal waarnemingen voor Dataset 1.

<i>Dataset 2</i>	
<i>n₀: voorkenniswaarde, n: # waarnemingen</i>	<i>Standaard deviatie interval- breedte (200 simulaties)</i>
<i>Nieuw n₀ = 100, n = 100</i>	<i>298.831</i>
<i>Nieuw n₀ = 67, n = 133</i>	<i>327.109</i>
<i>Nieuw n₀ = 153, n = 47</i>	<i>188.801</i>
<i>Huidige model n = 200</i>	<i>396.186</i>

Tabel 12: Standaard deviatie van de intervalbreedte voor 200 simulaties bij verschillend aantal waarnemingen voor Dataset 2.

<i>Dataset 3</i>	
<i>n₀: voorkenniswaarde, n: # waarnemingen</i>	<i>Standaard deviatie interval- breedte (200 simulaties)</i>
<i>Nieuw n₀ = 100, n = 100</i>	<i>136.382</i>
<i>Nieuw n₀ = 67, n = 133</i>	<i>176.803</i>
<i>Nieuw n₀ = 153, n = 47</i>	<i>92.151</i>
<i>Huidige model n = 200</i>	<i>256.525</i>

Tabel 13: Standaard deviatie van de intervalbreedte voor 200 simulaties bij verschillend aantal waarnemingen voor Dataset 3.

Aan de standaard deviaties te zien is het nieuwe model met weergegeven n_0 -en n -waardes minstens zo betrouwbaar als het huidige model (Mits de a-priori parameters juist zijn ingevuld).

Voor het nieuwe model met minder waarnemingen is de gemiddelde intervalbreedte vergelijkbaar met de intervalbreedte van het huidige model (bij gegeven n_0). Ook de betrouwbaarheid van de uitkomst is minstens zo hoog als die bij het huidige model. Op basis van deze gegevens lijkt het erop wanneer er goede voorkennis bestaat dat het nieuwe model gebruikt kan worden en dat er minder waarnemingen gedaan kunnen worden.

9 Conclusie

Om een realistisch model voor de voorraadcontrole te ontwikkelen waarbij eventueel de steekproefomvang verkleind kan worden, is allereerst bekeken of het mogelijk zou zijn een Bootstrap methode te ontwikkelen voor de voorraadcontrole, zodat de aanname van normaliteit van de residuen niet meer nodig is. We hebben gezien dat met een Bootstrap methode bij de huidige steekproefomvang de maximale fout (en de onzekerheid over de totale fout) hoger wordt dan bij de huidige methode. Dit is echter geen reden om te stellen dat Bootstrap geen goede methode is voor dit probleem. Om te bepalen of de Bootstrap een goede methode is voor dit probleem, zal verder onderzoek moeten worden gedaan. Als blijkt dat dit het geval is, kan wellicht een Bayesiaanse Bootstrap methode ontwikkeld worden, zodat ook voorkennis meegenomen wordt in de evaluatie. Omdat het aantonen dat de Bootstrap werkt behoorlijk veel tijd en energie zal kosten, is er besloten dit niet te doen tijdens dit onderzoek.

Naast de Bootstrap is een Bayesiaans model ontwikkeld op basis van het huidige model dat toegepast kan worden bij de voorraadcontrole. Een groot voordeel van een Bayesiaans model ten opzichte van het huidige model is dat omdat vooraf al bepaalde verwachtingen gespecificeerd zijn, het geen probleem is als er nul fouten gevonden worden. Het Bayesiaans model kan voor alle mogelijke waarnemingen resultaten weergeven. Een correct Bayesiaans model geeft als uitkomst de kans op een bepaalde hypothese. Aan de hand van deze kans wordt bepaald of de hypothese wel of niet verworpen wordt. Bij Deloitte heerst er echter een grote voorkeur voor de uitkomsten 90% betrouwbaarheidsinterval van de totale fout en maximale fout. Daarom zijn ook deze waardes bepaald voor het nieuwe Bayesiaanse model. Dit is op dezelfde manier gedaan als bij de huidige methode. Vanwege de a-priori verdelingen is het echter niet aannemelijk dat het t-tabel element gebruikt dient te worden. Uit de resultaten blijkt dat het criterium $P(H_0 | \lambda_{post})$ ongeveer dezelfde resultaten oplevert als het criterium maximale fout. Daarom maakt het voor de praktijk in dit geval weinig verschil welk criterium er gebruikt wordt. Omdat bij Deloitte de voorkeur heerst voor het criterium maximale fout (met bijbehorend 90% betrouwbaarheidsinterval voor de totale fout) en omdat met dit criterium een goede vergelijking gemaakt kan worden tussen het huidige en het nieuwe model is er voor gekozen om voor dit criterium resultaten weer te geven. Wanneer er besloten wordt dit nieuwe model in gebruik te nemen wordt aangeraden het criterium $P(H_0 | \lambda_{post})$ te gebruiken omdat dit het correcte criterium is voor een Bayesiaans model.

