

Remembering G.E.P. Box: life, contributions and some personal experiences

Daniel Peña

Universidad Carlos III de Madrid

Universidad Politécnica de Cataluña
22 Noviembre, 2023



Agradecimientos

- A Pedro Delicado, y al Director del Departamento, Xavier Tort, por su invitación.
- A Albert Prat, que trajo por primera vez a George a Barcelona y con el que compartí una gran amistad y momentos inolvidables.
- A Alberto Luceño, por sus fotos y recuerdos de las felices estancias de George y Claire en Santander.



Outline

1. Introduction
2. Life and some personal experiences
3. Contributions
4. Pioneering Data Science
4. Conclusions



1. Introduction (I)

George Box is one of the five greatest statisticians of all times according to Kotz and Johnson book, *Breakthroughs in Statistics I and II*.

Number of contributions:

- **Fisher:** 3 (Foundation of MS,22; SMRW,25; DE, 26)
- **Box** 2 (EO,51; TS, 62)
- Hotelling 2 (T, 31; CC, 36)
- Neyman: 2 (HT, 33; Sampling, 34)
- Wald 2 (Sequential, 45; SD,49)

Note: After Vol III, Rao, Tukey and Smith are added with 2 contributions



- **Evolutionary operation (EO)** opened the field of **Industrial statistics**, adapting Fisher approach of experimental design for industry and moving the field from the control charts by Schewart, to process improvement.
- His work with Jenkins on **Time series** has a huge influence in statistics and econometrics and changed the way to model dynamic dependent data.

On the Experimental Attainment of Optimum Conditions

G.E.P. Box and K.B. Wilson
Imperial Chemical Industries
Blackley, Manchester, UK

Some Statistical Aspects of Adaptive Optimization and Control

G.E.P. Box
University of Wisconsin, Madison
G.M. Jenkins
Imperial College, London



2. George Box's life

18 October, 1919 – 28 March, 2013

and

Some personal memories
(*1976 - 2013*)



Periods of George Box's life

1. Youth and Education (28 years, 1919-1947)
2. Working in Industry (8 years, 1948-1956)
3. Princeton (3 years 1956-1959)
4. U Wisconsin-Madison (32 years 1960-1992)
5. Emeritus, Center for Q&P Madison (21 years, 1992-2013)



2.1 Youth and Education (1919-1947)

(I)

- Born was as the last child into a working family
- He was born colorblind
- At 16 started working as assistant in a chemical company and studied an external degree in Chemistry at London University
- Second World War, he is 20 and enrolls in the Army.
- He was sent to Chemical Defense Experimental Station near London.



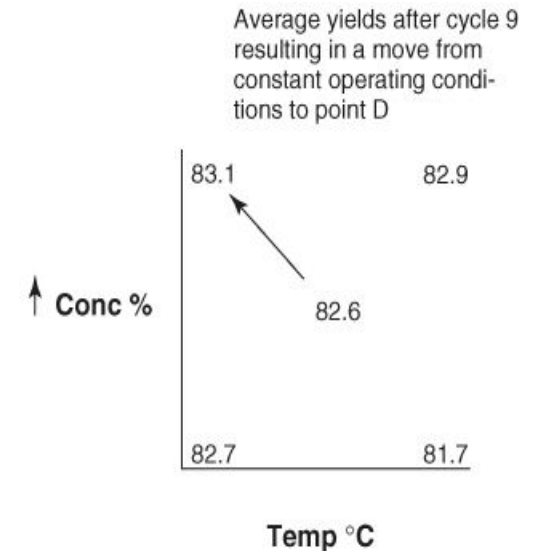
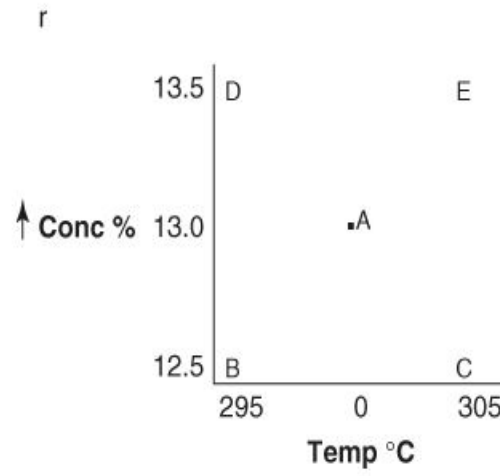
2.1 Youth and Education (1919-1947) (II)

- Need statistics for the experiments on the effect of poison on animals and starts reading Fisher.
- Sent by the Army to ask Fisher's advice on experimental design.
- After the war, B. Government's grant for soldiers admitted in Universities. Box to studies statistics under E. Pearson. B.Sc Statistics University London 1947, (18 months instead of 3 years)
- Met Jessie in the Army and married, 1945.



