

Boekbesprekingen

Redactie

Alex J. Koning

postadres: Vakgroep Toegepaste Economie
Erasmus Universiteit Rotterdam
Postbus 1738
3000 DR Rotterdam
telefoon: 010-4081268/59
fax: 010-4527746
e-mail: koning@few.eur.nl

Besproken boeken in Kwantitatieve Methoden nr. 58

Venables, W.N., Ripley, B.D.

Modern applied statistics with S-PLUS

Winkelmann, R.

Econometric analysis of count data

Koosis, J.

Statistics. A self-teaching guide. Fourth edition

Embrechts, P., Klüppelberg, C., Mikosch, T.

Modelling extremal events for insurance and finance

Niesing, J.

Simultaneous component and factor analysis methods for two or more groups

Dudewicz, E.J. (ed)

Modern digital simulation, II

Brink, W. van den

Statistiek oefeningen

Gourieroux, C.

ARCH models and financial applications

Firebaugh, G.

Analyzing repeated surveys

Chow, S.L.

Statistical significance. Rationale, Validity and Utility

Venables, W.N., Ripley, B.D.

Modern applied statistics with S-PLUS

1997, Statistics and computing,

Springer-Verlag, Berlin,

540 blz, DM 88.00, ISBN 0-387-98214-0.

Dit boek is in de eerste plaats een gids die laat zien hoe statistische technieken geïmplementeerd kunnen worden in S-PLUS. Voor de volledigheid zij vermeld dat S-PLUS een commerciële computerapplicatie is, bedoeld voor statistische dataverwerking. Met name de grafische mogelijkheden zijn zeer uitgebreid. S-PLUS behelst ook een programmeertaal die objectgeoriënteerd is. Uw zelfgedefinieerde objecten en methodes worden naadloos in de programmeertaal opgenomen. In de programmeertaal kunnen formules uit met name de lineaire algebra vrijwel rechtstreeks tot uitdrukking worden gebracht. Er zijn ook faciliteiten om te communiceren met gecompileerde FORTRAN- en C-code. Kortom een paradijs voor een toepassingsgericht statisticus die zonder al te veel rompslomp zijn eigen methodes kan programmeren en inbedden in de structuur die S-PLUS biedt.

De mijns inziens meest toepasselijke aanprijzing vond ik op de achterkaft: "Het boek is een gids voor het gebruik van S-PLUS voor het uitvoeren van statistische analyses en voorziet zowel in een inleiding voor het gebruik van S-PLUS als in een cursus in moderne statistische methodes".

Het boek is bedoeld voor hen die S-PLUS gaan gebruiken. De lezer wordt aan de hand genomen bij het min of meer grondig analyseren van een groot aantal databestanden van zeer uiteenlopende aard, die overigens vaak aan de vakliteratuur ontleend zijn. Het boek biedt voldoende informatie voor het ontwikkelen van afgebakende applicaties in de vorm van een bibliotheek met hulp-bestanden. Het boek lijkt me ook geschikt als basis voor een cursus computergebruik in de statistiek, voor de onderwerpen waar de gebruiker al mee vertrouwd is.

Het adjectief *modern* in de titel duidt op het computer-georiënteerd zijn in met name twee aspecten. Ten eerste de verfijnde grafische mogelijkheden die computers bieden en die behulpzaam kunnen zijn bij het onderzoeken van datacollecties. Ten tweede de rekenkracht van computers die rekenintensieve methodes bruikbaar maken. Het templat voor de hoofdstukken is het geven van een samenvatting van een gebruikte techniek mede om de gebruikte notatie bekend te maken, en vervolgens zijn implementatie in S. De inleidingen zijn zeer goed voorzien van meestal recente literatuurverwijzingen. Het helpt volgens de auteurs als de lezer al bekend is met een gebruikte methode, maar het schijnt niet noodzakelijk te zijn. De auteurs rapporteren het succesvolle gebruik van enkele hoofdstukken voor gespecialiseerde cursussen in statistische methoden. Besproken onderwerpen uit de statistiek zijn onder meer dichtheidsschatten en bootstrap-methodes, het (gegeneraliseerde) lineaire model, robuuste statistiek, niet-lineaire modellen, projection pursuit, neurale netwerken, survival analyse, multivariate analyse (cluster analyse, discriminant analyse), tree-based methodes, tijdreeksen, ruimtelijke statistiek (kriging), classificatie.

Onmisbaar bij het boek is de beschikbaarheid van S-PLUS 3.3 of recenter. Ook onmisbaar zijn de gebruikte S-PLUS bibliotheken en datacollecties, te vinden op het World Wide Web, en eenvoudig(!) te installeren. Bovendien zijn er een aantal elektronische documenten die aspecten bespreken die de auteurs graag zien als aanvulling op het boek, zg. on-line complements. Deze zijn minder geschikt om in het boek zelf opgenomen te worden omdat ze te zeer van het computer platform en de versie van S-PLUS zouden afhangen. Hieronder vallen aspecten die niet echt meer een statistische inhoud hebben, zoals het binden van gecompileerde FORTRAN- en C-code aan S-PLUS alsof het deel uitmaakt van S-PLUS. Al deze bestanden zijn vrij verkrijgbaar via het World Wide Web.

De auteurs geven veel concrete informatie over de werking van S-PLUS. Dit stelt de lezer in staat om te begrijpen dat bepaalde S-PLUS constructies kunnen leiden tot een tijdrovende uitvoering. Daarbij worden alternatieve (minder intuïtieve) constructies aangegeven. De informatie heeft niet altijd een positieve uitwerking. De presentatie van de ins en outs van specificatie van lineaire (en ook gegeneraliseerd lineaire en niet-lineaire) modellen, komen mij zo gecompliceerd over dat ik bang zou zijn een gevonden uitkomst verkeerd te interpreteren. Als toegepast gebruiker zou ik eventuele gevolgtrekkingen, zoals predictie en model-selectie, dan verder ook helemaal over laten aan S-PLUS.

Vergeleken bij de eerste editie van 1994, is er de evidente wijziging dat de tekst en de online bestanden zijn aangepast aan S-PLUS 3.3 en later. Met name is de nieuwe 'Trellis'-versie van grafiekafhandeling behandeld. Verder is menig hoofdstuk meer of minder gewijzigd, wat geleid heeft tot een uitbreiding met 100 pag. Het is duidelijk dat de auteurs de afgelopen jaren met hun boek bezig zijn geweest.

