

Statistical Audit

Een overzicht van onderzoek over de afgelopen 10 jaar

Auteurs¹: H. H.W. Kloosterman RE RA

Drs. G.B. Broeze

Leden van de projectgroep Research Statistical Auditing van het Limperg Instituut.

(Bestemd voor MAB; niet geplaatst).

Inleiding

Onder de vlag van de projectgroep Research Statistical Auditing van het Limperg Instituut² is onderzoek verricht naar verschillende mogelijkheden om resultaten uit steekproeven te evalueren. Wij geven een overzicht van dat onderzoek en duiden het belang ervan tegen de achtergrond van het vakgebied accountantscontrole.

De projectgroep Research Statistical Auditing heeft als doelstelling onderzoek te laten uitvoeren om de toepassing van statistische methoden en technieken in auditing te verbeteren. De projectgroep is multidisciplinair. Hij bestaat uit leden die zich bezighouden met het toepassen en adviseren op het gebied van Statistical Auditing en leden die zich voornamelijk bezig houden met wetenschappelijk onderzoek naar en het toepassen van statistiek.

In dit artikel besteden wij aandacht aan de onderzoeksprojecten van en namens de Stuurgroep die zich hebben gericht op de volgende vragen:

- Hoe kan materialiteit of controletolerantie voor een verantwoording als geheel worden toegedeeld aan een deel van die verantwoording.
- Hoe kunnen onderzoeksresultaten op deelpopulaties worden vertaald naar resultaten over de samengestelde populatie.
- Is het wiskundig aantoonbaar dat de Stringer Bound conservatief is.
- Zijn er methoden om het conservatisme van de Stringer Bound te reduceren.
- Onder welke voorwaarden is het mogelijk een uitspraak in monetaire eenheden te doen op basis van een postensteekproef.

Omdat accountants het feitelijk gebruik van steekproeven in de financiële controle nog maar beperkt toepassen beginnen wij dit artikel met een beknopte inleiding Statistical Audit. In deze inleiding proberen wij al de onderzoeksvragen op hun plaats te laten

¹ De auteurs danken Drs. P.C. van Batenburg voor de substantiële bijdrage die hij aan dit artikel heeft geleverd.

² Het Limperg Instituut is het interuniversitair onderzoeksinstituut op het gebied van bedrijfseconomie en accountancy.

vallen. Daarna laten wij de resultaten van de genoemde projecten zien. Per project geven wij het belang voor de effectiviteit en de efficiency van de controle aan.

Statistical audit

Statistical Audit beschrijft de controle met behulp van statistische methoden en technieken. Wij gaan er in deze beknopte beschrijving van uit dat de audit op een financiële verantwoording plaatsvindt. Als voorbeelden van financiële verantwoording noemen wij de jaarrekening, de financiële rekening van een overheidsorgaan en belastingaangiften.

Transacties

Een financiële verantwoording is geaggregeerd uit de vastleggingen van de transacties van de betreffende organisatie. De accountant controleert de verantwoording. Hij wil dan weten of alle transacties zijn vastgelegd (kan niet met steekproeftheorie worden beschreven), of alle relevante eigenschappen van die transacties zijn vastgelegd (ook hier biedt steekproeftheorie geen soelaas) en of de waarden van die eigenschappen van de vastgelegde transacties juist zijn vastgelegd (en dit is het onderzoeksgebied van statistical audit!). Een financiële verantwoording kan schattingen bevatten, zoals bijvoorbeeld afschrijvingen of voorzieningen. In dit artikel abstraheren wij daarvan. Het vakgebied Statistical Audit geeft met behulp van statistiek het onderzoek naar de juistheid van de gegevens over de transacties vorm. Wij gaan er daarbij vanuit dat het juistheidsvraagstuk kan worden beantwoord door de controle van de vastgelegde gegevens ('Ist'-positie) met de brongegevens ('Soll'-positie).

Steekproeven op geldseenheden

De statistical auditor vertaalt vervolgens deze vastleggingen van transacties in gepresenteerde geldeenheden. Immers als hij zo ver is gekomen kan hij het controleproces uitdrukken in kwantitatieve grootheden zoals materialiteit en controletolerantie (als maatstaven van onnauwkeurigheid) en steekproefrisico of accountantscontroleerisico (als onzekerheidsmaatstaf).

In statistical auditing pleegt onder materialiteit te worden verstaan de onnauwkeurigheid die de gebruiker net niet³ op het verkeerde been zet⁴.

³ Voor de verdere omschrijvingen is 'net wel' misschien beter.

⁴ Arens & Loebecke schrijven in *Auditing, an integrated approach*; 3rd ed., blz 41: 'A misstatement in the financial statements can be considered material if knowledge of the misstatement would affect a decision of a reasonably intelligent user of the statements.' Dat kan men geld zowel als niet-financieel interpreteren. Arens en Loebecke definiëren hiermee 'materieel' (of materialiteit als 'material misstatement') vanuit de optiek van de gebruiker. Zij geven daarmee de financiële norm aan die het publiek aan een financiële verantwoording stelt.

Controletolerantie is de onnauwkeurigheid die de onderzoeker toestaat. Statistical Audit beschijft het steekproefrisico of accountantscontrolerisico als de kans op ten onrechte goedkeuren van een slechte populatie⁵.