Voor het ontwikkelde model is aangetoond dat het vergelijkbare resultaten geeft als het huidige model wanneer er geen voorkennis beschikbaar is. Daarnaast is laten zien dat bij gunstige voorkennis, de intervalbreedte versmald wordt en de spreiding van de mogelijke resultaten kleiner is vergeleken bij het huidige model bij gelijke steekproefomvang. Ook is onderzocht of er bij het nieuwe model met een kleinere steekproefomvang voldaan kan worden. Daarvoor zijn de nauwkeurigheid (dus intervalbreedte voor de totale fout) en de betrouwbaarheid (spreiding van de resultaten) van het huidige model en van het nieuwe model met kleinere steekproefomvang vergeleken. Op basis van deze vergelijking lijkt een kleinere steekproefomvang mogelijk.

Op basis van dit onderzoek lijkt gesteld te kunnen worden dat wanneer goede voorkennis beschikbaar is het nieuwe model gebruikt kan worden en dat daarbij met minder waarnemingen kan worden volstaan. Deze uitspraak verdient echter nog wat voorzichtigheid.

Ten eerste zijn de resultaten slechts gebaseerd op drie datasets. Meer onderzoek is nodig om meer zekerheid te verkrijgen over de resultaten.

Ten tweede zijn de resultaten gebaseerd op simulaties uit de 200 al gecontroleerde items en niet uit de totale steekproefpopulatie. Een onderzoek waarbij gebruik gemaakt kan worden van een integraal gecontroleerde dataset. Zou meer zekerheid kunnen bieden over de juistheid van het nieuwe model. Daarbij kan namelijk uit de gehele steekproefpopulatie getrokken worden. Bovendien kunnen de resultaten vergeleken worden met de werkelijke totale fout. Tijdens dit onderzoek is alleen maar vergeleken met de resultaten van het huidige model.

Ten derde moet er goed bij stil gestaan worden dat de werking van het nieuwe model zeer afhankelijk is van de juistheid van de a-priori parameters en daarmee dus van de gebruikte voorkennis. Hiermee staat of valt het model.

De algemene conclusie van dit onderzoek luidt daarom dat met de positieve onderzoeksresultaten de eerste stap naar een nieuw (Baysiaans) model gezet is. Echter voordat het ook daadwerkelijk ingevoerd zou kunnen worden in de voorraadcontrole dient het model meer getest te worden, zodat meer zekerheid over de juistheid van het model kan worden gecreëerd.

Literatuur

- [1] Berger, J.O., 1985, *Statistical Decision Theory and Bayesian Analysis*, Springer.
- [2] Broemeling, L.D., 1985, *Bayesian Analysis of Linear Models*, Marcel Dekker, Inc.
- [3] Cochran, W.G., 1977, *Sampling Techniques*, Wiley.
- [4] Geertman, F., 1994, *Over de toepasbaarheid van regressieschatter en verschilschatter in de accountantscontrole*, Intern Document Deloitte
- [5] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 2004, *Bayesian data analysis*, Chapman & Hall/CRC.
- [6] Van de Genugten, B.B., 1986, *Inleiding tot waarschijnlijkheidsrekening en wiskundige statistiek*, Universiteit van Tilburg
- [7] De Gunst, M.C.M, van der Vaart, A.W., 2001, *Statistische Data Analyse*, Vrije Universiteit, Amsterdam.
- [8] Iverson, G.R., 1984, *Bayesian Statistical Inference*, Sage University Paper series on Quantitative Application in the Social Sciences, 07-043.
- [9] Kleijnen, J.P.C., Kriens, J., Timmermans, H. and Van den Wildenberg, H., 1988, *Regression sampling in statistical auditing*. FEW 306, Universiteit van Tilburg.
- [10] Kriens, J. en Peterse, J.J.M., 1986, *Toepassingen van de regressieschatter in de accountantscontrole*. Kwantitatieve methoden 20.
- [11] Krijkamp, R.W.G., 1994, *Statistische modellen ter ondersteuning van de controleaanpak van Deloitte & Touche*. Universiteit van Tilburg.
- [12] Lew, T.R., 1998, *Regressieschatter, Literatuurstudie naar de praktijkaannames voor toepassen van normaliteit*, Vrije Universiteit, Amsterdam
- [13] Moore, D.S., McCabe, G.P., 1994, *Statistiek in de Praktijk*, Academic Service.

