Statistical Science
1986, Vol. 1, No. 1, 106–113

Opera Selecta Boxi

Seymour Geisser

A review of *The Collected Works of George E. P. Box*, (George C. Tiao, ed.) Belmont, CA, Wadsworth Publishing Co., 1985. Volume I, xiv + 657 pp, $42.95. Volume II, xiv + 710 pp, $44.95.

1. INTRODUCTION

The Collected Works of G. E. P. Box edited by G. C. Tiao features 69 articles out of a totality of 120 articles and 6 books attributed to him and coauthors during the period 1947–1984. All of his books and more than ⅔ of his papers were collaborative efforts with a wide variety of statisticians, probably a larger number than any other statistician during the last 40 years—so much for the statistics.

Without a doubt the responsibility for the prominence of the Wisconsin Statistics Department, many of whose members or now former members have been his major collaborators, is mainly due to the efforts of Box. Most frequent as coworkers have been G. C. Tiao, N. R. Draper, G. Jenkins, J. S. Hunter, W. G. Hunter, G. Ljung, B. Abraham, and J. F. MacGregor. This is not to gainsay the important papers he wrote with D. R. Cox, P. W. Tidwell, S. L. Andersen, I. Guttman, K. B. Wilson, H. L. Lucas, D. A. Pierce, and a number of others.

Clearly, Box exhibits an enormous capacity for simultaneously inspiring and working closely with a number of different researchers on a variety of statistical issues; no mean feat, given the history of statistical egos, polemics, and assorted petty quarrels. For example, from the late 1950s to the mid-1960s, he must have been working more or less during the same period with Jenkins on control problems and time series, with Tiao on Bayesian inference, with J. S. Hunter on factorial designs, with Draper on response surfaces, with D. W. Behnken on rotatable designs, with W. G. Hunter on modeling, with G. S. Watson on robustness, and with Tidwell and Cox individually on transformations. Indeed, before anything else is said, one crucial role Box has played is as "The Great Collaborator," of course not of the Quisling variety. Later we shall hear of him as "The Great Communicator" (in the sense of mastery of exposition rather than actor transmuted into President).

Seymour Geisser is Professor and Director, School of Statistics, University of Minnesota, 270 Vincent Hall, 206 Church Street, S.E., Minneapolis, MN 55455.

The *Collected Works* appear in 2 volumes divided into 5 parts. Volume I, Part 1 contains 16 papers on statistical inference, robustness, and modeling strategy, while Part 2 features 14 papers of experimental design and response surface methodology. In Volume II, the remaining parts are: 3, time series analysis and forecasting; 4, distribution theory, transformation of variables and nonlinear estimation; 5, application of statistics. Each part is prefaced with an introduction by a distinguished figure in the field giving his view of the motivation and highlights of the more important papers presented. Appropriately distributing a large number of an individual's works into a few mutually exclusive categories presents difficulties, but here a sensible allocation was made. Within each part the papers are arranged chronologically by publication date. Several minor exceptions may be noted. In Part 1, the last paper is not in its proper chronological order (also true of the 10th paper in Part 3) and the 4th paper appears to have been better placed in Part 2. A further quibble is in regard to the 7th paper in Part 4 as with a few other papers, they could just as easily have been placed in others parts.

It would be presumptuous to believe that these works exhibit the totality of Box's contributions because obviously his career is far from over. In this sense the *Collected Works* is premature since by no means have his research efforts abated. In view of this fact the editors missed the rare opportunity of allowing the scientist to comment on his own work, discussing what he believed most important, and perhaps how he arrived at some of his ideas and the connection between works that may seem to us quite disparate. This would also have permitted him to correct mistakes, misprints, etc. of one kind or another in the text. The articles are photocopies of the originals and hence large variations in font, texture, and typography are evident.

2. INFERENCE AND ROBUSTNESS

The papers in Part 1 are preceded by an erudite summary as only S. M. Stigler can render. He traces the growth of Box's original conception of robustness to his more mature views later on (7, 13, 22).* This certainly needs to be contrasted with the current industry it spawned. I recall, when first hearing about someone who claimed his method or analysis was robust, picturing a portly gentleman of the Colonel Blimp variety oblivious to all that one might learn from the data keenly intent only on the fact that his nominal significance level was approximately correct. Later on, when the frequentist industry had sufficiently proliferated, I imagined a procedure to be robust if it could find the center, whatever that might mean, of any of a set of differing, perhaps topologically misshapen, fat-tailed cheeses, replete with varied sized holes. But it is informative to understand Box's original view and how it progressed. He simply stated that a statistical criterion that retained its sensitivity to changes in the factor of presumed interest but was more or less impervious to extraneous perturbations was robust and on that account useful. This came to be known as criterion robustness. There is a presumption in this view that irrespective of the perturbations the criterion retains its factor of interest—a fact that may not be the case.