2.2 Industry : Statistician at ICI (1948-1956)

- Box and Wilson (1951) "On the experimental attainment of optimum conditions" *JRSSB*
- Ph.D. Statistics, 1952, London University
- Evolutionary Operation and Fractional Factorial Designs



Evolutionary Operation and Industrial Statistics at ICI

Instead of a variable at a time move all of them jointly, as suggested by Fisher, with factorial designs. With many variables, apply fractional factorial designs

Make experiments in a continuous way improving the process when it is working. *A process gives a product, but also information about how to improve it*

Study the response surface as a non linear, may be quadratic function, and simplify the analysis using linear transformations of the variables.

Established Industrial Statistics, started by Walter A. Shewhart, of Bell Telephone Laboratories in 1924 with control charts, adding to *process control* by observations, *process improvement* by experimentation.



Invited to North Carolina (1953-54)

- His landmark paper with Wilson in JRSS (first rejected and published thanks to Barnard) lead to Stu Hunter, graduate student there of G. Cox, to promoted his invitation.
- Went to NC on one year leave from ICI.
- Visited Princeton and met John Tukey.
- Enjoyed living in the US with his family (Jessi, Henry).



2.3 Princeton : Director of Statistical Research Group (1956-59)

- Industrial Statistics and Experimentation with Bill Hunter
- Control and forecasting time series with Jenkins
- Divorced Jessie and married Joan Fisher
- With S. Hunter and C. Daniel founded *Technometrics*, 1959.

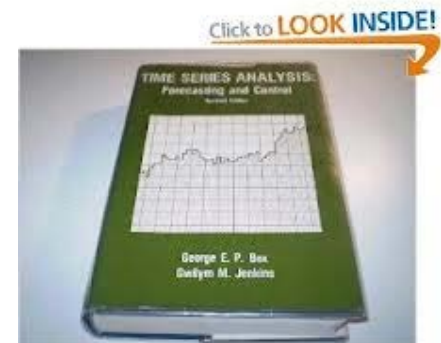


Plate 11. Mrs. Fisher 1938, with daughters, in order of age, Margaret (top right), Joan (bottom right), Phyllis (top left), Elizabeth (bottom left), Rose standing beside her chair, and June in her lap.

2.4 U Wisconsin-Madison (1960-1992)

1st period: 1960/70

- "Some statistical aspects of adaptive optimization and control" (with Jenkins, 1962), *JRSSB*
- "An analysis of transformations" (with Cox, 1964), *JRSSB*.
- Evolutionary Operation-A statistical method for process improvement, 1969 (with Draper).
- Time Series Analysis Forecasting and Control (with Jenkins, 1970) Holden-Day.



U Wisconsin-Madison(1960-1992)

1st period:1960/70

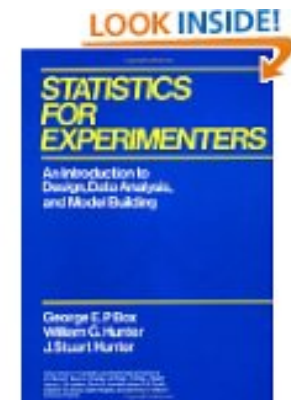
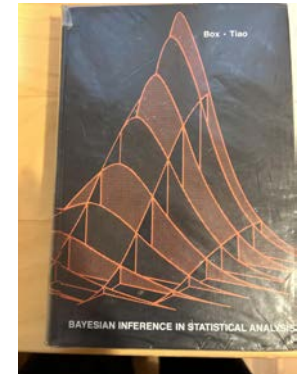
- As Director of the Department for 8 years attract many good people: D. Rubin, G. Tiao, B. Hunter, I. Guttman, N. Draper, G. Wahba, and many others
- No experience with teaching: *"I Always looked at George Tiao's face"*
- Joint appointments of statistics and other departments: engineering, economics, ...
- Advice: *"Get excellent people; concentrate on the people and not so much on the subject"*



2.4 U Wisconsin-Madison (1960-1992)

2nd period: 1970/80

- *Bayesian Inference in Statistical Analysis*, 1973 (with G.C. Tiao).
- *Statistics for Experimenters*, 1977 with W.G. Hunter and J.S. Hunter, (more than 165,000 copies sold, traducido al español en UPC).
- The statistic department UWM starts being more theoretical and George moves to the Mathematical Research Center.
- To promote applications starts the *Monday Night Beer and Statistics* seminar in the basement of his home in Madison.



2.4 U Wisconsin-Madison (1960-1992)

2nd period: 1970/80

- With G.C. Tiao, "Intervention analysis with applications to economic and environmental problems" *J. Amer. Stat. Assoc.*, 1975.
- With G.C. Tiao, "A canonical analysis of multiple time series" *Biometrika*, 1977.