Als de accountant het juistheidsonderzoek wil inrichten met behulp van steekproeven op geldeenheden, wanneer heeft hij dan voldoende werk verricht om (in beginsel) te kunnen goedkeuren? Als het onderzoek redelijke zekerheid biedt dat de gebruiker niet op het verkeerde been wordt gezet. Het element 'redelijke zekerheid' zien we terug in het accountantscontrolerisico (of het complement: Assurance). De feitelijke onnauwkeurigheid die met behulp van statistiek uit de bevindingen wordt berekend, de maximale fout, geeft aan of de gebruiker al dan niet op het verkeerde been kan worden gezet. Als bij de planning ervan is uitgegaan dat de steekproef geen fouten mag bevatten, wat meestal wordt geadviseerd, en de steekproef blijkt inderdaad geen fouten te bevatten is de maximale fout gelijk aan de controletolerantie. Daaruit mag volgen dat de controletolerantie ten hoogste de materialiteit mag bedragen.

Is er sprake van de controletolerantie of zijn er meer?

Een van de onderzoeksprojecten gaat over het al dan niet toedelen van controletolerantie. Dat betekent dat de vraag luidt in termen van statistical audit: is er bij een onderzoek naar een financiële verantwoording sprake van één populatie guldens, of moet er rekening gehouden met afzonderlijke deelpopulaties (vastleggingen van) transacties?

Samenvoegen van steekproefresultaten op deelpopulaties

Er is veel onderzoeksenergie gestoken in de vraag of en hoe steekproefresultaten uit deelpopulaties kunnen worden geaggregeerd. Dit onderzoek is vooral van belang als afzonderlijke delen blijken te kunnen worden goedgekeurd en de som der delen niet hun voorbeeld lijkt te kunnen volgen.

Conservatieve berekeningswijzen: grote steekproeven

Tot nu toe hanteert men bij de berekening van de maximale fout een aantal conservatieve berekeningswijzen⁶. Als de omvang van de steekproef tot stand komt door

⁵ Vaak laat men de voorwaarde 'van een slechte populatie' weg. Dat weglaten maakt het moeilijker de betekenis van het complement, vaak Assurance' genoemd, te beschrijven. Sommigen vinden dat 'Assurance' terecht goedkeuren van een goede populatie is. Dat is vreemd nu het complement is 'de kans op terecht niet goedkeuren van een slechte populatie'.

⁶ Wij plegen uit te gaan van 5% steekproefrisico en de Poissonverdeling. We staan geen fouten toe in de steekproef. De R-factor uit de Poissonverdeling is dan ongeveer 3. Als de controletolerantie (MP) 450.000 is dan wordt de omvang van het te hanteren interval (MP/R

nul fouten in de steekproef toe te staan, pleegt men hem al gauw aan de grote kant te vinden. Als het mogelijk moet zijn meer fouten toe te staan dan wordt de hoeveelheid werk wel als erg, misschien te, groot ervaren.

Bij steekproeven in geldeenheden doet zich het fenomeen voor dat niet de hele post waartoe de getrokken geldeenheid behoort fout hoeft te zijn. We kunnen dan de fout in de post uitdrukken in de boekwaarde, de Ist-positie, van de post. De gedeeltelijke fout noemen we ook wel tainting of taint. De gedeeltelijke fout t_i definiëren we als:

$$t_i = \frac{BV - AV}{BV}$$

Hierbij geldt:

BV = Book Value (boekwaarde = Ist-positie)

AV = Audit Value (auditwaarde = Soll-positie).

Deze tainting is nu de mate waarin de getrokken gulden fout is.

Als er sprake is van meerdere fouten, gedeeltelijke fouten en gehele fouten, is door K.W. Stringer een methode bedacht om de statistische bovengrens te berekenen. Deze methode is uitgewerkt door Leslie, Teitlebaum en Anderson⁷ en door hen TARS (Tainted Attribute Random Sampling) genoemd. Wij verwijzen naar bijlage 1 waarin de berekeningswijze is toegelicht.

De Stringermethode werd na introductie door velen argwanend bekeken. De vraag was of wel de nominale zekerheid (bijvoorbeeld ten hoogste 5% kans op ten onrechte goedkeuren van een slechte populatie) wordt gehaald. Simulatiestudies hebben daarna aangetoond dat de methode conservatief is.

De vraag is of analytisch⁸ kan worden aangetoond dat de methode leidt tot te hoge bovengrenzen en daardoor te grote steekproeven.

Zijn er methoden om conservatisme Stringer Bound te reduceren

Wij gaan ervan uit dat de materialiteit van een financiële verantwoording vaststaat⁹ en er op basis van een daaruit afgeleide controletolerantie een steekproefomvang is ontworpen. Als het steekproefonderzoek fouten laat zien, dan heeft het zin een zo laag mogelijke bovengrens, de maximale fout, te berekenen waarbij goedkeuring nog tot de mogelijkheden behoort. De goedkeuring houdt dan in dat de stakeholders nog steeds

= 450.000 / 3 =) 150.000. Als de populatie 150.000.000 groot is wordt de steekproefomvang 1000. Deze omvang wordt als (te) groot ervaren.

⁷ Leslie, Teitlebaum and Anderson, Dollar Unit Sampling, Pitman 1979

⁸ Analytisch is hier bedoeld als: met behulp van wiskundige analyse.

⁹ Wij beschouwen de materialiteit als een exogeen gegeven. Het gegeven komt in beginsel niet uit de koker van de accountant of van het bedrijf maar van de stakeholders.

met redelijke zekerheid mogen aannemen dat het bedrag van de fout in de verantwoording niet het bedrag van de materialiteit overschrijdt.

