Journal of Official Statistics
Vol. 4, No. 1, 1988, p. 79–80
© Statistics Sweden

Special Note

International Symposium on Operationalization and Research Strategy Amsterdam, September 8–9, 1988

Organizing Committee

Prof. dr. J. de Jong-Gierveld (chairperson)
Dr. J.J. Hox (secretary)
Prof. dr. R. Pawson (RC33)
Prof. dr. P.G. Swanborn
Dr. W. Jansen

On Thursday and Friday, September 8 and 9, 1988, the Research Committee on "Conceptualization and Research Design," institutional member of the Social and Cultural Sciences Section of the Netherlands Organization for the Advancement of Pure Research (ZWO), is organizing an International Symposium on "Operationalization and Research strategy."

The symposium is cosponsored by the Research Committee on "Logic and Methodology" (RC33) of the International Sociological Association.

The Symposium

The topic of the Symposium is the interrelationship between the operationalization of central research concepts and general research strategy, incorporating c.q. including the changing function of theories and of mathematical modeling.

In recent years ideas about the role of theories in empirical research have taken at least two directions: some scientists firmly support verbal theory as a heuristic instrument for analyzing networks of central theoretical concepts, others adhere to formalized theory and strongly data oriented approaches for developing measurement instruments.

The primary goal of the symposium is to bring together experts from both fields to discuss the state of the art and further developments. The symposium also aims at bringing together researchers from the Netherlands and abroad to exchange ideas.

Program

The symposium will consist of three main parts. The first part gives an introduction to the problem area. In the second part, the so-called data or model approach will be the central theme. The so-called nomological approach will fulfill a central role in the third part of the symposium.

At present, the preliminary list of speakers includes: D. Alwin (ISR, USA), H.M. Blalock, Jr. (Un. of Wash., USA), J. de Leeuw (UCLA, USA), S. Levy (Inst. of Soc. Res., Israel), R. Pawson (Un. of Leeds, UK), K.D. Opp (Un. of Hamburg, BRD), P. Schmidt (Un. of Giessen, BRD), J. Dessens, W. Jansen, P.G. Swanborn (RUU), E.E. Roskam (Un. of Nijmegen), J. de Jong-Gierveld (Free Univ.), J.J. Hox, G.J. Mellenbergh (Un. of Amsterdam).

Location

The symposium will be held at the Free University, De Boelelaan 1105, Amsterdam. The university is located within a five minutes' walk from the railroad station Amsterdam-South. From there it takes only ten minutes to reach Schiphol (Amsterdam Airport) by train. More explicit directions will be available later.

Registration

The registration fee is Dfl. 150 (Dutch guilders). The fee will cover organizational costs, conference facilities, coffee and tea during breaks, and the conference proceedings.

Registration forms, including payment, should be sent to:

Dr. J.J. Hox
Begripsvorming en Research Design
Prinsengracht 227
1015 DT AMSTERDAM
The Netherlands
postal giro account 2743948
(E-mail: U00005 at HASARA5)

Payment can also be made by sending a cheque for Dfl. 170 (including cheque costs) to Dr. J.J. Hox at the address above. Late registrations (after April 30) should contact Dr Hox to insure participation.

This symposium has been financially supported by:
The Royal Netherlands Academy of Arts and Sciences (KNAW)
University of Amsterdam, Faculty of Educational Sciences
Free University, Amsterdam, Faculty of Social Sciences
University of Utrecht, Faculty of Social Sciences
CBS, Voorburg/Den Haag
DSWO, Leiden
ITS, Nijmegen
IVA, Tilburg
NIDI, Den Haag
NISSO, Utrecht
SEO, Amsterdam
SISWO, Amsterdam
Burke-Inter/View, B.V., Amsterdam

Journal of Official Statistics
Vol. 4, No. 1, 1988, pp. 81–90
© Statistics Sweden

Book Reviews

Books for review are to be sent to the Book Review Editor Jan Wretman, Statistical Research Unit, Statistics Sweden, S–115 81 Stockholm, Sweden.

Dalenius, T., Elements of Survey Sampling. The Swedish Agency for Research Cooperation with Developing Countries (SAREC), Stockholm, 1986, ISBN 91 8682604-2, vi + 284 pp., (Distributed by Statistics Sweden, S-115 81, Stockholm, Sweden).

Elements of Survey Sampling is a collection of outlined notes intended for use in the instruction of the elementary principles and techniques of survey sampling. It is not a self-contained text for either classroom teaching or self-instruction of survey sampling students. Rather, it is a course outline which focuses on topics of sampling theory and methods that are essential for the training of survey practitioners. In the author's own words, "... the prime purpose of these notes is to provide a *guide* to the literature on survey sampling, just as a map and a travel folder may serve as a guide to an unknown territory."

Seasoned instructors of survey sampling are aware of changes that have occurred over time in the composition of student audiences. With the proliferation of surveys as a research tool and increasing specialization of survey methodology, today's beginning students tend more and more to be nonstatisticians in the social sciences, public health, demography, business, and other fields, who are interested primarily in a basic introduction to survey sample design. These individuals may work extensively in survey development and the analysis of survey data, but most will never be responsible for the actual sample design and selection. A related but somewhat different situation occurs in developing countries where students and researchers may have only a limited opportunity to study survey sampling theory and methods. These students may be best served by a basic course that stresses the fundamentals of survey practice.