Many Honors

- 1972: Wilks Memorial Medal
- 1974 Fellow of American Academy of Arts & Sciences
- 1978 President, American Statistical Association
- 1979 President, Institute of Mathematical Statistics



Some personal memories 1970/80

J. R. Statist. Soc. A,
(1973), 135, Part 3, p. 337

337

Some Comments on a Paper by Chatfield and Prothero and on a Review by Kendall

By G. E. P. BOX and G. M. JENKINS

University of Wisconsin, Madison

University of Lancaster

IN a recent paper read to a meeting of the Society, which neither of us was able to attend, Chatfield and Prothero (1973) applied the “Box–Jenkins” approach to the forecasting of a particular series and were dissatisfied with the results they obtained. They included in their paper a large number of general comments, remarks and conclusions on the topic of forecasting which they believed to be supported by extrapolation from their unsatisfactory experience with one particular series. In fact, as was pointed out at the Society’s meeting, their difficulties arose because they applied the wrong transformation to the data. Their conclusions are not, therefore, supported by their example.

We have written this note because we believe many of their comments to be misleading. In particular, the authors thought that they found in their anomalous results support for views, expressed by M. G. Kendall (1971) in a review of our book, which appeared in this *Journal*. We take this opportunity to comment on these views.

Some personal memories 1970/80

- Meeting in Cambridge in Time series 1976.
(Arthur Treadway introduced BJ methods in Spain, Seminar in EOI)
- First International Meeting held in Valencia (Spain) May 28 to June 2, 1979.
(Jose Miguel Bernardo invited George Box to Spain for the first time to the 1st Bayesian Meeting)
- First course of George Box and George Tiao in Spain, Madrid. 1980.
(DP invited GB to give a course in Madrid and he came with GTiao. Agustín Maravall knew them from UWM and helped in the course)



2.4 U Wisconsin-Madison (1960-1992)

3rd period: 1980/92

- Leader of the Quality Movement (Center for Quality and productivity Improvement, 1985, UWM)
- *Empirical Model-Building and Response Surfaces*, 1987 (with N.R. Draper).
- In 1988 the American Society for Quality started *Quality Engineering* and GB often published there to help the journal.

Many honors

- 1985: **Elected Fellow of The Royal Society**;
- 1989: Youden Prize;
- 1989: Deming Medal;



Some personal memories 1980/92

- George Box in Barcelona invited by Albert Prat, 1982.
- DP invited by GB as Visiting Professor, Mathematics Research Center, U. Wisconsin-Madison 83/84.
- Xavier Tort was studying a Master at UWM.
- Work with GB on last time series paper (with DP) "Identifying Simplifying Structure in Time Series" (1987), *JASA*.
- Attended the *Monday Night Beer and Statistics* George's seminar, work with Kitchell, 1984.
- George divorced from Jane.



Science
AAAS



Universidad
Carlos III de Madrid
www.uc3m.es

Periodicity of Extinctions in the Geologic past: Deterministic Versus Stochastic Explanations

Author(s): Jennifer A. Kitchell and Daniel Pena

Source: *Science*, Nov. 9, 1984, New Series, Vol. 226, No. 4675 (Nov. 9, 1984), pp. 689-692

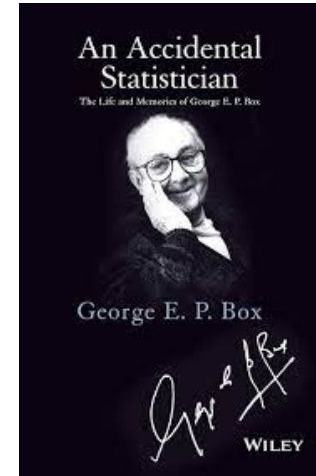
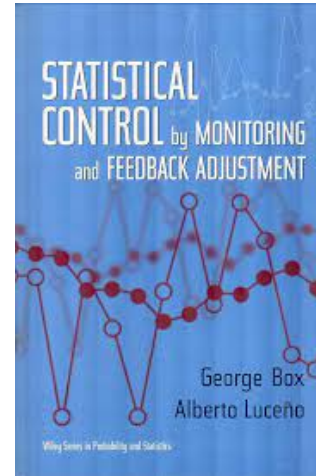
Some personal memories 1980/92

- Attended 65th Birthday GB, 1985. Meeting 25th years Department Statistics Madison. George with Claire.
- Box focus on Quality and Taguchi ideas at Center for Quality and Productivity with Bill Hunter.
- GB invite Albert Prat as Visiting Professor UWM 85 o 86?.
- In 85/92 George made several visits to Madrid and Barcelona (including AlbertP and TinaR wedding in 1986). Xavier Tort translates to Spanish SE and Josep Ginebra go for Ph.D. at UWM



2.5 Emeritus (1992-2013)

- Statistical Control by Monitoring and Feedback Adjustment, John Wiley & Sons, 1997 (with A. Luceño).
- An Accidental Statistician, March 2013.