Steken op posten uitspraak in geldeenheden

De literatuur op het gebied van Statistical Audit staat bol van het steken op posten en het trekken van conclusies in geldeenheden. Wij noemen standaardwerken als die van Herbert Arkin¹⁰, Miklos A. Vasarhelyi en Thomas W. Lin¹¹ en Dan M. Guy¹². Leslie, Teitlebaum en Anderson stellen dat de accountant die steekproeven wil uitvoeren en een goede schatting in geldeenheden wil maken gewoon steekproeven op geldeenheden moet trekken. In de praktijk blijkt dat er omstandigheden zijn die het toepassen van steekproeven op geldeenheden danig bemoeilijken. Daarnaast zijn er praktische omstandigheden die het mogelijk maken een postensteekproef te trekken en een redelijk nauwkeurige uitspraak te doen in geldeenheden. Over de laatste categorie is al heel wat in de literatuur verschenen. Wij vatten een overzicht ervan samen in bijlage 2.

Het onderzoek dat de projectgroep heeft doen plaatsvinden heeft zich gericht op steekproeven op populaties waarin sprake is van geringe foutdichtheden. Zeg maar populaties waar bij voorkeur op geldeenheden zou zijn getrokken.

De onderzoeken

Wij behandelen achtereenvolgens de onderzoeken waar de leden van de projectgroep bij zijn betrokken:

- Toedelen controletolerantie voor een verantwoording als geheel aan een deel ervan
- Samenvoegen van onderzoeksresultaten op deelpopulaties tot resultaten die gelden voor de samengestelde populatie
- Wiskundige aantoonbaarheid conservatisme Stringer Bound
- Methoden reductie conservatisme Stringer Bound
- Uitspraken in monetaire eenheden op basis postensteekproeven

¹⁰ Herbert Arkin; Handbook of Sampling for Auditing and Accounting; third edition, 1984, McGraw-Hill Inc, Colorado Springs

¹¹ Miklos A. Vasarhelyi, Thomas W. Lin; Advanced Auditing, Fundamentals of EDP and Statistical Audit Technology; 1988, Addison Wesley, New York

¹² Dan M. Guy, Douglas R. Carmichael, O. Ray Whittington; Audit Sampling, an Introduction; fourth edition, 1998, John Wiley & Sons Inc, New York

Toedelen controletolerantie

De stuurgroep noemt dit project het 'Allocatieproject'. Prof. J.H. Blokdijk RA heeft over de toedeling van controletoleranties¹³ geschreven dat toedelen van de tolerantie voor de verantwoording als geheel niet nodig is. Lang is gedacht dat bij het toedelen van controletolerantie aan deelverantwoordingen, de controletolerantie in gulden per deelverantwoording scherper genomen moet worden dan die voor de gehele verantwoording. De belangrijkste vaststelling in dit project is, dat dit type toedeling in veel gevallen vervangen kan worden door 'geen toedeling', dat wil zeggen tot een toedeling die voor elke deelverantwoording dezelfde controletolerantie (in gulden) toekent als die voor de hele verantwoording geldt. Accountants spreken doorgaans en een oordeel uit over 'de verantwoording als geheel', en een oordeel over 'de samenstelling ervan'. Dat kan betekenen dat de accountant op sommige delen een kleinere onnauwkeurigheid wil toestaan. Dit houdt in dat de accountant een verzameling onderzoeken heeft verricht op deelpopulaties en de controletolerantie niet voor alle deelpopulaties gelijk hoeft te zijn. Doet zich die situatie voor dan vormt de grootste controletolerantie van de delen, de tolerantie voor de populatie als geheel.

Blokdijk geeft aan dat bij een nulfoutenverwachting de planning conform het vorenstaande kan plaatsvinden. Als dan inderdaad de successievelijke steekproeven foutloos blijken, verloopt de evaluatie conform de planning. De (grootste) controletolerantie is de maat voor de bovengrens van de fouten in het totaal van de onderzochte deelpopulaties, ofwel de populatie als geheel.

De problematiek wordt complex zodra er in steekproeven op deelpopulaties fouten worden aangetroffen.

De belangrijke winst door de uitkomst van dit onderzoek is dat een instrument als de wortelformule overbodig is geworden. Dat instrument heeft gezorgd voor rekenintensieve begeleiding bij het toepassen van steekproeven in de controle. De vereenvoudiging door het niet meer hoeven toerekenen van controletoleranties maakt dat de toepassing van steekproeven op monetaire eenheden in beginsel breder ingezet kan worden.

Samenvoegen onderzoeksresultaten

De stuurgroep noemt dit combinatieprobleem het 'Samenvoegingsprobleem'.

Een van de gestelde vragen is wat de invloed is op het oordeel over de populatie als geheel, als in het onderzoek op de afzonderlijke deelpopulaties toleranties van

¹³Prof. J.H. Blokdijk RA; Toedeling van de controletolerantie bij de accountantscontrole. MAB juli/augustus 1996. Blz. 350 e.v.

verschillende omvang zijn gebruikt en er sprake is van steekproeven waar fouten in voorkomen.

Steele en Van Batenburg

Een brandend deelprobleem ter zake is de situatie dat alle deelpopulaties afzonderlijk kunnen worden goedgekeurd maar dat een overall evaluatie niet tot goedkeuring leidt. Dit is een dilemma dat door Neter, Kim en Graham¹⁴ en door Boritz, Zhang en Aldersley¹⁵ op schrift gezet is. Neter c.s. schrijven dat de bevindingen uit alle deelpopulaties op een hoop moeten worden gegooid voordat de beslissing kan worden genomen of goedkeuring mogelijk is. De publikatie van A. Steele en P.C. van Batenburg¹⁶ stelt dat de oplossing van Neter c.s. informatie weggooit. Steele en Van Batenburg bieden een oplossingsrichting voor dit probleem. De oplossingsrichting wordt gezocht in informatie-equivalenten. Dat komt erop neer dat de uitkomst van een steekproef met k fouten kan worden omgerekend naar een uitkomst met nul fouten. De auteurs staan deze methode alleen toe als de deelpopulaties heterogeen zijn (de fouten in deelpopulatie A kunnen niet in deelpopulatie B optreden). Als niet aan die voorwaarde voldaan is worden de deelpopulaties als samenstellers van de grotere populatie gerekend; de fouten rekent men in dat geval, net als Neter c.s. toe aan de samengestelde populatie. Steele en Van Batenburg betogen verder nog dat wil de accountant niet (althans beheerst) tot ten onrechte goedkeuren van de (totale) populatie komen, heterogeniteit moet worden aangetoond en homogeniteit als aanwezig moet worden verondersteld.

