

George Box: An Interview for the International Journal of Forecasting

Daniel Peña

Universidad Carlos III de Madrid

George Box was born in England in October 18, 1919. His education includes a PhD. in Statistics and a D.Sc. both from London University. He started his professional life working as statistician for eight years (1948- 1956) for a large Chemical Company (ICI) in the UK. During his stay with ICI he took a year's leave of absence as visiting professor of the University of North Carolina. Later he moved to Princeton, as Director of the Statistical Research group, and then to Madison, Wisconsin, to create the Department of Statistics in 1960. In a few years and under his leadership the Department of Statistics of the University of Wisconsin was ranked among the best Departments of Statistics in the world. In 1980 he was appointed Vilas Research Professor of Statistics, UW. In 1986 he created, with Bill Hunter, the Centre for Quality and Productivity Improvement that has had a pioneering role in the Quality Movement in the USA and elsewhere. He became Emeritus Professor in the University of Wisconsin-Madison in 1992.

George Box has published nine books and more than 200 research papers. The importance of his contribution to Science has been recognised by receiving many awards, among them The Gold Medal of the Royal Statistical Society, The Shewart medal from The American Society for Quality Control, The American Society for Quality Control Brumbaugh Award, the Wilks memorial medal, The Youden Price, The Deming Medal, The British Empire Medal, and the Byron Bird Award for Excellence in Engineering Research, among others. He was elected a member of the American Academy of Arts and Science in 1974, President of the American Statistical Association in 1978, President of the Institute of Mathematical Statistics in 1979 and in 1985 he was elected a Fellow of the Royal Society. He has Honorary Doctor's degrees from the University of Rochester, Carnegie Mellon University, and the Universidad Carlos III of Madrid.

Readers interested in the work of George Box are advised to read the interview with him made by Morris DeGroot (1987). George Tiao edited his collected works up to 1984 in two volumes (Tiao et al, 1995). A third volume, Tiao et al (2000), including his work on Quality will appear soon.

The following conversation took place in Chicago, October 1999, the day after the party to celebrate his 80th birthday.

Q: Let us start talking about the origins of your work in time series and forecasting. The first paper you wrote with Jenkins (Box and Jenkins, 1962) has been considered as a breakthrough in statistics (Johnson and Kotz, 1992). How do you become interested in time series?

A: When I studied statistics at University College, after the war, I remember having a course on time series. It was all very mathematical and I didn't see what application it had to anything. At ICI I was mostly working on experimental design. However, there was a group there called Intelligence Department that forecasted monthly sales. I remember asking them for records of what their monthly forecasts had been and what in fact had happened. It was clear that the one step ahead forecast errors were not random and that they were overreacting. So I went back over the data and just took some moving averages. These clearly did a lot better than what they were currently doing. I don't think at the time exponentially weighted averages had been invented, but I had been thinking about putting more weight on the last points of the average. However soon after that, I left ICI.

In the United States, I did quite a bit of consulting for the chemical industry, and one of the problems they mentioned to me went like this. For this particular process there was a relationship approximated by a quadratic curve which connected yield Y with temperature X . The curve was drifting because the catalyst was decaying, and it was impossible to predict which way it would go. The problem was to cause the process to automatically change temperature to follow the moving maximum. My idea was to inject a sine wave about set point into the temperature, varying it by a small amount, then looking for this sine wave in the output by multiplying by a second generated sine wave and integrating the result. If this integral was positive then you needed to move forward, if negative to move back. I tried to get them to build an automatic optimiser of this kind at Princeton in the Chemical Engineering Department but without success. I discussed this problem with Gwilym Jenkins when he came on a short visit. I had thought about this in a deterministic kind of way but

Gwilym pointed out that, because of the dynamics of the system, the sine wave would be changed in phase and amplitude and that it would also be necessary to allow for autocorrelated noise in the system. About that time I moved from Princeton to Wisconsin and there, with the help of Olaf Hougen, the “grand old man” of Chemical Engineering, we went ahead with this project. We got some money from NSF and, after Gwilym Jenkins came to visit us in Wisconsin, we made rapid progress and eventually built the reactor and got the thing to work. We eventually realised that our automatic optimisation was a particular case of feedback control. Later we saw that the kind of control we were talking about was really related to forecasting, since we were arranging that the forecast of the deviation from target would be cancelled out by the action that was taken. So, by this time, we had become very interested in the modelling of time series and of dynamics (one of the series in Box and Jenkins was in fact obtained from the optimising reactor that we built).. Gwilym had a good deal of practical experience from the time he spent working on aircraft design.