1997 Brumbaugh award

More Honors

- 1993: Gold Medal of the Royal Statistical Society
- 1993, 97, 2007, 2010: American Society for Quality Control Brumbaugh Award ;
- Honorary Doctorates (Rochester, 75; Carnegie Mellon 89 and Universidad Carlos III de Madrid, 95).



Some personal memories, (1992-2013)

1995



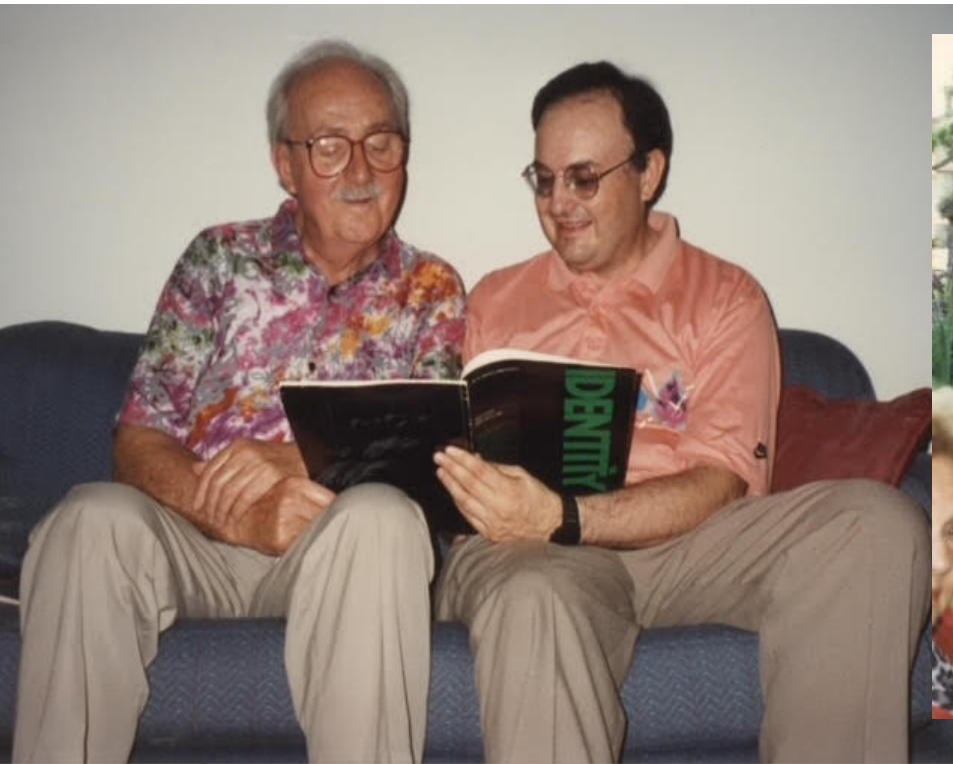
Some personal memories, (1992-2013)

1995



Some personal memories (1992-2013)

- Alberto Luceño visited Madison, 1991/92
- GB and Claire visited Santander in 95/96 and 98/99 with Alberto and Marian.
- Josep Ginebra finished Ph.D UWM. Albert Prat died, in 2005. .



1999: Birthday party in Chicago

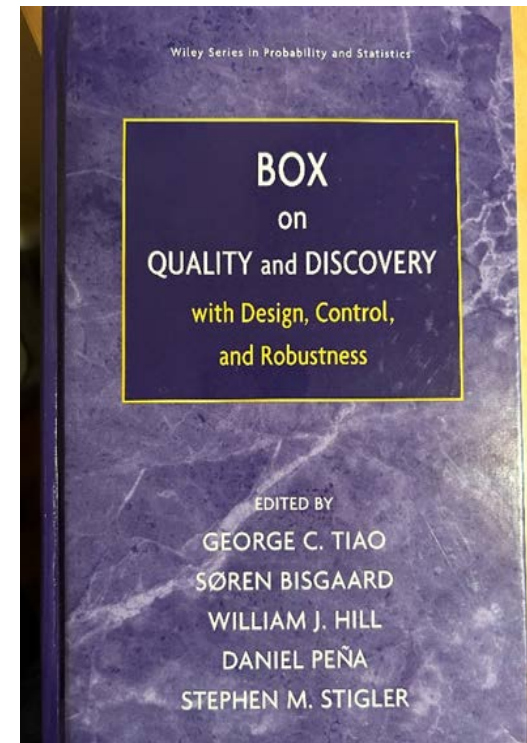
George Box: An interview with the International Journal of
Forecasting