*dr H.W.M. Hendriks
Vakgroep Wiskunde
Katholieke Universiteit Nijmegen*

Winkelmann, R.

Econometric analysis of count data

1997, Springer-Verlag, Berlin,

xiv+195 blz, DM 128.00, ISBN 3-540-62201-2.

Een van de minder bekende maar desalniettemin toch belangrijke ontwikkelingen in de econometrie gedurende de afgelopen vijftien jaren is de analyse van modellen voor telgegevens geweest. In deze modellen wordt *het aantal* gebeurtenissen dat optreedt in een bepaald interval geanalyseerd. Modellen voor telgegevens zijn nauw gerelateerd aan modellen voor duurgegevens, maar specificaties die analytisch handig zijn in de context van duurgegevens zijn dat niet altijd in de context van telgegevens, en omgekeerd. Het boek van Winkelmann geeft een overzicht van de huidige stand van zaken in de literatuur over de analyse van telgegevens. Het grootste gedeelte is gewijd aan theorie, en op het einde wordt aandacht besteed aan toepassingen:

hoofdstuk 1 General Introduction**hoofdstuk 2** Probability Models for Event Counts**hoofdstuk 3** Econometric Modeling — Basic Issues**hoofdstuk 4** Econometric Modeling — Extensions**hoofdstuk 5** Further Topics**hoofdstuk 6** Application to Labor Mobility**hoofdstuk 7** General Conclusion

Het boek bevat tenslotte nog twee appendices over kans-genererende functies en tabellen met empirische resultaten.

Het boek veronderstelt een gedegen kennis van maximum-likelihood theorie, en van enige basisnoties betreffende semiparametrische methoden. Het is derhalve slechts geschikt om gebruikt te worden in een doctoraal keuzevak of postdoctorale cursussen. Helaas geeft de auteur zelf niet precies aan welke voorkennis hij heeft verondersteld bij het schrijven van dit boek, noch wat precies zijn doelgroep is.

Uit de inhoudsopgave krijgt men de indruk dat veel verschillende modellen aan de orde zullen komen, en die indruk is correct. De auteur start met de bespreking van het Poisson-regressie model. Dit model is erg restrictief, aangezien het veronderstelt dat het gemiddelde van de telvariabele gelijk is aan de variantie van de telvariabele. Het grootste gedeelte van het boek bespreekt dan ook alternatieve modellen die deze aanname niet maakt. Zo komen in hoofdstuk 4 modellen aan de orde voor niet-geobserveerde heterogeniteit, afhankelijke telprocessen, endogene selectie, en ge-generaliseerde variantie. Zo af en toe wordt aangegeven waar een bepaald model of een bepaalde toets goed voor is, aan de hand van een (veelal uiterst gestyleerd) economisch voorbeeld. Aan het einde van het boek (hoofdstuk 7) schetst de auteur een soort algemene onderzoeksstrategie die wordt aanbevolen om telgegevens te analyseren. Eerst wordt een Poisson regressiemodel geschat, waarna de verdelingsaanname wordt getoetst. Indien gewenst kunnen de parameters en standaardfouten robuust worden geschat. In een volgende stap wordt een model geschat dat minder restricties legt op de relatie tussen het gemiddelde en de variantie. Als het mogelijk is kan ook informatie worden gebruikt die buiten de geanalyseerde dataset wordt gevonden (met name het modelleren van de gebeurtenis $Y = 0$ is soms een probleem). Het beste model wordt tenslotte gevonden door geneste of niet-geneste toetsen uit te voeren. Hoewel de auteur zelf deze onderzoeksmethodologie toepast in hoofdstuk 6, is het mij niet duidelijk dat een dergelijke aanpak altijd werkt. Elke dataset en elk onderzoek heeft specifieke eigenaardigheden, en die worden niet optimaal aangepakt door een dergelijk standaardrecept te volgen.

Het is een modern gefabriceerd boek: de auteur heeft zelf het zetwerk (in $\text{\LaTeX}$) verzorgd en dat is door de uitgever vermenigvuldigd en gebonden. Uit het resultaat

blijkt duidelijk dat geen drukker naar het resultaat heeft gekeken: de hoeveelheid wit rondom de vergelijkingen varieert, wiskundige symbolen in de hoofdttekst staan niet allemaal cursief, en het ongelukkig geplaatste **Assumption 2** (p. 57) doet pijn aan de ogen. Gezien de prijs van het boek had er meer tijd en zorg aan de typografie besteed moeten worden, en had de uiteindelijke tekst gecontroleerd moeten worden door een editor.

Er zijn—voor zover mij bekend—weinig alternatieven voor dit boek op de markt. Misschien komt Andersen et al. (1993) misschien nog wel het meest dichtbij, maar de vereiste wiskundige voorkennis voor dat boek is aanmerkelijk hoger dan voor dit boek. Het boek geeft een brede inleiding tot de theorie van modellen voor telgegevens, en is in die zin waardevol. Helaas heb ik wel het gevoel dat het hoofdzakelijk een bespreking is van een grote verzameling modellen, een coherent kader wordt niet duidelijk geschetst en dat is een gemiste kans.

Verwijzingen:

- [1] Andersen, P.K., O. Borgan, R.D. Gill and N. Keiding (1993) *Statistical Models Based on Counting Processes*, Heidelberg: Springer Verlag.

dr R.H. Koning
Vakgroep Econometrie
Vrije Universiteit Amsterdam

Koosis, J.

Statistics. A self-teaching guide. Fourth edition

1997, John Wiley & Sons Ltd., Chichester,
ix+278 blz, £ 13.99, ISBN 0-471-14688-9.

Het boek is geschreven voor iedereen die door middel van zelfstudie de basis van de statistiek wil leren. Volgens de tekst op de kaft is het kunnen "lezen" van formules de enige vereiste om het boek te kunnen gebruiken.

In het boek worden de traditionele onderwerpen die voor een inleiding in de statistiek van belang zijn, beschreven.

Hoofdstuk 1 vormt een inleiding in de beschrijvende statistiek. Het maken van histogrammen en het berekenen van enkele kengetallen komt aan de orde.