For some time, there has been a need for elementary texts and pedagogical materials which are theoretically sound but emphasize survey sampling practice. Williams (1978), Kalton (1983), and Stuart (1984) are recent examples of monographs that introduce new students and researchers to the fundamentals of sample design methods. In *Elements of Survey Sampling*, Dalenius has taken a different route by providing not a self-contained text but an annotated course outline which defines a model for the introductory course in survey sampling.

The author organizes the subject into eight major parts. Part 1 is devoted to basic survey definitions and concepts required in the design of a sample-based survey. Important discussion of sampling frames and the basic taxonomy of probability sampling methods follows in Part 2. Parts 3 to 5 cover those aspects of the theory and applications of element, cluster, and multi-stage sampling

that a practitioner would need. Nonsampling errors are introduced in Part 6, followed by Part 7 in which the author stresses the importance of considering both sampling and nonsampling errors in the total survey design. The notes conclude with Part 8 containing six case studies that are selected to illustrate the principles outlined in the preceding sections.

The basic topic by topic outline of the notes is very good for a one semester course or training seminar in survey sample design, although instructors should expect to augment many sections with additional material. The notes address the elements of sampling theory and methods in a way that is consistent with real problems and contemporary survey practice. For example, the chapter on cluster sampling moves directly to sampling of unequal size clusters and treats sampling of equal size clusters as a special case. Important topics such as unequal probability sampling, weighting, poststratification, and the role of nonsampling errors in total survey design are covered; valuable special topics such as the selection of samples of time dependent processes are also dealt with.

For the most part, the statistical presentation is clear and easy for the new student to follow. Occasionally, the presentation is so compendious that some chapters are likely to be more useful for the instructor than for the student. Two examples are the special chapters on the intraclass correlation coefficient and the use of conditional expectations in derivations. By the nature of the outline-like form, the composition of the book is uneven. In many chapters, topics are developed in a smooth, readable fashion much like a standard text. Other chapters are terse and rely heavily on references and parallels to previous sections. Exercises included at the end of each chapter and a supplemental set of exercises provided in the appendix are not complicated and appear well designed to test the student's knowledge of the notes and for stimulating discussion in the classroom.

Elements of Survey Sampling provides a good basic outline for an introductory course in the fundamentals of survey sampling and is a useful addition to the survey sampling instructor's library. However, as a course text it should be used in conjunction with supplemental notes and selected readings from journals and more advanced texts. In many respects, this is not a disadvantage, since individual instructors almost invariably will want to adapt the course syllabus to the specific needs of their students.

References

Kalton, G. (1983): Introduction to Survey Sampling. Sage Publications, Beverly Hills, CA.

Stuart, A. (1984): The Ideas of Sampling. Charles Griffin and Company, High Wycombe.

Williams, B. (1978): A Sampler on Sampling. John Wiley & Sons, New York.

Steven G. Heeringa
University of Michigan
Ann Arbor, MI
U.S.A.

Mosteller, F. and Wallace, D.L., Applied Bayesian and Classical Inference: The Case of the Federalist Papers. Springer-Verlag, New York, 1984 (The first edition of this book, *Inference and Disputed Authorship: The Federalist*, was previously published by Addison-Wesley Publishing Company, Inc., Reading, MA, in 1964), ISBN 0-387-90991-5 (Springer-Verlag, New York), ISBN 3-540-90991-5 (Springer-Verlag, Berlin), xxxviii + 303 pp., $ 32.30.

1. Introductory Comments

The reviewed book describes a complex study, in which Bayesian and non-Bayesian methods of classification have been employed on a large textual data set, for the purpose of statistically resolving a problem of disputed authorship. More specifically, the data consists of seventy-seven newspaper articles, called the *Federalist Papers*, published in 1787–1788 by Alexander Hamilton, James Madison and John Jay. Five of these papers are known to be written by Jay, forty-three papers by Hamilton and fourteen by Madison. Three papers are known to be written jointly by Hamilton and Madison. The authorship of the other twelve papers is disputed. The problem is to decide, for each of the disputed papers, whether it was written by Hamilton or

by Madison. In addition to these *Federalist Papers*, there are other articles written by Hamilton and by Madison, that could be used as a training set. It is not easy to decide who is the author of a disputed paper by obvious measures like: length of paper, content, style, and structure of sentences. The articles of Hamilton and Madison vary in length from 900 to 3 500 words. Other stylistic features are similar and are inappropriate to the purpose of discrimination. On the other hand, the frequency of occurrences of certain *function words*, e.g., "this," "an," "would," "while," "whilst," were found to constitute good characteristics for classification. Contextual words are not used in order to avoid the need to assess evidence historically. Thus, a considerable part of the study is devoted to the question of selecting the appropriate function words for discrimination.

Modeling the probability distributions for the counts of the discriminators (function words) and selecting the prior distributions for the parameters of these probability distributions is one of the book's basic problems. In addition, computing the posterior odds (or log-odds) of Hamilton versus Madison for each of these words is another major task. The known papers and other works of Hamilton and Madison are used as training samples for the numerical determination of prior parameters. These values are then applied to the disputed papers to obtain the Bayesian degree of evidence (posterior odds) concerning the authorship. There are, however, many technical difficulties associated with this research program, and various simplifying assumptions are required. Some of these assumptions (like independence, stability over time) are tested but it is impossible and not necessary to test every assumption.