Later on he introduced a Bayesian view of robustness (24, 28, 43) wherein one assesses the sensitivity of an inference about a parameter θ conditional on a "grudging and judicious elaboration" of the current model to a set of models wherein θ retains its physical interpretation. In Box's example the standard model is neatly encapsuled in a larger one by the introduction of a discrepancy parameter β, which for a particular value $\beta = \beta_0$, yields the original model. The posterior distribution of θ conditional on β and data $D = (y_1, \cdots, y_n)$ may be examined to determine the effect of varying the discrepancy parameter. When the effect is minimal, the older model could be retained and the inference "judiciously" subsumed in $P(\theta \mid D, \beta_0)$, i.e., based on the usual model since this distribution will represent a more parsimonious description and generally a tighter set of high probability values for θ. However, when this is not the case, robust estimation is "grudgingly" provided by the marginal posterior distribution $P(\theta \mid D) = E_\beta P(\theta \mid D, \beta)$. Box had now removed the board from under those agile surfers who *frequent* every new wave.

In commenting on Box's Bayesian robustification, Barnard (1980) noted that when the effect is minimal, one had a robust sample. In terms of likelihoods, if

$$\frac{L(\theta_1 \mid D, \beta_0)}{L(\theta_2 \mid D, \beta)} \doteq K(\theta_1, \theta_2)$$

varied little with alternative values of β for almost all admissible pairs (θ_1, θ_2), then the sample represented by D was robust.

Similarly in a Bayesian context, if for all interesting θ

$$\sup_\beta | P(\theta \mid \beta_0, D) - P(\theta \mid \beta, D) | < \delta$$

for example, where δ is "small" enough to suit one's purpose, then D should qualify as a robust sample with respect to the estimation of θ. Further and most importantly, if Y is either a future observable or some function of a set of them from the process that generated the data D, and again if for all interesting β

$$\sup_\beta | P(y \mid \beta_0, D) - P(y \mid \beta, D) | < \delta,$$

then D is a robust sample for the prediction of Y. Note that we could qualify this by using only particular values of y because there may be situations where our interest is very specific. For example, our interest might be focused on the computation of the chance that Y will exceed a specific threshold value y_t, say. It could turn out that for a specific set of values of $y \in \mathscr{Y}$ under interest

$$P(y \mid \beta_0, D) \doteq P(y \mid \beta, D) \quad \text{for all} \quad \beta,$$

then for this purpose D is specifically robust. However, for other values of y, this may be far from the case and D would not be completely robust. Clearly at $y = \pm\infty$, D is always specifically but meaninglessly robust.

I stress predictive robustness here because of its potential for being different from parametric robustness in certain cases; i.e., lack of parametric robustness need not imply lack of predictive robustness. Further if $\theta = (\eta, \tau)$, the sample could be robust with respect to η but not with respect to τ, and it may or may not be robust with respect to y. Not being robust with respect to y is however a clear indication of the failure of robustness with respect to θ. The property of predictive robustness is, I believe, the most important and useful of all robustness criteria. A more encompassing notion of a predictively robust sample would involve a prescribed lack of variation when both the prior and the likelihood (i.e., the entire model) are jointly perturbed from some standard.

Tracking Box's path through robustness is in many ways similar to tracing through his other research concerns. They are informed by his changing but flexible inferential philosophy. They all begin with an adherence to frequential theory (3, 7, 11, 12), then succumb to the influence of Fisherian ideas (in this case permutation or randomization tests (13)), and finally conform to a rather flexible Bayesian approach (24, 32) (Box and Tiao, 1973), as he becomes more

* Articles by Box reprinted in the *Collected Works* will be designated by the number in the list at the end of this article.

enmeshed in real technical and scientific problems. His latest efforts appear to reflect the fact that he now is being persuaded by the value of predictive arguments in research work. This is demonstrated by his more recent opus (43) where he used the predictive distribution for criticizing a model.