Vervolgens wordt in hoofdstuk 2 ingegaan op de voor de statistiek belangrijke begrippen populatie en steekproef. Een en ander wordt uitgelegd door veelvuldig gebruik te maken van voorbeelden.

Hoofdstuk 2 vormt de afsluiting van de inleiding in de beschrijvende statistiek. In de volgende hoofdstukken ligt de nadruk op de verklarende statistiek.

Hoofdstuk 3 beschrijft de theorie van het schatten van de parameter(s) van een kansverdeling. De verdelingen waarvoor schatters (en betrouwbaarheidsintervallen) van de parameter(s) worden gegeven zijn de normale en de binomiale verdeling. Ook in dit hoofdstuk wordt weer veel gebruikt gemaakt van voorbeelden om de theorie te verduidelijken.

Als vervolg op het schatten van de parameter(s) van een kansverdeling wordt in hoofdstuk 4 ingegaan op het toetsen van hypothesen.

Na een stuk algemene theorie over het toetsen van hypothesen, beschrijft de auteur daarna een aantal standaard toetsen, te weten: de t-toets voor twee steekproeven (zowel gepaard als ongepaard) en de F-toets voor twee steekproeven.

Nadat deze toetsen besproken zijn, gaat de auteur verder met de situatie waarin meer dan twee steekproeven zijn betrokken. In hoofdstuk 6 wordt dan de een-weg variantieanalyse besproken.

Tot nu toe zijn slechts technieken behandeld waarbij slechts een variabele een rol speelt. De auteur vindt dat een beperking en laat door middel van een hoofdstuk over regressieanalyse zien dat die beperking kan worden opgeheven. Tevens wordt ook nog ingegaan op de twee-weg variantieanalyse.

De auteur heeft een boek willen schrijven dat geschikt is voor zelfstudie. Door de heldere opbouw en de uitgebreide uitleg denk ik dat dat het boek daar zeer geschikt voor is. Hierbij is het prettig dat ieder stuk theorie gevolgd wordt door een aantal vragen daarover zodat de lezer direct kan nagaan of hij/zij de theorie begrepen heeft.

Een ander sterk punt van het boek is dat bij iedere statistische techniek melding wordt gemaakt van de onderliggende veronderstellingen. Ook wordt bekeken hoe kan worden nagegaan of aan de onderliggende veronderstellingen wordt voldaan. Zo wordt er bijvoorbeeld veelvuldig op gewezen dat om een een-weg variantieanalyse te mogen uitvoeren, de populatievarianties gelijk moeten zijn.

De auteur richt zich op het praktisch gebruik van statistiek. Hij realiseert zich dat de berekeningen de nodige tijd vragen. Echter, tegenwoordig beschikt bijna iedereen over een computer met spreadsheetsoftware. De auteur geeft daarom in afzonderlijke paragrafen uitleg over statistische functies binnen Excel en Lotus.

Concluderend: een aardige inleiding in de statistiek. Een sterk punt bij de theorie is dat bij iedere statistische techniek duidelijk de onderliggende veronderstellingen worden beschreven. Praktisch gezien is de beschrijving van de relevante statistische functies in Excel en Lotus een pluspunt.

*drs M.B. Nijhuis
Akzo Nobel*

Embrechts, P., Klüppelberg, C., Mikosch, T.

Modelling extremal events for insurance and finance

1997, Applications of mathematics,

Springer-Verlag, Berlin,

644 blz, DM 118.00, ISBN 3-540-60931-8.

For an audience of which a considerable percentage lives below the sea-level, one does not have to motivate the analysis of extremal events. Only the highest flood levels are used to determine dike levels in the Netherlands. The application of the methods treated in the current monograph by Dutch hydraulic engineers has been a major success

story for statistical extreme value theory. More recently statistical extreme value theory has found wide application by risk managers in the financial industry for the so-called Value-at-Risk issue, and it has been a proven tool in the insurance industry for the ruin problem in claim analysis.

The book is the first in the area which strikes a proper balance between mathematical rigor and scope on the probability side, and the statistically oriented applications for the practitioner. For the analytically trained hydraulic engineers, risk managers and actuaries the book provides a rich source of the currently available methods of and knowledge about extreme value analysis. The authors mostly rely on financially oriented examples, but the coverage of methods is such that anyone using extreme value technique should consult the book. Because of its mixture of theory and statistical applications it is also a welcome teaching device for upper undergraduate and first year graduate courses in the mathematics curriculum. Necessary backgrounds are courses in analysis and probability theory. For the interested researcher it provides a very good introduction to the area and due to its voluminous coverage it is also a very handy source book. The book fits between the somewhat older theory books by Leadbetter, Lindgren and Rootzen (1983) and Resnick (1987), and the book by Reiss and Thomas (1997) which is oriented towards statistical application. The book covers a number of recent theoretical developments, but as the authors note themselves, it is entirely focused on the uni-variate analysis.

The contents of the treatise-like book are broadly as follows. The starting chapter covers the classical ruin problem in claim analysis with emphasis on the class of sub-exponential and heavy-tailed distributions. Both in finance and insurance data happen to be heavy-tailed, and therefore these distributions figure dominantly in the book. The second chapter is devoted to classical LLN, CLN and LIL results. Extreme value analysis is treated in chapters 3, 4 and 5, where point processes are used as a basic tool for the analysis. Statistical methods are covered in chapter 6. Chapters 7 and 8 are devoted to special topics such as time series analysis with heavy-tailed data, the extremal index, ARCH processes and stable process. The appendices provide background on convergence issues, sub-exponentiality and regular variation.

The reviewer is himself a user and modest contributor to the area. Nevertheless, the book was very helpful in providing a general overview of the area, fitting in the gaps, mostly in the area of insurance, in which I lacked knowledge. Some of the statistical methods such as the ratio of maximum to some plots I was unaware of. The book is written in a clear and enthusiastic manner, which will certainly motivate students. The book treats several recent topics such as ARCH process, where light-tailed innovations nevertheless result in a heavy-tailed unconditional distribution through variance clustering. But it also necessarily has to leave some topics aside, among which the said multivariate analysis is the most glaring. In chapter 6 on statistical methods one would perhaps also have liked to see a comparison of the different techniques. For example some analysis of efficiency might have organized the different methods in a somewhat more coherent manner than is presently the case. At the time of writing it was not known how to choose the number of extreme realizations in practice. Ne-

vertheless, such is essential for any of the statistical techniques presented, and the practically oriented reader has to be aware of this problem. But this is nit-picking. It is an exceptional book, recommended to practitioners, students and researchers alike!