When we looked at several business and economic series all of them seemed to be non-stationary. We realised that it was such series that Holt and Winters and other people were trying to forecast in operations research. For such series it was very reasonable to forecast with a weighted average in which the weights fell off geometrically. So we started to think of what this could mean in terms of dynamic and stochastic models and this was the beginning of the ARIMA models. We knew that exponentially weighted moving average (EWMA) would provide the best forecast if this series followed what we later called an IMA model - a moving average stationary in its first difference. Also there was a great deal of work that had already been done by Wald and others, particularly about autoregressive series, so we put this all together. I had met Gwilym in the second half of 1959 and our first paper appeared in 1962, so that things had been moving very fast, and I don't think at that point we had got it quite right.

Q: At what point of time did you decide to write the book?

A: Our work was being supported by the office of Air Force Research and published as a series of technical reports. One day, much to my surprise, Gwilym said to me that, we should not just publish reports, we should write a book. I was surprised at this idea at first and didn't really think that many people would be interested in it; but we started thinking along those lines.

Q: When was this? 63? 64?.

A: About 1963. Gwilym came for a year's visit to Madison. At that time he was beginning to have some serious problems with what was later diagnosed as Hodgkin's Disease. So that from then on,

until he died, Gwilym was more and more sick, going through periods when he would get very ill, and then making temporary recoveries. He was extremely courageous and continued his research and his lecturing until the end of his life.

Gwilym lived in a beautiful house about 4 miles north of Lancaster in England. I started to visit him there every summer and we worked together. The previous owner of Gwilym's house had kept servants so I had the maid's quarters on the second floor. I had a room to sleep in and a room to work in. (although the maid wasn't there any more). There was a long hallway and at the other end of the hallway was Gwilym's office. We would make some progress, then he would come to see me or I would go to consult him and we went on like that. We had a number of specific problems. There was a "golf course problem" named because we used to discuss it as we walked around the edge of the local golf course.

Q: What was "the golf course problem"?

A: It was a problem developing what are now called bounded adjustment charts. We assumed a quadratic loss from a process being off-target, and a fixed cost to adjust the process. So the problem was how could you devise a scheme for minimising overall cost such that if the process wanders too far from target you are going to stop and readjust it. We solved it eventually using dynamic programming supposing the last observation had just gone over the limit and working backwards from there. An important question was what would be the run lengths with various choices of the parameters before the process needed to be adjusted. We found an approximation which was fairly good, but subsequent research has found better ones. Another problem was the "jam jar" problem which involved the relation between differential equation models and difference equation models.

Q: So the way that the book advanced was from meetings during the summers?

A: Yes, pretty much.

Q: And during wintertime?

A: I would send a tape recording to him and if there were equations, I would write them on a piece of paper, fold the paper around the tape and send it air-mail. He would do the same. We would sometimes listen to tapes that we had made a year before. We would hear ourselves talking about a problem, now solved, that at that time we hadn't known how to solve, and we always thought, "Why didn't we see that before?".

Q: Do you have these records?

A: There may be some somewhere. I haven't looked for them.

Q: What were the problems that were hardest to solve? The things that stopped you going ahead?

A: Seasonal time series was one. At that time I was doing some work for Arthur D. Little, the well-known consulting firm, and they were using exponential discounting but with polynomial models that were not producing very good forecasts. We found the airline series in the book by R. G. Brown. There was a twelve monthly pattern and it was a developing pattern. After this we thought about seasonal differences and, in particular, the “airline model”. This had twelve roots on the unit circle and we realised we were getting a big bang for our buck, because this model containing only two parameters could produce a pattern which could develop and change as new data became available. In other words, sine waves at all the various frequencies were being individually updated. Most of the time it seemed to work remarkably well.

Q: How do you solve the problem of fitting ARIMA models?.

A: I was, and am, a great friend and admirer of George Barnard. In 1959 Gwilym had been visiting the department at Stanford for a year. After a short time he wrote to George Barnard saying he was not happy there. So George wrote me at Princeton. In particular he said “Gwilym is very good on time series and I would accept his judgement even before John Tukey’s.” When I showed this to John he said “I think we should get this guy here.” So that is how Gwilym and I met. We had both learned a lot about likelihood from George Barnard and it seemed natural to use likelihood for estimation. At that time I was getting more and more into Bayes, but I thought of likelihood as being a timid man’s way of doing Bayes. For samples of the size you need to estimate time series parameters it didn’t make much difference anyway.

Q: Was it hard to devise the algorithm with the back forecasting and everything?.