That Box did not earlier recognize the value of predictive distributions perhaps stems from the fact that he dealt largely with problems in the physical sciences where the error was not as inherent in the sampling unit as in the biological and social sciences. By this I mean that in many (not all) problems in the physical sciences, meaningful physical entities can be established and presumably when enough important factors are included the error that remains is to a large degree a function of the measuring device. In the biological and social sciences the material under investigation is subject to inherent variation irrespective of the accuracy of measuring instruments and hence requires a much stronger emphasis on observables and inferences about them. Box enjoys a middle ground between these extremes. He asserts that in a relationship $\eta = \phi(x_1, \ldots, x_p)$, all physical variables, say, where the x's are controllable and measurable essentially without error, there is still "error" involved in the response variable for repetition of the experiment at the same set of x's not due to inadequate measuring devices, but to uncontrolled factors. Or, to put it another way, it is due to the imperfection of the postulated model. The relationship ϕ is governed by a set of parameters $\beta_0, \beta_1, \cdots, \beta_k$. Since we have the capability or potentiality of making as many observations η at all values of the x's within our interest, ad infinitum, both the parameter representing experimental error and the set $\beta_0, \beta_1, \cdots, \beta_k$ may have some physical meaning as representing constants of the tentative model depending of course on its approximative value. The fact that there is still experimental error even in the presence of "perfect" measurement reflects the fact that the model is only approximately adequate.

This situation is generally different in the softer sciences even where measurements can be made essentially without error, but relationships are either vague, complex, or completely unknown. Often the data consists of the varied responses of a sample of individuals, and our inference is to some aspect of the population of potentially observable individuals that our sample represents, which is invariably finite. We may be interested in the response of a randomly selected new individual from this finite population or some function of the next M in a future sample, e.g., the fraction that lie in a certain interval.

Even in the cases that absorbed Box's attention an important way for achieving better models is to compare their predictive capacities. As he sooner or later came to realize, conventional hypothesis testing was inadequate to cope with this issue. Sorting among rival models was much more sensibly treated in the Bayesian framework; but this would always require a great deal of prescience concerning the totality of alternatives to be entertained and one's prior probability about the potential truth of each of them. Few serious scientists appear to work that way. Box then essentially fused a Bayesian predictive model with current scientific operating procedure to entertain a provisional model as the source for the generation of a data set when no alternative appeared as yet on the horizon. This would be checked or criticized by a predictive significance test which calculated the probability of the set consisting of those points in the sample space the probability density of which was no larger than the observed sample, rather than some ordering of the sample points as to their discrepancy from the hypothesis. The latter being in line with the more standard Fisher-Barnard-Cox view of significance tests. Box's procedure (43) although leading to logical conundrums, *cf.* Geisser (1985), is not rescued by ordering sample points as Cox (1980) suggested. It is nevertheless an excellent operational procedure and will make no less progress than a logically perfect one that is either too difficult to apply or is in peril of being seriously misapplied.

In other words we have now a further extension of "robustness," whether Box intended it or not. Just as a robust analysis is one that can resist most perturbations that might occur, so it is with Box's predictive model criticism.

What has been always somewhat curious is the fact that Box has not stressed predictive distributions for inference about observables. It appears to me that his whole Bayesian philosophy is oriented in that direction. An example of this is his views on prior distributions of parameters which he claims ought to depend on the experiment. In other words the Boxian model parameters in many instances are fathered by the experiment and may not have a physical existence outside of the experimental setup. Then rightly, the prior distribution of such a "parameter" may depend on the likelihood the experiment induces. Why then should it be of primary interest to draw a conclusion about such an entity which may be totally artifactual or at best the real status of which may be murky, rather than a potential observable or some function of a future set of them? The "parameter" only comes to life, as it were, as a limiting value of an interesting function of potential observables. It is curious that Box's extraordinary sense of experimentation and understanding of models has not liberated him further from being engrossed with the estimation of these entities.

A response that he might make, not unjustifiably, is that he is interested in a series of experiments whereby a theory, or more likely a process, which was inadequately understood at first could, after several iterations, be much better understood. This would then lead to a sensible working model which would cover the main theoretical aspects of the problem. Without in anyway disagreeing with such a laudible enterprise (aptly explained in a variety of appealing diagrams in several of his papers), I would point out that, at least, when such a model was established, then its main purpose was the prediction or even the control or regulation of observables. Also, during the iterative process there was ample scope for the use of predictive distributions and predictive sample reuse techniques for assessing the adequacy of the succession of provisional models. Just as our hierarchial Bayesian colleagues must at some point cease their regress in prior hyperparametric assumptions, so too does this iterative modeling scheme need to be put to use. On the other hand there is no quarreling with success and unarguably Box is an eminently successful statistical scientist.

