Statistical Science
2002, Vol. 17, No. 3, 357–378

A Conversation with Emanuel Parzen

H. Joseph Newton

Abstract. Emanuel Parzen was born in New York City on April 21, 1929. He attended the Bronx High School of Science, received an A.B. in Mathematics from Harvard University in 1949, an M.A. in Mathematics from the University of California at Berkeley in 1951 and his Ph.D. in Mathematics and Statistics in 1953, also at Berkeley. He was a research scientist at Hudson Labs, Physics Department of Columbia University, from 1953 to 1956 and an Assistant Professor of Mathematical Statistics at Columbia from 1955 to 1956. In 1956, he moved to Stanford University, where he stayed until 1970, at which time he joined the faculty at the State University of New York at Buffalo, where he served first as Leading Professor and Chairman of the Department of Statistics and then as Director of Statistical Science. In 1978 he moved to Texas A&M University as a Distinguished Professor, a post he currently holds. He has been a Fellow at Imperial College London, at IBM Systems Research Institute and at the Center for Advanced Study in the Behavioral Sciences at Stanford, as well as a Visiting Professor at the Sloan School of MIT, the Department of Statistics at Harvard and the Department of Biostatistics at Harvard. In 1959 he married Carol Tenowitz. They have two children and four grandchildren.

Professor Parzen has authored or coauthored over 100 papers and 6 books. He has served on innumerable editorial boards and national committees, and has organized several influential conferences and workshops. He has directed the research of many graduate students and provided advice, encouragement and collaboration to students and colleagues around the world. To honor these contributions, he has been elected a Fellow of the American Statistical Association, of the Institute of Mathematical Statistics and of the American Association for the Advancement of Science. In 1994, he was awarded the prestigious Samuel S. Wilks Memorial Medal by the American Statistical Association.

The following conversation took place at Texas A&M on May 17, 2000.

Newton: Manny, you have had a long and distinguished career and I have heard you talk many times about the role of luck and skill in determining success. Could you talk a little bit about that?

Parzen: Yes, I'd like to. I've been very successful in the education I've received and the jobs that I've had, and I can't believe it is because I'm smarter than others. I think it is because I am luckier than others. Many people have said to me, "Of course, you've been very lucky." So the question is, what is luck? I give a formal definition of luck which is the Yiddish word *Mazel*, in which *M* stands for place, *z* for time and *l* for speech. So, luck is being in the right place at the right time and saying the right things. There is, however, a rational element of luck of which people making choices in their careers should be aware. People in the academic world often do well, not because they are smarter than others, but because they have chosen somehow sexier fields to research. People who have accomplished much are highly rewarded and esteemed because what they are doing happens to be at the frontier of interest. So the issue of luck in academics

H. Joseph Newton is Professor, Department of Statistics, and Dean, College of Science, Texas A&M University, College Station, Texas 77843-3143 (*e-mail: jnewton@stat.tamu.edu*).

is the question of, in general, good taste in the choice of problems for research.

EARLY LIFE

Newton: Anyone who knows you can immediately tell that you grew up in New York City. What was your family like and what impact has New York had on your life?

Parzen: Let me first talk about coming from New York. People from New York are different from other people (it has been pointed out to me) because they have high self-esteem and high energy. In Texas people have told me, "You do everything faster than us; you talk faster, think faster and move faster." There is a characteristic of people from New York that was pointed out to me in 1949 when I entered California at Berkeley. When we have a conversation, our first words when we respond are, "You're wrong!" People are aghast, as they have never heard anyone say that to anyone. That is just a New York mannerism, I think. Richard Feynman writes that as a young physicist at Los Alamos, Niels Bohr met with him daily because he said, "You're the only one who tells me I am wrong."

I am the youngest of nine children; my father was a rabbi. I was born at home in Harlem. At the age of three we moved to the Bronx. I went to Yeshiva, which is a Jewish parochial school. The one advantage I had was that, thanks to my numerous siblings and my parents, there was always a lot to read in the house, so I was always up on current affairs even from a young age and had a broad knowledge of facts and always loved learning new information.

Newton: What was the Bronx High School of Science like?

Parzen: I entered the Bronx High School of Science the year that a special high school in New York City called Townsend Harris closed down and thus Bronx Science was getting a lot of transfer students. I, too, had transferred as a sophomore from Yeshiva. Because I was very interested in current events, I was very involved in social studies and led an organization called the Forum which brought in many outside speakers, congressmen and so on. Consequently, people who knew me then are surprised I turned out to be a mathematical type because they were sure I was going to be a lawyer.

I was student body president in my last semester and I tell the story that my campaign slogan was, "My opponent Bill has personality, and I have character." People told me that what I actually had was poise. I assumed poise is knowing enough so you are confident in how to act in any social situation. The one thing that people claimed showed that I have poise was that, whenever I walked into a room in anyone's home or anywhere, the first thing that I did was head toward the bookcases and start reading the titles on the books. So then it was said, "That guy has poise." Instead of standing around looking lost, I was at home. I would go right to the bookcase and then when someone said things were starting, I'd turn around and say "Let's begin."

HARVARD AND PROBABILITY AND STATISTICS

Newton: After Bronx Science, you went to Harvard. How did you become interested in probability and statistics?

Parzen: A series of things happened that I claim determined my future. The first thing to understand is that I was lucky in choosing to major in math. In 1945, colleges like Harvard would only admit a limited number of students from public high schools. Only two of us (out of a very brilliant class) were admitted from Bronx Science. My classmate feels that he made a big mistake in going to Harvard because he majored in physics and received very little attention from his department. I majored in mathematics, halfway between my two real interests, physics and economics. As a result, I got a lot of attention from the faculty. For example, there was a tutorial system. Every week you met with a math faculty member and discussed something. This had the drawback that I had the problem of finding topics to discuss that interested me. I happened to find a secondhand book (in a bookstore in Harvard Square) that was just published, but available for only $4. It was Harald Cramér's *Mathematical Methods of Statistics*, which is now being reprinted and is described in the ads as "a synthesis of English and American statistical theory and Russian and French probability theory." I was only a sophomore, but I had the opportunity of studying something with a tutor and I decided to choose the book by Cramér. This was the first thing that determined my future.

The second incident that happened involved one of the world's greatest mathematicians, Saunders Mac Lane. He felt that statistics was important, and as a kind of a one-time experiment in his life offered an introductory course in statistics. However, no one wanted to be the grader. I hadn't had the course, but I volunteered in order to get into the job of being a grader of math papers. So I was a grader in my first statistics course. After that, I had a course in mathematical statistics from Fred Mosteller early in his career at Harvard, and he has been a good friend of mine for 50 years.

Finally, I had to write a senior honors thesis and I was very interested in the philosophical aspects of life. Probability is a subject of both the philosophical and mathematical sciences, so I chose to write a bachelor's honors thesis on the philosophical and mathematical foundations of probability. Thus, when I went to graduate school at Berkeley, I went as a math student, but oriented toward probability.

Newton: You were in graduate studies at Berkeley from 1949 to 1953. Who were some of the people at Berkeley at that time?

Parzen: First, you have to understand that I came to Berkeley at a very historic moment: the "Year of the Oath." There was a lot of faculty fervor, constant meetings and protests because the faculty were being compelled by the regents to sign a loyalty oath (which eventually the Supreme Court ruled unconstitutional). The math department was unique on the campus in believing that since we were all in this together, the teaching assistants should be considered faculty and join the faculty meetings in order to help with the fight. So that brought a lot of closeness.

When I became a teaching assistant at the University of California at Berkeley, I was only 20 and a first-year student. It was extremely rare for a first-year student to become a teaching assistant, so I naturally asked them how it happened. They said, "Well, you have wonderful letters of recommendation; in particular from George Mackey, who very rarely writes a good letter of recommendation." They treated me like a star. I didn't realize that teaching assistants rarely taught calculus. Not only did I want to teach calculus but I wanted to teach the first semester of calculus my first semester, calculus II my second, and so on. I relearned college math by teaching it for two years in the correct order. I was very much part of the establishment from age 20.

Indeed, in order to get a drink in California you had to be 21. I didn't have a driver's license, but I would show my teaching assistant card and they were sure that you had to be 21 to have a teaching assistant card, and they had never met anyone who had been a teaching assistant who wasn't 21, so they assumed I could be served.

One person with whom I had close contact was the man who built the Department of Mathematics at Berkeley, Griffith C. Evans. He shared my interest in mathematical economics and was also graduate advisor when I entered. I attended his seminars in mathematical economics. My second contact was with Charles Morrey, who was the head of the department, and whose research assistant I was in my third year there. I wrote my thesis answering a question of his. A third influence was Anthony P. Morse, a great mathematical expert on measure theory. I took several courses and seminars on measure theory with Morse. He was very wealthy; he was a member of the Morse family who invented the Morse code.

Newton: How about on the statistics side?