Daniel Peña*

Universidad Carlos III de Madrid, Madrid, Spain

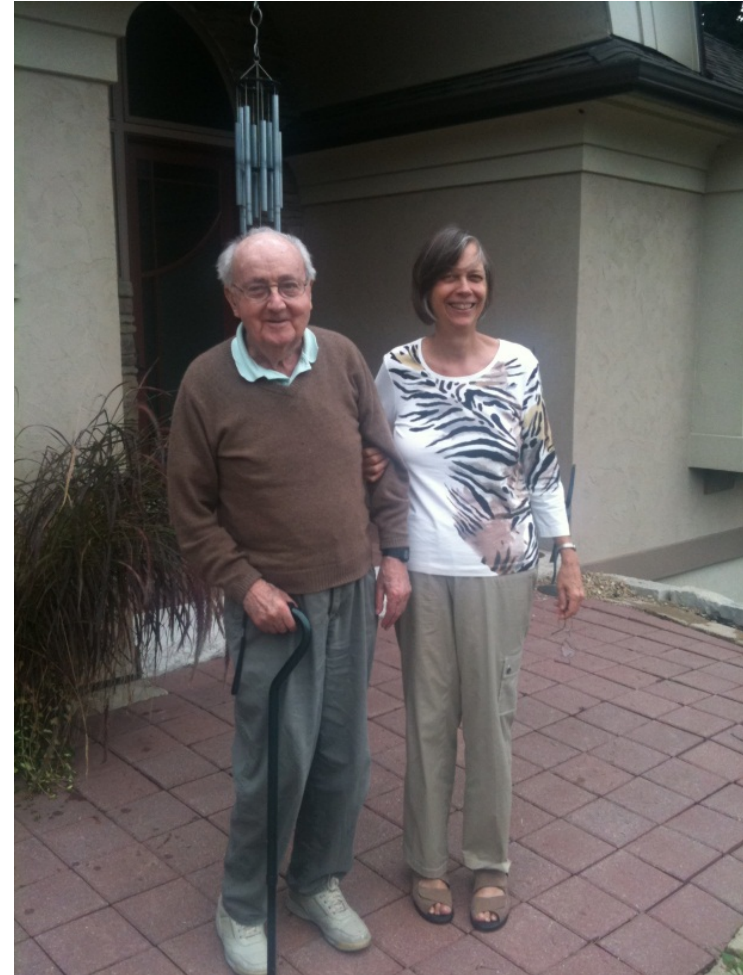


George Box on the left and George Tiao on the right.



Some personal memories (1992-2013)

- Last visit to George Box: He received me with the Ph.D gown.
- I drove from Chicago where I participated in Tiao's 75th birthday in 2008 (George was 87 and in very good mood)
- I talked to him after this by phone several times and also with Claire for his last book *An accidental statistician*.





“All models are wrong,
but some are useful”
George E.P. Box

Obituary: George Box (1919-2013) “We remember not only his brilliance, but also his wit, skits, and storytelling”

23rd April 2013

George Edward Pelham (G.E.P.) Box, “George”, or “Pel” (for those very close to him), passed away on Thursday, March 28, 2013 at home at the age of 93. He was one of the greatest statistical minds of the 20th century, shared his wisdom with many and will be missed by many.





Educación

'IN MEMORIAM'

George Box, estadístico

Considerado uno de los científicos más influyentes y citados en la segunda mitad del siglo XX

DANIEL PEÑA

07 abr 2013 - 00:11 | Actualizado: 09 ABR 2013 - 02:08 CEST



George Box falleció el pasado 28 de marzo a los 93 años de edad en Madison, Wisconsin. Nacido en Inglaterra en 1919, comenzó trabajando como estadístico industrial en una gran compañía

Por todo ello se considera uno de los científicos más influyentes y citados en la segunda mitad del siglo XX. Además, era una persona admirable: gran profesor, excelente conversador y polemista con fino sentido del humor, comprometido con las causas sociales y la paz, adalid del pensamiento crítico en la ciencia y en la vida, y generoso y desinteresado con su tiempo y sus conocimientos. Tuvo mucha relación con España, a través de las universidades de Madrid, Barcelona y Santander. Así, en 1995 fue nombrado doctor honoris causa por la Universidad Carlos III de Madrid. Su muerte es una gran pérdida para todos los científicos. Descanse en paz.



3. Contributions to Statistics

Norwegian Academy of Science
Drammensvein 78
NO-0271
Oslo
Norway

Madrid 10 November 2005

To whom it may concern:

This letter is to nominate Prof. George E. P. Box for the Abel Prize in Mathematics. In our opinion George Box is one of the leading figures in inventing mathematical models that are useful in science and the most influential statistician in science at large of the second half of the XX century. His contributions cover a wide spectrum of topics, including statistical inference, robustness, time series analysis, design of experiments, response surface methodology, evolutionary operations and statistical quality methods. Some of his key contribution in these areas are as follows:

1. His first important contribution is in developing Experimental Design for Industrial applications. He invented Evolutionary Operation, a very original and powerful procedure to improve continuously the working conditions of an industrial plant by using experimental design. This procedure was completely revolutionary at the time and has had a great impact especially in the chemical industry. He continued the work made by R. A. Fisher and was the main figure in developing mathematical ideas for experimental design in industry, including surface response methods.