In the following sections we discuss the Bayesian methodology of the present study. Some non-Bayesian (frequentist or classical) methods have also been used in the book, but this review is focused only on the Bayesian procedures.

2. *Abstract Structure of the Main Study*

Let $\tilde{\theta}_l$ be a parameter assigned to the lth *unknown* paper ($l = 1,\ldots,L$), $\tilde{\theta}_l = 1$ for Hamilton, and $\tilde{\theta}_l = 2$ for Madison. Let $\tilde{Y}_l$ be the data associated with the lth paper (frequency counts of selected function words), and let $\tilde{X}$ represent the data associated with the *known* papers. Let $\underline{\mu}$ be a vector of parameters which specify the probability distribution functions. The following are basic assumptions on the conditional and prior densities (p. 112):

$$(A.1)\quad p(\tilde{X},\tilde{Y}_1,\ldots,\tilde{Y}_L \mid \theta_1,\ldots,\theta_L,\underline{\mu}) = p(\tilde{X} \mid \theta_1,\ldots,\theta_L,\underline{\mu}) \prod_{l=1}^{L} p(\tilde{Y}_l \mid \theta_1,\ldots,\theta_L,\underline{\mu});$$

$$(A.2)\quad p(\tilde{Y}_l \mid \theta_1,\ldots,\theta_L,\underline{\mu}) = f_l(\tilde{Y}_l \mid \theta_l,\underline{\mu}),\ l = 1,\ldots,L;$$

$$(A.3)\quad p(\tilde{X} \mid \theta_1,\ldots,\theta_L,\underline{\mu}) = g(\tilde{X} \mid \underline{\mu});$$

$$(A.4)\quad p(\underline{\mu},\theta_1,\ldots,\theta_L) = h(\underline{\mu})\,\Pi\,(\theta_1,\ldots,\theta_L).$$

According to (A.3) and (A.4), the posterior probability distribution function (p.d.f.) of $\underline{\mu}$, given $\tilde{X}$, is $h(\underline{\mu} \mid \tilde{X}) \propto g(\tilde{X} \mid \underline{\mu})h(\underline{\mu})$. Thus, after determining $h(\underline{\mu} \mid \tilde{X})$ one can compute the posterior odds for any assignment $(\theta_1,\ldots,\theta_L)$ of the L disputed papers ($\theta_l = 1,2$). Let $\theta_l^1 = (1 + \theta_L) \pmod 2$, $l = 1,\ldots,L$. The posterior odds for the given assignment is

$$\frac{P\{\tilde{\theta}_1 = \theta_1,\ldots,\tilde{\theta}_L = \theta_L \mid \tilde{X},\tilde{Y}_1,\ldots,\tilde{Y}_L\}}{P\{\tilde{\theta}_1 = \theta_1^1,\ldots,\tilde{\theta}_L = \theta_L^1 \mid \tilde{X},\tilde{Y}_1,\ldots,\tilde{Y}_L\}} =$$

$$= \frac{P\{\tilde{\theta}_1 = \theta_1,\ldots,\tilde{\theta}_L = \theta_L\}}{P\{\tilde{\theta}_1 = \theta_1^1,\ldots,\tilde{\theta}_L = \theta_L^1\}}$$

$$\times \frac{E\left\{\prod_{l=1}^{L} f_l(\tilde{Y}_l \mid \theta_l,\underline{\mu}) \mid \tilde{X}\right\}}{E\left\{\prod_{l=1}^{L} f_l(\tilde{Y}_l \mid \theta_l^1,\underline{\mu}) \mid \tilde{X}\right\}},$$

where, according to the above assumptions,

$$E\left\{\prod_{l=1}^{L} f_l(\tilde{Y}_l \mid \theta_l, \underline{\mu}) \mid \tilde{X}\right\}$$

$$= \prod_{l=1}^{L} f_l(\tilde{Y}_l \mid \theta_l, \underline{\mu}) h(\underline{\mu} \mid \tilde{X}) d\underline{\mu}.$$

Thus, the posterior odds is a product of the *prior odds* and the *odds-factor*, which is obtained by averaging likelihood functions with respect to posterior distribution of $\underline{\mu}$ given $\tilde{X}$.

N words are selected. Let $\tilde{X} = (X_1, \ldots X_N)$ and $\tilde{Y}_l = (Y_{l1}, \ldots, Y_{lN})$. Similarly, the parametric vector $\underline{\mu}$ is partitioned into N components $\tilde{\mu}_n$. It is assumed that

$$g(\tilde{X} \mid \underline{\mu}) = \prod_{n=1}^{N} g(X_n \mid \tilde{\mu}_n)$$

$$f_l(\tilde{Y}_l \mid \theta_l, \underline{\mu}) = \prod_{n=1}^{N} f_l(\tilde{Y}_{ln} \mid \theta_l, \tilde{\mu}_n).$$

This assumption means that the random variables associated with different function words are *conditionally independent*.

Thus, the study was performed in three stages. In stage 1 the posterior distributions $h(\tilde{\mu}_n \mid X_n)$, $n = 1, \ldots, N$, where determined. In stage 2 the odds-factors for the assignment of authorship for each disputed paper were determined. In stage 3 prior odds and odds-factors were combined.