*prof. dr C.G. de Vries
Tinbergen Instituut
Erasmus Universiteit Rotterdam*

Niesing, J.

Simultaneous component and factor analysis methods for two or more groups

1997, M & T series,

DSWO Press, Leiden,

236 blz, fl 48.00, ISBN 90-6695-134-6.

This book is a dissertation about the relative performance of various methods for simultaneous principal component analysis (SCA). The dissertation was supervised by Jos Ten Berge and Henk Kiers. I begin this review with an introduction to SCA, followed by a brief description of the study.

Let x denote an observed $n \times p$ data matrix with zero means. Principal component analysis (PCA) is used to find a $p \times q$ weight matrix W ($q < p$) yielding uncorrelated and standardized linear combinations $Z = xW$ that optimally (in a least-squares sense) summarize the information contained in x . These linear combinations are the components scores. Let $x_{\text{pred}} = PZ$ be a prediction for x in this regression, where P is a $p \times q$ matrix of regression weights. The matrix P is called the pattern matrix. The weight matrix differs from the pattern matrix in the sense that the weights describe the regression of the component scores on x , while the pattern matrix describes the regression of x on the component scores. A third matrix is the structure matrix of correlations between components and variables. SCA concerns simultaneous principal component analysis in multiple groups. Each of the forementioned matrices (the weight matrix, the pattern matrix, or the structure matrix) can be constrained to be equal across groups, allowing for a different interpretation of the components. The main theme of the book is the comparison of these three kinds of SCA by means of Monte Carlo techniques. The first chapters are about the SCA method. The third and fourth chapter provide background to the simulation study. The results are discussed and summarized in the remaining chapters. The discussion of the SCA method is terse and purely technical. It is clear that the reader is expected to be conversant with factor analysis. Important issues such as standardization, are given short shrift and other developments in the area of multigroup PCA and not mentioned. In contrast, there is an excellent and detailed discussion of the simulation study.

The present book is of definite interest to anyone who wishes to conduct a simulation study involving exploratory factor analysis. Methodological issues are hardly discussed and the reader is expected to have a detailed knowledge of exploratory factor analysis. Those that are not conversant with SCA are referred to Ten Berge (1993,

DSWO Press), or Kiers and Ten Berge (1994, *Br. J. of Math. and Stat. Psy.*).

*Th. Bechger
Vakgroep Psychonomie
Vrije Universiteit Amsterdam*

Dudewicz, E.J. (ed)

Modern digital simulation, II

1997, American series in mathematical and management sciences,
American Sciences Press, Syracuse, New York,
502 blz, \$ 135.00, ISBN 0-935950-42-7.

Readership: researchers and graduate students with an interest in data analysis and in simulation models based on data.

In practice the following situation often occurs. A researcher has a (univariate or multivariate) data set and he/she wants to fit a model to the data and then use the model for simulation purposes. In the present book there are four papers devoted to this type of situation.

In the first paper "The Extended Generalized Lambda Distribution (EGLD) system for fitting distributions to data with moments" (E.J. Dudewicz and Z.A. Karian) the authors provide tables and algorithms to fitting a Generalized Lambda Distribution (GLD) to a given data set. The $GLD(\beta(1), \beta(2), \beta(3), \beta(4))$ is a four parameter family of distribution functions (d.f.) defined through its inverse $G(y)$:

$$G(y) = \beta(1) + (y\beta(3) - (1 - y)\beta(4))/\beta(2).$$

In order to have a valid d.f. the parameters have to satisfy certain restrictions. This representation of the GLD-model is very convenient for simulation studies. To fit a specific GLD the method of moments can be used. The resulting set of four equations usually is rather complex and one is forced to use numerical methods to solve these equations for the parameters $\beta(i), i = 1, \dots, 4$. In the paper the authors provide tables which can be used for practical purposes. With these tables the authors extend previous available tables and describe algorithms to cover a large part of the possible $(s^2(3), k(4))$ -plane, where $s(3), k(4)$ denote the theoretical skewness and kurtosis of the GLD. In some cases they also use a Generalized Beta Distribution (GBD) as an alternative model. A GBD is also a four parameter family defined by the random variable $Y = \beta(1) + \beta(2)X$ where X has a $BETA(\beta(3), \beta(4))$ -distribution.

In the second paper "A bivariate GLD using Placketts method of construction" (N.B. Beckwith and E.J. Dudewicz) the authors fit bivariate GLD to bivariate data sets and to bivariate d.f.. The model they use was proposed by Plackett in 1965 and is of the following form: given a parameter $p > 0$, the bivariate d.f. H is the solution to

$$p = H(1 - F - G + H)/(F - H)(G - H)$$

where F and G are the marginals d.f. The parameter p is a type of measure of association between the marginals. Given a bivariate data set, the authors first fit each of the marginals to a (possibly different) GLD and then use a broad range of p -values to find "the best" p and bivariate d.f. H . A similar procedure can be used to fit a bivariate GLD to theoretical bivariate d.f. such as the bivariate Normal d.f., Gumbel's bivariate Exponential type I d.f., the bivariate Cauchy d.f. etc. Whenever possible, the method of moments is used. Otherwise the method based on percentiles is used.

In the third paper "Comparing censusPES and censusPLUS when there is the possibility of undercounting" (T. Wright and L. Bates) the authors present the results of a large Monte Carlo simulation study. In their paper the authors compare different methods for the integration of different sample results in the case where in each of the samples there is the possibility of undercounting.

In the last paper "Statistical resampling methods in biology: a case study of beaver dispersal patterns" (L. Sun and D. Müller-Schwarze) the authors compare the bootstrap (B), the generalized bootstrap (GB) and the jackknife resampling method using beaver dispersal distance data. In GB the authors fit a GLD (as in the first paper) and then use the corresponding GLD to do the usual bootstrap calculations. In the beaver example, the three resampling methods provided similar (and good) results. The GB earns a positive point because GB seems to be more consistent in performance than the ordinary bootstrap.