2. The second contribution is on the methodology of building statistical models. He invented and developed the idea of robustness, created the now standard procedures for transforming variables, extended the scope of Bayesian Statistics, developed new tools for model diagnosis and had a key role in the integration of the classical and Bayesian schools of thought in Statistical Inference.



3. The third key contribution is time series analysis and forecasting. The methods he developed between 1962 and 1972 with G. Jenkins for analyzing and forecasting time series, commonly called the Box-Jenkins approach, are now learnt by all students of statistics, economics and engineering as part of their standard curriculum, and are thought at the Ph. D. levels for students of Psychology, History, Sociology, Chemistry and Biology. Very few scientific developments of the last 50 years have had such a big impact over the whole scientific community. To give an idea of his influence in economics, today all major central Banks and Statistical agencies are analysing economic data and forecasting them by using the methods he developed with his co-authors.



4. Finally the fourth major contribution is to the Statistical Quality field. Professor Box has had a key role in showing the importance of statistical thinking in improving quality and productivity, and he has developed many statistical tools and procedures to improve quality and productivity in industry.

Please below find the name and addresses of ten specialists from different countries who may be contacted to obtain information about George Box:



Prof. T. W. Anderson,
Department of Statistics,
Stanford University, Stanford, CA 94305-4065
twa@standord.edu

Professor David Cox
Department of Statistics, University of Oxford,
1 South Parks Road, Oxford, OX1 3TG, UK
Work Fax: 01865 272595
david.cox@muffield.oxford.ac.uk

Professor Herman Chernoff
Department of Statistics
Harvard University
Cambridge, MA 02138
Telephone: 617-495-5462
Fax: 617-496-8057
chernoff@hustat.harvard.edu

Prof. Manfred Deistler
Institut für Konometrie and System Theory
Technische Universität Wien
Wien, Austria

Professor Bradley Efron
Department of Statistics
Sequoia Hall
390 Serra Mall
Stanford University

Prof. Clive Granger
Nobel price winner in Economics
Department of Economics
University of California San Diego
La Jolla, California (United States)
Homepage: <http://weber.ucsd.edu/Depts/Econ/>
Email: webecon@weber.ucsd.edu
Phone: (858) 534-3383
Fax: (858) 534-7040
Postal: 9500 Gilman Drive, La Jolla, CA 92093-0508

Prof. Frederick Mosteller
Professor Emeritus
Roger I. Lee Professor of Mathematical Statistics.
Department of Statistics
Harvard University
Science Center
One Oxford Street
Cambridge, MA 02138-2901
USA

Prof. Donald B. Rubin,
John L. Loeb Professor of Statistics & Chairman
Department of Statistics
Harvard University
Science Center
One Oxford Street
Cambridge, MA 02138-2901
USA
Telephone: (617) 495-5496
Fax: (617) 496-8057
rubin@stat.harvard.edu

Prof. Stephen M. Stigler
Ernest DeWitt Burton Distinguished Service Professor
Statistics and the College;

Prof. Arnold Zellner,
H.G.B. Alexander Dist. Serv.
Professor Emeritus of Economics and Statistics
Grad. Sch. of Business, U. of Chicago

Sincerely yours,

Daniel Peña
Universidad Carlos III de Madrid

George C. Tiao
University of Chicago



4. Pioneering Data Science

Wegman (1992): BJ anticipated the attitude and approach to data analysis that we think of as contemporary data analysis

- 1) Statistics is learning from data by iterative experimentation.
- 2) Learning from a combination of models, instead of a single one, correct model.
- 3) Data reduction by linear transformations.
- 4) Use transformation for non linearity.
- 5) Estimate non linear responses by an iterative gradient algorithm.

