Pak. J. Statist.
2011 Vol. 27(4), 329-335

KEN BREWER'S CONTRIBUTIONS TO SURVEY STATISTICS AND THE FOUNDATIONS OF STATISTICAL INFERENCE

Muhammad Hanif
National College of Business Administration and Economics, Lahore, Pakistan
Email: drhanif@ncbae.edu.pk

K.R.W. (Ken) Brewer joined the Sampling Section of the Commonwealth Bureau of Census and Statistics in 1954 two years after its founding by E.K. Foreman. Ken had been introduced to basic statistics through lectures by G.A. Barnard while studying for a B.Sc. (Physics) at Imperial College, London. His practical education in survey statistics came from working in the small group around Ken Foreman whose responsibilities for sample design and analysis were rapidly expanding. In a matter of months he was being given samples to design and analyse on his own, with minimal supervision. By the late 1950s, when the bulk of the then Sampling Techniques Branch was busily engaged in setting up an operational Labour Force Survey, Ken, with three others to help him, was shouldering responsibility for all the remaining survey work. As the Branch's activities returned to a less hectic state, Ken's interests turned in a more theoretical direction.

Specifically, he became absorbed with the problem of how best to tackle sampling with unequal probabilities without replacement. His first two published papers in the *Australian Journal of Statistics*, Brewer and Undy (1962) and Brewer (1963a), were on that topic. Shortly afterwards Ken came out with the most original of his early contributions to sampling theory, Brewer (1963b). He had been concerned with an entrenched reliance on "design-based", or as he later described it "randomization-based", sampling inference. In that paper he pointed out several areas in which "model-based" (also referred to as "prediction-based") inference was more useful. Those five were, in the words that he used at the time:

"(1) An expression for the conditional variance of a ratio estimator, subject to a particular sample of population units having been selected

"(2) The optimum probabilities of selection of the individual units in the population

"(3) The likely accuracy of a ratio estimate based on the largest population units deliberately selected (a "partial collection")

"(4) The extent of the diminution of the variance of the ratio estimator obtained by sampling without replacement (with total probability of selection proportional to size or some function of size) instead of with replacement (with the probability of selection at each draw proportional to size or the same function of size)

"(5) An extension of the expressions and estimates for conditional variance to two-stage sampling. (These formulae can easily be generalized to any desired number of stages.)"

During the interval between Ken's 1963 paper and the publication of Royall (1970), Ken published nothing more on this, later to become highly controversial, topic. Early on this was because he had become interested in the fundamentals of statistical inference, but later and far more seriously it was because of the decision made in 1965 to conduct a Census of Population in the Territory of Papua New Guinea (then under Australian administration) for which the Commonwealth Statistician had been given responsibility.

Ken was appointed as the Commonwealth Statistician's Personal Representative for the conduct of that Census, the bulk of the work for which was being supervised by the Territory Bureau of Statistics, and carried out in conjunction with its Departments of Native Affairs, of Agriculture, Stock and Fisheries, and of Public Health. The entire period between the decision to proceed with the Census and the commencement of the collection proper was only eleven months. An account of that extremely demanding activity, which Ken has always admitted could never have been brought to a successful conclusion without extensive help from Ken Foreman, can be found in Brewer and Whittington (1969). Ken was to pay a more general tribute to the role in the development of Australian official survey statistics along Hansen lines played by his Bureau supervisor in his 2005 E. K. Foreman lecture, published as Brewer (2005).

A useful by-product of the Territory Census was the interest it fostered in the selection of sample units with unequal probability without replacement, which is obviously more efficient but considerably more difficult to handle than "with replacement". As mentioned above, Ken's first three papers on the theory of survey sampling had been on sampling without replacement, but treatments of the same topic also appeared later in Brewer (1967), Brewer and Hanif (1970) and Brewer (1975).