3. *Parametric Modeling: Likelihoods and Priors*

The parametric model for the number of occurrences of a given word in a block is the negative binomial p.d.f.

$$f_{nb}(x; \mu, \varkappa) = \frac{\Gamma(\varkappa + x)}{\Gamma(\varkappa) x!} \left(\frac{\mu/\varkappa}{1 + \mu/\varkappa}\right)^{x}$$

$$\times \left(\frac{1}{1 + \mu/\varkappa}\right)^{\varkappa}, \quad x = 0, 1, \ldots$$

where μ is the rate of occurrence and $\delta = \mu/\varkappa$ is called the non-Poissonness index. Thus, the expected number of occurrences of a given function word, in a string of size ω is $\omega\mu$, and its variance is $\omega\mu(1 + \omega\delta)$. The p.d.f. with the parameter ω assumes the form

$$f_{nb}(x; \omega\mu, \varkappa, \delta)$$

$$= \frac{\Gamma(\varkappa + x)}{\Gamma(\varkappa) x!} (\omega\delta)^{x} (1 + \omega\delta)^{-(x+\varkappa)} .$$

Let $\tilde{X} = \{x_{nij}; n = 1, \ldots, N; i = 1,2; j = 1, \ldots, J_i\}$, where $i = 1 \leftrightarrow$ Hamilton, $i = 2 \leftrightarrow$ Madison; J_i is the number of known papers of the ith author, and N is the number of words. The p.d.f. of $\tilde{X}$, given $\{\mu_{ni}, \delta_{ni}\}$; $n = 1, \ldots, N$; $i = 1,2$, is:

$$p(\{x_{nij}\} \mid \{\mu_{ni}, \delta_{ni}\})$$

$$= \prod_{n=1}^{N} \prod_{i=1}^{2} \prod_{j=1}^{J_i} f_{nb}(x_{nij} \mid \omega_{ij}\mu_{ni}, \varkappa_{ni}, \omega_{ij}\delta_{ni}).$$

For the prior distributions of the parameters, they consider the reparametrization:

$$\sigma_n = \mu_{n1} + \mu_{n2} ,$$
$$\zeta_{ni} = \log(1 + \delta_{ni}), \; i = 1,2 ,$$
$$\xi_n = \zeta_{n1} + \zeta_{n2} ,$$
$$\tau_n = \frac{\mu_{n1}}{\mu_{n1} + \mu_{n2}} ,$$
$$\eta_n = \zeta_{n1}/(\zeta_{n1} + \zeta_{n2}) .$$

Let $\underline{\gamma}_n = (\sigma_n, \tau_n, \xi_n, \eta_n)$, $n = 1, \ldots, N$. It is assumed that:

i. $\underline{\gamma}_n$ are independent across words;

ii. (ξ_n, η_n) and (σ_n, τ_n) are independent for each n;

iii. the prior measure of σ_n is a constant (diffused quasi-prior);

iv. the conditional distribution of τ_n, given σ_n, is a symmetric beta $(\beta_1 + \beta_2\sigma_n, \beta_1 + \beta_2\sigma_n)$; η_n is distributed like beta (β_3, β_3); ξ_n has a gamma distribution with a scale parameter (β_4/β_5) and shape parameter β_5. The set $(\beta_1, \ldots, \beta_5)$ is called a set of *underlying constants* (hyperparameters).

Notice that the prior distribution assumed are for the parameters $\underline{\gamma}_n$ across words. In other words, it is assumed that for groups, or sets of N words, the set $(\beta_1,\ldots,\beta_5)$ of underlying constants is the same. The vectors of parameters $\underline{\gamma}_1,\ldots,\underline{\gamma}_N$ could be considered as conditionally independent and identically distributed (i.i.d.) random variables.

4. *Estimation of Underlying Constants*

In a purely Bayesian framework, the statistician has to determine a-priori values for the underlying constants $(\beta_1,\ldots,\beta_5)$. In an applied study, of the nature discussed in the book, the researcher would generally like to estimate the (prior) underlying constants from the $\tilde{X}$ data. Section 4.5 of the book addresses this problem. It is a very difficult problem to estimate the underlying constants from the data $\tilde{X}$. In the book, a step-wise process of estimation, using the method of moments, is described. The sensitivity of the log odds to variations in the values of the underlying constants is studied in Section 4.5. The overall conclusion from this sensitivity analysis is that "a 12 percent allowance in log odds for ill-determined priors is generous."