Parzen: On the statistics side, I had very close interaction with Michel Loève, who taught probability theory. Interestingly, the people who first taught measure theory in the University of California were the statisticians. When Jerzey Neyman came to Berkeley, he wanted to bring in advanced probability. The math group did not teach measure theory. Essentially, Neyman fought for the right to teach courses like that; he negotiated with the head of the math department for a document that he calls his Magna Carta Liberatum: his freedom to teach what he wanted even though it overlapped with what the mathematicians may have thought was theirs.

So I had close contact with Loève and very close contact with Neyman. I was very fortunate that I was a teaching assistant in mathematics, not statistics, so I wasn't under Neyman's control, because he obviously wanted to control the actions of the people who worked for him. I was able to relate to him as a person with similar intellectual interests to mine, but was not under his financial control. And then, I had close contact with Eric Lehmann, who in a certain sense was trying to advise me from being too much under the sway of Neyman.

At the time I was finishing up my thesis, Lucien Le Cam was finishing his thesis. He was a year before me, and I adapted a lot of his work and extended it to uniform convergence.

Newton: What was your dissertation about?

Parzen: My dissertation was likely one of the hidden gems of the field of statistics. It was about uniform convergence. It started because I was Morrey's research assistant, and he was trying to derive the equations of hydrodynamics from statistical mechanics with the idea of approximating a continuum as the limit of an ensemble of n particles as $n \to \infty$, and he wanted the approximation to be uniform in time. Stimulated by that problem, I extended it to comprehensive research on uniform versions of the theorems of probability and statistics. So my Ph.D. thesis was "On Uniform Convergence of Families of Sequences of Random Variables." Now that is an interesting title! The man who hired me for my first job was a physicist, William Nierenberg, who was the outside member of my Ph.D. committee. I went to the physics department office to put my thesis into his box. I happened to hang around there, and he showed up with some students and took the thesis out of his box and read the title to the students with a big laugh, "Can you imagine something like this? How useful can it be, 'on uniform convergence of families of sequences of random variables'?" It had three chapters. Chapter 1 concerns characteristic functions converging uniformly in a parameter. In statistics we don't have a single probability distribution, but rather parameterized families of distributions; not P but P_θ. So you would like to know whether the theorems are uniform in this unknown probability distribution P_θ. The first task is to get the conditions for the uniform convergence of characteristic functions of the parameter to imply uniform convergence of the distribution functions. To accomplish that, and in particular to get necessary and sufficient conditions, I had introduced the right kind of definitions.

Chapter 2 was on uniform-in-a-parameter probability limit theorems, central limit theorems, weak and strong laws of large numbers and the Borel–Cantelli lemma. The key contribution I made was figuring out how to define uniform-in-a-parameter concepts for the strong law of large numbers and Borel–Cantelli lemma. It is a case where if you have the right definition, just follow other people's work and prove the theorem. Chapter 3 was on uniform-in-a-parameter mathematical statistics, maximum likelihood estimation and the Glivenko–Cantelli theorem.

I should mention students I knew in Berkeley. Besides Le Cam, two other fellow students who around 1953 got their Ph.D.s with Professor Loève were Julius Blum and Leo Breiman.

INTEREST IN TIME SERIES

Newton: How did you become interested in signal detection and time series?

Parzen: Again, I was lucky in my choice of first job, which changed my career from probabilist to time series analyst and system communication engineer, doing research on statistical inference on stochastic processes and statistical spectral analysis. I chose to work at a physics laboratory for personal reasons. The Hudson Laboratory was in New York, where I wanted to live with my mother, and it also gave me a deferment from the Korean War (a motivation shared by a lot of the physicists at the laboratory). I was hired because the physicist on my Ph.D. committee, William Nierenberg, took the job as director of the Hudson Labs and thought that working there would be a good thing, which it was. The problems of the lab were connected with signal processing. One of the jobs they were doing was studying the ocean. They wanted to know what the ocean sounded like. They were measuring noise in the ocean and analyzing spectra of the noise. I decided that this was a problem where we needed more theory.

Another problem that I worked on was passive detection of submarines. Usually you send out a signal and you can determine if something is there, but of course the signal you send out could be heard by the person it was sent out to find, so it alerts them that you are there. The question was whether it was possible to detect submarines just from the noise they were making, that is, by listening without sending out any signal. We devised a method that was essentially estimating the fourth cumulant of the stochastic process. You had the noise of the submarine in the presence of the Gaussian noise in the ocean, so you modeled the noise of the submarine through the propellers—called cavitation noise—and detected the submarine by detecting the presence of a non-Gaussian component in the noise. We actually had trials of submarines and it seemed to work.

Newton: Who were the people at the time who were working in stochastic processes?

Parzen: On the mathematical side, the people who in 1953 were doing work on what you might call the foundation of statistical inferences on stochastic processes were all in Sweden. In particular, there was Ulf Grenander, whom I knew about because he actually had lectured at Berkeley when I was there. I didn't have too much contact with him, but I started studying his thesis. He worked on spectral analysis and detection of signals. Well, this field that Grenander (1950) started became, of course, the foundation of statistical communication theory and signal processing. All the engineers were trying to figure out what this was about, and I happened to get there first because I had the submarine detection problem at Hudson Labs. Another person doing leading work in stochastic processes at that time was Peter Whittle.

Let me make a general remark. A lot of mathematicians became attracted to signal processing theory and

FIG. 1. *International Statistical Institute meeting, Paris 1961: on right side, starting with second person, Ingram Olkin, Dorothy Gilford, Emanuel Parzen, Carol Parzen, Ellen Chernoff, Judy Chernoff; on left side, starting with second person, Betty Scott, Jerzy Neyman, Ann Durbin, Jim Durbin, Miriam Chernoff, Herman Chernoff, Ed Deming is also in the picture.*

essentially turned into engineers, even becoming professors of engineering. I think what was different about me was that I became interested in the same problems, but retained my identity as a statistician working in a statistics department. When I would go to signal processing meetings, I was the only person there from a statistics department.

MODERN PROBABILITY AND ITS APPLICATIONS

Newton: While that was going on, you were doing work that led to your 1960 book, *Modern Probability Theory and Its Applications*, which has had a huge influence on generations of statisticians and nonstatisticians alike. How did the book come about?

Parzen: I was teaching at Columbia University, which at that time had the world's greatest electrical engineers in its electrical engineering department—Lofti Zadeh in particular. They were spending half of their time teaching probability because probability was an important part of their approach to electrical engineering. They recognized that this was very inefficient and that it would be much better to have a course in probability that the students in electrical engineering could take. I happened to be there and I created such a course and started teaching it. I have to emphasize that the first day I started teaching introductory probability I had hundreds of students, because there were all these engineers wanting to study signal processing and therefore probability. When I went to Stanford in 1956, I started a course in probability and again I had immediately hundreds of students who were interested in working in industry.

So that's essentially why I wrote the book. My job as a graduate student at Berkeley had been to fill in the symbols in the manuscript of Loève's probability book, so I was absolutely familiar with that book cover to cover. Thus, I told my students that I would give them in an elementary way the contents of the most advanced book on probability that had just been published by Loève.

Newton: Why do you think the book has been so popular?

Parzen: I think there are several reasons. Number one, engineers have told me "Your book purports to be a mathematics book, but I can tell it's an engineering book." Somehow it had that flavor. I did teach engineers, so the book is very readable. Students love to read it and at the same time it is encyclopedic. To this day, I have never failed to find a topic in the index, whether it be cumulants or medians or anything else. It is chock full of information, plus it has tables of solutions to occupancy problems and other problems and tables of distributions.

I can boast that at Stanford 500 students per year took the course; the bookstore manager told me he never saw a used copy of my book. It was a book that people wanted to hold on to. I also should point out that my book has a minicourse in advanced probability in its last two chapters. Many people have told me, including Russian probabilists, "You're so young and your book is so profound." The proofs I have of the theorems of advanced probability that in many other books take 20 pages are compressed into a few pages.

Another reason that book was written was that when I started writing it, I gave it to the typing pool that the Stanford department had, and a wonderful lady, Mary Alice McComb, began typing it. She would come to my office door every morning and say, "Do you have more for me to type? Otherwise, they will put me back in the typing pool." No Mary Alice not that! So I began writing furiously so that Mary Alice would have something to type and not go back in the typing pool.

FIG. 2. *International Statistical Institute meeting*, *Paris 1961*: *Sam Greenhouse*, *Sir Ronald A. Fisher*, *unknown*, *Carol Parzen*, *Ingram Olkin*, *Emanuel Parzen*.

YEARS AT STANFORD

Newton: In 1956, you went to Stanford, where you stayed for 14 years. How did you happen to go to Stanford?

Parzen: Again, I want to emphasize that, as in all my jobs, I was recruited. The Department of Statistics at Stanford was started in 1948. It was built to eminence by Albert Bowker and in the official 50th anniversary history of the Stanford Department they write that by 1956 the department had reached eminence. That was the year Sam Karlin and I arrived. Bowker wanted to hire the best young probabilists and went around the country interviewing people about a list of the best young probabilists, and I was number one on his list. That was the story of the way I got to Stanford and again I was given star treatment (until Lincoln Moses became Head in 1964).