The 1970s, with the Territory Census safely behind him, were a period during which Ken was once again able to turn his attention to theoretical statistics. The paper by Foreman and Brewer (1971) consolidated advances that those two authors had been making in sampling theory up to that time. They were chiefly concerned with harmonizing the properties of "randomization-based" and "prediction-based" inference. The eventual goal was to be able to use both approaches simultaneously, but the achievement of this continued to be elusive. In the meantime Brewer and Mellor (1973) made a light-hearted case for taking both of these (and especially prediction-based inference) seriously. A similar style of writing was later used in Brewer and Hanif (1983), Brewer, Early and Hanif (1984), Hanif and Brewer (1980), Brewer and Hanif (1991, 1992 and 1994) and in Brewer (2002a).

During the later 1970s and early 1980s, as this topic was becoming increasingly adversarial, several more papers appeared that Ken had co-authored with one or more of E.K. Foreman, R.W. Mellor, D.J. Trewin, A.J. Scott, E.W.H. Ho and M. Hanif. All these papers were variations on the theme that both approaches were useful, but when combined invariably one of them (usually if not always the prediction-based) was seen as subordinate to the other. A comparison of half a dozen possible approaches was presented in Brewer and Särndal (1983), and the question as to how nearly model-based prediction and design-based estimation could ever be reconciled was raised explicitly in Brewer, Hanif and Tam (1988).

The decision taken in 1991 to hold an International Conference on Establishment Surveys (ICES) in Buffalo, N.Y. in 1993 was instrumental in bringing these matters to a head. The possibility of a complete reconciliation between randomization-based and prediction-based methods of survey inference was by then a hot issue, and Ken was recognized as being one of the most active researchers in that area. He had retired from the Australian Public Service in March 1992 in order to do research at the Australian National University, so when Phillip S. Kott, as editor for Part B of the envisaged monograph – the Part that was specifically dealing with Sample Design and Selection, asked him to write on that topic he decided that he would do his best to bring those two ideas together in time for the Conference.

This was no small ambition, but by the time Ken set out for the Conference in June 1993 he was confident he had indeed found the answer. The paper he presented was duly accepted for publication and appeared in *Business Survey Methods*, the 1995 book that contained a selection of the papers at that Conference, as Brewer (1995).

Immediately following the Conference, however, Ray Chambers carried out a test of Ken's suggested estimator and found it wanting. Ken took note of this, confirmed that it did indeed need changing, and requested permission to alter the paper before publication, but found that the Conference officials placed greater emphasis on the integrity of the publication as being an accurate account of proceedings. As it happened Ken was also preparing a paper for *The Pakistan Journal of Statistics* and took that opportunity to report on the relevant deficiency and how to remedy it. The result was published as Brewer (1994).

A more comprehensive discussion of the problem and how it could be completely resolved later appeared as the *Survey Methodology* article listed below as Brewer (1999b). Ken there described his method as "Cosmetic Calibration" after Särndal and Wright's (1984) article in the *Scandinavian Journal of Statistics*, which described "a cosmetic estimator" as one that was readily interpretable both as a design-based and a model-based estimator. In Brewer (1999b) Ken took the opportunity to define a "cosmetically calibrated estimator" more narrowly as one for which the randomization-based and the prediction-based forms were identical. A fuller account of the controversy between advocates of purely randomization-based and purely prediction-based sampling inference can also be found in Brewer and Gregoire (2009).

Ken's retirement from the Australian Public Service had freed him to devote more of his time to statistical research. Initially his interests had predominantly been in survey sampling, but even before his 1999b Cosmetic Calibration article had effectively eliminated the distance between randomization and prediction inference, he had been turning much of his attention towards statistical inference and hypothesis testing.

As he then saw it the Frequentist approach in which he had been brought up was incapable of using the tools of Bayesian inference for hypothesis testing, largely because the latter either demanded a subjective prior (which by definition could vary from individual to individual) or else, if they used an objective prior it was always one that signally failed to provide a convincing result in limiting circumstances. It was largely this failure that had given rise to the "Lindley Paradox" of Lindley (1957), which demonstrated that Frequentists and Bayesians, faced with the same evidence, could come to contradictory conclusions.