The present 'book' provides an overview of an interesting field of sciences. It provides new and promising methodology concerning problems in modelling. Each of the authors has written a clear paper and some of the authors provide detailed C programs of the algorithms discussed in the text.

It would be interesting to compare the method of moments and the method of percentiles when fitting a GLD-model. Also a discussion about the stability of the models would be welcome. I don't see why the authors wait to write a real book about univariate and multivariate GLD and applications.

prof. dr E. Omey
EHSAL

Brink, W. van den
Statistiek oefeningen

1997, Boom, Meppel,
116 blz, fl 34.50, ISBN 90-5352-168-2.

Dit boek hoort bij de bekende driedelige serie van dezelfde auteur, samen met P. Koele, getiteld Statistiek (deel 1: datareductie, deel 2: theorie en deel 3: toepassingen). De oefeningen bestaan uit 173 driekeuze vragen, afkomstig van oude propedeutische statistiekentamens voor psychologiestudenten. De oefeningen staan geordend naar de hoofdstukken van de drie boeken. Daarnaast staan er nog zes oefentoetsen in van elk 20 driekeuze vragen. Elke toets beslaat min of meer de gehele stof. Het boek bevat

naast correcte antwoorden ook korte uitwerkingen van alle vragen. In die uitwerkingen wordt vaak verwezen naar de drie boeken.

Doordat alle vragen zijn beantwoord door grote aantallen studenten heeft de auteur inzicht in de kwaliteit van de opgave. Die kwaliteit wordt uitgedrukt in twee maten. Enerzijds is dat het percentage studenten dat de vraag correct heeft beantwoord. Dat geeft een beeld van de moeilijkheidsgraad van de opgave. Anderzijds wordt ook per vraag de correlatie weergegeven tussen de itemscore en de toetsscore, voor alle studenten. Dat getal geeft aan of de vraag statistiekkennis op de juiste wijze meet.

Het zal duidelijk zijn dat dit oefeningenboek moeilijk te gebruiken is zonder de drie theorieboeken. Althans als we het hebben over studenten. Voor docenten is het ook los van de serie een nuttig boek. Het biedt een grote hoeveelheid korte, helder geformuleerde vragen. De onderwerpen zijn voor een breed publiek te gebruiken.

Na al deze positieve geluiden ook een negatief geluidje. Bladerend door de opgaven viel mijn oog –begin december– op het bekende Sinterklaasprobleem (vraagstuk 56). Hoe groot is de kans dat het trekken van lootjes fout gaat bij drie personen. Bij de uitwerkingen stond het correcte antwoord ($2/3$) als volgt afgeleid. De kans dat het lootjes trekken goed gaat wordt daar als volgt berekend: eerst moet persoon 1 het goed doen (kans $2/3$), vervolgens moet persoon 2 het goed doen, kans $1/2$. Persoon 3 doet het dan automatisch goed. Het product van deze kansen is $1/3$. Het antwoord is goed, maar de redenering deugt niet. U ziet waarschijnlijk wel waarom. Zo niet dan moet U er William Feller maar eens op naslaan. Overigens was dit de enige misser die ik heb ontdekt.

*dr J.J.M. Bedaux
Afd. Theoretische Biologie
Vrije Universiteit Amsterdam*

Gourieroux, C.

ARCH models and financial applications

1997, Springer series in statistics,

Springer-Verlag, Berlin,

305 blz, DM 88.00, ISBN 0-387-94876-7.

Many financial time series, such as returns on exchange rates and stock market indices, exhibit volatility clustering, i.e., large (small) absolute realizations tend to be followed by large (small) realizations of either sign and periods of turbulence alternate with periods of relative tranquility. Given the importance of volatility, as a measure of risk, for financial theory and decision making, accurate forecasts of future volatility are critical for the implementation and evaluation of, e.g., asset pricing theories. Therefore, it is not surprising that over the last decade quite some effort has been made to come closer to understanding the properties and determinants of the volatility process.

The class of Autoregressive Conditional Heteroskedastic [ARCH] models is by now the most widely used model for this purpose. As noted in the introductory chapter of this book, the number of papers written on this particular class of models has

been growing exponentially and already runs into the hundreds. This book intends to bring together some of the major theoretical results on the statistical aspects of ARCH models which have been derived over the last fifteen years, as well as to present an analysis of the application of ARCH models in financial problems such as the construction of efficient and hedging portfolios.

The book roughly consists of two parts. The first part, which comprises of chapters 2-6, focuses on the statistical aspects of ARCH models. Chapter 3 introduces the basic univariate ARCH and Generalized ARCH [GARCH] models and discusses the properties of the associated time series. Chapter 4 presents the necessary ingredients for estimation of GARCH models by (pseudo) maximum likelihood, properties of the resulting estimates, and procedures for testing for the presence of ARCH effects. Chapter 5 discusses some extensions; in particular, ARCH models as approximations to continuous time processes, nonlinear variants of the basic GARCH model, and the integrated GARCH model. Chapter 6 focuses on multivariate extensions of the GARCH model and compares several variants of these models.

The second part of the book, consisting of Chapters 7-9, provides a thorough introduction into basic financial theory. Mean-variance efficiency, hedging, performance measures, factor models, Arbitrage (Pricing) Theory, and the Capital Asset Pricing Model are all discussed. Although these theories can be found in many finance textbooks, the added value of this part of the book is that it presents the theory in a lot of (mathematical) detail and points out quite a few pitfalls and intricacies involved.

Although the book in general is well written and quite enjoyable to read, some critical comments should be made. First and most important, the two parts of the book are not linked up as tightly as might be expected. In particular, the examples in the second part of the book on how to incorporate the notion of time-varying volatility into standard financial theory are quite concise and superficial. Similarly, given that the number of different subjects which are addressed is large while the number of pages is limited, most subjects are described rather briefly. It would not have been harmful if some extra pages were spent to provide more intuition and more elaborate examples to illustrate the theoretical results. Finally, the references (which are presented subject-by-subject) are not always up to date. Papers which have been published quite some time ago are still listed as working paper, while other, very recent publications are already listed as such!

Concluding, the book presents a lot of different aspects of ARCH models and their application in financial theory and decision making in a concise and very readable way. As such it can serve both as a useful starting point for students and practitioners wishing to quickly acquire knowledge in this area, and as a reference guide for established researchers and experienced users of ARCH models.