Newton: Who were some of the people with whom you worked?

Parzen: The people I interacted with at Stanford were first Bowker, then Chernoff, then Karlin and many others. Let me say that the atmosphere at Stanford in 1956 was unbelievably friendly. We would have department parties after every football game and were very supportive of each other. By "we" I mean myself, Herman Chernoff, Herb Solomon (who was Head from 1959 to 1964), Ted Anderson, Ingram Olkin, Sam Karlin, Tom Cover, Ken Arrow (in economics), Pat Suppes, Tom Kailath (in electrical engineering), Gene Golub (in computer science), George Forsythe (also in computer science, a big supporter of me) and Joe Keller (in mathematics). The joint Berkeley–Stanford statistics colloquiums were held twice a semester and provided fantastic parties and opportunities for interaction with everybody.

Newton: During that time a remarkable number of things happened in your career. For example, your stochastic process book was published. How did that come about and why has this book had such an influence?

Parzen: At Stanford I developed Statistics 116, "Introductory Probability," which used my textbook *Modern Probability Theory*, and I also developed introductory stochastic processes courses, Statistics 217 and 218. Now you need to understand that this was part of a larger controversy in statistics about how to teach probability and statistics. Traditional statisticians felt that you should have a single year course that had probability taught as part of a statistics course, and did not support the concept of an introductory probability course. There were national workshops that were held on this question. It was very controversial among statisticians to have a separate probability course. I argued that there were in fact two groups: those who wanted to go on to statistics and those who wanted to go on to stochastic processes. Again, coming from my contact with the engineers, I knew that they wanted to study stochastic processes. Eventually, the engineers studied a famous book by Papoullis that was a combined course on stochastic processes and probability. I was interested in the whole curriculum and I proposed that, under our quarter system, the first quarter would be probability out of my book, and then we would have two quarters on stochastic processes, which was of great interest to students in operations research and electrical engineering. We would also have two quarters on statistical inference, which essentially could start off without teaching probability, as that material would have already been covered in the course on probability.

Over the years, many scientists and engineers have told me how important to their careers were the techniques they learned from my book on stochastic processes. This was a time when the synthesis of ideas and modeling techniques that they provided were highly relevant and influential to the problems being researched by many scientists and engineers. Students in my course on stochastic processes reported that they understood that, to solve problems, the trick was to choose a variable on which to condition. They lamented that there seemed to be no general procedures for choosing the conditioning variable. My goal was to teach creativity—not how to quickly complete the solution, but how to start it and proceed step by step. Students who take engineering stochastic processes courses learn a lot of material as a cookbook, and their experience is that they have no idea how to start when they encounter an unfamiliar problem.

Newton: It was also during this time that your pioneering work on statistical spectral analysis appeared. What led you to your work on optimal spectral density estimators?

Parzen: To answer that I need to first talk about the impact of spectral estimation. In 1960, I organized a session at the national statistics meetings on spectral estimation that was eventually published in *Technometrics* in 1961. It consisted of contributions by Gwilym Jenkins and myself, along with discussion by John Tukey and Roy Goodman. I think these four papers did a lot to spread the word about spectral estimation to applied scientists, economists and engineers. It is interesting to note that Jenkins's paper was called "General Considerations in the Analysis of Spectra" and mine was called "Mathematical Considerations in the Estimation of Spectra," so the two papers represented the British style and the American style. I had many economists tell me that they couldn't understand Jenkins' paper, even though it was supposed to be the practical, down-to-earth paper, but that they fully understood my paper. Really, it got them going. Statisticians debate a lot about the issue of mathematics and statistics, and whether we should be problem-solving oriented or general-theory (theorem proving) oriented. I regard our spectral analysis papers as a counterexample to the claim that we lead people astray unless we emphasize the problem-solving view. It is obvious that there are people who would find a British style "general considerations" paper more understandable, but there were many who found that the American style "mathematical considerations" more understandable. So I feel that we were very successful in spreading the word.

Newton: What was the basis of your spectral analysis ideas?

Parzen: My first paper in 1955 on statistical spectral analysis considered both continuous and discrete parameter time series. This was lucky because it gave me insights needed to obtain general formulas for the bias and mean square error of kernel spectral estimators. My spectral estimation theory was based on the concept of a covariance approach to spectral estimation proposed by Norbert Wiener in his 1930 theory of generalized harmonic analysis. At a meeting in 1967, he told me that he was afraid that his 1930 paper contained nothing new and that physicists knew it all. I assured him that in no way was this true. In the 1980s, we learned that the immortal physicist, Albert Einstein, had proposed the basic ideas of spectral estimation in 1914 in response to a question from a friend about analyzing meteorological time series by smoothing a periodogram. This is called the direct approach. Einstein had not taken Wiener's covariance approach, called the indirect approach, which provides the most rigorous approach to both the probabilistic and statistical theorems of spectral representation analysis. The direct method of spectral estimation by smoothing a periodogram became computationally competitive and attractive in the 1960s with the introduction of the fast Fourier transform (FFT) by Tukey and Cooley. This contribution was recently named one of the top ten algorithms having the greatest influence on the development and practice in science and engineering in the 20th century. The popularity of the FFT began with the recognition that it revolutionized computing times of time series covariances.

My progress in spectral estimation theory in 1955 was due to my asymptotic evaluation of the bias of kernel spectral estimators in terms of the differentiability of the unknown density and the kernel's characteristic exponent r, which is intuitively equivalent to $c+1$, where c is the number of vanishing moments of the kernel function. The mean square error of spectral estimators was shown to decrease as $n^{-2r/(2r+1)}$. For nonnegative spectral estimators the exponent r is at most 2, for which we obtain the celebrated $n^{-4/5}$ rate of mean square error attainable by nonnegative spectral estimators with kernels guaranteeing positive estimators. The often quoted advice that mean square error depends mainly on bandwidth is true for kernels with the same exponent r. This fact is very important as refutation of the criticism that research on kernels is an example of how mathematical research is not beneficial to statistics.

Newton: What led to the Parzen window?

Parzen: I proposed in 1957 the very popular (in applications) Parzen window ($r=2$) for kernel spectral estimation. Competitors were the Bartlett window ($r=1$) and the Tukey window (not necessarily nonnegative). In the 1960s, my econometrician friends (such as Mark Nerlove) would say that if you want to obtain publishable results from spectral estimation use the Parzen window; its graphs of estimators were most interpretable, especially for cross-spectral analysis. Scientists, particularly oceanographers, reported that using the Parzen window reduced "ghosts" (spurious modes due to aliasing).

The theory of statistical spectral analysis provided a role model for theories of function estimation and semiparametric inference which are being applied in economic data analysis.

Because there was controversy among statisticians about criteria for good spectral estimators, in Parzen (1958) I proposed that the appropriate response of a methodologist was: "In a situation such as this, it would appear that in first approaching the problem one should obtain as many theorems as possible. One's criteria as to what constitutes a solution may change but the theorems endure, as statements of incontrovertible facts which may or may not be relevant to the problem at hand." A major result in my 1958 paper—that I regret not exploiting empirically—is the optimal kernel $k(u)=1/(1+u^{2r})$ which was shown by Cogburn and Davis in 1974 to be equivalent to fitting splines. Any modern collection of tools for spectral analysis must include function approximation by splines and wavelets. I regard my 1958 paper as secretly influential because its methods of proof were adapted in several Ph.D. theses to find optimal estimators for various function estimation problems.

Newton: Can you comment on the fact that many people know very little about frequency domain time series methods?

Parzen: In the 1970s, and even now, the only part of the theory of time series analysis known to many applied researchers was time domain methods, especially the highly applicable ARIMA modeling methods introduced by George Box and Gwilym Jenkins. I still urge spectral domain thinking; it is useful even when one is not looking to measure cycles but only measuring signal in noise (such as in intervention analysis) to take advantage of the fact that a time domain model of the noise is an adequate fit if its spectrum has the right shape. Mandelbrot in the 1960s told me that R/S research on economic time series with long tails and infinite second moments was intended to "slay the spectral analysis dragon." I assured him that his ideas help us amend, extend and strengthen (but not discredit) the application of spectral analysis to real economic data. I am confident that important applications of spectral methods to economics will yet emerge in the analysis of financial time series.

REPRODUCING KERNEL HILBERT SPACES

Newton: In the late 1950s and early 1960s you also introduced the idea of reproducing Hilbert space representations of estimators in time series. Do you remember how you thought of this and could you say why this idea is so useful in so many areas of science and engineering?