Ken, however, refused to accept that the Bayesian analysis of this situation would necessarily require a subjective prior. In Brewer (2002) (another *Pakistan Journal of Statistics* article) he quoted extensively two Bayesian statisticians, Lavine and Wolpert (1995), as follows:

"Before asking whether [O'Hagan's (1995)] solution [to the problem as to how to avoid using a subjective prior] is sensible we want to ask first whether any solution is sensible.

"Improper priors are often used in the hope that their posteriors approximate well the posterior that would have resulted from any well-thought-out prior. We typically reason that 'my prior is flat compared with the likelihood', 'there is much more information in the likelihood than in the prior' and therefore 'my posterior is well approximated by the posterior from a convenient improper prior.' Then we adopt the improper prior and invest our effort more productively in other aspects of the analysis.

"But approximation of the prior is not the same as approximation of the Bayes Factor. When O'Hagan considers an improper prior without considering which proper prior's posteriors he is hoping to approximate then we should bear in mind all proper priors with posteriors similar to that from the improper prior. If those priors were to yield roughly similar Bayes factors, then it would be possible to associate a Bayes Factor (or a small range of Bayes Factors) with the improper prior... If, however, there are priors that yield posterior distributions similar to that from the improper prior, but that yield vastly different Bayes Factors, then the specification of a Bayes Factor for the improper prior is problematic at best."

Lavine and Wolpert next presented a range of proper uniform priors that yielded roughly the same posterior distributions as O'Hagan's improper prior (uniform over the real line) but with Bayes Factors that ranged from 10^{-10} up to infinity. They effectively concluded that no Reference Bayesian hypothesis test could be devised for the problem.

Ken, however, refused to accept that finding a stable Bayes Factor was the only way out of the problem. He went back to the original exchanges between Lindley and Bartlett in 1957. There the original "Lindley Paradox" had actually been stated by Lindley as follows:

"An example is produced to show that if H is a simple [null] hypothesis and x the result of an experiment, the following two phenomena can occur simultaneously:

"(i) a significance test for H reveals that x is significant at, say, the 5% level;

"(ii) the posterior probability of H, given x, is, for quite small prior probabilities of H, as high as 95%.

"Clearly, the common-sense interpretations of (i) and (ii) are in direct conflict."

What is still commonly referred to as "the Lindley Paradox", however, is actually one stated by Bartlett only a few months later, namely that when H [the null hypothesis] was precise, and the variance of the prior distribution over the parameter space of $\bar{H}$ [the complementary alternative hypothesis] was infinite, the probability of H being true was unity for every finite value of x. Getting around what Bartlett described as this "silly answer" has arguably remained a perplexing problem for half a century.

Subjective Bayesians get around it by refusing to consider any limiting case, but there are other statisticians who would regard this as an unduly heavy price. (As to its attribution, however, Ken, in his 2002b article, pointedly refers to this problem as "the *Bartlett* Paradox".) More importantly, however, Ken then also offered another way to resolve it, namely to replace the unstable Bayes Factor by a stable Reference Posterior Odds.

This required a departure from the conventional use of probabilities in that context in favour of what Ken described as "Lebesgue-style" measures. Those measures were defined as obeying all the probability laws other than those that constrain them either to sum or to integrate to unity. In Brewer (1999c) he used such measures to produce a modified version of the Schwarz (1978)-inspired Bayesian Information Criterion which was (in its single parameter version) exactly equivalent to the limiting case of a Bayesian Hypothesis Test.

This problem has been a fruitful area for continuing research. New results are due to be published this year by Ken with collaborators Genevieve Hayes and A.N. Gillison that argue if the actual or implied null hypothesis (H) is diffuse, Fisher's use of his p-statistic to distinguish between what is significant at the 95% level and what is not so significant is broadly justified, but when H has a precise location, the requirement for 95% significance becomes, as has been argued by Berger and Delampady (1987), Berger and Sellke (1987) and Sellke, Bayarri and Berger (2001), something like $p = 3$. The mixture of foundational reflection, active collaboration, rigour and scientific application that this illustrates are recurring themes over Ken's remarkably productive career.