5. *Modal Approximations*

The log odds-factor for all N function words found in a disputed paper, of size ω_0, is

$$\lambda = \sum_{n=1}^{N} \{ \log \int\int f_{nb}(x_n \mid \omega_0\mu_{n1}, \varkappa_{n1}, \omega_0\delta_{n1}) \, h(\mu_{n1}, \delta_{n1} \mid \tilde{X})\, d\mu_{n1} d\delta_{n1}) - \log \int\int f_{nb}(x_n \mid \omega_0\mu_{n2}, \varkappa_{n2}, \omega_0\delta_{n2}) \, h(\mu_{n2}, \delta_{n2} \mid \tilde{X})\, d\mu_{n2} d\delta_{n2} \} .$$

The expectation (double integration) of $f_{nb}(x \mid \omega\mu, \varkappa, \omega\delta)$ with respect to the bivariate posterior distribution of (μ, δ), given $\tilde{X}$, cannot be performed analytically. Instead of using numerical integration the authors decided to approximate $E\{ f_{nb}(x_n \mid \omega_0\tilde{\mu}_{ni}, \tilde{\varkappa}_{ni}, \omega_0\delta_{ni}) \mid \tilde{X} \}$ by $f_{nb}(x_n \mid \omega\hat{\mu}_{ni}, \hat{\varkappa}_{ni}, \omega_0\hat{\delta}_{ni})$, where $(\hat{\mu}_{ni}, \hat{\delta}_{ni})$ is the *mode* of the corresponding posterior distribution. Such modal approximations have been used in the literature of Bayesian analysis. The present study devotes a considerable portion (Section 4.6) to the study of the effectiveness of the modal approximation, and to the adjustment of the approximation to obtain better results. In particular, the Laplace Integral Expansion is utilized to obtain an adjustment appropriate for the negative binomial model. The correction is of order $1/m$, where m is the sample size. A study of the effectiveness of the modal approximation for five of the most important words suggests that the approximation overstates the log odds, and a 15 % reduction was indicated. The reader may find in this part of the study technical details of interest, which are potentially applicable in other studies of similar nature.

6. *Logarithmic Penalty*

The question is how to evaluate, or compare, different methods of probability predictions.

Suppose that a trial has k alternative outcomes. Let $q_{ij}(\phi)$ denote the probability prediction of the ith outcome at the jth trial, by method ϕ; $q_{ij}(\phi) \geq 0, \sum_{i=1}^{k} q_{ij}(\phi) = 1$. In the present study $k = 2$, $i = 1$ for a decision to ascribe paper j to Hamilton, and $i = 2$ if the paper is ascribed to Madison. The index j will range over the known papers of Hamilton and of Madison, the penalty should increase as $q_{ij}(\phi)$ deviates from 1. Thus, the penalty for using method ϕ is defined as

$$U(\phi) = \sum_{j\varepsilon \{48 \text{ Hamilton test papers}\}} (-\log q_{1j}(\phi)) + \sum_{j\varepsilon \{48 \text{ Madison test papers}\}} (-\log q_{2j}(\phi)).$$

We denote by $H(\phi \mid \phi')$ the expectation of $U(\phi)$ according to $q_{ij}(\phi')$, i.e.,

$$H(\phi \mid \phi') = \sum_j [q_{1j}(\phi')(-\log q_{1j}(\phi)) + q_{2j}(\phi')(-\log q_{2j}(\phi))].$$

Differences of $U(\phi)$ from its expectation are measured by the standardized score

$$Z(\phi \mid \phi') = \frac{U(\phi) - H(\phi \mid \phi')}{\sqrt{V\{U(\phi) \mid \phi'\}}}$$

Applying this log penalty score, Poisson and negative binomial models, with different sets of underlying constants, have been compared.

7. *Concluding Remarks*

The present book is difficult to read and difficult to review. It is not the type of "theorem-proof" book, which is simpler to review. To assist the reader, the authors provided an Analytic Table of Contents. Following the title of each section or subsection is a description of the content of the section. This Analytic Table of Contents serves several important functions. It provides a detailed synopsis of the book, a short summary of each section, explanation of the various data tables, and guidelines for theoretical development. This is a nice and helpful feature which other authors should be encouraged to follow.

This book presents a very complex project of Bayesian inference, with a very nonconventional data base. There are many subjective components in defining the pertinent variables, in modeling their distribution functions and the corresponding prior distributions, and in applying inferential techniques. A reader who has had experience with similar types of studies may offer alternative suggestions for the analysis of the data. However, without a thorough understanding of the data and its characteristics, alternative modeling would be of no value. The authors obviously had penetrating insight into the particular problem, and presented their research in a masterful manner. The book is a very well written scholarly work and I have found it highly gratifying reading. I strongly recommend it as a text for graduate seminars on Bayesian inference in applied statistical research.

Shelemyahu Zacks
State University of New York
Binghamton, NY
U.S.A.

Rausser, G.C. (Ed.), New Directions in Econometric Modeling and Forecasting in U.S. Agriculture. North-Holland Publishing Company, New York, 1983, ISBN 0-444-00764-4, xiii + 830 pp., $85.00.

Econometric models and analyses have a long tradition in agricultural statistics in both the United States and Sweden. The development and application of agricultural models have often led developments in econometrics. Starting with single equation models for the analysis of the supply and demand of a particular agricultural commodity, the models have expanded to include systems of simultaneous equations spanning groups of commodities and several subsectors of an agricultural system, including links between different subsectors.

The early 1970s saw a faultering in economic development, with sharp increases in the prices of food and oil. During this period, it became obvious that the then current econometric models no longer gave good results. Forecasts and policy simulations were far off the mark, and contradicted the reality concurrently observed. The major drawback of these models seemed to be that they ignored the international links. They were constructed under the assumptions valid for agriculture in the U.S., i.e., a closed system. In retrospect, a model assuming an open system seems much more relevant.