3. EXCURSIONS ON RESPONSE SURFACES

The second part is introduced by B. H. Margolin who, in a careful and thorough manner, explains Box's explorations of response surfaces. Boxian experimental designs grew out of and then departed from its Fisherian roots because of the necessity of accommodating differing scientific and technical needs. Rather than assessing the effects of various factors in a multifactorial comparative trial, the problem Box faced was to devise schemes for efficiently determining the optimal conditions for the output of some industrial process which depended on k controllable quantitative variables. Here instead of designing a single experiment, we have a sequence of such trials each depending on its predecessors. Standard factorial designs were employed for estimating first derivatives in given subregions which would indicate which further subregions to explore. This steepest ascent method is continued until the vicinity of a stationary point is achieved; hopefully one yielding an optimal response. For a more detailed exploration of this near stationary region entirely new "composite" designs were devised to estimate higher derivatives. The two sources of error inherent in such studies are measurement error and the bias due to whatever difference there is between an assumed response function and the actual one. The accuracy of the estimates of the derivatives will be determined by the arrangement of the experimental points.

Box investigated the optimal estimation of the constants of a planar regression surface depending on the k quantitative controllable variables subject to homogeneous measurement error where $N > k$ combinations of the levels of the variables would be chosen. He showed that the minimum variance property for the optimal design is invariant under rotation. This property can be used to reduce bias, eliminate certain systematic effects without losing efficiency, and allow the usual normal theory tests to be exact, independent of the distribution of the observations by a suitable randomization scheme. The basic ideas of response surface designs and methodology were exploited, developed, and extended in a series of papers (5, 6, 9, 10, 14, 16–21, 26, 30) to include nonlinear response functions, Bayesian design criteria, and the accommodation of potential bias in an assumed response surface by the construction of appropriate designs. Younger statisticians, disheartened by referees' comments, may derive some comfort from the fact that the first paper in the above series, when submitted to be read before the research section of the Royal Statistical Society, was rejected by one of the referees.

Margolin properly intimates that Box's work on response surface methodology was, by itself, sufficient to keep a statistician of the first rank busy for a quarter of a century. The same, of course, may be said and is for at least several other of these research areas that were developed during that same period.

4. TIME SERIES

Box's efforts in time series are succinctly summarized by C. W. J. Granger, who remarks that Box's early work, brought together and amplified in his "landmark" book, Box and Jenkins (1970), "had a widespread, immediate and dramatic effect on the modeling and forecasting of time series." And so it did. This book revised and expanded in a second edition (1976) included the bulk of their output in this area up until that time.

An appealing feature of the work was their efforts at building parsimonious and applicable stochastic models for a sequence of dependent discrete observables in the time rather than the frequency domain. (Box rightly emphasizes observable inference rather than *spectral* inference, or more bluntly the substance rather than the *spectre*.) This would enable one to investigate the process underlying the series or at least provide the simplest flexible smoothing function that best represented the series consonant with whatever one knew about the process. Optimal forecasting of future observables from such a series would then naturally follow. The ARIMA models developed and employed by Box, capable of handling nonstationary and seasonal time series (23, 31, 33, 34, 41), were extended to represent relationships between several such series with a view toward simultaneous forecasting and forecasting future values using its previous values (42, 45) and those of a related series. Procedures were devised for fitting and checking transfer function models and designing optimal control schemes (35, 37).

In pollution problems, a new situation arose whereby the effect of some external shock (in this case a pollution control law) needed to be taken into account in the time series. Thus arose intervention analysis (29, 37–39).

Throughout his work Box emphasized the importance of the iterative strategy of model identification, efficient estimation of the model parameters, and assessing the adequacy of the model's fit by means of diagnostic checks. If the fit were demonstrated to be inadequate, one looped through this process recursively until one attained a model suitable for forecasting (or control). To Box this was the scientific method, just as the search for an optimum of a response surface required sequential alteration of design was the scientific method. A difficulty, if the procedure is used incautiously or robot-like, is the possibility of being unable to extricate oneself out of an interminable loop; for example, when the ARIMA program is deficient for the task.

His development of the various theories and methods for time series was only exceeded by what was perhaps his principal achievement in this area. This turned out to be the superb organization of all these components into a coherent program that indelibly marked his work. Of course the fact that, in most instances, provision was made for a ready availability and easy implementation of his efforts was no doubt responsible for much of the pervasive popularity his work enjoyed.

What might there be to criticize in this program? At least two possibilities strike one. Instead of attempting to pay attention to the actual mechanism in the time series one resorts to essentially black box (ARIMA-BOX) techniques which represent a lower order of scientific inquiry. In many cases this is not a serious objection because the underlying mechanism is often either so ephemeral or so complex (certain economic time series, for instance) as to defy discovery in time for any appropriate inference, decision, or action. Further, the crucial issue for Box was forecasting (or control); and if that can be accomplished efficiently or near optimally by these techniques, that should suffice for most practical purposes.

A second criticism is one that entirely permeates this review and is most pronounced in this part primarily because it underlies both ultimate goals of time series analysis—forecasting and control. Where are the predictive distributions of future values? At best we are given "predictive" means and variances or distributions on the assumption that all the parameters are known; this is not sufficient. Is Box not a Bayesian that his inferential posture is incapable of such results? Not so! Although a good deal of Box's time series work was in the classical frequentist estimative mode, he also provided posterior distributions of real and imagined parameters in no small measure.