References – to papers (co-)authored by Brewer and referred to in the article

1962 Brewer, K.R.W. and Undy, G.C. Samples of two units drawn with unequal probabilities without replacement. *Austral. J. Statist.*, 4(3), 89-100.

1963a Brewer, K.R.W. A model of systematic sampling with unequal probabilities. *Austral. J. Statist*., 5(1), 5-13.

1963b Brewer, K.R.W. Ratio estimation and finite populations: some results deducible from the assumption of an underlying stochastic process. *Austral. J. Statist*., 5(3), 93-105.

1967 Brewer, K.R.W. A note on Fellegi's method of sampling without replacement with probability proportional to size. *J. Amer. Statist. Assoc*., 62, 79-85.

1969 Brewer, K.R.W. and Whittington, A.J. The 1966 sample population census of the Territory of Papua and New Guinea. *Rev. Internat. Statist. Assoc.*, 37(2), 149-165.

1970 Brewer, K.R.W. and Hanif M. (1970). Durbin's new multistage variance estimator. *J. Roy. Statist. Soc.*, Sr. B, 32(2), 302-311.

1971 Foreman, E.K. and Brewer, K.R.W. The efficient use of supplementary information in standard sampling procedures. *J. Roy. Statist. Soc.,* Ser. B, 33(3), 391-400.

1973 Brewer, K.R.W. and Mellor, R.W. The effect of sample structure on analytical surveys. *Austral. J. Statist*., 15(3), 145-152.

1975 Brewer, K.R.W. A simple procedure for sampling π *pswor*. *Austral. J. Statist*. 17(2), 166-172.

1977 Brewer, K.R.W., Foreman, E.K., Mellor, R.W. and Trewin, D.J. Use of experimental design and population modelling in survey sampling. *Bull. Internat. Statist. Inst*. XLVII, 173-190.

1977/78 Hanif, M. and Brewer, K.R.W. Comparison of Poisson sampling with some selection procedures for unequal probabilities. *J. Mathematics*, Punjab Univ, 10-11, 33-40.

1978 Scott, A.J., Brewer, K.R.W. and Ho, E.W.H. Finite population sampling and robust estimation. *J. Amer. Statist. Assoc*., 73, 359-361.

1979a Brewer, K.R.W. Seasonal adjustment of ARIMA series. *Economie Appliquée*, XXXI(1), 7-22.

1979b Brewer, K.R.W. A class of robust sampling designs for large scale surveys. *J. Amer. Statist. Assoc*., 74, 911-915.

1979/80 Hanif, M. and Brewer, K.R.W. Generalization of the Horvitz-Thompson estimator. *J. Mathematics*, Punjab Univ, 12-13, 23-34.

1980 Hanif, M. and Brewer, K.R.W. Sampling with unequal probabilities without replacement, a review. *Internat. Statist. Rev*., 48 (3), 317-366.

1983 Brewer, K.R.W. and Särndal, C-E. Six approaches to enumerative survey sampling. *Incomplete data in survey sampling*, 3, Session VIII, 363-368. Academic Press.

1983 Brewer, K.R.W. and Hanif, M. *Sampling with unequal probabilities*, Sprinter Verlag, New York, USA.

1984 Brewer, K.R.W., Early, L.J. and Hanif, M. Poisson, Modified Poisson and Cllocated sampling. *J. Statist. Plann. Infer.,* 10, 15-30.

1985 Hanif, M. and Brewer, K.R.W. Percentile regression and some related topics. *Scientific Report Published by the Australian Bureau of Agriculture and Resource Economics.* Canberra, Australia.

1986 Brewer, K.R.W. and Hanif, M. A further taxonomy of linear sample survey estimators. *Pak. J. Statist.*, (*J. App. Statist.*) 2(3), 21-30.