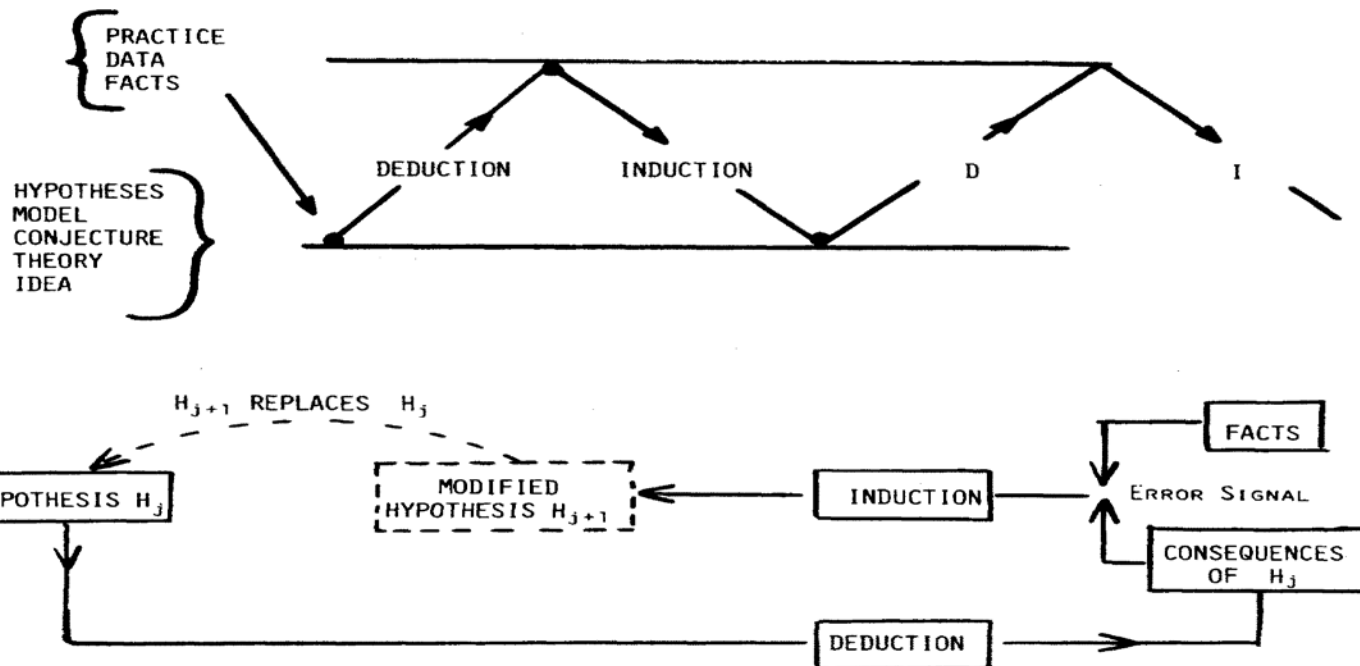


4. 1 Statistics: Learning from data by iterative experimentation

Science and Statistics

GEORGE E. P. BOX*

A. The Advancement of Learning *A(1) An Iteration Between Theory and Practice* *A(2) A Feedback Loop*



4.2 From single model to several models

- Box was a pioneer in introducing several models in statistics, for response surfaces and for outlier analysis.
- First in ICI. Process change with time and we have to adapt the conditions to this. The response model is different in different areas of the explanatory variables. EVOP is designed to find the maximum output on changing conditions.
- Today, in our phone or in social networks automatic design of experiments using EVOP ideas are used to influence our behaviour.
- Analysis of data heterogeneity leads to outlier analysis, mixtures of models and clusters. Pioneer with Tiao (1968) in outlier detection with mixtures from a Bayesian point of view. Big data are heterogeneous
- Changes over time lead to time series integrated process: some derivative is stable or stationary.