Schapendonk-Maas

Binnen het kader van de projectgroep Research Statistical Auditing heeft mevrouw Helma Schapendonk-Maas¹⁷ onderzoek gedaan naar onder meer het combineren van bevindingen uit verschillende (deel-)populaties. Mevrouw Schapendonk heeft door middel van simulatiestudie aangetoond dat bij enkelvoudige steekproeven de evaluatie volgens Stringer Bound, EVA-methode en 'my-dollar-right-or-wrong' allemaal (macro

¹⁴ John Neter, Hyo Seuk Kim, Lynford E. Graham; On Combining Stringer Bound for Independent Monetary Unit Samples from Several Populations; 1984, Auditing A Journal of Theory and Practice, Vol. 4, no 1

¹⁵ J. Efrim Boritz, Ping Zhang, Steve Aldersley; On Combining Evidence from Subpopulations into a Composite Conclusion; 1993 Contemporary Accounting Research, Vol. 10 no 1, 227-245.

¹⁶A. Steele and P.C. van Batenburg; Aggregation of risk and the allocation of materiality: more on combining evidence from sub-populations into a composite conclusion. Paper verstrekt bij 21st annual congress of the European Accounting Association, Antwerpen ,april 1998.

¹⁷Helma Schapendonk-Maas, Betrouwbaarheidsbovengrenzen en steekproeftechnieken in de accountancy, februari 1998; Research Memorandum, Limperg Instituut

gezien) voldoende coverage hebben. In de coverage-score is niet de accountantsvraag begrepen of en in hoeveel gevallen er sprake geweest zou zijn van ten onrechte goedkeuren. Die vraag is wel beantwoord t.a.v. de Stringer Bound en EVA-methode in het artikel van P.C. van Batenburg, A. Lucassen en H. Moors¹⁸. Het onderzoek naar de combinatie van Stringer Bounds ingeval van controle van meerdere deelpopulaties wijst in de richting van meerdere valide (in de context van accountantscontrole) evaluatiemethoden.

Conservatisme Stringer Bound; wiskundig onderzoek

Dr. M.C.A. van Zuijlen en Dr. G. Pap hebben meerdere keren gepubliceerd over de theoretische basis onder de Stringer Bound. Het eerste onderzoek ging over hoe conservatief de Stringer bovengrens is als aangenomen wordt dat de taintings uniform verdeeld zijn¹⁹. Een ander onderzoek beschrijft het gedrag van de Stringer bovengrens in het geval de steekproefomvang onbepaald groot wordt²⁰. Beide schrijvers hebben een derde paper over het onderwerp gepubliceerd samen met Drs. N.G. de Jager RA²¹. Uit de publicaties blijkt dat de Stringer Bound asymptotisch conservatief blijkt te zijn voor berekende betrouwbaarheden²² groter dan of gelijk aan vijftig procent. 'Asymptotisch' betekent 'voor een steekproefomvang oplopend naar oneindig'. Conservatief betekent in dit verband dat in feite een te hoge bovengrens wordt berekend, zodat een goedkeurend oordeel met een grotere dan de 'normatieve' betrouwbaarheid wordt gegeven. Dit leidt tot meer werk dan strikt nodig zou zijn om tot een goedkeurend oordeel te komen. Opmerkelijk is de bevinding dat voor betrouwbaarheden onder de vijftig procent het conservatisme omslaat in optimisme, dat wil zeggen dat een uitspraak gedaan wordt met een betrouwbaarheid die ligt onder de normatieve. Betrouwbaarheden onder de vijftig procent kunnen voorkomen doordat sommige vormen van risicoanalyse dergelijke lage betrouwbaarheden uitrekenen en het voorkomen van fouten toestaan. Die vormen van risicoanalyse dienen derhalve in heroverweging te worden genomen.

¹⁸A. Lucassen, H. Moors en P.C. van Batenburg, Modifications of the Stringer Bound: a simulation study on the performance of audit sampling evaluation methods, *Kwan titatieve methoden*, nr. 51, 1996, pp. 17-25

¹⁹ M.C.A. van Zuijlen en G. Pap; On the Stringer Bound in case of Uniform Taintings; 1995, *Computers Math. Applic.* 29, 51 – 59.

²⁰ M.C.A. van Zuijlen en G. Pap; On the Asymptotic Behaviour of the Stringer Bound, 1994, *Research Memorandum*, Katholieke Universiteit Nijmegen.

²¹ N.G. de Jager, G. Pap, M.C.A. van Zuijlen, Facts, phantasies and a new proposal concerning the Stringer bound, *Computers Math. Applic.*, vol 33, no. 5, 1997, pp. 37-54. Ook verschenen als *Research Memorandum* van het Limperg Instituut te Amsterdam (september 1996).

²² Betrouwbaarheid betekent hier steeds: de kans op ten onrechte goedkeuren van een slechte populatie.