1988 Brewer, K.R.W., Hanif, M. and Tam, S-M. How nearly can model-based prediction and design-based estimation be reconciled? *J. Amer. Statist. Assoc*., 83(1), 128-132.

1990 Brewer, K.R.W. and Hanif, M. Design-based model-based sampling inference. *Proceedings of the Second Islamic Countries Conference on Statistical Sciences* (ICCS-II), Vol. IV, 970-994.

1990 Brewer, K.R.W. and Tam, S-M. Is the assumption of uniform intra-class correlation ever justified? *Austral. J. Statist.,* 32(3), 411-423.

1991 Brewer, K.R.W. and Hanif, M. Umar and Salim go population modelling. *Pak. J. Statist*., 7(3) B, 9-16.

1994 Brewer, K.R.W. Survey sampling inference: some past perspectives and present prospects. *Pak. J. Statist*., 10(1) A, 213-233.

1994 Brewer, K.R.W. and Hanif, M. Has sample survey undergone a scientific revolution? Paper presented as invited speaker during *Second Iranian Statistical Conference*, 1994.

1995 Brewer, K.R.W. Combining design-based and model-based inference. *Business Survey Methods*, Chapter 30. Wiley, New York, 589-606.

1999a Brewer, K.R.W. Design-based or prediction-based inference? Stratified random vs stratified balanced sampling. *Int. Statist. Rev.*, 67, 35-47.

1999b Brewer, K.R.W. Cosmetic Calibration with Unequal Probability Sampling. *Survey Methodology*, 25(2), 205-212.

1999c Brewer, K.R.W. Testing a Precise Null Hypothesis using Reference Posterior Odds. *Revista de la Real Academia de Ciencias* (*Review of the Spanish Royal Academy of Sciences* Special English language edition entitled *Bayesian Methods in the Sciences*), 93(3), 303-310.

2002a Brewer, K.R.W., *Combined survey sampling inference: Weighing Basu's Elephants*. Arnold: London.

2002b Brewer, K.R.W. The Lindley and Bartlett Paradoxes. *Pak. J. Statist*., 18(1), 1-13.

2005 Brewer, K.R.W. Anomalies, probings, insights: Ken Foreman's role in the sampling inference controversy of the late 20th century. *Austral. N. Z. J. Statist*., 47(4), 385-399.

2009 Brewer, K.R.W. and Gregoire, Timothy G. Introduction to Survey Sampling. Chapter 1 of *Handbook of Statistics* 29A, *Sample Surveys: Design, Methods and Applications*, D. Pfeffermann and C.R. Rao (Eds), Elsevier.

A LIST OF OTHER STATISTICAL PUBLICATIONS REFERRED IN THIS ARTICLE

1957 Lindley, D.V., A Statistical Paradox. *Biometrika*, 44, 187-192.

1957 Bartlett, M.S. A Comment on D.V. Lindley's Statistical Paradox. *Biometrika*, 44, 533-534.

1970 Royall, R. M., On finite population sampling theory under certain linear regression models. *Biometrika*, 57, 377-387.

1978 Schwarz, G. Estimating the dimension of a model. *Ann. Statist*., 6, 461-464.

1984 Särndal, C-E. and Wright, R.L. Cosmetic form of estimators in survey sampling. *Scand. J. Statist*. 11, 146-156.

1987 Berger, J.O. and Delampady, M. Testing precise hypotheses (with discussion). *Statistical Science,* 2, 317-152.

1987 Berger , J.O. and Sellke, T. Testing a point null hypothesis: the irreconcilability of p-values with evidence (with discussion). *J. Amer. Statist. Assoc.,* 82, 119-139.

1995 O'Hagan, A. Fractional Bayes factors for model comparison (with discussion). *J. Roy. Statist. Soc.,* B 57, 99-138.

1995 Lavine, M. and Wolpert, R. Comment in the discussion of O'Hagan (1995), 132.

2001 Sellke, T., Bayarri, M.J. and Berger, J.O. Calibration of *p*-values for testing precise null hypotheses. *Amer. Statistician,* 55, 62-71.