*D.J. van Dijk
Tinbergen Instituut Rotterdam
Erasmus Universiteit Rotterdam*

Firebaugh, G.

Analyzing repeated surveys

1997, Quantitative applications in the social sciences,
Sage, Thousand Oaks, California,
vii+72 blz, £ 7.95, ISBN 0-8039-7398-5.

'Analyzing Repeated Surveys' door Glenn Firebaugh is voer voor sociologen. Het boek behandelt adequaat en down-to-earth een viertal modellen om verandering te bestuderen. In alle gevallen worden de gebruikte modellen geïllustreerd met voorbeelden, waarvan de uitkomsten ook telkens weer een adequate illustratie van het betoog vormen. Zoals gezegd is de oriëntatie sociologisch, en de methoden zijn dan ook bij uitstek geschikt om veranderingen op macro niveau te onderzoeken; de repeated surveys zijn niet noodzakelijkerwijs identieke samples op verschillende meetmomenten.

Na een introductie waarin nog eens de basics betreffende repeated surveys vs. panel surveys, gross and net change en nog wat andere definities behandeld worden, passen in hoofdstuk 2 leeftijd, periode en cohort effecten de revue. De door de auteur genoemde strategieën om met het identificatie probleem om te gaan zijn niet vernieuwend, en staan ook enigszins buiten de lijn in de rest van het boek. In hoofdstuk 3 wordt trend analyse behandeld, ook weer voor het macroniveau. In hoofdstuk 4 behandelt de auteur de ontbinding van verandering in een deel veroorzaakt door verandering van de individuen in de onderzochte groep, en een deel dat is toe te schrijven aan de met de tijd veranderde samenstelling van de onderzoeksgroepen (zgn. cohort replacement). In hoofdstuk 5 worden de onderzoeksvragen verder verdiept met behulp van modellen waarmee onderzocht kan worden aan welke veranderingen in het niveau danwel het effect van verklarende variabelen de geobserveerde veranderingen zijn toe te schrijven; het gaat daarbij overigens telkens om situaties met twee meetpunten. In hoofdstuk 6 tenslotte wordt het changing-parameter model voor de analyse van veranderingen op individueel niveau behandeld, waarbij de effecten van de verklarende variabelen als functies van de tijd worden gemodelleerd.

Het boekje is goed leesbaar, de uiteenzetting is vrijwel zonder uitzondering helder. In een zeer beperkt aantal bladzijden wordt een respectabel volume aan modellen en technieken behandeld, met allerlei aanwijzingen voor analyse en interpretatie. In de behandeling van de voorbeelden wrekt zich enigszins het feit dat de tekst is ontstaan uit de lecture notes van een cursus. Het is allemaal wel erg direct op het voorbeelden en formule niveau: voor de lezer die het allemaal eens voorgerekend wil zien om het daarna zelf te gaan gebruiken is dat prima. Als je als noviet het boek als overzichtswerk wilt lezen krijg je denk ik toch op een gegeven moment dat je door alle voorbeelden de grote lijn en de verbanden niet meer ziet. Bijvoorbeeld bij de behandeling van het gebruik van constraints in de periode/leeftijd/cohort kwestie had het wat mij betreft af en toe wel wat abstracter en conceptueler gemogen. Een tweede klein punt van bezwaar voor de niet-Amerikaanse lezer (maar dat kan de auteur niet aange-rekend worden) zijn de voorbeelden die allemaal erg 'Amerikaans' zijn: vrijwel alle voorbeelden gaan over trends in anti-black prejudice of over verschillen tussen Demo-

crats en Republicans, kwesties waarbij de Nederlandse lezer zich misschien toch wat minder kan voorstellen.

*dr C.C.J.H. Bijleveld
Afdeling Onderzoek-WODC
Ministerie van Justitie*

Chow, S.L.

Statistical significance. Rationale, Validity and Utility

1996, Sage, Thousand Oaks, California,
xi + 205 blz, £ 13.95, ISBN 0-7619-5205-5.

Het boek "Statistical Significance" van Sui L. Chow is een interessant boek. Het is geen standaard tekstboek waarin een aantal statistische technieken de revue passerend. In tegendeel, dit boek richt zich geheel op één statistische techniek, namelijk de N.H.S.T.-procedure, dat wil zeggen: de nul-hypothese-significance-test procedure. Deze procedure is door een aantal (psychologische) onderzoekers sterk bekritiseerd. De auteur stelt zich nu ten doel deze kritiek te inventariseren en te analyseren.

Het eerste hoofdstuk bevat een litanie van kritiek die tegen de N.H.S.T.-procedure is ingebracht. Op bijna ieder onderdeel van de procedure is kritiek uitgeoefend. Nadat de schrijver heeft vastgesteld dat het begrip "statistische significantie" betrekking heeft op het verwerpen van de nulhypothese met behulp van een van te voren vastgestelde fout van de eerste soort α , kan hij overgaan tot de formulering van de bezwaren.

Critici vinden het een bezwaar dat het begrip "statistische significantie" geen stabiele eigenschap van de data is, maar afhangt van een willekeurig criterium, namelijk de fout van de eerste soort α . Daarnaast zijn zij van mening dat het niet erg informatief is om alleen maar te weten dat een resultaat statistisch significant is omdat deze uitspraak niets zegt over de kans dat H_1 waar is. Het is bovendien niet duidelijk wat de conclusie in veel psychologisch onderzoek moet zijn indien statistische significantie niet bereikt wordt. Tevens wordt opgemerkt dat de nulhypothese als punthypothese nooit waar kan zijn hetgeen tot een methodologische paradox leidt. Voeg daarbij de bezwaren dat significantie geen informatie verschaft over het substantiële of praktische belang van het onderzoek zelf en dat geen echte informatie over de parameter (bv. μ) wordt verkregen, dat de overschrijdingskans p gemakkelijk verkeerd kan worden opgevat, dat de neiging bestaat de fout van de tweede soort β te verwaarlozen en de procedure verantwoordelijk kan worden gesteld voor "a simple-minded trial and error approach to experimentation" en men heeft al een lange lijst van bezwaren te bespreken.