De publicaties leiden niet tot een analytisch bewijs voor de Stringer Bound, wel tot een analytisch bewijs voor het conservatisme, en daarmee de geoorlooftheid, van de Stringer Bound, mits het steekproefrisico niet groter wordt dan 50%.

Methoden reductie conservatisme Stringer Bound

Er zijn meerdere onderzoeken verricht naar het reduceren van het conservatisme van de Stringer Bound. Van Zuijlen, Pap en De Jager²³ hebben in hun artikel een richting aangeduid waar de reductie van het conservatisme gezocht kan worden. In 1993 heeft G.H. Van uitkomsten van onderzoek gepresenteerd waarbij hij met bootstrapping²⁴ probeert het conservatisme te beperken. Verder noemen wij het artikel van Lucassen, Moors en Van Batenburg²⁵ die de taintings in een andere volgorde sorteren. Wij eindigen deze paragraaf met de bevindingen van mevrouw Schapendonk.

M.C.A. van Zuijlen, G. Pap en N.G. de Jager

In 1996 hebben Van Zuijlen, Pap en De Jager²⁶ het wiskundig bewijs van het asymptotisch gedrag van de Stringer Bound aangevuld met een voorstel tot aanpassing van de Stringer Bound. Het toepassen van dit voorstel levert bij evaluatie een reductie van de bovengrens op. Het artikel heeft als verdienste dat het behalve het theoretisch bewijs van het conservatisme van de Stringer Bound ook een praktisch handvat biedt voor de praktijk. Het artikel sluit namelijk af met de algorithmes om de aanpassing van de Stringer Bound in een spreadsheet te verwerken.

²³N.G. de Jager, G. Pap, M.C.A. van Zuijlen, Facts, phantasies and a new proposal concerning the Stringer bound, *Computers Math. Applic.*, vol 33, no. 5, 1997, pp. 37-54. Ook verschenen als Research Memorandum van het Limperg Instituut te Amsterdam (september 1996).

²⁴ Bootstrapmethoden komen neer op het veelvuldig steekproeven trekken op de getrokken steekproef. Dit zijn methoden waarbij geen aanname wordt gedaan over de kansverdeling van de populatie waaruit de steekproef is getrokken. De methoden baseren zich geheel op de frequentieverdeling van de elementen (in dit geval de taintings van een guldenssteekproef) in de steekproef (de zogenaamde 'empirische verdeling').

²⁵A. Lucassen, H. Moors en P.C. van Batenburg, Modifications of the Stringer Bound: a simulation study on the performance of audit sampling evaluation methods, *Kwan titatieve methoden*, nr, 51, 1996, pp. 17-25

²⁶N.G. de Jager, G. Pap, M.C.A. van Zuijlen, Facts, phantasies and a new proposal concerning the Stringer bound, *Computers Math. Applic.*, vol 33, no. 5, 1997, pp. 37-54. Ook verschenen als Research Memorandum van het Limperg Instituut te Amsterdam (september 1996).

G.H. Van

In 1993 heeft G.H. Van²⁷ uitkomsten van simulatieonderzoek gepresenteerd. Hij heeft hierbij bootstraptechnieken gebruikt. Onderwerp van onderzoek zijn op taintings gebaseerde schatters in een steekproef op monetaire eenheden.

Van heeft onderzocht of zogenaamde bootstrapmethoden de Stringer Bound (of liever: een van taints afgeleide foutbovengrens) minder conservatief kunnen maken. Hij zoekt als het ware die betrouwbaarheidsbovengrens voor de schatting, die hoort bij de gevraagde betrouwbaarheid.

De methode vraagt om intensief rekenwerk. De technische uitwerking ervan voert te ver voor dit artikel. Ter illustratie: een steekproef met een omvang van 50 levert een verzameling van mogelijke steekproeven van omvang 50, met teruglegging getrokken, van 50^{50} stuks! Door 'slechts' 1000 (of een ander groot aantal) steekproeven uit de gegeven steekproef te trekken, kan echter de gevraagde grootte worden benaderd. Dit is wat in het algemeen in bootstrapmethoden gebeurt: 'resampling' teneinde de 'gewenste betrouwbaarheid' te handhaven.

Van kwam tot de conclusie dat voor bepaalde typen verdeling van de taintings en voor een niet al te klein aantal fouten in de steekproef (meer dan vier) een zogenoemde studentized bootstrapschatter niet te conservatief en ook niet te optimistisch is voor het schatten van de betrouwbaarheidsbovengrens.

Het onderzoek heeft conservatisme van de Stringer Bound bevestigd. Naast de bootstrapgrenzen, is ook steeds een Stringer Bound uitgerekend. Hiervan konden eveneens de eigenschappen worden nagegaan. Bijna steeds bleek het feitelijke overdekkingspercentage, d.w.z. het percentage intervallen dat de werkelijke fout bevat veel hoger dan het berekende. Vaak bijna 100% als 95% betrouwbaarheid als uitgangspunt genomen was.

Verder bleek dat de studentized bootstrapschatter onbetrouwbaar is als er veel taintings zijn met een omvang 1. In de simulatiestudie is gebleken dat dit leidt tot overdekkingspercentages die onder de nominale betrouwbaarheid liggen, m.a.w. de betrouwbaarheidsgrenzen vallen in deze situatie ook met de gebruikte bootstrap methoden te optimistisch uit.

Het onderzoek heeft ruimschoots bevestigd dat van alle typen verdelingen die zijn onderzocht, de intervallen berekend met de normale verdeling veel te optimistisch zijn: in bijna alle gevallen lag het overdekkingspercentage rond 10% onder de normatieve betrouwbaarheid. Er kwamen uitschieters voor met verschillen tot 30% (bij een steekproefgrootte van 50).