- [14] Steele, A., 1992, *Audit risk and Audit Evidence, The Bayesian approach to statistical auditing*. Academic press limited.
- [15] Stuart, T.R., 2005, *Guam: General Unified Assurance Model*. Intern document Deloitte
- [16] De Vos, A., *De godsdienstoorlog der statistici*,
<http://staff.feweb.vu.nl/avos/ArtikelenNL/Godsdiens/Godsdiens.pdf>
- [17] Zellner, A., 1971, *An introduction to Bayesian inference in Econometrics*, John Wiley & Sons, Inc.
- [18] Zhu, M., Lu, A.Y., 2004, *The Counter-intuitive Non-informative Prior for the Bernoulli Family*, Journal of Statistics Education, volume 12, number 2.
- [19] *UNICON Uniform Controleproces, De controle-aanpak van Deloitte & Touche, Handleiding steekproeven*, Intern document Deloitte.

Appendix

Paragraaf 5.3

Om te bepalen wat de grens materialiteit is waarbij de nulhypothese wel of niet verworpen mag worden hebben we voor een aantal waarden van de materialiteit de Bootstrap uitgevoerd. Op basis hiervan is de grens bepaald.

<i>Mat</i>	<i>Situatie 1</i>		<i>Verwerpen situatie 1 ($T_n < -t$)?</i>	<i>Situatie 2</i>		<i>Verwerpen situatie 2 ($T_n > -t$)?</i>	<i>Verwerp H_0</i>
	<i>-t</i>	<i>T_n</i>		<i>-t</i>	<i>T_n</i>		
50.000	-1,489	0,323	Nee	2,15	2,07	Nee	Nee
60.000	-1,535	0,149	Nee	2,14	2,24	Ja	Nee
130.000	-1,632	-1,068	Nee	2,29	3,45	Ja	Nee
140.000	-1,558	-1,242	Nee	2,37	3,63	Ja	Nee
160.000	-1,651	-1,590	Nee	2,,34	3,97	Ja	Nee
165.000	-1,673	-1,676	Ja	2,34	4,06	Ja	Ja
170.000	-1,688	-1,763	Ja	2,38	4,15	Ja	Ja

Tabel 14: Bootstrap toets voor verschillende waarden van materialiteit op dataset 1.

<i>Mat</i>	<i>Situatie 1</i>		<i>Verwerpen situatie 1 ($T_n < -t$)?</i>	<i>Situatie 2</i>		<i>Verwerpen situatie 2 ($T_n > -t$)?</i>	<i>Verwerp H_0</i>
	<i>-t</i>	<i>T_n</i>		<i>-t</i>	<i>T_n</i>		
1.220.000	-5,12	-1,59	Nee	1,29	3,606	Ja	Nee
1.300.000	-4,824	-1,816	Nee	1,291	3,823	Ja	Nee
1.500.000	-4,250	-2,250	Nee	1,368	4,256	Ja	Nee
2.000.000	-3,345	-3,33	Nee	1,446	-1,916	Ja	Nee
2.100.000	-3,26	-3,55	Ja	1,50	5,56	Ja	Ja
2.200.000	-3,12	-3,77	Ja	1,47	5,77	Ja	Ja
2.500.000	-2,89	-4,42	Ja	1,55	6,43	Ja	Ja

Tabel 15: Bootstrap toets voor verschillende waarden van materialiteit op dataset 2.