Given the disappointing performance of econometric models in the 1970s, it became imperative to look for new directions. To this end, a series of annual conferences focusing on the modeling of the U.S. agricultural sector were held between 1976 and 1980. The volume under review is an edited collection of selected papers from the five sequential conferences.

In this volume, the editors present six areas characterized by new directions and developments:

I. Supply Response and Demand Analysis.
II. Expectation Formation Patterns.
III. Qualitative Econometrics.
IV. Agricultural Trade Analysis.
V. Government Policy Analysis.
VI. Forecasting Methods, Evaluations, and Model Use.

The first part contains six papers. One of them presents an application and the others are reviews of the current research and some new developments.

The first chapter in Part I by G.W. Ladd deals with analysis, in particular the incorporation of product characteristics in econometric models. Ladd reviews developments starting with the studies by Waugh (1928). In the next chapter, M. Hanemann deals with another problem in demand analysis, that of modeling the demand for commodities when the consumer's choices are influenced by the quality of the products. Two groups of models are explicitly evaluated, namely, the Generalized Lancaster model and the Houthakker-Theil model. These models are also based on work by Waugh (1928).

Chapter 4, by P. Berck and G.C. Rausser, reviews new developments in the modeling of the structure between the production of and retail of agricultural goods. The introduction of different grades and branches into the models, together with consumer uncertainty with respect to quality, are new features. In their work, Berck and Rausser also discuss testing for pure or imperfect competition in the structure. The next chapter by R. Weaver deals with an application of supply response functions on data from North and South Dakota. These models are based on different theories of choice.

Starting from a distinction between normative versus positive economic analysis, J.A. Hallam, R.E. Just, and R.D. Pope in the last chapter of Part I discuss the analysis of risk in agriculture. Their approach utilizes the positive economic analysis in terms of primal and dual approaches, the former due to Marshak and Andrews (1944), the latter due to Shephard (1970); McFadden (1978), and Diewert (1974).

The incorporation, specification, and operational use of unobservable variables is the theme of the three chapters of Part II. In its first chapter, a treatment of rational expectations in linear, simultaneous econometric models is given by J.-P. Chavas and S.R. Johnson. The article treats the subject without any applications but with a clear focus on methodological aspects valuable for applied work.

In the next chapter, an application of price-expectations in a corn market model is presented by A. Subotnik and J.P. Houck. They present a simultaneous system in six estimated equations and one identity, using quarterly data. The last chapter by M.H. Jameson discusses statistical testing as applied to the U.S. hog price cycle. The hypothesis under testing states the use of rational expectations by hog producers in their breeding decisions. The three chapters of this part constitute a well-balanced mix of theoretical developments and applications.

Part III starts with Chapter 10 where qualitative econometrics is defined as the problem of analyzing and estimating models where the observed dependent variables are not completely continuous. In this chapter, R.G. Chambers and R.E. Just present an overview of qualitative econometrics with special emphasis on agriculture. The analysis of data and models of agricultural research can benefit much from the new econometric techniques.

This part of the monograph concludes with an illustrative application. G.C. Rausser and C. Riboud analyze the U.S. wheat and corn market where government support of prices justifies the use of disequilibrium models. In the article, a comparison is made between conventional 2SLS estimation and a modified tobit technique.

The inclusion of the international market in what is called open economy models spurred revision of the modeling and analysis of the U.S. agricultural sector. The introduction of open systems is the theme of the first chapter of Part IV on agricultural trade analysis. R. Thompson and P. Abbott give an account of recent developments. Although they give attention to possible topics for future research, their article is based on past and present findings. They include a list of over 100 references. The second and last chapter of this part could also have appeared under the heading "Qualitative Econometrics." It is an application of disequilibrium techniques on the U.S. beef import market written by R.G. Chambers, R.E. Just, L.J. Moffitt, and A. Schmitz.

The fifth part of the monograph deals with new developments in governmental policy analysis. The first two chapters deal with problems concerning the analysis of stocks for the wheat market (Burt, Koo, and Dudley) and the grain market (Gardner). The conventional demand and supply equations are supplemented with specific information on the storage capacity and, in the latter article, with

demand and supply equations for storage. The chapter by Burt, Koo, and Dudley develops a storage optimization model in the form of a modified Markov process dynamic programming system. In Gardner's work, a special feature is the discussion of governmental storage as well as private (speculative) storage.

In Chapters 16 and 17, the links between the domestic and the international economies are once again illustrated in empirical models. The first of the two chapters, by R.E. Just and J.A. Hallam, focuses on price stabilization. After a review of new theoretical developments and their applications in econometric modeling, a model of the U.S. wheat market is developed. The model is estimated and analyzed with special emphasis on nonlinearities and residual characteristics with consequences for the flexibility of the model. The final simulation suggests large welfare gains from price stabilization, with more benefits for foreign concerns than individual domestic concerns.

In Chapter 17, a medium-sized econometric model with 51 behavioral equations and 36 identities is developed to evaluate the interaction between sector policies, especially the agricultural sector, and the general economy. This model focuses on the effects of different exogenous shocks on the above-mentioned interactions. The authors are J.W. Freebairn, G.C. Rausser, and H. de Gorter.