Parzen: I remember very well. Let me first introduce the background. In the 1940s Hilbert space methods were introduced by Kolmogorov and Loève to study the probabilistic structure of time series. A time series $X(t)$ with finite second moments has a covariance kernel $K(s,t)$ which is positive definite. Therefore, there exists a Hilbert space, denoted $H(K)$, of functions $f(t)$, with suitable inner product $\langle f,g\rangle$ such that $\langle f, K(*,t)\rangle = f(t)$ for all points (observation times) t and functions f in the Hilbert space $H(K)$. $H(K)$ is called a reproducing kernel Hilbert space (RKHS) because of the way the kernel reproduces (represents) the value of f at a point t. It is natural to introduce RKHS to solve the optimization problem of finding the function f of minimum norm $\|f\|$ in a Hilbert space H with inner product $\langle f,g\rangle$ when f is subject to the constraint $\langle f, g(t)\rangle = b(t)$. Current research by statisticians on functional inference makes extensive use of RKHS, as is shown in the excellent book of Ramsay and Silverman (1997).

In the 1950s, the problem was to develop the foundations of linear statistical inference on time series—more precisely, regression analysis of continuous parameter time series. The engineering custom was to choose a coordinate system in which to frame the problem, where the problem might be Fourier series expansion, eigenfunction expansion, or discrete sampling or interpolation. In 1957, lecturing on this engineering literature (especially the classic book by Lanning and Battin, 1956), I had the inspiration that RKHS representations could be used to give a coordinate free presentation of statistical estimators in terms of RKHS inner products, whose computation would be the final step in the analysis.

Newton: Can you give me an example?

Parzen: Suppose somebody was trying to model data collected by airplanes flying at various heights. You have discrete heights and continuous times. The modeler would work on the problem of estimation and would create pages and pages of symbols with summations over the heights and integrals over the times. My opinion is that it is impossible not to make a mistake if you kept on writing formulas with summations and integrals, but you could replace all the summations and integrals with inner product symbols and achieve an economy of thought that would reduce the chance of error. Then at the end you would get the form of the answer and implement the various inner products by actual summations and integrals.

In addition, suppose you tried to detect a signal in the presence of noise. What you do is take the signal and expand it and do the formal calculations, and say, "Here is my estimator." Once you know the answer (the estimator), then you can just verify that it has properties the answer is supposed to have under these conditions. So, the method of doing the calculations, the hard math of actually computing the answer, is easier if you have an easy way (soft math) to verify that it is the right answer. So, that is why I think that RKHS is a very beautiful thing. More importantly, it led to the concept of regression analysis of continuous parameter time series which turns out to help solve a very central regression design problem: Given a parametric family of functions on the interval zero to one, what is the minimum number of points that you need to observe of the function in order to compute estimators of the parameters that are almost as efficient as if you had observed the whole function? Well, this problem turns out to be equivalent to the problem of computational complexity. How do you compute an integral with the fewest computations that in fact give you practically the right answer? This was solved by Sacks and Ylvisaker, who used the framework of reproducing kernels to do it. Randy Eubank applied these seminal results in his 1979 thesis on representing a sample by a few order statistics.

In the 1950s abstract reproducing kernels started. I started it in statistics, and the numerical analysts started it in numerical analysis. Statisticians should know about this history and the many applications.

Newton: How did you know about reproducing kernel Hilbert spaces?

Parzen: The mathematics department at Berkeley didn't have the usual Ph.D. comprehensive exams; instead, they had the equivalent of a minor thesis. You were handed a few words and you would have to find out mathematically what those words represented. You lectured on that topic to your committee. The committee then asked questions on anything they wanted to, but it began with a presentation of a lecture on the minor thesis that you were assigned—that was intended to be chosen in a field that wasn't your main field of interest. Many students took years to fulfill this requirement. The abstract concept of reproducing kernels was introduced in France in the early 1940s by Aronszajn, who published his first comprehensive paper in English in 1950. Around that time I was given my minor thesis topic, which just said "reproducing kernels." My first problem was to find out what this meant. I told my fellow graduate students, "I have to find out what reproducing kernels are and I can't seem to find any literature on it." One of my fellow graduate students was married to a French woman and he found a reference to the first Aronszajn paper in French. Now the fascinating thing about reproducing kernels is that they arose out of a thesis by Stephen Bergmann, who was said to speak all the languages of the world equally badly. He came to Berlin in 1920 and took a course in several complex variables, and when he finished writing up what he thought was his version of the lecture notes, it turned out to be his thesis. From 1920 to 1940, the idea of the kernel function as a way of studying the various problems of differential equations and several complex variables developed among applied mathematicians. But, instead of having a general theory, they would say, "Well, we are going to follow the steps we had in our previous paper that introduced the kernel with the following property and so on and so forth." Aronszajn had the inspiration that we could take all these ideas and by abstracting them prove general versions of properties found in individual cases. So he developed the abstract Hilbert space theory of reproducing kernels to essentially unify 20 years of a certain school of applied mathematicians on the kernel function method of doing applied mathematics. That was 1940 and led to his 1950 comprehensive paper. So I studied this for the sake of passing my comprehensive exam, but I studied this whole literature and learned all about kernel functions.

I regard myself as being very lucky to have learned about kernel functions for my comprehensive exam. Then in the year 1952, the great mathematician Solomon Bochner decided to spend a year slumming among statisticians and I ate lunch with him every day and took his course on Fourier analysis and Bochner's lemma, which I used in my spectral density estimation paper and my theoretical work on signal processing. For those interested in the details, a collection of 20 of my papers from 1956 to 1964 is published in my 1967 book *Time Series Analysis Papers*.

I think the main thing about my books and my papers is that they do not just present isolated facts; they try to be a comprehensive presentation. If people need facts, they are all there. In the reviews of my book on stochastic processes, for example, very applied people would say that it has a wonderful presentation of Markov chains, but it is really too mathematical for an applied person. But when they take an example of data in which they are interested and go through the material with their example, it all works and makes sense. I think that it is beautiful that people who have examples can learn from the example point of view, and report, "Oh this is very pure, but it really works."

Newton: You mentioned your probability density estimation work. How is that related to your spectral density estimation work?

Parzen: I was aware that people should study nonparametrically the idea of probability density estimation and that was my goal in my research. I knew that I should have an application of the problem of estimating a probability density, so I chose the estimation of the mode. This explains the title of my 1962 *Annals of Mathematical Statistics* paper (Parzen, 1962). Essentially, I took all the knowledge that I had from my previous experience of probability limit theorems and maximum likelihood and deriving the mean square error properties of the spectral density estimator, and combined it and wrote a paper about the mathematical theory of probability density estimates. Other statisticians such as Akaike, Rosenblatt, Bartlett (the British say "Well, Bartlett really had these ideas"), Fix and Hodges, Watson and Leadbetter had certainly done pioneering work in this area. Nevertheless, in spite of the fact that many people deserved credit, engineers who began applying these estimators often used the name "Parzen estimators" and quoted my 1962 paper as their statistical reference because it was a comprehensive presentation of the theory and applicability of probability density estimators.

Let me talk about a very important related area in theory and application: nonparametric regression which estimates the conditional expectation of Y given X. I believe this might be the right answer to the wrong problem from the point of view of practical applications. Often it is more relevant to estimate the conditional quantile function of Y given X and this is the emphasis of my current research.

Newton: How applicable do you think kernel density estimation and nonparametric regression have become?

Parzen: How applicable probability density estimation is has long been a topic of discussion. People have claimed it's not an objective theory that can be routinely used by applied people. Now I have two reactions to that: First, it's good that applied people should realize that they need to consult statisticians. It is not really true that even introductory statistical methods are applicable without consulting a statistician. Second, having gone to many workshops on probability density estimation and studying its vast literature, I feel that researchers in the field do not have a consensus on how to do it. I regard this as part of the problem of the enormous gap between what is known by the discipline of statistics and what is applied in the profession of statistics. I think goodness-of-fit testing is another example. When we list open problems at the frontier of research, we need to add the problem of filling this gap—to enable methods that we theorists know have merit to be more widely and easily used by applied people.

YEARS AT BUFFALO

Newton: In 1970, you moved to the Statistics Department at State University of New York at Buffalo, where you stayed until 1978. How did that happen?

Parzen: I spent 1969–70 in New York City on a Fellowship from IBM. At that time, Marvin Zelen, who had been my close friend since 1958, had moved to Buffalo. I felt that perhaps I wanted to move my family to an Eastern lifestyle. Marvin was a very compatible colleague so again I didn't apply but was recruited to take a job at Buffalo.

Newton: I joined the department in 1971 as a graduate student and remember the environment as being very exciting but tense.

Parzen: We, in a certain sense, encountered all of the problems of developing a statistics department. At Buffalo I developed my mantra that I advise everybody who is trying to put together a statistics department to learn—namely, that you have to find good people in the university who are willing to sing: "We do better research and our students get better jobs because the statistics department is here." I think we were very successful at that in my time at SUNY Buffalo. Unfortunately, although some of our statistics faculty recognized that statistics is an evolving field and wanted to move on to new emerging frontiers, others (in my opinion) wanted to teach only what they learned in graduate school and required a lot of work from students that didn't really have much intellectual payoff. I felt that we should liberate the students by giving them the freedom, for example, of taking one of their classes pass/fail to concentrate on other courses that interested them more. I would rather see a student get three A's and a pass than four B's.