²⁷ G.H. Van: Bootstrapmethoden in auditing; TU Delft / Algemene Rekenkamer, Den Haag, maart 1993

De conclusie uit het onderzoek luidt dan ook dat een steekproef op geldseenheden in beginsel opgezet moet worden op basis van de Stringer Bound, daar in het algemeen vooraf slechts een gering aantal fouten wordt verwacht. De evaluatie van de populatie kan worden uitgevoerd met de Stringer Bound met als alternatief de Studentized Bootstrap methode. Die methode leidt tot een lagere bovengrens waardoor mogelijkwerwijs goedkeuring nog mogelijk is.

Lucassen, Moors, Van Batenburg

A. Lucassen, H. Moors en P.C. van Batenburg²⁸ hebben onderzoek gepubliceerd naar een alternatief voor de Stringer Bound. Zij hebben (simulatie)onderzoek gedaan waarbij de taintings op een andere wijze zijn gesorteerd dan de voorzichtige wijze die Stringer heeft genoemd. De Stringer Bound wordt gecreëerd door de taintings in aflopende volgorde te sorteren. Lucassen, Moors en Van Batenburg creëren de bovengrens door de taintings te sorteren op aflopende volgorde van het foutbedrag in geld.

Dit levert een lagere bovengrens op en een voldoende 'coverage', dat wil zeggen de nominale kans op ten onrechte goedkeuren wordt niet overschreden.

Ook bij deze methode geldt als voordeel dat de lagere bovengrens tot goedkeuring kan leiden, waar de Stringer Bound aangeeft dat goedkeuring (nog) niet mogelijk is.

Helma Schapendonk-Maas

Het onderzoek van mevrouw Schapendonk-Maas²⁹ omvat wiskundig onderzoek naar het gedrag van alternatieve grenzen voor de Stringer Bound. Zij heeft onder meer gevarieerd met de volgorde van taintings.

Eén deel van haar studie is gegaan over het asymptotische gedrag van deze grenzen.

Dit leverde opmerkelijke resultaten op: met de 'Inverse taint order' (ITO³⁰) wordt een grens gevonden die asymptotisch de betrouwbaarheidsongergrens is; de 'Random taint order' (RTO) convergeert asymptotisch naar de verwachting (is het populatiegemiddelde) van de taints.

Het onderzoek bevestigt dat de Stringer Bound en de methode die Lucassen, Moors en Van Batenburg beschreven conservatief te zijn. Het onderzoek falsifieert twee andere mogelijke sorteervolgorde.

²⁸A. Lucassen, H. Moors en P.C. van Batenburg, Modifications of the Stringer Bound: a simulation study on the performance of audit sampling evaluation methods, *Kwantitatieve methoden*, nr. 51, 1996, pp. 17-25

²⁹Helma Schapendonk-Maas, *Betrouwbaarheidsbovengrenzen en steekproeftechnieken in de accountancy*, februari 1998; Research Memorandum, Limperg Instituut

³⁰ De 'Inverse Taint Order' is volgorde in oplopende grootte van de taintings. Dit is 'inverse' ten opzichte van de Stringer Bound, waar de taintings in volgorde van afnemende grootte van de taintings zijn gesorteerd.

Samenvatting onderzoek reductie bovengrens bij steekproeven op geldeenheden

Het bovengenoemde onderzoek heeft tenminste vier mogelijkheden laten zien om betrouwbaarheidsbovengrenzen bij steekproeven op geldeenheden te berekenen. De Stringer Bound zelf, de methode van Lucassen, Moors en Van Batenburg, de Studentized Bootstrap en de door de Jager, Pap en Van Zuijlen aangepaste Stringer Bound leiden alle niet tot een grotere kans op ten onrechte goedkeuren dan was begroot. De bootstrapmethode vraagt het meest intensieve rekenwerk, de andere methoden kunnen doorgaans op eenvoudige wijze in de bestaande toepassingen worden ingepast.

Steekproef op posten uitspraken in geld

In september 1997 heeft Arnold Willemsen in 'On determining an upper confidence bound for the total error amount in audit populations'³¹ de resultaten neergelegd van simulatieonderzoek naar het gedrag van een directe schatter en enkele bootstrapmethoden. Het onderzoek gaat over het vertalen de bevindingen van postensteekproeven naar uitkomsten in geldeenheden. Het onderzoek wijst uit dat geen van de onderzochte methoden met aanvaardbare resultaten de conversie van een postentrekking naar een uitspraak in geldeenheden kan uitvoeren. Dat betekent dat postensteekproeven alleen op een aanvaardbare kunnen worden gebruikt om een schatting in geldeenheden te geven als methoden als de verschil- en regressieschatter worden gebruikt.

Ter illustratie geven wij een aantal bevindingen.

In zijn studie vond Willemsen overdekkingspercentages (coverages) die 10% tot 20% onder de normatieve betrouwbaarheid lagen. Dat werd op voorhand verwacht van de methode 'directe schatter'. Dat dergelijke percentages ook gelden bij de bootstrapmethode ervoer de projectgroep als een tegenvaller.

Willemsen heeft ook schatters met hulpvariabelen gebruikt (de techniek ervan maakt gebruik van de postgrootte als hulpvariabele; hoe een en ander uitwerkt gaat te ver voor dit artikel). De daarmee bereikte betrouwbaarheidsbovengrenzen voor het fouttotaal leverden weliswaar betere dekkingspercentages op dan de directe schatter. Tegen de verwachting in leidde dit niet tot een dekking die groter of gelijk is aan de nominale betrouwbaarheid.