A: The problem, of course, with any kind of time series is that what happens at the beginning depends on what had happened before you started. However, we realised our models were reversible, and since the transients would die out quickly, a natural thing was to forecast what you didn’t know by starting at the other end. We found that this method converged very rapidly when we tested it on a number of series and we later looked at it from a mathematical point of view.

Q: How did you manage to do the computation?. Did you have students to write the programs?. I remember that at that time, the maximisation of the likelihood took a lot of time.

A: George Tiao and I got some people working on programming at Madison, and this work was further developed under the direction of Professor Lon Mu Liu at Statistical Computing Associates in Chicago. Gwilym developed a program independently at the University of Lancaster. He had started a

Department of Systems Engineering there which I think was quite remarkable. He and his students developed close associations with local industry and worked on projects with them. The department was very successful and brought in a great deal of money to the University. But I think people in other departments became jealous, and there were many difficulties. So Gwilym decided to leave and set up his own company (Gwilym Jenkins and Partners) where his programs were still further developed.

The other thing about the book was that when Gwilym was at Stanford, he had met a man called Murphy who had started a new publishing company.

Q: Holden Day?.

A: Yes, Holden Day. Well, while at Stanford, Mr. Murphy had asked Gwilym to be an editor for their time series books. So we decided that our book should be published by them. Unfortunately, they were quite reluctant to pay any royalties and we found the only way to get paid was to threaten to sue them. One day Gwilym asked me to call his lawyer in San Francisco about this, and I asked him how he had found a lawyer in San Francisco and he said through the British Consulate there. So I called up the lawyer, who in an extremely strong English accent said "Oh, Professor Box, I'm terribly glad you called me. Gwilym said you would". I said he didn't sound to me like an American lawyer, and he said, "Of course, I used to be in England, but I took the Bar Exams and I am working over here now". He arranged for a permanent pending legal action and he would threaten to use it whenever we didn't get paid. Later Holden Day went bankrupt and I am glad to say the third edition of the book, with Gregory Reinsel as a co-author, was taken up by an excellent publisher - Prentice Hall.

Q: What was the kind of reaction that you received when it was first published?. Did many people object to the new ideas that were in the book, or do you think it was generally accepted?. What was, in general, the reaction to the book at the beginning?

A: As I recall, what reaction there was, tended to be negative. Some people said it wasn't rigorous enough; others said there was nothing new in it. However, I have found that initially original work is inevitably met with some hostility. For example, my first paper on response surfaces and my first paper in which the word "robustness" appeared for the first time were both extremely difficult to get published. I think new ideas upset people.

Q: Were you surprised by some of the comments or criticisms of the book, or not? For instance, the book had a big impact in economics and business. But I suppose at the beginning, people thought in a different way. For instance, I remember that there was some discussion about differencing. Many

people objected that differencing was not the right way to do it. And now, as you know, in econometrics cointegration and unit roots are a big business.

A: Econometric models have a lot in common with the kinetic models used in describing complex chemical systems. Gwilym and I both had some experience with this. For example, several quite different theoretical kinetic models were put forward by different sets of chemical engineers and chemists to explain the production of ozone, and the subsequent pollution of the atmosphere. The problem was that all the models contained very large numbers of parameters which it was quite impossible to estimate from the available data. This was rather like economic modelling. We thought that a good approach might be to start from the data and produce a dynamic stochastic model empirically, and *then* try to relate this to theoretical mechanisms that could be identified. At one time we had a joint project along these lines with the economists, but I am sorry to say nothing came of it.

Q: Another question. At that time, in the seventies, you were also interested in Bayesian statistics, you were doing research in this area. Didn't you try to incorporate Bayesian ideas into the forecasting?

A: Gwilym wasn't against Bayes, but he was less enthusiastic than I was. However so far as parameters estimation was concerned it made very little difference.

Q: How did intervention analysis come about?

A: People concerned with air pollution in Los Angeles approached George Tiao and I. They had a tremendous amount of data on air pollution with many different pollutants. We worked closely with them and particularly with an excellent chemist called Hamming. We knew that in an attempt to deal with pollution the authorities had introduced various kinds of laws –for example, outlawing the burning of leaves–. They didn't want to believe that what was going on was really due to what came out of the back end of an automobile, but our friend Hamming was certain that this was the explanation. One critical circumstance was a law that was passed which, from a certain date onwards, made it illegal to sell gasoline which contained more than a certain proportion of hydrocarbons. We developed intervention analysis initially to look for a possible change associated with this and we found it. It was closely related to a paper that George Tiao and I had published about a change in mean in a IMA time series. The difference in means was estimated by a linear combination of the data, the weight function for which consisted of two exponentials back to back. This was very sensible because what happened just before and just after the event were obviously of most importance. It was about this time that we started saying to people who didn't believe in the preponderance of non-

stationary time series, “Well, unless you believe that something that happened three years ago is just as important as what happened today then you don’t really believe in stationarity”.