Thus, we confronted all the issues that statistics departments are still confronting in designing their curriculum. The problem in graduate school was always phrased as breadth versus depth. Students would study certain traditionally required courses of the statistics curriculum because they needed it for breadth, but I felt that breadth was interfering with depth.

I also took the attitude from my engineering friends that the statistics curriculum should be such that students should be able to start research after, say, two years of study. Under further consideration, we developed the concept of statistical science which was based on my contact with Neyman. Neyman basically felt that there were mathematical statisticians and there were applied statisticians, a word used by him to describe people doing routine statistics. What we needed was a third kind of statistician who had an appreciation of mathematical statistics, but wanted to work on real problems and wouldn't just use the methods of routine applied statistics. The debate is whether such people can be trained or do they train themselves. I felt that we could train such people in the concept of statistical science.

The phrase "statistical science" is widely used nowadays, but we actually had a precise definition of it. A statistical scientist was a statistician who wanted to do evolving science but was also trained in what I will call "core." "Core" is the discipline of statistics, as opposed to "outreach," which is the profession of statistics. A statistical scientist tries to blend core and outreach, and go beyond the usual statistical methods used by the applied statistician to evolve new methods, thus adding to development of the core at the same time as they solve applied problems.

Anyhow, the SUNY Buffalo experiment had a problem that unfortunately started the year I came—they really did not have funds to support their vision. We put in a proposal to the university to develop statistics that was rated as the number one proposal by the university administration because it combined attainability with great goals. In other words, many departments would say, "give us two million dollars and we will do something," but rather than saying "give us a lot of money and we'll spend it," we said we can incrementally achieve great things with this plan. Unfortunately, none of these plans were implemented. As a result, we had a situation where some faculty had a lot of funding, and some faculty didn't. The faculty who didn't wanted a tax so that those with money would benefit those without money. Obviously, the people with money didn't like that idea. Anyhow, those were the kinds of tensions, but we felt we had a strong program.

Newton: How did you leave Buffalo?

Parzen: In 1977, Marvin Zelen was attracted to go to the Dana Farber Cancer Institute and the Harvard School of Public Health. We had developed a biostatistical laboratory at Buffalo as a separate activity of the faculty of statistical science, in order to be able to accomplish this goal of building a strong research program. This group was so strong they were recruited by Harvard, and then in 1977 you and I were approached by Texas A&M to help strengthen the statistics program there. The issues that led to our exodus from Buffalo were the day-to-day issues indicated above. In addition, although SUNY Buffalo had claimed that it wanted to become the Berkeley of the East, it in fact didn't have any money, and ultimately took the attitude that it had overinvested in statistics and didn't want to support statistics. We felt that there was not much point in trying to work in a system that didn't seem to want to support statistics.

Newton: So it's kind of sad that there are no longer statistics degrees awarded at SUNY Buffalo. Do you think that ...?

Parzen: I would sing my mantra. I would say that there is a pattern of statistics programs trying to be eliminated by vice presidents who feel that the university has overinvested in statistics. The only way for that not to happen is for there to be a constituency that supports a department of statistics. In every university, there are people not in statistics departments who are happy to teach statistics but don't support the concept of a separate department of statistics because they don't want to have to utilize the services of that department. The only way that they will utilize it is to convince them to sing the song: "We do better research and our students get better jobs, because statistics is here."

But, I think that you can't get people to sing that song if the statistics department is always at risk. There are a lot of people who want "to eat their lunch." And a vice president or dean who decides that he can make other people happy with the resources he could obtain from statistics will do so. It has happened at many universities. So, my advice to every statistics program is to develop a constituency. Unless you do that, you won't have a statistics department.

MORE ABOUT TIME SERIES

Newton: Let's return now to time series analysis. What are your recollections about the development of the so-called Box–Jenkins method?

Parzen: Well, very fond memories. In order to explain the development of the Box–Jenkins method I have to go back into my own history. I spent the summer of 1958 in England on my way to the International Congress of Mathematicians at Edinburgh (where I networked with Maurice Priestley, Maurice Bartlett and C. R. Rao). Deciding that summer to become an expert in statistical time series, I networked in London with the great Maurice Kendall, who recommended that I interact with Gwilym Jenkins. I went to his office at Imperial College and said, "I'm Manny Parzen; you're Gwilym Jenkins. Let's talk." I invited him to be a visiting faculty member at Stanford in 1959–1960. On his way to Stanford in 1959, Jenkins visited John Tukey in Princeton, and on his way home in 1960 he visited George Box in Wisconsin. At Wisconsin, he and Box began the world-renowned Box–Jenkins approach to time series analysis, first published as a paper in 1962 and exposited in a book in 1970. With support from the Office of Naval Research, I spent 1961–62 at Imperial College, London, visiting Jenkins. George Box told me around that time that he had saved Gwilym Jenkins for statistics; after his year at Stanford, Gwilym was very discouraged and said that statistics was too hard the way they practice it at Stanford.

There is no question that Box–Jenkins is a very valuable methodology; it arose out of practical problems. Box was interested in problems of process control. In those days a control process was analogous to knobs on a machine that has to be adjusted to differentiate, integrate and be in proportion. Jenkins began teaching Box how to justify this: that there was an underlying model, the ARIMA model, that this method of control implied. The ARIMA model is the model that explains why the methods of what they called proportional control, differential control and integral control worked. From process control, Box and Jenkins moved into the forecasting context. I had close interaction with Box and Jenkins throughout this period, discussed their work at various meetings and always tried to put it in the broader context of autoregressive modeling, but Box refused to incorporate anything about criteria for model selection, either my CAT or Akaike's AIC. He took the attitude that automatic model selection techniques were too dangerous and would be misused. So, he was very happy with Box–Jenkins model building and didn't integrate model selection criteria into their methodology.

I should mention here that Peter Whittle was always upset that Box and Jenkins did not refer to his work or acknowledge that he had developed the theoretical

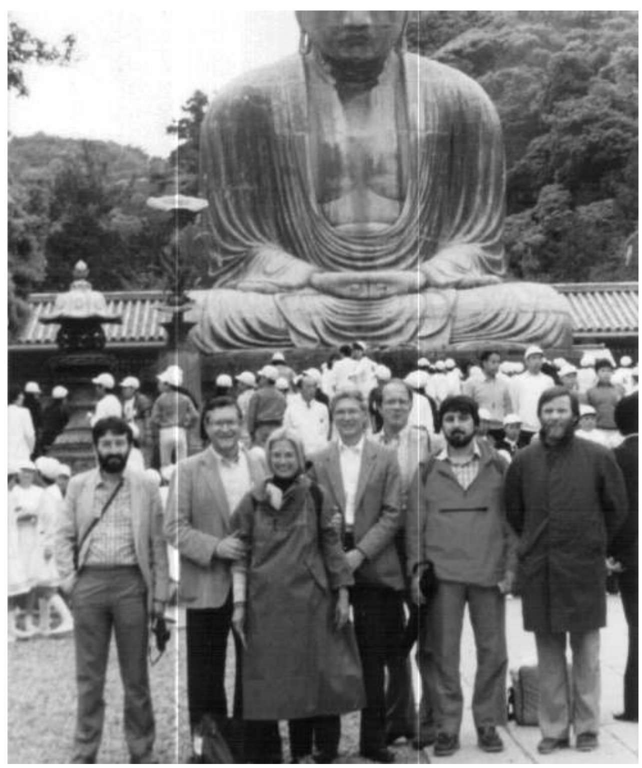

FIG. 3. *U.S. Time Series Analysts and Smoothers, Tokyo, 1984: Victor Solo, Emanuel Parzen, Grace Wahba, Wayne Fuller, Bob Shumway, Bill Cleveland, David Brillinger.*

FIG. 4. *Statistical Time Series Analysis U.S.–Japan Joint Seminar, Hawaii January 25–29, 1993*: (*first row*) *David Brillinger*, *Manny Parzen*, *Ruey Tsay*, *George Tiao*, *Joe Newton*; (*second row*) *unknown*, *David Stoffer*, *Bob Shumway*, *Will Gersch*, *Roger Brockwell*, *Clive Granger*.

framework for the ARIMA models in his influential book *Prediction and Regulation*.

Box and Jenkins wrote a theory of time series analysis which became a success. But you have to recognize that they are role models of how to spread knowledge, particularly Jenkins. Box would tell me that they would have these short courses and people would bring their data. Jenkins would sit at a computer terminal all day modeling their data, showing them how to model it. He was really the practitioner par excellence.

Newton: You mentioned model selection methods, which in the ARIMA context means selecting the order of the ARIMA model. Can you talk a little about that, and in particular about your relationship with Akaike?