This part of the monograph concludes with a review by G.C. Rausser, E. Lichtenberg, and R. Lattimore on recent developments in modeling governmental behavior in terms of endogenous variables. Four different paradigms are discussed: The liberalist-pluralist framework, the theory of the state framework, the theory of economic regulation, and the theory of rent-seeking interest groups and conflict resolution.

New methodology in forecasting, evaluation, and model construction is the topic of the last part of the monograph, Part VI (Chapters 19 to 23). A. Zellner has written a chapter which reviews the research on statistical methods for construction, analysis, and use of econometric models. Zellner expects Bayesian analysis to offer solutions to many finite-sample problems. One main point of this chapter is that the synthesis of traditional econometric model-building with time series analysis techniques will lead to an improved modeling algorithm.

The next chapter by G.C. Rausser, Y. Mundlak, and S.R. Johnson deals with parameter inconstancy, defined as the problem of structural change. Starting with an exposition of the problems and methods developed in this fast growing area of research, the authors develop new techniques for the estimation and forecasting of single and simultaneous equation systems. The following chapter, "Composite Forecasting in Commodity Systems," written by S.R. Johnson and G.C. Rausser is a review of the general research and new developments. One specific point of this chapter could be described as the development of links between on one hand the objective methods of analysis, say the formulation of an econometric model, estimation and simulation with passively generated data, and on the other hand the subjective methods. The latter could be exemplified by experts' judgemental forecasts. The composite methods, as the third developed alternative, use the two former components in analyzing forecasting systems rather than individual forecasting models.

The second to last chapter is devoted to discussions of general principles of policy modeling, and the last chapter deals with a case study. First, G.C. Rausser and R.E. Just formulate and extend their views on ten principles for policy modeling in agriculture. The case study is that of "Agriculture Canada" where S.R. Johnson, H.B. Huff, and G.C. Rausser summarize and evaluate the experience of the Marketing and Economics Branch of Agricultural Canada in its work on a large scale econometric model. The comprehensive monograph is appropriately concluded with a set of recommendations for use by policy analysts and forecasting units in their work on incorporating modern econometric thinking and modeling into agricultural research.

One weakness sometimes encountered in monographs is a lack of coordination between the different papers. In this volume, efforts have been made to prevent this inconsistency. This work has of course been much facilitated since the editor, G.C. Rausser, is the author or coauthor of not less then eight chapters.

This volume could have been improved by a less space-consuming treatment of the references in the text. With so many authors writing on somewhat overlapping themes, it is natural that many references are duplicated.

Maybe the editor should have chosen to present a complete, self-contained list of all references in the text. This might have saved some space, but, more important, such a list in itself gives a good illustration of the scope of the new developments. It gives a lot of information about both published and unpublished material that constitute the basis for the new methods, and it also tells you what has been left out. To get such a picture from 23 different, overlapping lists is not so easy.

Concluding this review, I would say that the general statistician may not find this work very interesting, but the general econometrician has much to gain from its contents. However, some new developments in methods and methodology are better presented in recent text-books such as Kmenta (1986) and Maddala (1986). The econometrician working in the agricultural field or for that matter, any econometrician interested in applications will certainly appreciate this volume. Its strength is the applications and the illustrations of new methodology in an area of great practical interest, namely, the agricultural sector.

References

Diewert, W.E. (1974): Applications of Duality Theory. Frontiers of Econometrics, Vol 2, eds. M. Intriligator and D. Kendrick. North Holland Publishing Company, Amsterdam.

Kmenta, J. (1986): Elements of Econometrics. Second Edition, Wiley, New York.

Maddala, G.S. (1986): Econometrics. McGraw-Hill, New York.

Marshak, J. and Andrews, W. (1944): Random Simultaneous Equations and the Theory of Production. Econometrica, 12, pp. 143–205.

McFadden. D. (1978): Cost Revenue and Profit Functions. Production Economics: A Dual approach to Theory and Applications, eds. M. Fuss and D. McFadden, Vol 1: The Theory of Production. North Holland Publishing Company, Amsterdam.

Shepard, R. (1970): Theory of Cost and Production Functions. Princeton University Press, Princeton, New Jersey.

Waugh, F.V. (1928): Quality Factors Influencing Vegetable Prices. Journal of Farm Economics, 10, pp. 185–196.

Lennart Bodin
University of Örebro
Örebro
Sweden

Rystedt, B. (Ed.), Land-Use Information in Sweden: Applications of New Technology in Urban and Regional Planning and in the Management of Natural Resources. Swedish Council for Building Research, Stockholm, 1987, ISBN 91-540-4665-3, 139 pp., SEK 70.00. (Distributed by Svensk Byggtjänst, Box 7853, S-103 99 Stockholm, Sweden.)

This book is a collection of essays by various authors describing state-of-the-art techniques used in Sweden to collect and process land-use information. The book was written as a basis for the seminar "New Techniques to Collect and Process Land-Use Data," held in Gävle, Kiruna, and Stockholm, Sweden, in January 1987.

The first essay is an introduction where the authors present the problem of natural resource planning (physical planning) in Sweden. Sweden is a large country relative to its population. However, the uneven geographic distribution of the population causes problems of meeting the demand for water, land, and other resources. The necessity of having adequate planning data for managing these natural resources was recognized in the mid-1960s. The government initiated a major project called "national physical planning." As part of the project, information was collected through county inventories. These became the cornerstone of the first phase of the national physical planning. An important outcome of this project was that the planning process spurred a dialogue among all levels of governmental planning agencies. The authors conclude their essay by formulating the desirable properties of an EDP-based information system to support physical planning.