Het lezen van dit eerste hoofdstuk was voor mij een merkwaardige ervaring omdat een aantal bezwaren mijn mening voorkwamen uit een gebrek aan statistische kennis en een ander aantal duidelijk meta-statistisch van aard waren. Ik was dan ook benieuwd naar het vervolg van het boek.

Het moet worden gezegd dat de schrijver met zorg al de bezwaren in beschouwing neemt om ze vervolgens te weerleggen dan wel te relativiseren. Wat het boek naar

mijn mening interessant maakt waardoor het zich verheft boven het niveau van een gemiddeld tekstboek zijn de hoofdstukken drie, vier en vijf. Hier levert de auteur een duidelijke bijdrage aan de bestaande gebruikelijke literatuur.

In hoofdstuk drie maakt de schrijver terecht bezwaar tegen de te eenvoudige wijze waarop, in de gebruikelijke leerboeken, de N.S.H.T.-procedure aan de orde wordt gesteld. Terecht maakt hij onderscheid tussen een drietal verschillende experimentele situaties: "The utilitarian, the clinical and the general experiment". Vervolgens wordt duidelijk gemaakt dat empirisch wetenschappelijk onderzoek heel wat meer is dan alleen maar het toepassen van een statistische techniek. Het betoog van de schrijver is in de hoofdstukken drie, vier en vijf statistisch-wetenschapsfilosofisch-methodologisch van aard. De plaats van de statistiek in het totaal van empirisch wetenschappelijk onderzoek wordt duidelijk gemaakt aan de hand van een uitgebreid voorbeeld dat vooral psychologen direct zal aanspreken. De auteur laat zien hoe de onderzoeker soms alleen via een reeks van voorwaardelijke uitspraken tot de statistische hypothese kan geraken.

Gestart wordt met de "substantive hypotheses" die nog geheel in het abstracte vlak ligt en dus in het geheel niet toetsbaar is. Om tot een toetsbare hypothese te komen gaat de onderzoeker over tot een aantal logische implicaties. Allereerst de "research hypothesis"; hieruit wordt de "experimentele hypothese" afgeleid en deze laatste hypothese leidt tot de "statistische hypothese". Deze benadering stelt de auteur en slaat tal van interessante en verhelderende opmerkingen te maken met betrekking tot de N.H.S.T.-procedure. Het boek wordt besloten met een uitstekend samenvattend hoofdstuk. Wat zijn de belangrijkste conclusies waartoe de auteur komt? We zullen ze puntsgewijs noemen:

1. Er is niets mis met de N.H.S.T.-procedure. Zij speelt een beperkte maar belangrijke rol in het nagaan of de data geïnterpreteerd kunnen worden in termen van toevalsinvloeden.
2. Empirisch onderzoek is veel meer dan statistiek.
3. Theorie bevestiging is meer dan het toetsen van statistische hypothesen.
4. Er bestaat een fundamenteel verschil tussen experimenteel en niet-experimenteel onderzoek.

Wat is mijn slot conclusie? Het boek levert zonder enige twijfel een belangrijke bijdrage aan de bestaande literatuur. Zij doet dit door haar wetenschapsfilosofisch-methodologische karakter. Het boek is naar mijn mening net altijd even makkelijk leesbaar, maar dat kan komen omdat ik zelf geen psycholoog ben. Ik vond het licht merkbaar dat de auteur van origine een psycholoog is en geen statisticus van professie.

Ik beveel het boek van harte aan alle empirische onderzoekers uit de sociale wetenschappen.

*prof. dr J. Koerts
Vakgroep Toegepaste Economie
Erasmus Universiteit Rotterdam*

Binnengekomen boeken

Hieronder volgt een overzicht van de recente uitgaven die sinds de vorige aflevering van *Kwantitatieve Methoden* bij de redactie zijn binnengekomen. Mocht er belangstelling bestaan voor het bespreken van een boek, dan kan contact opgenomen worden met de boekbesprekingsredacteur. Indien het boek dan nog niet vergeven is, krijgt de kandidaat het boek gratis thuisgezonden. Hiervoor dient dan binnen een half jaar een recensie te worden teruggezonden. Een aantal titels is niet fysiek toegezonden, maar kan door ons besteld worden bij de uitgever. Dit heeft consequenties voor de levertijd van deze titels.

Prabhu, N.U.

Foundations of queueing theory

1997, International series in operations research and management science,
Kluwer, Dordrecht,
224 blz, fl 190.00, ISBN 0-7923-9962-5.

Lootsma, F.A.

Fuzzy logic for planning and decision making

1997, Applied optimization,
Kluwer, Dordrecht,
208 blz, fl 185.00, ISBN 0-7923-4681-5.

Zhengyan, L., Chuanrong, L.

Limit theory for mixing dependent random variables

1997, Mathematics and applications,
Kluwer, Dordrecht,
344 blz, fl 240.00, ISBN 0-7923-4219-4.

Ramakrishnan, R., Stuckey, P. (eds)

Constraints and databases

1997, Kluwer, Dordrecht,
192 blz, fl 240.00, ISBN 0-7923-8045-2.

Migdalas, A., Pardalos, P.M., Värbrand, P.

Multilevel optimization: algorithms and applications

1997, Nonconvex optimization and its applications,
Kluwer, Dordrecht,
406 blz, fl 320.00, ISBN 0-7923-4693-9.

Cela, E.

The quadratic assignment problem: theory and algorithms

1997, Combinatorial optimization,

Kluwer, Dordrecht,

304 blz, fl 220, ISBN 0-7923-4878-8.

dos Santos Gromicho, J.A.

Quasiconvex optimization and location theory

1997, Applied optimization,

Kluwer, Dordrecht,

240 blz, fl 190.00, ISBN 0-7923-4694-7.

Lebart, L., Salem, A., Berry, L.

Exploring textual data

1997, Text, speech and language technology,

Kluwer, Dordrecht,

280 blz, fl 190.00, ISBN 0-7923-4840-0.

Wakayama, T., Kannapan, S., Khoong, C.M., Navathe, S., Yates, J.

Information and process integration in enterprises. Rethinking documents

1997, The Kluwer international series in engineering and computer science,

Kluwer, Dordrecht,

440 blz, fl 230.00, ISBN 0-7923-8059-2.

Coelli, T., Prasada Rao, D.S., Battese, G.E.

An introduction to efficiency and productivity analysis

1997, Kluwer, Dordrecht,

296 blz, fl 215.00, ISBN 0-7923-8060-6.