Parzen: Well that is another interesting story. I was the first person to invite Gwilym Jenkins to the United States and of course he then met Box and Tukey. And I was the first person to invite Akaike to the United States. Akaike and I met and became lifelong friends in 1965 in Japan at a seminal U.S.–Japan Joint Seminar on "Applications of Stochastic Processes to Engineering" whose story is told in "A conversation with Hirotugu Akaike" by David Findley and myself (Findley and Parzen, 1995). We organized two U.S.–Japan joint seminars on "Statistical Time Series Analysis" (Tokyo 1984, Hawaii 1993).

In 1972, Akaike pioneered the Akaike information criterion (AIC), which is an information criterion for statistical model identification. I was stimulated to develop a criterion for autoregressive time series model identification which I called CAT, an abbreviation for criterion autoregressive (AR) transfer function; its idea was to minimize the integrated mean square error of approximation of an infinite order AR by a finite order AR. My luck was being able to express both bias and variance in terms of the innovation variances of successive order AR models. Although derived from

FIG. 5. *Hirotugu Akaike*, *Manny Parzen*, *Maurice Priestley*.

different philosophies, the graphs of AIC and CAT in ARIMA modeling practice have similar shapes and choose the same estimated order. My 1974 paper describing CAT, "Some Recent Advances in Time Series Modeling," appears in the same engineering journal as Akaike's highly cited paper (Parzen, 1974; Akaike, 1974). My 1974 paper also describes the theory of inverse correlations (Fourier transform of the inverse of the spectral density, introduced by Bill Cleveland, whose investigation inspired my study of orthogonal polynomials on the unit circle that enabled me to derive the formula for CAT).

Newton: In the time you were at Buffalo you also did some work on multiple time series modeling. Can you describe some of that please?

Parzen: The paper I remember the most as being a comprehensive paper was a follow-up to the work I did on multiple time series analysis in 1966 in which I announced to the members of the multivariate analysis statistical fraternity that I had reduced the problem of multiple time series analysis to the problem of multivariate analysis. You tell me how you would analyze multivariate data, and I would tell you how I would analyze multiple time series. They of course said that they didn't know what I meant. If you compute canonical correlations or any other multivariate statistic, they can tell you its distribution, but they weren't familiar with the idea of model building for multivariate data. Basically, that is what we were trying to do—develop a strategy for modeling multiple time series. Parzen (1969) was where I first discussed various autoregressive approaches and sketched out the theory of autoregressive spectral density estimation. The sketch was filled in and completed by a brilliant Ph.D. student of mine at Stanford, Ralph Kromer, who unfortunately never published his comprehensive thesis.

When we were turning to the problem of multiple time series analysis, one of my first goals was to obtain the analog of the univariate CAT criterion for multiple time series, and I was also motivated in joint work with you to develop the idea of periodically correlated time series, which is another approach to modeling multiple time series by somehow treating them as a series of linked univariate time series. This research was motivated by the application of modeling atmospheric ozone, which was a controversial topic.

At that time you and I also began the idea of developing a user-friendly timeseries program called TIMESLAB and multiple TIMESLAB. In my opinion, multiple time series is still not as automatic a topic as univariate time series. My own view of the matter is that you should take an information approach. If you want to model the time series $Y(t)$ in terms of the related explanatory time series $X(t)$, you should compute the information about $Y(t)$ in various subsets of the past of Y, the past of Y and X, the entire series X and the past of Y or other combinations of values of Y and X, and determine the value of adding or deleting these subsets to the modeling of Y. I feel that these are important topics but they don't yet lead to as comprehensive a theory of multiple time series as I think we have for univariate time series.

TIME AT TEXAS A&M

Newton: In 1978, you moved to Texas A&M as a Distinguished Professor and continued some of your work in time series as well as work in quantile function theory. Let me start by asking you what it was like to work with H. O. Hartley, the founder of the department at Texas A&M?

Parzen: Well, I think Hartley was a truly great man; I felt he was a very lonely man, a victim of the European idea of being the only leading professor. He should have continued to be a professor but he retired because he didn't want to be just a professor but also the head of a department (which was not possible at age 65 then at Texas A&M).

I had many wonderful conversations with Hartley. I remember that, in my first few years at Texas A&M, he called me into his office the day he got a letter from Henry Daniels saying he (Daniels) had just discovered that Herman O. Hartley was the same person as the author of "A Foundation Paper of Correspondence Analysis," which was published in England in 1935. I don't believe the fact that Hartley created the foundational ideas of correspondence analysis is as widely known as his other research. Did you know about his correspondence analysis work?

Newton: Yes, I'm trying to remember what his name was in his youth.

Parzen: It was Hirschfield, which was changed to Hartley.

Newton: Continuing with your work on time series analysis, you became interested in forecasting nonstationary time series and participated in a very famous forecasting competition.

Parzen: We were developing our own methods as alternatives to the Box–Jenkins methodology. The difference between our methods and Box–Jenkins methods is that (in my opinion) Box–Jenkins was a closed system. If you tried Box–Jenkins methods and

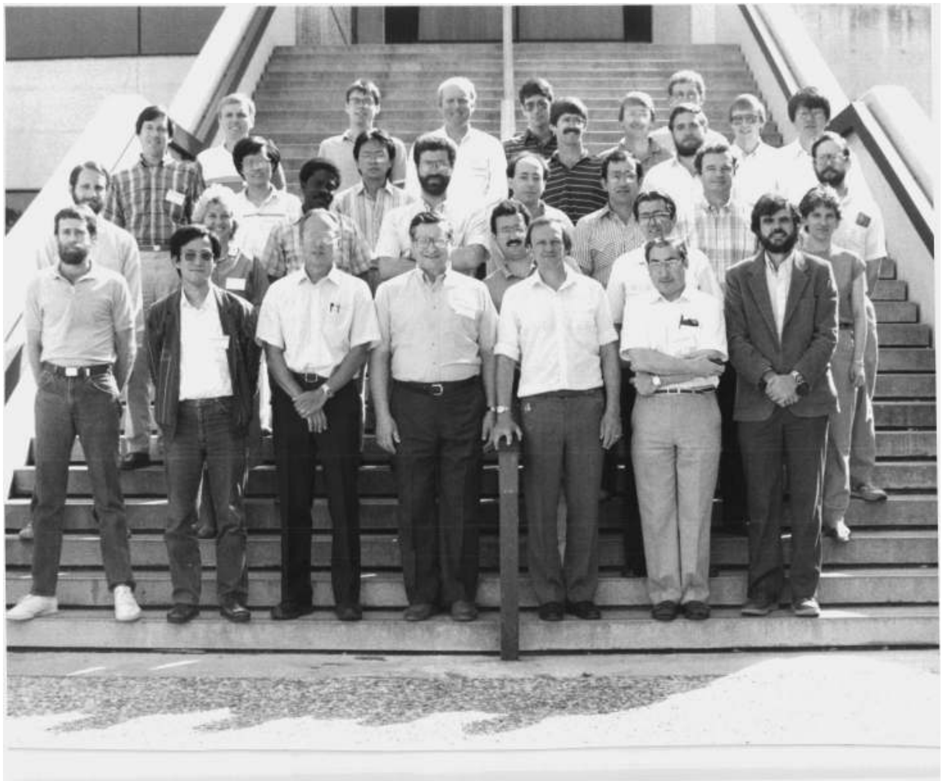

FIG. 6. *Time Series Analysis and Signal Processing Cutting Edge Advanced Seminar, held in Austin and College Station, Texas, May 23–June 4, 1986: T-shirt presented to Professor Parzen proclaims "CATs, Quantiles, Windows, and RKHS. No place but Texas" and "Long Memories of the Cutting Edge of Time Series"; (first row) Randy Eubank, Ritei Shibata, Richard Jones, Manny Parzen, Craig Ainsley, Maurice Priestley and Joe Newton.*

they didn't seem to work, there wasn't any suggestion as to how to go on from there. We were interested in developing a method of time series analysis that was open ended and didn't run out of ideas.

Forecasting has an intellectual community of its own with its own meetings and publications and, of course, its own annual conferences attended by people who are in the forecasting business. I once asked a leading forecaster, "What do you promise the client?" He said, "Oh, I never promise the client anything. All I promise them is that if they hire me, they will do 50% better than they are doing now."

In order to somehow compare various forecasting schemes, a group at a great business school in France led by Makridakis created a database of economic time series, some quarterly, some annual and some monthly, and asked people to forecast 18 time periods ahead. You and I jointly entered this competition and our results were a winner. We had total mean square error over 18 periods that was less than any other. The method that we imposed is similar to ARIMA but in ARIMA you take differences as the first step. Instead, we automatically regressed the data on one lag, not necessarily the most recent lag, and allowed that one lag AR to be nonstationary. Then we took the result of that and fit a stationary ARMA, where the ARMA order was chosen by AIC. So we called AR for the first step and ARMA for the second step, thus giving something we called ARARMA.