Q: One paper that I liked very much is the canonical analysis paper that you wrote with George Tiao. This paper includes the key idea of finding linear combinations of nonstationary time series that are stationary, that is the idea of cointegration. How did you come to the idea of looking for linear combination of multivariate time series?.

A: My own association with multivariate analysis came about because my education in formal statistics didn’t end when I left University College. When I was a statistician in ICI, I turned out a great deal of work for the company. As a result my boss said if I wanted to go to meetings or lectures, or anything like that, I should just go. So I went to the afternoon lectures by M. S. Bartlett who was then at Manchester University. In particular, I took a course with him on multivariate analysis, which included canonical analysis. I found Bartlett’s lectures were extremely clear, particularly since he used n-dimensional Geometry which I have always found very helpful.

Q: After time series, you became interested in quality. You have been making contributions to the quality field in the last fifteen years. What was the driving force that lead you to quality after time series?

A: Sometime in the middle of the 1980’s I remember seeing a slide that said, “What do these things have in common?” – it included automobiles, cameras and every kind of technological gadgets. The answer had become that for each of these products the United States had lost fifty percent of its market to the Japanese in the last five years. In particular, the United States automobile industry had become astonished and embarrassed by the clever designs and narrow tolerances of their Japanese competitors. A number of senior executives in the Ford Motor Company and other industries visited Japan and discovered that one reason for the superiority of their products was the wide application of Experimental Design. In particular, they found, that an engineer, Dr. Taguchi, had used statistically planned experimental arrangements to carry through what he called parameter design. Both here and in the United Kingdom I think, the higher ups in industry were oversold on Taguchi. They started to suggest that all the statistics we had been doing in the past was worthless and if you weren’t using “Taguchi methods” then you should be fired. So one thing we thought we should do was to study Taguchi’s ideas. We found that, although these were not original, they had certainly stimulated tremendous interest in statistics in our industries. So we started to make careful studies of what was good and not so good about Taguchi’s ideas. We examined them in detail and presented a number of

research reports. In particular we found that some of the methods that Taguchi was advocating were inefficient or even totally wrong and we suggested more efficient, and usually simpler, alternatives. I worked very closely with Bill Hunter in this enterprise, and we together set up the Centre for Quality and Productivity Improvement in the department of Engineering at the University. One day Bill surprised me by telling me he had just been to see the Mayor of Madison about making Madison a quality city. The mayor said he was prepared to give it a try. One of his biggest problems at that moment was the city automobile maintenance garage. They had a big fleet of police cars and garbage trucks that went there to be maintained and repaired. It seemed to take an inordinately long time to get them fixed. Bill Hunter and Peter Shultes went down to the garage and talked to the mechanics. They said, "Look, we do a good job down here, and all we get is people blaming us". So Bill and Peter said "Why don't you just keep tabs on how long a car sits out in the back lot waiting to come in; how long it is before it gets into the shop; how long it takes to get spare parts; how long it takes you to actually fix it; and how long does it sit on the front lot waiting to be picked up. They kept records for about five or six weeks. It became obvious that the longest period of all was for the police to come and pick up their own automobiles. The shortest time was the time the mechanics took to fix them. So the mayor was very impressed with this, and wanted quality methods applied to all kinds of other things. Soon after this, Bill taught a course in the evening so that people from the local hospitals, banks as well as industry could come and take the course. Many of those people are now leaders encouraging quality improvements in their various fields. So since that time we have been making a careful study of the many different aspects of quality, and up to now, the Centre has published over one hundred and fifty technical reports, most of which have been published in various journals.

Q: Is there any learning that people interested in forecast can get from the quality field?

A: Well, it certainly is true that the correct way of looking at quality control charts is as a plot of a time series. Also time series analysis can show how feedback control can be adapted for SPC. In addition, one needs good forecasts to run a quality company, but the people in the boardroom have often used only elementary ideas –how did what happened this year compare with what happened last year– and so on. Better methods for comparison and forecasting have lead to better decisions and hence to better quality.

Q: Lately, you have come back to the problem of control with the book that you wrote with Alberto Luceño. What have you learned about control that was not in the Box Jenkins book?