De conclusie uit dit onderzoek moet zijn dat postensteekproeven om een foutbovengrens in een financiële populatie te schatten niet toereikend zijn tenzij de

³¹A. Willemsen; On determining an upper confidence bound for the total error amount in audit populations; Research memorandum

methoden als de verschil- en regressieschatter worden gebruikt en aan de voorwaarden om die te gebruiken is voldaan.

Samenvatting

Wij hebben de context geschetst voor het toepassen van technieken op het gebied van statistical auditing. In deze bijdrage hebben we ons vooral gericht op de onderzoek waarbij de leden van de Stuurgroep Statistical Audit van het Limperg Instituut bij betrokken zijn geweest. Al het gepresenteerde onderzoek vond plaats op het deelgebied van de mathematische steekproef. De belangrijkste conclusies vinden wij:

- Voor het plannen van onderzoek is het niet nodig om de controletolerantie (uitgedrukt in monetaire eenheden) voor afzonderlijke deelpopulaties te verkleinen, tenzij op zichzelf staande eisen bij deelverantwoordingen dat nodig maken. Dit is een enorme efficiëntiewinst en een ingrijpende vereenvoudiging ten opzichte van het toepassen van complicerende zaken als wortelformules en dergelijke.
- Om resultaten uit steekproeven op afzonderlijke deelpopulaties te combineren tot een steekproef op het geheel zijn er meerdere valide methoden toepasbaar. De methoden zijn alle conservatief genoeg.
- Van de Stringer bound is het conservatisme aangetoond voor grote(re) steekproeven en voor betrouwbaarheden van minstens 50%.
- Er zijn voor de steekproef op geldeenheden foutbovengrenzen gevonden die onder condities niet het conservatisme van de Stringer bound hebben en toch als valide kunnen worden aangemerkt.
- De berekening van de bovengrens blijkt al gauw aan de conservatieve kant te zijn. Het verdient daarom aanbeveling om alle berekeningswijzen te hanteren die voldoende 'coverage' (95%) hebben. Het is aanvaardbaar daarvan de laagste bovengrens te kiezen voor de beantwoording van de vraag of de goedkeuringsgrens al dan niet wordt overschreden.
- Er zijn voor de postensteekproef foutbovengrenzen in geldeenheden gevonden die onder condities een verbetering betekenen t.o.v. de gebruikelijke methoden; helaas zullen ze echter nog steeds niet leiden tot de vereiste dekking. De verschil- en regressieschatter blijven, onder de voor die methoden gelden voorwaarden, het beste alternatief.

Bijlage 1

Wat is de Stringer Bound

De Stringer Bound is een berekening van de foutboven grens waarbij geprobeerd is de informatie over alle gevonden fouten in de schatting tot uitdrukking te laten komen. De Stringer Bound methode gebruikt daartoe zogenaamde taintings. Controle leidt tot de vaststelling van een administratieve waarde (Ist) y_i en de juiste waarde (Soll) x_i . Ist heet ook wel 'Bookvalue' = BV en Soll heet ook wel 'Auditvalue' = AV. In elke aangewezen post wordt de foutfractie bepaald door de 'tainting': $t_i = \frac{BV - AV}{BV}$

De puntschatting voor de foutkans wordt:

$$\text{(som van alle taintings) / steekproefomvang, of: } \frac{1}{n} \sum_{i=1}^n t_i = \frac{1}{n} \sum_{i=1}^k t_i$$

Dit betekent dat het aantal fouten is vervangen door de totale foutfractie. Alleen de taintings ongelijk nul spelen nog een rol bij de berekeningen.

De foutboven grens wordt bepaald door voor iedere afzonderlijke tainting (ongelijk 0) na te gaan hoe sterk deze fout de factor R' ophoogt in de formule:

foutboven grens = R' / steekproefomvang,

waarin R' wordt afgeleid uit de tabel van foutboven grenzen uit de Poisson (of soms binomiale of hypergeometrische) verdeling. Die afleiding gaat als volgt:

- als er geen fouten worden gevonden is R' gelijk aan R(0)
- het vinden van de i-de fout hoogt R' op met $t_i \times [R(i) - R(i-1)]$
De overstap van foutfracties naar foutbedragen gebeurt weer door vermenigvuldiging met de omvang van de populatie in geld, of door gebruik te maken van het steekproefinterval;
- puntschatting = (totaal taints) x steekproefinterval
- foutboven grens = $[R(0) + (\text{totaal ophogingen } t_i \times [R(i) - R(i-1)])] \times \text{steekproefinterval}$

Omdat de reeks $[R(i) - R(i-1)]$ afloopt, is het belangrijk om te weten in welke volgorde de fouten geëvalueerd worden. Stringer meldde dat als de taintings dalend werden ingevoerd, de uitkomst het hoogst zou worden. Dit is lang gelezen als een gebod om de fouten in volgorde van aflopende tainting in te voeren.

Bijlage 2

Generalisatie schattingsmethoden

We onderscheiden:

- methoden om de omvang van een fout vast te stellen en
- methoden om een doelparameter te schatten.

In de accountantscontrole dient op grond van de bevindingen de (mogelijke) omvang van een fout te worden bepaald. Hiervan is bij voorkeur de geschatte omvang gelijk aan nul en de bovengrens kleiner dan de materialiteit. Het statistisch basismodel spreekt in zo'n geval over het schatten van een fout. Met die schatting vindt te zijner tijd een toetsing plaats.