I should emphasize that prediction is regarded as a time series analysis problem and is a very controversial subject. For example, if you want to forecast 18 steps ahead, should you use the method that is best for one step ahead and then the one that is best for two steps ahead and so on, or should you just build a model and use it to forecast one through 18 steps ahead?

We essentially built our forecast 18 steps ahead by building a model. One thing we learned was how important the idea of memory length is. If you identify a series as short memory, you can identify a model (fit an ARMA) by automatic model identification techniques. But if it's not stationary (of a kind we call short memory, which essentially means that the spectral range is more or less finite), one must transform nonstationary to stationary in a nonunique way. For example, if you had hourly data of electricity demand, there might be a daily, monthly or weekly period, so you could take perhaps a 168- or 24-hour difference. My experience is you want to do the longest lag difference first and leave more latitude, but in general what you want to do is the simplest transformation of the data that will make it just barely a short memory time series so that you can then model this by automatic modeling techniques. As an exercise in the transformation of a nonstationary to a stationary series, we modeled municipal water use demand in Texas and we regressed water use on population and found that this was a transformation whose residuals were a stationary time series.

So my philosophy is that anything you can do to transform your original observations to a residual which is just barely stationary is the way to handle the problem of nonstationary time series, and that is what we did in the ARARMA method where the initial transformation to stationarity was a one-lag AR whose coefficient could be greater than 1.

I should point out that we didn't really push our method because we had no desire to join the forecasting community, because forecasting is a full-time profession. Basically you have to become a salesman for your technique. I have gone to forecasting meetings and every talk seems to present a problem and a discussion of how well the speaker's method did. If you ask them, "Well, did you try this other method?" They say, "No, I want to use my method. It worked very well so why do I need to try any other method?"

QUANTILE METHODS

Newton: Since the late 1970s, you have championed the idea that methods based on quantile functions can be a more useful way to study random variables than those based on their density or mass functions. Why do you feel that this is true?

Parzen: Well, as usual there is a story of how I was led to being interested in something. I was lecturing on the total time on test statistic at Buffalo. The total time on test statistic is used to test that observations are exponential based on the property of the exponential distribution that if you observe a process at an arbitrary time and measure how long until it dies, which is called its residual life, then the residual life distribution is also exponential. Thus the residual life distribution is the same as that of the observations.

So it was an example of testing whether a distribution function (cdf) F equals a distribution function G, and it turns out that a superior way of testing the equality of two distributions or two probabilities is not to take their difference but to take their ratio, which in the case of distribution functions is $G(F^{-1})$, which would be uniform if $F = G$. I was led to rephrasing the problem of testing the identity of two distributions by asking whether the comparison distribution function $G(F^{-1})$ was equal to the identity distribution function.

To study the comparison distribution function, it is natural to study the quantile function F^{-1}, so called because, for any number p in the interval $[0, 1]$, F^{-1} evaluated at p gives the pth quantile of the distribution defined by F. Now, the derivative of the quantile function, called the sparsity function by Tukey, is nonnegative, and this caused me to draw upon my background in time series analysis. I have always felt that time series are the underworld of probability because the probability density is a nonnegative function, the spectral density is a nonnegative function and so there exist mathematical analogies between probability based on density functions and time series based on spectral density functions.

Now we had another nonnegative function—the derivative of the quantile function, which instead of being called the sparsity function I called the quantile density function. It turns out that, just as the raw spectral density is exponentially distributed, the raw sample quantiles (i.e., the spacings between order statis-

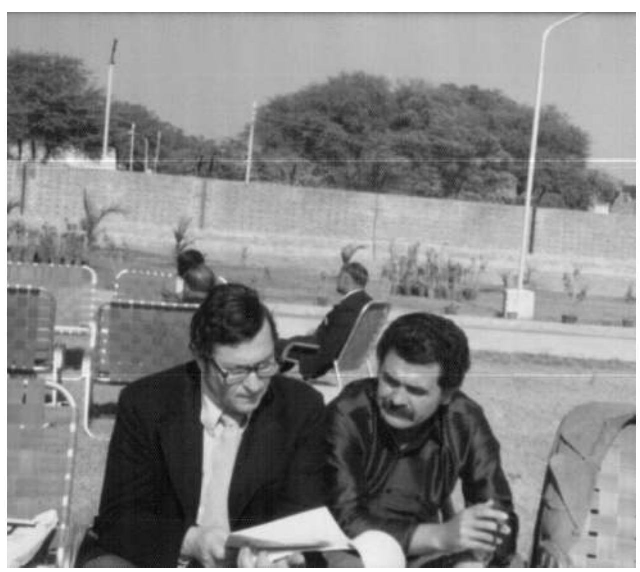

FIG. 7. *Manny Parzen and Miklos Csörgő, 1979 International Statistical Institute meeting in New Delhi, discussing applications to quantile data analysis of strong approximation limit theorems for quantile processes.*

FIG. 8. *Ph.D. students and colleagues of Emanuel Parzen celebrating his 60th birthday in 1989 in College Station, Texas: Don Ylvisaker, Grace Wahba, Joe Newton, Marcello Pagano, Randy Eubank, Manny Parzen, Will Alexander, Marvin Zelen, Scott Grimshaw. Recent distinguished Ph.D. students include Cheng Cheng and Todd Ogden.*

tics) are also exponentially distributed. So the mathematical properties of the spectral density are equivalent to those of the sparsity function. Not only that, but the reciprocal of the sparsity function (which I call the density quantile function) arises naturally as the probability density function evaluated at the quantile function; also the reciprocal of the spectral density is the inverse spectral density. It thus turns out that an idea which seems purely mathematically motivated is natural because there is a duality of functions: the sparsity function and the density quantile function are analogous to the spectral density and the inverse spectral density. Furthermore, the mathematical properties of the natural sample estimator of the function $G(F^{-1})$ under the null hypothesis that $F = G$ are those of a Brownian bridge. So you have an analogy between testing for white noise in the time series case and testing homogeneity of populations in the distribution function case.

Well, now you get the quantile function and you realize that the quantile function was being made popular by John Tukey and his five-number summary, and not only that but when you start writing down formulas for these functions, the quantile density, the density quantile and so on, they are all very beautiful formulas. It is a set of facts that statisticians neglected. I complained to the leaders of the field such as Peter Bickel, my peers, and said "Why aren't people celebrating these facts?" There are wonderful properties of the quantile function and the various functions derived from it, especially $G(F^{-1})$, which I called the comparison distribution function, used for comparing two continuous distributions.

To compare two discrete distributions I found a natural way. This led to what I consider a breakthrough idea that says if you are comparing two distributions which are initially continuous, you want to do that by computing the comparison distribution. And the comparison distribution for samples (i.e., the estimator of the population comparison distribution) is the comparison distribution of the two discrete distributions which are the sample distributions of the data.

This insight led to a very beautiful theory which extends the theory of nonparametric statistics. You have the development of the Wilcoxon statistic and then the formalization of studying its properties. The work of Pyke and Shorack said we can actually derive the results of the Chernoff–Savage theorem from a general theorem about convergence in distribution of a sequence of stochastic processes. But they used this just for theorem proving; they didn't think of using this stochastic process as something you would actually look at as a statistic to study and actually use for data analysis. It turned out that the stochastic process that they were defining was a very ugly process. It didn't really lend itself to a truly computable representation of linear rank statistics as a function of the process.

So I studied the various ways in which you could get more applicable data analysis versions of that stochastic process and in my opinion the best version was a continuous process: namely, the comparison distribution of the sample distributions.

Newton: Are you still working in this area?

Parzen: Yes, because I feel we can unify many of the standard results of classical nonparametric and parametric statistics by expressing them in terms of quantile functions and mid–distribution functions. Any formula you have is worth looking at in the quantile domain. In fact, this has led me to a definition of what an average is. How do you compute the mean of a set of numbers? You don't just add them up and divide by n. First, you arrange them in increasing order, and then add them up. So I say statistics is arithmetic done by sorting before adding. Any time you have a statistic that is a sum, the question is what is the additional information that comes from taking the things you are summing and forming their quantile function. For example, the mean of any data set is the integral of its quantile function.

This is another example of what I call the gap between what is known and what is applied. Many students have written beautiful theses in the quantile domain at Texas A&M to fill this particular gap, not just my students but, for example, the students of Wehrly and Longnecker.

People often accuse me of saying that you should do everything in the quantile domain. No, not at all. Obviously you should use ordinary methods based on distribution functions in some circumstances. I will use an analogy to time series. In time series, we have a choice of domains—the time domain and the frequency domain—and the reason that is good is that we don't know in advance which is the most parsimonious way to look at data. The same thing is true of statistical methods; we have a choice of the distribution domain and the quantile domain and that is good because we don't know in which domain nature will be most parsimonious. For example, quantiles are much simpler for handling censoring. This example again illustrates that there are many beautiful things you can do in the quantile domain. For example, a neglected fact is that a confidence interval for a quantile function can be computed without doing plug-in estimation of the asymptotic variance of the estimated quantile function.