<i>Mat</i>	<i>Situatie 1</i>		<i>Verwerpen situatie 1 ($T_n < -t$)?</i>	<i>Situatie 2</i>		<i>Verwerpen situatie 2 ($T_n > -t$)?</i>	<i>Verwerp H_0</i>
	<i>-t</i>	<i>T_n</i>		<i>-t</i>	<i>T_n</i>		
<i>100.000</i>	<i>-4,11</i>	<i>-1,59</i>	<i>Nee</i>	<i>1,57</i>	<i>5,37</i>	<i>Ja</i>	<i>Nee</i>
<i>120.000</i>	<i>-3,88</i>	<i>-2,29</i>	<i>Nee</i>	<i>1,64</i>	<i>6,07</i>	<i>Ja</i>	<i>Nee</i>
<i>150.000</i>	<i>-3,98</i>	<i>-3,33</i>	<i>Nee</i>	<i>1,77</i>	<i>7,11</i>	<i>Ja</i>	<i>Nee</i>
<i>170.000</i>	<i>-4,04</i>	<i>-4,03</i>	<i>Nee</i>	<i>1,85</i>	<i>7,81</i>	<i>Ja</i>	<i>Nee</i>
<i>171.000</i>	<i>-3,99</i>	<i>-4,06</i>	<i>Ja</i>	<i>1,85</i>	<i>7,84</i>	<i>Ja</i>	<i>Ja</i>
<i>175.000</i>	<i>-4,05</i>	<i>-4,20</i>	<i>Ja</i>	<i>1,90</i>	<i>8,00</i>	<i>Ja</i>	<i>Ja</i>
<i>180.000</i>	<i>-3,97</i>	<i>-4,38</i>	<i>Ja</i>	<i>1,89</i>	<i>8,16</i>	<i>Ja</i>	<i>Ja</i>

Tabel 16: Bootstrap toets voor verschillende waarden van materialiteit op dataset 3.

Paragraaf 8.2.3.4

Bewijs dat de a-priori verdeling een geconjugeerde verdeling is voor dit model:

$$\begin{aligned}
 f_{post}(\mu, \sigma^2, p | \lambda_{post}) &= f_{post}(\mu, \sigma^2, p | \lambda_{prior}, e) \propto f_{prior}(\mu, \sigma^2, p | \lambda_{prior}) \cdot f(e | \mu, \sigma^2, p) = \\
 &= f_{prior}(p | \lambda_{prior}) \cdot f_{prior}(\mu, \sigma^2 | \lambda_{prior}) \cdot f(e | \mu, \sigma^2, p) = \\
 &= \left(\frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\alpha_0-1} (1-p)^{\beta_0-1} \right) \cdot \left(\frac{1}{\sqrt{2\pi\sigma^2/k_0}} \exp^{\left\{-\frac{1}{2\sigma^2/k_0}(\mu_0 - \mu)^2\right\}} \cdot \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma^2)^{-\frac{(\nu_0+1)}{2}} \exp^{\frac{-\nu_0\sigma_0^2}{2\sigma^2}} \right) \cdot \\
 &\left((1-p)^{n-k} p^k \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^k \exp^{\frac{-1}{2\sigma^2}((k-1)s_{e_i \neq 0}^2 + k(\bar{e}_{e_i \neq 0} - \mu)^2)} \right) = \\
 &K \cdot p^{\alpha_n-1} (1-p)^{\beta_n-1} \cdot \frac{1}{\sigma} \cdot (\sigma^2)^{-\frac{(\nu_n+1)}{2}} \exp^{\left[-\frac{1}{2\sigma^2}[\nu_n\sigma_n^2 + k_n(\mu_n - \mu)^2]\right]}
 \end{aligned}$$

Waarbij:

$$K = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \cdot \frac{1}{\sqrt{2\pi/k_0}} \cdot \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} \cdot \left(\frac{1}{\sqrt{2\pi}} \right)^k$$

$$\mu_n = \frac{k_0}{k_0 + k} \mu_0 + \frac{k}{k_0 + k} \cdot \bar{e}_{e_i \neq 0}$$

$$k_n = k_0 + k$$

$$\nu_n = \nu_0 + k$$

$$\nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (k-1) \frac{k S_{e_i \neq 0}^2}{T^2} + \frac{k_0 k}{k_0 + k} (\bar{e}_{e_i \neq 0} - \mu_0)^2 \rightarrow \sigma_n^2 = \frac{\nu_0 \sigma_0^2 + (k-1) \frac{k S_{e_i \neq 0}^2}{T^2} + \frac{k_0 k}{k_0 + k} (\bar{e}_{e_i \neq 0} - \mu_0)^2}{\nu_0 + k}$$