A: One thing that was taken much further in the Box and Luceño book is the study of constrained control. It is easy to show that control which gives minimum mean square error of the output is often quite impractical because it may require a tremendous amount of manipulation at the input. However, you can show that for a very small increase in mean square error at the output (sometimes only two or three percent) you can often get a huge reduction in the amount of manipulation you need. Another thing that is different about Box and Luceño is that instead of saying, let's solve the mathematical problem of optimal control, we rather said let's take some of the simple form of control which we know have worked in industry, and appropriately adapt these for SPC. We show in the book that tremendous simplification can be obtained by this approach with almost no loss of efficiency.

Q: George, it would great if some other people could learn from your experience in discovery and learning, so what kind of advice can you give to young people who are starting out in research and are trying to discover new things in statistics?

A: As undergraduates, students are encouraged to sit with their mouths open and their teachers pour in "knowledge" for several years. Then they graduate. But then, when they become graduate students, they are expected to do something totally original. They have been regularly fed and now they suddenly have to go out and feed themselves and they haven't been taught how to do it. Undergraduate education should give more opportunities for student to use their creativity. Because of inexperience new graduate students tend to start off trying to solve a problem in full generality with all the bells and whistles. I tell my students don't try to get a general solution at once. Start out with $n = 1$, and $m = 2$. Once you can really understand the problem in its simplest form, then you can begin to generalise.

A story I have often told my graduate students is about a man who was very tall and who had a little son of about four years old. They were walking down to get a newspaper at the corner, and the father suddenly realised that the little boy was having to run to keep up with him. So he said, "Sorry, Tommy am I walking too fast?" And the little boy said, "No, Daddy. I am."

Very often people miss the point because there is nothing less obvious than the obvious. It's best if you try to forget everything you know and try to think of the problem from first principles. It says in the New Testament "Except ye be as little children ye shall not enter the kingdom of heaven". If you don't approach problems in this way, you tend to get caught in the tramlines, and think what everyone else has already thought, and you don't come up with anything new.

I also think that Mathematics alone is not an adequate basis for Statistics. Mathematics is primarily concerned with proving certain propositions. Given certain assumptions, is this true or isn't it? But Statistics ought to be concerned with discovery, which is about finding out things which were not in the original model. Also I believe it is application that most often suggests new theory, and not the other way around. Albert Einstein said that many people thought that he came up with the idea of relativity from pure theory, but that this was untrue and that the theory of relativity was based on observation.

Because of the necessity to continually change the model as one's understanding develops, scientific investigation can never be coherent in the sense that mathematical statisticians would like to make Statistics. This is the reason that Gwilym and I were at pains to emphasise the iterative nature of model building.

Q: When something really new comes to you, usually what is the way that it happens? Have you been thinking about the problem for a long time, or sometimes does it happen suddenly?

A: I think it usually doesn't happen by sitting down and working at one's desk. I think the subconscious mind goes on trying to figure things out when we are not aware of it. You might be in the shower or taking a walk and you suddenly see what might be the best way to tackle the problem. Also I like to work with someone else. My work would not have been possible without my students and co-authors. Particularly Stu Hunter, Bill Hunter, Gwilym Jenkins, George Tiao, Daniel Peña and Alberto Luceño. When I have worked with these and other colleagues I have felt that the sum of our efforts has always been greater than the parts. Bouncing ideas around, discussing, arguing and so forth greatly catalyse the process of learning and discovery.

Q: Apart from work, what do you like to do?

A: These days I like to go for walks in the country. It could be in Wisconsin where I live, or it could be in Northern Spain, or almost everywhere. We tend to forget that we are basically very primitive people. Now there is no evidence that the mind has changed very much in the last 10,000 years so getting back to our environment, to the sea, the fields, the hills, the mountains, is a very fundamental need for all of us.

OK, George, Thank you very much. I think this is a good point to stop.

References

Box, G. E. P. and Luceño, A. (1997) *Statistical Control by Monitoring and Feedback Adjustment*. Wiley

Box, G. E. P. and Jenkins, G. M. (1962) "Some Statistical Aspects of Adaptive Optimisation and Control. *Journal of Royal Statistical Society B*, 24, 297-243.

DeGroot, M. H. (1987) "A conversation with George Box." *Statistical Science*, 2,3, 239-258

Johnson, N. L. and Kotz, S. (1992) *Breakthroughs in Statistics, Volume II*. Wiley

Tiao, G. C. et al. (1985) *The Collected Works of George E. P. Box*. Wadsworth

Tiao, G. C. et al. (2000) *George Box on Quality and Discovery*. Wiley