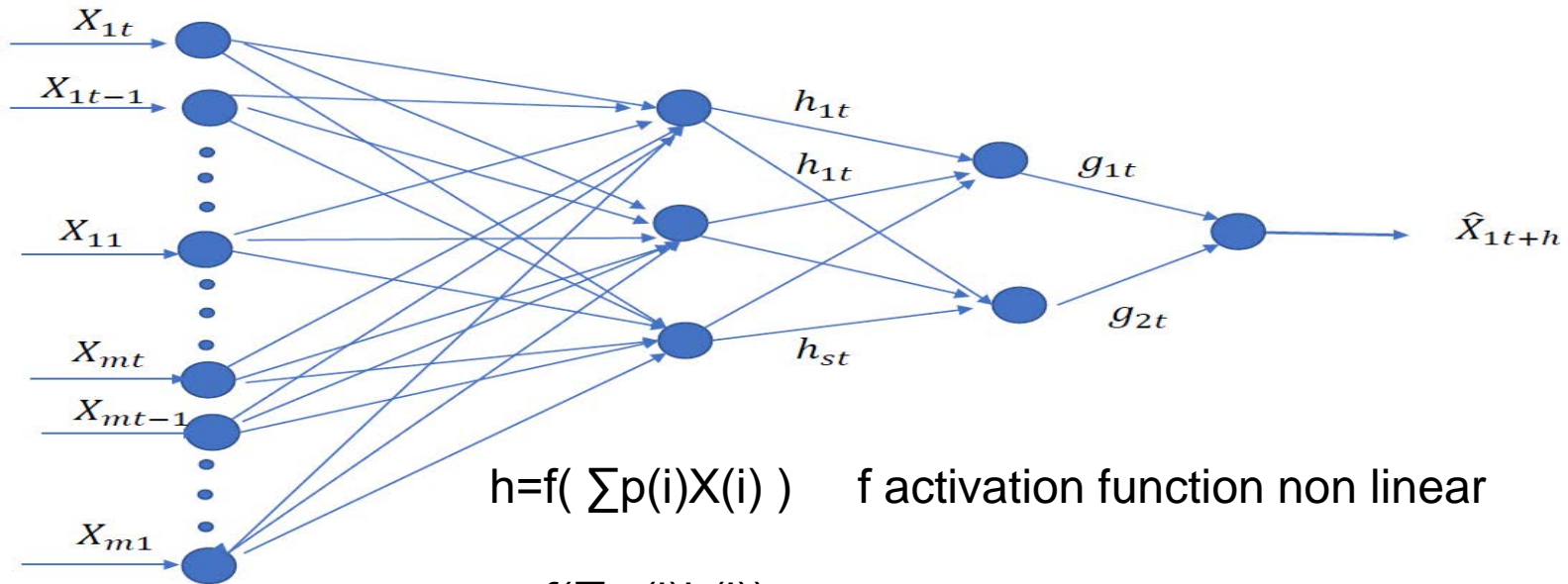


4.3 Data reduction by linear combinations (LC)

- Hotelling introduced LC for static data, PC.
- In time series BT linear combinations of maximum predictability (BT, Bka. 1970)
- Factor models for forecasting large sets of time series (PB, JASA, 1987)
- Today LC of many variables is the basis of Neural Networks and Deep Learning
- Linear combination of predictions



A Dynamic Factor model is a NN with one layer and linear response



$$h = f(\sum p(i)X(i)) \quad f \text{ activation function non linear}$$

$$g = f(\sum w(i)h(i))$$

Prediction = $g(\sum a(i)g(i))$ Usually f for Class, linear for Prediction



4.4 Model non linearity by a transformation

- Instead of $y = \sum_{i=1}^p b_i x_i$

think of $g(y) = \sum_{i=1}^p b_i x_i$ where $g(y)$ belongs to Box-Cox transformations. Then,
 $y = g^{-1}(\sum_{i=1}^p b_i x_i)$.

Same idea in NN instead of $h = \sum_{i=1}^p b_i x_i$
use $h = g(\sum_{i=1}^p b_i x_i)$ where g is the activation function



4.5 Non linear response and non linear estimation by gradient descent

BW'52 proposed a non linear response function and a gradient descent algorithm to reach the maximum with EVOP.

BJ'61 generalized this idea for dynamic systems that change over time and to estimate ARIMA models.

The key idea of estimation of NN is back propagation based on a Least squares gradient algorithm and back propagation going back from the network.

BJ proposed a first way to do back propagation to estimate the unobserved residuals of the model



Conclusions (I)



- *Great influence in shaping applied statistics*

Bayesian and Likelihood are both needed; Estimation and Diagnosis and model cheking,

Iterative model building,

Data transformations (with Cox),

Data heterogeneity and Robust procedures (with Tukey),

Bayesian analysis (with Jeffreys and Tiao),

Mixtures of Models (with Tiao),

ARIMA time series (with Jenkins) and Multiple time series (with Yule, Quenouille and Tiao).



Conclusions (II)



- *Pioneer in methods for big data analysis:*

Deal with non linearity with gradient estimation, transformation and model changes

With many variables use linear combinations,

Emphasize model for prediction,

Learning by iterative divide and conquer strategies



Conclusions (II)

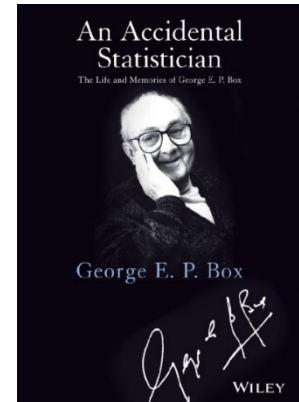


Data oriented teaching of statistics

- Teach Statistics as a scientific discipline, emphasize project work and avoid mathematics for the sake of it.
- Teach Statistics as the key tool for scientific learning and apply it to all stages of the process.



References



- DeGroot, M. H. (1987). A conversation with George Box. *Statistical Science*, 2(3), 239-258.
- Peña, D. (2001). George Box: an interview . *International Journal of Forecasting*, 17(1), 1-9.
- Peña, D. (2002). The major contribution of Professor George EP Box to applied statistics. *Chemometrics and intelligent laboratory systems*, 63(1), 5-6.
- J. Stuart Hunter (2013) In Memoriam: George E. P. Box, *Technometrics*, 55:2, 119-120.
- Peña,D. (2013). In Memoriam: George Box, estadístico. *El País*, 7 abril 2013.
- Ljung, G. M., Ledolter, J., & Abraham, B. (2014). George Box's contributions to time series analysis and forecasting. *Applied stochastic models in business and industry*, 30(1), 25-35.