As previously indicated, he did come around in more recent times to considering predictive distributions, but mainly for model criticism and not for what they are best designed for, actual prediction, inference, comparison, decision, etc. I have no alternative but to infer that he kindly left something undone so that others might enjoy the effort in completing the task—or he is at it now.

5. DISTRIBUTIONS AND FUNCTIONAL RELATIONSHIPS

The set of papers in Part 4 is a potpourri of Box's work in the derivation of distributions, transformations, functional relationships, and nonlinear models. In a lengthy and effusive introduction, I. Guttman presents a comprehensive summary of the 13 papers in this part.

Several of Box's most important papers appear here. For example, the first one (3) derives the distribution of a class of modified likelihood ratio test statistics occurring in multivariate analysis the moments of which are of general specified form. He shows that the distribution function can be written in terms of an asymptotic series involving χ^2 distributions of successively higher degrees of freedom. The coefficients are such that one need only use a suitable number of terms to calculate the tail probability of the statistic with sufficient accuracy under the null hypothesis.

The next two papers, representing half of the selected papers that Box published in the *Annals of (Mathematical) Statistics* and a third of the six papers out of his total of 120 (one wonders if this is not a statement about one of our most "prestigious" journals or G. E. P. Box which I leave the reader to decide), deal with theorems on the distribution of quadratic forms and their application to the null distribution of the F test in the analysis of variance. His particular interest here is to determine the effect that departures from certain standard assumptions have on the test. He shows that moderate variance heterogeneity has only a modest effect for equal group sizes in one way analysis of variance situations, but that unequal group sizes have a larger effect. In a two way analysis of variance he showed that serial correlation between the columns can induce large discrepancies in the nominal significance level for testing the equality of column means.

An extension of Box's work to repeated measurement designs involving several groups was made by Geisser and Greenhouse (1958). They suggested an F test the degrees of freedom of which were reduced by the appropriate factor ϵ calculated from Box's work, which then was estimated from the sample covariance matrix among the repeated measurements assumed to be multivariate normal. This $\hat{\epsilon}$ F-test continues to serve as a popular alternative to a full scale multivariate test (Collier et al., 1967; Wilson, 1975); under most reasonable alternatives it has greater power. It has also been shown to be useful in growth curve situations even with incomplete data (Schwertman, 1978; Schwertman et al., 1985). The estimator $\hat{\epsilon}$ has also recently been shown to provide a locally best invariant test of whether the standard analysis of variance F-test in repeated measurement situations is appropriate (Grieve, 1984).

Box returned to this work a quarter of a century later to analyze analysis of variance situations with autocorrelated observations using a Bayesian approach (44).

A series of papers on transformations involved a response modeled as $Y_i = f(x, \theta) + e_i$ for x a set of known variables and θ a set of unknown parameters that can be fitted by least squares when the e_i were independently and normally distributed with constant variance. His first efforts were transformations on the set x, in terms of powers, logs etc. to reduce the function $f(x, \theta)$ to as simple a form as possible (25), usually a linear function when the true relationship was unknown.

Soon thereafter came the famous Box-Cox paper (27) in which nonlinear transformations were examined for the elements of the vector

$$Y^{(\lambda)} = (Y_1^{(\lambda)}, \cdots, Y_n^{(\lambda)})',$$

$$Y_i^{(\lambda)} = \begin{cases} Y_i^{\lambda} & \text{for } \lambda \neq 0 \\ \log Y_i & \text{for } \lambda = 0 \end{cases}$$

such that for some unknown λ,

$$E(Y^{(\lambda)}) = X\theta + e$$

where X is a known matrix and θ a vector of unknown parameters and the vector $e \sim MVN(\mathbf{O}, \sigma^2 I)$.

Maximum likelihood and Bayesian methods were presented for the estimation of λ. Once $\hat{\lambda}$ was established, a standard analysis conditional on that value might proceed. A second Box-Cox offering (46) in this vein was a rebuttal to a paper by Bickel and Doksum (1981). The latter showed that the joint estimates $\hat{\lambda}$ and $\hat{\theta}$ can be highly correlated even when the error variance was small, so that the marginal variances of $\hat{\theta}$ can be considerably larger than the conditional variance given λ. Box and Cox argued that when θ depends on λ and λ is poorly determined, the units in θ are not comparable. It seems to me that the whole brouhaha revolved around the wrong issue. The proper focus should have been the effect on the prediction of Y after transforming back from $Y^{(\lambda)}$ so that a standard metric is established for comparison. In such situations the most important issue is the prediction of future values of Y, since hardly anyone would believe in the mechanistic validity of models that are simultaneously linearized, normalized, and homoscedastisized with rather bizarre powers of Y; unless there were unimpeachable scientific reasons for their acceptability. This kind of modeling is basically a convenient way of reasonably approximating relationships that posses adequate predictive power.