$$\alpha_n = \alpha_0 + k$$

$$\beta_n = \beta_0 + n - k$$

Paragraaf 8.3.2

In dit onderzoek zijn een tweetal soorten a-priori parameters gebruikt. Allereerst de parameters waar geen voorkennis in opgenomen is. Deze worden gebruikt om een vergelijk te maken tussen het nieuwe model en het huidige model. Ten tweede a-priori parameters waarin wel voorkennis is opgenomen. Deze voorkennis is bij dit onderzoek alleen gehaald uit data uit het verleden. Wanneer in het voorgaande jaar een steekproef is gedaan, kunnen de a-priori parameters aan de hand van die resultaten ingevuld worden. Wanneer geen steekproef in voorgaande jaren gedaan is en men wil de a-priori parameters toch op basis van voorgaande jaren invullen, kunnen de volgende vragen helpen bij het invullen van de parameters:

1. *Wat was de ingeschatte geprojecteerde fout van het voorgaande jaar? -- > PF*
2. *Wat was de ingeschatte maximale fout van het voorgaande jaar? --> MF*
3. *Wat is de fractie fouten die u verwacht in de populatie? -- > p*
4. *Hoeveel waarde denkt u te kunnen geven aan de zojuist beantwoorde vragen. Met andere woorden: aan hoeveel waarnemingen vindt u de zojuist gegeven kennis gelijk staan? -- > n₀*

De a-priori parameters kunnen nu als volgt ingevuld worden:

- $\mu_0: \frac{PF}{p \cdot T}$
- $\sigma_0^2: \left(\frac{MF - PF}{t_{\alpha, v} \cdot T \cdot p} \right)^2 \cdot n_0 p$
- $k_0: n_0 \cdot p$
- $v_0: (n_0 \cdot p) - 2$
- $\alpha_0: n_0 \cdot p$
- $\beta_0: n_0 \cdot (1 - p)$

Waarbij:

T : het aantal items in de populatie van het voorgaande jaar.

$t_{\alpha, v}$: het t-tabel element voor onbetrouwbaarheid α en aantal vrijheidsgraden v . $\alpha = 0,05$ en $v = v_0$

Natuurlijk is historische dat niet de enige informatiebron waaruit voorkennis gehaald kan worden. Een andere belangrijke bron van informatie is expertkennis. Het probleem hierbij is vaak dat het moeilijk is deze informatie te kwantificeren. Daarom volgen nu een aantal vragen die gesteld kunnen worden om het kwantificeren van de informatie te vergemakkelijken:

1. *Wat is de fractie fouten die u verwacht bij deze populatie?*
2. *Wat is de verwachte totale fout volgens u in deze populatie?*
3. *Hoe groot acht u de variantie van de verschillen die ongelijk aan nul zijn?*
4. *Hoe zeker bent u over de zojuist gegeven antwoorden in termen van aantal waarnemingen.*
Met andere woorden: aan hoeveel waarnemingen vindt u uw kennis over deze populatie gelijk staan?

Met behulp van deze vragen kan op de volgende manier een schatting voor de a-priori parameters gemaakt worden:

- μ_0 : antwoord vraag 2 / (antwoord vraag 1 * aantal items in populatie). (Het aantal items in de populatie wordt bekend verondersteld).
- σ_0^2 : antwoord vraag 3. Vraag 3 is nogal een lastige vraag. Vooral voor niet wiskundig ingestelde mensen is het vaak niet mogelijk deze vraag te beantwoorden. Om toch een redelijke schatting voor σ_0^2 van de expert te kunnen verkrijgen zou de expert verzocht kunnen worden een soort van tekening van een normale verdeling te maken van de fout. Aan de hand daarvan is het wellicht mogelijk een schatting te maken.
- k_0 : antwoord vraag 1 * antwoord vraag 4.
- v_0 : (antwoord vraag 1 * antwoord vraag 4) - 2.
- α_0 : antwoord vraag 1 * antwoord vraag 4.
- β_0 : (1 - antwoord vraag 1) * antwoord vraag 4.

De a-priori parameters baseren op een combinatie van informatie uit historische data en informatie van experts is ook mogelijk. Het meest logische is om ze te combineren op de manier waarop de a-priori parameters en de parameters uit de waarnemingen ook gecombineerd worden.