Rapcsák, T.

Smooth nonlinear optimization of R^n

1997, Nonconvex optimization and its applications,

Kluwer, Dordrecht,

388 blz, fl 295.00, ISBN 0-7923-4680-7.

Fang, S.-C., Rajasekera, J.R., Tsao, H.S.J.

Entropy optimization and mathematical programming

1997, International series in operations research and management science,
Kluwer, Dordrecht,
360 blz, fl 250.00, ISBN 0-7923-9939-0.

Glover, F.W., Laguna, M.

Tabu search

1997, Kluwer, Dordrecht,
408 blz, fl 295.00, ISBN 0-7923-9965-X.

Parthasarathy, T., Dutta, B., Potters, J.A.M., Raghaven, T.E.S., Ray, D., Sen, A.

Game theoretical applications to economics and operations research

1997, Theory and decision library C,
Kluwer, Dordrecht,
316 blz, fl 315.00, ISBN 0-7923-4712-9.

Moeschlin, O., Grycko, E.,

Experimental stochastics

1998, Springer-Verlag, Berlin,
206 blz, DM 130, ISBN 3-540-14619-9.

Prokhorov, Y.V., Shiryaev, A.N.

Probability theory III. Stochastic calculus

1998, Encyclopaedia of mathematical sciences,
Springer-Verlag, Berlin,
256 blz, DM 158.00, ISBN 3-540-54687-1.

Bridges, D.S.

Foundations of real and abstract analysis

1997, Graduate texts in mathematics,
Springer-Verlag, Berlin,
345 blz, DM 79.00, ISBN 0-387-98239-6.

Gordon, H.

Discrete probability

1997, Undergraduate texts in mathematics,
Springer-Verlag, Berlin,
255 blz, DM 68.00, ISBN 0-387-98227-2.

Rudas, T.

Odds ratio in the analysis of contingency tables

1998, Quantative applications in the social sciences,
Sage, Thousand Oaks, California,
88 blz, £ 8.99, ISBN 0-7619-0362-3.

Saris, W.E., Münnich, A. (eds)

The multitrait-multimethod approach to evaluate measurement instruments

1995, Eötvös University Press, Budapest,
280 blz, ISBN 963-463-001-4.

Saris, W.E., Veenhoven, R., Scherpenzeel, Anette C., Bunting, B. (eds)

A comparative study of satisfaction with life in Europe

1995, Eötvös University Press, Budapest,
299 blz, ISBN 963-463-081-2.

Saris, W.E., Kaase, M. (eds)

Eurobarometer measurement instruments for opinions in Europe

1997, ZUMA Spezial Nachrichten Band,
Zentrum für Umfragen, Methoden und Analysen, Mannheim,
181 blz, ISBN 3-924220-12-3.

Jaccard, J.

Interaction effects in factorial analysis of variance

1998, Quantative applications in the social sciences,
Sage, Thousand Oaks, California,
112 blz, £ 8.99, ISBN 0-7619-1221-5.

Mitrani, I.

Probabilistic Modelling

1997, Cambridge University Press, Cambridge,
x+223 blz, £ 16.95, ISBN 0-521-58530-9.

Davison, A.C., Hinkley, D.V.

Bootstrap methods and their application

1997, Cambridge series in statistical and probabilistic mathematics,
Cambridge University Press, Cambridge,
x+582 blz, £ 24.95, ISBN 0-521-57471-4.

Zopounidis, C., Dimitras, A.I.

Multicriteria decision aid methods for the prediction of business failure

1998, Applied optimization,
Kluwer, Dordrecht,
196 blz, fl 180.00, ISBN 0-7923-4900-8.

40.00

Tuy, H.

Convex analysis and global optimization

1998, Nonconvex optimization and its applications,
Kluwer, Dordrecht,
350 blz, fl 2, ISBN 0-7923-4818-4.

Jacoby, W.G.

Statistical graphics for visualizing multivariate data

1998, Quantitative applications in the social sciences 120,
Sage, Thousand Oaks, California,
112 blz, £ 6.99, ISBN 0-7619-0899-4.

Hawkins, D.M., Howell, D.H.

Cumulative sum charts and charting for quality improvement

1998, Statistics in engineering and physical science,
Springer-Verlag, Berlin,
270 blz, DM 98.00, ISBN 0-387-98365-1.

Oproep voor boekbesprekers

De redactie van Kwantitatieve Methoden is steeds op zoek naar mensen die bereid zijn een pas verschenen boek te lezen met het doel daarover een oordeel te geven. Zo'n recensie verschaft de leden van de VVS en overige lezers van Kwantitatieve Methoden inzicht in de kwaliteiten van het boek en verschaft nuttige informatie bij een overwogen aanschaf. Voor de schrijvers biedt een kritisch oordeel van het publiek waarvoor het boek geschreven een mogelijkheid tot verdere verbetering. De uitgever heeft uiteraard meer baat bij een lovende recensie. Als tegenprestatie mogen de recensenten de besproken boeken behouden.

Behalve de algemene boeken over statistiek of besliskunde, waar veel mensen zinvolle dingen over kunnen zeggen, behandelen veel boeken een specialistisch onderwerp. Als wij, soms op de gok, iemand een dergelijk boek toesturen, blijkt een enkele keer dat 'het boek precies aansluit bij het onderzoek' van de recensent. Dat is prettig, omdat veel mensen daarbij voordeel hebben. De recensent, omdat hij/zij het nieuw verschenen boek op het onderzoeksgebied ter beschikking heeft, en de auteur(s) omdat er een terzake kundig oordeel wordt gegeven. Om nu de succeskans van het toesturen van een boek te vergroten, willen wij graag in contact komen met zoveel mogelijk collega statistici, kansrekenaars en besliskundigen. Stuur het onderstaande antwoordstrookje naar het adres van de boekbesprekingsrubriek als u interesse heeft om in de toekomst een boek te bespreken. Als er dan in de toekomst wellicht een gloednieuw boek op uw onderzoeksgebied binnenkomt hebben wij een grotere kans op een perfecte match.

Ik ben graag bereid een boek op mijn interessegebied te bespreken.

Naam	
Adres	
Telefoon	
e-mail	
Interessegebieden (graag nauwkeurig omschrijven)	