6. ON THE APPLICATION OF STATISTICS

The last part features 10 papers on the application of statistics. The informative introduction by R. D. Snee extolls Box's expository clarity. We are also instructed that with regard to statistics Box was an autodidact, and this, in conjunction with his masterly facility for presenting graphical paradigms, was mainly responsible for his formidable skill as the great statistical communicator. So much for his formal training (B.A., Ph.D.) at University College, London.

As a result of a wartime project, several papers (1, 2), displaying his practical statistical acumen, deal with the effects of phosgene and mustard gas on laboratory animals. Another paper (8) involves an investigation of an automatic machine for testing pigment strength in the chemical industry, and a duo of papers (38, 40) detail statistical studies of Los Angeles smog data wherein he popularized intervention analysis, used to analyze the effect of the occurrence of an event on a time series. This was a topic (without the felicitous term) he had studied theoretically some 10 years previously (29).

In still another paper (36), he leans heavily on his involvement with environmental problems such as smog to offer his views on scientific experimentation by diagramming the flow and interaction of hypotheses, models, experimental data, deduction, induction, etc. I was struck here by his near analogy of the almost perfect correlation between storks' nests and human births with the "correlation" between smoking and lung cancer. This was an uncalled for comparison and I trust that after some reflection he no longer seriously believes, if he ever did, that both "relationships" enjoy equivalent evidential standing.

A paper (15) on "evolutionary operations" or EVOP, another Boxism patterned on natural selection, presented a very sensible way to run a manufacturing process, or to quote Box, "A process should be run as to generate product plus information on how to improve the product." Perhaps even more generally he might now add "and to implement the improvement based on a cost-benefit analysis." At present, natural selection is being altered to take account of "punctuated equilibrium," so that one can only wonder if "punctuated or perhaps punctuated interventionary operations" is the new order of the day. (PIOP, one hopes is not restricted to the sky.) Designed for manufacturing processes, EVOP has not much influenced statistics or statisticians, which probably is a source of chagrin to its creator. It's actual impact on industry is unknown to me.

His paper (with less fanciful appellations) on growth and wear curves (4), on the other hand, is probably one of the more heavily cited in the statistical literature. Here he initiated the practical groundwork for analysis of variance problems involving repeated measurements and growth curves, and followed this up later with theoretical work on the distribution of quadratic forms which was previously discussed.

7. THE BOXIAN VIEW

Unlike mathematicians, many distinguished statisticians such as Box often take some time to attain their full creative powers. Early in their careers, there is a fair display of technical virtuosity. Later, they develop deeper insights into important statistical issues. Some are led into philosophical paradigms for the foundations which at least overtly they believe to be unflawed. Being without blemish may require being devoid of relevance, or if relevant, impossible to execute. Indeed, the best of statistical ideas often engender disturbing paradoxes or counterexamples that render them suspect as to their capability of being generalized into a fault-free inductive system. The history of the logical foundations of statistics is replete with such failed panaceas.

For Box, a preoccupation with foundational issues was never an overriding concern, although he could very vigorously and trenchantly defend his eclectic view which combined such disparate notions as: the Bayesian approach; prior distributions depending not only on the likelihood (Box and Tiao, 1973) and not only on the sample size (Box and Tiao, 1968), but even on the observations themselves (27); randomization and permutation tests (13, 30); predictive tests of significance (43); and the concept of power (3, 7, 22). Attacked from all sides of the ideological fence he, in an apian manner, disarmed friendlier critics with mellifluous argument and subdued hostile ones with stinging wit.

His work indicates that he was quick to perceive what the important practical problems were and readily devised informative statistical paradigms accompanied by sensible (not necessarily final) solutions for them, uninhibited by a strict adherence to any one inferential ideology. Hence, the principal characteristic of the Boxian approach is best summed up as pragmatism par excellence.

Our perusal of Box's contributions indicated simultaneous progress on many different fronts with the bulk exhibiting a rare combination of theoretical and methodological aspects geared toward solving problems in technology and applied science. One recent feature discerned was an increased emphasis on predictive distributions and the analysis of observables. If we add to this his evident interest in EVOP, it would not be too surprising to see Box getting involved in areas such as quality assurance, manufacturing engineering, automated processes, experimental therapeutics, and adverse drug reactions using and expanding such notions as control, regulation, feedback, and near optimization. An excellent although limited initiative into a few of these aspects using observable or predictive analysis appears in the work of Aitchison and Dunsmore (1975).