The next essay discusses the role of Statistics Sweden in the generation and dissemination of planning statistics. The authors state that the production of "small-area" planning statistics must be obtained from "all-inclusive investigations" or "comprehensive registers." The authors go on to describe the primary registers available in Sweden. These lists include the Central Population Register, the Central Business Register, and the Central Real Property Register. In addition, registers of important data items such as population change, income, car ownership and employment are also available. The authors explain how unique identifiers are used in these registers to create a cohesive statistical data base.

In addition, Statistics Sweden designed a key-code system to produce small area statistics. The key-code is a six digit hierarchical code designating a statistical area. This code is assigned to each unit of real property within a statistical area.

The third essay discusses the production and uses of digital geographic information in Sweden. Digital geographic information is produced along with the development of official maps in scales ranging from 1:10 000 to 1:1 000 000. In addition to topographical and land coverage digital data, the National Land Survey in cooperation with other organizations produced data bases covering flight information, place names, and hydrography.

The authors also briefly describe the Swedish Land Data Bank System, the Real Property Tax Register, the Real Estate Appraisal System, the National Highway Data Base, and the Municipal Utility Data Base. The authors close this essay with a discussion of a government initiated project to survey existing spatially-referenced data bases. The goal of this project is to determine which agencies maintain the data bases, and then determine the prerequisites for coordination and joint use.

The fourth essay gives an introduction into the uses of satellite remote sensing in land-use planning. The author briefly describes the Landsat Thematic Mapper and SPOT High Resolution Visible Sensor. The essay mentions some potential uses such as updating maps, land-use change detection, and water quality analysis. The author points out that the use of satellite imagery in physical planning has been hampered by many technical problems. However, the simultaneous availability of geocoded satellite images and inexpensive processing systems are expected to create new opportunities.

The fifth essay describes the construction and use of the Swedish Land Data Bank System (LDBS). The LDBS is the automation of the real property registers in Sweden. It has been in operation for ten years and covers over 40 % of the real property units. The LDBS has increased the efficiency of the property and land register authority; allowed easy access to data increasing the efficiency of banks, brokers, and other real property data users; and enhanced the ability to compile, analyze, and present real property statistics.

The next two essays present specific local projects in land-use data processing. One essay outlines the development of a computer-based planning tool which tracks and forecasts housing trends in the city of Köping, Sweden. The system is based on a digitally-stored large-scale map and is maintained through links to ongoing city functions. Another essay describes the digital mapping system used in the city of Göteborg, Sweden. The author describes the interactive graphics system and the methods used to convert existing hard copy mapping materials to a digital form (manual digitalization and optical scanning).

The next three essays discuss regional planning. One essay describes how coordinate-based data are put into the Swedish LDBS and how they are used. The author describes the link between the LDBS and other Swedish registers where the real property designations are recorded. Another essay discusses research and development programs sponsored by the Swedish Association of Local Authorities. The programs entail more than 30 field projects such as large-scale mapping, maintenance of public utilities, office automation, and physical planning. The following essay describes the pooling of resources in the county of Norrbotten and the development of a regional Land Information Centre (LINFO). The LINFO Centre is jointly owned by the County Administration Board, the Regional Board of Forestry, Satellitbild AB (Satimage), and the Regional Land Survey.

The eleventh essay gives a discussion of the proposal for a new Swedish atlas. The proposed atlas would consist of 20 volumes and be produced during the period 1987 through 1996. The construction of the new atlas would be done in such a way as to generate a set of geographical data bases. In addition, an information system that uses the data bases for mapping and analysis would be developed.

This book is not a source of land-use data or of technical information on processing techniques. It is, however, useful as an overview of the current status of Swedish land-use data systems. It is also an informative reading on the experiences and evolution of a land-use data system. Readers interested in the field of land-use planning would find this book worth while.

Martin L. Holko
National Agricultural Statistics Service
Washington, D.C.
U.S.A.

Corrigenda

Gnoss, R. (1987): Letter to the Editor; Criteria for the Treatment of Nonresponse in Sample Surveys: A Reply to Bethlehem and Kersten (1985). Vol. 3, No. 4, pp. 477–483.

On p. 477 the contrast formula should read: $C = | \bar{Y}_r - \bar{Y}_{nr} |$.

On p. 478 the left side of formula (3) should read $S(\bar{y}_r)$;
the left side of formula (4) should read $| B(\bar{y}_r) |$;
and formula (6) should read $N \leq (t/k)^2 \{ (1/a-1)(1-\bar{Y}_r) / \bar{Y}_r \} / (1-Q)$. (6)

On p. 478, in the left column, 14 lines from the bottom, the y should be deleted.

On p. 483, in the third line, Y_r should read $\bar{Y}_r$.

Johnson, E.G. and King, B.F. (1987): Generalized Variance Functions for a Complex Sample Survey. Vol. 3, No. 3, pp. 235–250.

Exhibit 6, p. 246, was drawn incorrectly. The central rectangle of the histogram should show a frequency of 373 instead of 273, and the height of the rectangle and the vertical scale should be changed accordingly.

Printed in June 1988