Bij het schatten van een doelparameter hebben we het bijvoorbeeld over het taxeren van de omvang van de voorraad door middel van een steekproefgewijze opname ervan. Er bestaat een breed scala aan schattingsmethoden.

Wij geven een aantal hoofdlijnen:

- Statistisch basismodel (foutschatting):

Er bestaan methoden die zijn gebaseerd op de hypergeometrische of binomiale verdeling, niet zelden benaderd door de Poissonverdeling. Belangrijk hierbij is dat de populatie één onbekende parameter bevat, die zowel locatie als spreiding beschrijft.

De probleemstelling is dan dus het schatten van een onbekende foutkans p .

Andere methoden zijn gebaseerd op de normale verdeling of daarvan afgeleide varianten. Hierbij zijn altijd (minstens) twee parameters nodig om locatie en spreiding van de onderhavige kansvariabele te beschrijven. Algemene probleemstelling bij schattingsmethoden is dan dat van een kansvariabele het populatiegemiddelde μ_x en de spreiding σ_x onbekend zijn. Als schattingen m voor het gemiddelde en s voor de spreiding zijn gevonden, kan daarmee een betrouwbaarheidsinterval voor μ_x worden bepaald.

- vaststellen doelparameter:

Er bestaan methoden die zich richten op de onbekende juiste waarde van een populatie, en methoden die zich richten op de onbekende fout (als bedrag of als fractie) in die populatie. Eerstgenoemde zijn ook chronologisch eerst; met name Stringer (1963) heeft getracht accountants te bewegen tot het gebruik van laatstgenoemde methoden.

Vaststelling doelparameter

We gaan in op twee hoofdlijnen. Het berekenen van een 'directe schatter' en het berekenen van de doelparameter door middel van hulpvariabelen.

Directe schatter

Bij de eenvoudigste versie van de directe schatter wordt de kansvariabele x gezien als de gemiddelde onbekende juiste waarde in de populatie. Steekproefcontrole levert een aantal juiste waarden op, waaruit het gemiddelde en de spreiding worden geschat en een betrouwbaarheidsinterval kan worden geconstrueerd. Vermenigvuldiging hiervan met het aantal items in de populatie geeft een interval voor de onbekende totale waarde, waarna de puntschatting en de bovengrens voor het foutbedrag kunnen worden gevonden.

Deze methode is vaak, bij gebrek aan meer informatie, de enige mogelijkheid voor de accountant om een uitspraak te doen. De onnauwkeurigheid van het resultaat laat echter in veel toepassingen (door de grote waarde van σ_x) te wensen over. Daarnaast is er een methodologisch probleem, omdat het voor de hand ligt dat de gevonden puntschatting voor het foutbedrag ongelijk 0 zal zijn, zelfs wanneer er in de steekproefposten geen fouten zijn gevonden.

Schatters met hulpvariabelen

Hierbij wordt gebruik gemaakt van een hulpvariabele y , waarvan alle populatiewaarden en dus ook hun populatiegemiddelde μ_y en spreiding σ_y bekend zijn. Voor de hand liggend voorbeeld is dat y de administratieve, te controleren waarden, zijn van een post in een financiële verantwoording. In de steekproef worden gegevensparen (x_i, y_i) gemeten, als de controle van de ist-positie y_i uitmondt in een soll-positie x_i .

Als er een lineair verband tussen y en x van de vorm $x = \alpha + \beta y + \varepsilon$ verondersteld wordt, kan dit verband met behulp van regressieanalyse uit de gegevensparen (x, y) worden geschat. We spreken dan van de regressieschatter. De schatter voor μ_x is dan:

$$x_r = x' + b' (\mu_y - y')$$

waarin (x', y') het steekproefgemiddelde is, en b' de schatter voor β . De onnauwkeurigheid van deze schatting wordt uitgedrukt in de standaarddeviatie (s_r), en de betrouwbaarheidsbovengrens en de foutbovengrens kunnen daaruit worden afgeleid. De standaarddeviatie wordt bepaald door de spreiding in de waargenomen juiste waarden, en vergroot door de daaraan toegevoegde spreiding in de administratieve waarden, maar verkleind door hun onderlinge samenhang.

Naarmate de gegevensparen een grotere samenhang vertonen, zal de uiteindelijke spreiding in de schatter kleiner zijn dan zonder gebruik te maken van hulpvariabelen:

$$s_{r.}^2 = s_x^2 [1 - r^2(x, y)].$$

Bijzondere gevallen hiervan zijn

- de ratioschatter, waarbij α op 0 gesteld wordt,
- de verschillschatter, waarbij bovendien β op 1 vastgeprikt wordt. In dit geval komt het schatten van μ_x neer op het corrigeren van μ_y voor de gemiddeld in de steekproef aangetroffen afwijking tussen x en y .

De verschillschatter is dus ook te zien als een directe schatter waarbij de onbekende kansvariabele x wordt gezien als de fout in de populatie, net zoals de directe schatter weer een regressieschatter is waarin $\alpha = 0$ en $\beta = 1$ gesteld is en de hulpvariabele genegeerd wordt. Dit leidt tot een laatste variant.

- de regressieschatter waarbij de gemiddelde fout in de populatie wordt geschat uit een lineaire relatie met de ist-positie van de te controleren items daarin.

Deze methoden werken het beste als er veel fouten worden aangetroffen, die redelijk symmetrisch rond hun gemiddelde verdeeld zijn. Dan immers is de normaliteit van de te onderzoeken variabele het makkelijkst aannemelijk te maken. Bij weinig fouten, of bij een asymmetrisch verdeeld foutenpatroon, is voorzichtigheid bij de interpretatie van de uitkomsten geboden.