Once the difficult tasks of understanding the reasons and purposes for the collection of a data set and formulating an appropriate model are completed, the principal job of the statistician is calculating probabilities of observables that are unknown, conditional on known observables. The wave of the future in statistics is in calculating relevant probabilities for the future.

ACKNOWLEDGMENT

This work was supported in part by Grant GM25271 from the National Institutes of Health.

REFERENCES

AITCHISON, J. and DUNSMORE, I. R. (1975). *Statistical Prediction Analysis.* Cambridge University Press, Cambridge.

BARNARD, G. A. (1980). Contribution to discussion. *J. R. Statist. Soc. Ser. A* **143** (part 4) 404–406.

BICKEL, P. J. and DOKSUM, K. A. (1981). An analysis of transformations revisited, *J. Amer. Statist. Assoc.* **76,** 296–311.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control.* Holden-Day, Oakland, CA (2nd ed, 1976).

BOX, G. E. P. and TIAO, G. (1968). Bayesian estimation of means for the random effect model. *J. Amer. Statist. Assoc.* **63** 174–181.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA.

COLLIER, R. O., BAKER, F. B., MANDEVILLE, G. K. and HAYES, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. *Psychometrika* **32** 339–353.

COX, D. R. (1980). Contribution to discussion. *J. R. Statist. Soc. Ser. A* **143** (part 4) 410.

GEISSER, S. (1985). On the prediction of observables: A selective update. In *Bayesian Statistics 2* (BERNARDO, J. M., et al., eds). North-Holland Publishing Co., Amsterdam, pp 203–230.

GEISSER, S. and GREENHOUSE, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Ann. Math. Statist.* **29** 885–891.

GRIEVE, A. P. (1984). Tests of sphericity of normal distributions and the analysis of repeated measures designs. *Psychometricka* **49** 257–267.

SCHWERTMAN, N. C. (1978). A note on the Geisser-Greenhouse correction for incomplete data split-plot analysis. *J. Amer. Statist. Assoc.* **73** 393–396.

SCHWERTMAN, N. C., FLYNN, W., STEIN, S. and SCHENK, K. L. (1985). A Monte Carlo study of alternative procedures for testing the hypothesis of parallelism for complete and incomplete growth curve data. *J. Statist. Comp. Simulation* **21** 1–37.

WILSON, K. (1975). The sampling distribution of conventional, conservative, and corrected F ratios in repeated measurements designs with heterogeneity of covariance. *J. Statist. Comp. Simulation* **3** 201–205.

Articles in "Collected Works of George E. P. Box"

1. The effect of exposure to sublethal doses of phosgene on the subsequent L(ct)50 for rats and mice (with H. Cullumbine). *Br. J. Pharmacol. Chemother.* **2** 317–346, 1949.
2. The relationship between survival time and dosage with certain toxic agents (with H. Cullumbine). *Br. J. Pharmacol. Chemother.* **2** 27–37, 1947.
3. A general distribution theory for a class of likelihood criteria. *Biometrika* **XXXVI** (parts IXX and IV) 317–346, 1949.
4. Problems in the analysis of growth and wear curves. *Biometrics* **6** 362–389, 1950.
5. On the experimental attainment of optimum conditions (with K. B. Wilson). *J. R. Statist. Soc. Ser. B* **13** 1–45, 1951.
6. Multifactorial designs of first order. *Biometrika* **39** 49–57, 1952.
7. Non-normality and tests on variances. *Biometrika* **40** 318–335, 1953.
8. Pigment strength testing with the automatic muller (with M. T. Hobbs and P. North). *J. Oil Colour Chem. Assoc.* **XXXVI** 283–299, 1953.
9. A statistical design for the efficient removal of trends occurring in a comparative experiment with an application in biological assay (with W. A. Hay). *Biometrics* **9** 304–319, 1953.
10. The exploration and exploitation of response surfaces: Some general considerations and examples. *Biometrics* **10** 16–60, 1954.
11. Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect on inequality of variance and of correlation between errors in the two way classification. *Ann. Math. Statist.* **25** 290–302, 1954.
12. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects on inequality of variance and of correlation between errors in the two way classification. *Ann. Math. Statist.* **25** 484–498, 1954.
13. Permutation theory in the derivation of robust criteria and the study of departures from assumptions (with S. L. Andersen). *J. R. Statist. Soc. Ser. B* **17** (part 1) 1–34, 1955.
14. The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system (with P. V. Youle). *Biometrics* **11** 287–323, 1955.
15. Evolutionary operation: A method for increasing industrial productivity. *Appl. Statist.* **VI** 3–23, 1957.
16. Multifactor experimental designs for exploring response surfaces (with J. S. Hunter). *Ann. Math. Statist.* **28** 195–241, 1957.
17. A basis for the selection of a response surface design (with N. R. Draper). *J Amer. Statist. Assoc.* **54** 622–654, 1959.
18. Simplex-sum designs: A class of second order rotatable designs derivable from those of first order (with D. W. Behnken). *Ann. Math. Statist.* **31** 838–864, 1960.
19. Some new three level designs for the study of quantitative variables (with D. W. Behnken). *Technometrics* **2** 455–475, 1960.
20. The 2^{k-p} fractional factorial designs, part I (with J. S. Hunter). *Technometrics* **3** 311–351, 1961.
21. The 2^{k-p} fractional factorial designs, part II (with J. S. Hunter). *Technometrics* **3** 449–458, 1961.
22. Robustness to non-normality regression tests (with G. S. Watson). *Biometrika* **49** (parts 1 and 2) 93–106, 1962.
23. Some statistical aspects of adaptive optimization and control (with G. M. Jenkins). *J. R. Statist. Soc. Ser. B* **24** 297–343, 1962.
24. A further look at robustness via Bayes' theorem (with G. C. Tiao). *Biometrika* **49** (parts 3 and 4) 419–432, 1962.
25. Transformation of the independent variables (with P. W. Tidewell). *Technometrics* **4** 531–550, 1962.
26. The choice of a second order rotatable design (with N. R. Draper). *Biometrika* **50** (parts 3 and 4) 335–352, 1963.
27. An analysis of transformations (with D. R. Cox). *J. R. Statist. Soc. Ser. B* **26** 211–252, 1964.
28. A note on criterion robustness and inference robustness (with G. C. Tiao). *Biometrika* **51** (parts 1 and 2) 169–173, 1964.
29. A change in level of a non-stationary time series (with G. C. Taio). *Biometrika* **52** (parts 1 and 2) 181–192,1965.
30. Some aspects of randomization (with I. Guttman). *J. R. Statist. Soc. Ser. B* **28** 543–558, 1966.
31. Models for forecasting seasonal and nonseasonal time series (with G. M. Jenkins and D. W. Bacon). In *Spectral Analysis of Time Series* (HARRIS B, ed). John Wiley & Sons, New York, pp 271–311, 1967.
32. A Bayesian approach to some outlier problems (with G. C. Tiao). *Biometrika* **55** 119–130, 1968.
33. Some recent advances in forecasting and control (with G. M. Jenkins). *Appl. Statist.* **17** 91–109, 1968.
34. Some comments on a paper of Coen, Gomme, and Kendall (with P. Newbold). *J. R. Statist. Soc. Ser. A* **134** 229–240, 1971.
35. Some recent advances in forecasting and control, part II (with G. M. Jenkins and J. F. MacGregor). *Appl. Statist.* **23** 158–179, 1974.
36. Statistics and the environment. *J. Wash. Acad. Sci.* **64** 52–59, 1974.
37. The analysis of closed-loop dynamic-stochastic systems (with J. F. MacGregor). *Technometrics* **16** 391–398, 1974.
38. Analysis of Los Angeles photochemical smog data: A statistical overview (with G. C. Tiao and W. J. Hamming). *J. Air Pollution Control Assoc.* **25** 260–268, 1975.
39. Intervention analysis with applications to economic and environmental problems (with G. C. Tiao), *J. Amer. Statist. Assoc.* **70** 70–79, 1975.
40. Some empirical models for the Los Angeles photochemical smog data (with G. C. Tiao and M. S. Phadke). *J. Air Pollution Control Assoc.* **26** 485–490, 1976.
41. Analysis and modeling of seasonal time series (with S. C. Hillmer and G. C. Taio). *Proceedings of Conference on Seasonal Analysis of Economic Time Series* 309–344, 1976.
42. A canonical analysis of multiple time series (with G. C. Tiao). *Biometrika* **64** 355–365, 1977.
43. Sampling and Bayes' inference in scientific modeling and robustness. *J. R. Statist. Soc. Ser. A* **143** (part 4) 383–430, 1980.
44. Analysis of variance and autocorrelated errors (with G. Ljung). *Scand. J. Statist.* **7** (part 4) 172–180, 1980.
45. Modeling multiple time series with applications (with G. C. Tiao). *J. Amer. Statist. Assoc.* **76** 802–816, 1981.
46. An analysis of transformation revisited, rebutted (with D. R. Cox). *J. Amer. Statist. Assoc.* **77** 209–210, 1982.