Of course, this leads me to a goal that I really want to accomplish in the next few years: to sum up this work which represents the work I have been doing since I got to Texas A&M almost 25 years ago, sum it up in a book called *Statistical Methods Learning and Quantiles*. I don't quite understand why it is so hard for quantile data analysis to catch on.

I am optimistic that this will be expedited by my latest research on quantile–quartile $Q/Q(u)$ plots (to make box plots easier to interpret and to help identify distributions fitting data), estimation of conditional quantiles (by smoothly estimating quantile functions of middistribution transform of data) and comparison distributions [graphical statistics and limit theorems for their empirical processes, unification of nonparametric and parametric methods for con/crete (continuous and discrete) data].

In addition to a textbook *Statistical Methods Learning and Quantiles*, another book project I would like to accomplish is a textbook *Time Series Analysis Methods and Exponential Models*.

I'll give an example of a success of quantile data analysis. I looked at the original data that was published as the first data set of statistical data analysis by John Graunt. It was length of life and mortality rates in London, and I asked myself, "let me just see if these data are exponential." To my shock, it was exactly exponential. If you compute the comparison distribution and various statistics and plot them, you get a straight line. So I said to Steve Stiegler, "Do you know that John Graunt's data is exactly exponential?" He said, "Yes, that's how he created it." Graunt took the facts that 36% of the people died by 16, and some people lived to age 80, and he filled in a table on the number of deaths from 0–16, 16–26, 26–36 and so on, essentially using a rule that said of those alive in a period, $5/8$ are alive by the next period. It comes down to an exponential distribution. This is an example of how a quantile data analysis disclosed at a glance a fact that I don't think that most people in statistics are aware of—that the original data published by John Graunt were not real data.

CONFERENCES AND WORKSHOPS

Newton: Throughout your career, you have organized a large number of influential conferences and workshops. What are some of your recollections about these meetings and who were some of the more interesting people with whom you interacted?

Parzen: First, let me remark that I find it interesting that in the SIAM series of proceedings of the regional conferences, there are a conference on distribution functions, principal lecturer Jim Durbin, and a conference on quantile functions, principal lecturer Miklós Csörgő, and I happen to have been organizer of both. I think that it is an interesting coincidence that the two major research monographs on distribution functions and quantile functions owe their existence to my initiative.

Obviously, I hope all the conferences that I have been involved with have been very significant events in the lives of those who attended them. In addition to those two conferences on quantiles and distributions, in 1983 we had a "Symposium on Time Series Analysis of Irregularly Observed Data" that included all of the leading workers around the world. But again I can only say it was a very stimulating event for those there, and (of course) that since then the problem of irregularly observed time series has been an increasingly important one, so that the proceedings of the conference have been often cited.

We also had the idea for what we called the Time Series Analysis Bootcamp. We brought together young researchers for two weeks and had lectures on time series analysis and signal processing. Of course, I think those ideas are worth continuing; it just requires a tradition. You have to recognize in all honesty that organizing conferences is tough because people always want to know, "Why should you do it?" When we proposed an idea to CATS (Committee on Applied and Theoretical Statistics), some people on the committee said, "We understand the leading people in time series are Box and Jenkins, so why aren't they doing this?"

Newton: Gee, thanks!

Parzen: It didn't bother me, I was willing to do it because I think we will cast a much wider net. People will get a much deeper, broader education. That is what we are proposing; we are not just proposing Box–Jenkins. A very great event (in 1990 at the Institute for Mathematics and Its Applications) was a one-month workshop "New Directions in Time Series Analysis," an interdisciplinary workshop. The people who were there felt that they made many interesting connections. I think workshops advance the field and are worth doing.

We had a conference here at Texas A&M joining together economists and statisticians in 1996. In 1992, behind the scenes as a member of the Committee of Research Conferences in the Mathematical Sciences, I organized conferences on change-point analysis and on wavelets and their applications in statistics. So I have been again, *Mazel*, in the right place at the right time when the opportunities arose to do something that led to a workshop.

There are two kinds of statisticians: those who make their contributions in terms of research and those who make their contribution in terms of organizational ability. Some people imply that I am the kind of academic who contributes through organizational ability because I have organized many workshops. I have always thought that this view of my career is amusing. No, I want to make it by research contributions, but along the way I am happy to do organizational contributions as well.

THE PARZEN PRIZE

Newton: Another interesting thing that you did was in 1994: you and your wife Carol organized the Emanuel and Carol Parzen Fund and Prize for Statistical Innovation. Why did you do this, and has the prize had the effect that you had hoped it would?

Parzen: Let me preface my answer by giving my attitude on the future of statistics. We cannot plan or forecast the changing future of the discipline (core) and profession (outreach) of statistics. We should plan to convince students and young faculty to pursue a career in statistical science because one can do exciting and rewarding research as a statistician. For this we need to teach vision (the big picture of frontiers of research to know where to apply their talent) and technical power (expertise in mathematical, computational and applicable statistical theory). To attract the talent it deserves for its long-term health as a distinct discipline, statistical science must appear exciting and rewarding, and must pay attention to and applaud the achievements of professional statisticians.

To applaud and to promote the dissemination of statistical achievements, the Parzen Prize for Statistical Innovation is awarded (around April of even-numbered years) to North American statisticians who have made outstanding and influential contributions to the development of applicable and innovative statistical methods. The prize has been established to relieve the sparsity of prestigious awards and prizes that recognize outstanding careers in the discipline and profession of statistics. From 1994 to 2002 we have given five prizes: Grace Wahba, Donald B. Rubin, Bradley Efron, C. R. Rao and David Brillinger; all of whom of course are extremely highly recognized. I'd like to point out that Grace Wahba has been elected to the National Academy of Sciences.

In my view there is a lot of talent in statistics, and we have a very large number of outstanding people. The Parzen Prize provides a mechanism for helping

FIG. 9. *Jim Calvin, Joe Newton, Brad Efron, Manny Parzen, Carol Parzen, Michael Parzen at the ceremony awarding the 1998 Parzen Prize to Efron.*

us to appreciate the fact that we do have outstanding talent in statistics and statistical science. I feel the prize has worked very well. The prize winners we have had have all been very interested in receiving the prize, have established the credibility and the value of the prize, and I think we will continue to try to recognize achievement. I think we need to applaud people and this is essential for maintaining the health of statistics and to make people realize that it is a career that is exciting and rewarding.

Personally, I think statisticians do a very bad job of awarding prizes. I am involved in the American Statistical Association and have been awarded the Wilks Award. People who are getting the Wilks Awards are outstanding people and I think there should be much more attention, much more glitter to the awarding of these prizes. There is hardly any PR associated with them, even though this honor is one of the most prestigious prizes of the statistical profession. Let's face it, we don't get the newspaper attention that we should for the prizes that we give out. More prizes are being awarded and they aren't getting the publicity they deserve. I think that publicity for outstanding contributions in statistics is very important for the future health of the profession.

THE FUTURE OF STATISTICS

Newton: You have talked about what the future might be and we all know that the availability of massive computing power, and huge datasets, and data mining and things like that are having a big impact on statistics, but what do you think is the rightful role of statistics in science?

Parzen: Number one, statistics right now has been a small sample subject. The engineers are always telling me, "We don't need statistical inference." They have very large samples so their problems are problems of probability. What are large datasets? That is probability; that's a large sample. I think statistics is a changing field. There is financial mathematics, there's genomics, so a lot of statisticians are paying attention to other fields. What is it that historically links statistics with science? The problem of statistics, and the rise and fall of statistics departments, is that everybody is doing statistics and wants to do it, so why do you need professional statisticians?

To me, statistics is the process of technology transfer of methods that arise among psychologists or economists or among geophysicists, and so on. The technology transfer of these ideas is the broad field of statistics.

Statistics will obviously be changing, as I mentioned earlier in talking about the development of statistics departments. You can't have a statistics department that says a well-trained statistician is one who had the same graduate training I had 30 or 40 years ago. So we have got to learn how to be a changing field in order to have the optimal professional involvement in society. On the other hand, there has to be a discipline of statistics, for which we have to get support.

I happen to feel that the concept of data mining requires what I call statistical methods learning (unification), that is, knowing what the connection is between various statistical methods. As I mentioned about forecasting, there is a big tendency for people to use one method and say, "Okay, this is good." I have always argued that data is expensive and analysis is cheap. We should be prepared to investigate our data by various methods of analysis. There is no question that statistical computation is very important, and one thing I regret is that I haven't gotten more involved computing, though philosophically I have been involved since 1960. I do think, for many problems, raw computer power and just doing things computationally is very essential. On the other hand, we don't want statistics graduate study and practice to be entirely computational without theory. There is a tendency for people not to want to study theory because they want to emphasize computation. We obviously need to balance computation and theory. We have a very difficult philosophical problem, "How do the principles of small samples in statistics apply when you have large datasets?" I don't think we have thought deeply enough about that.

So, let's search for the fundamentals. One of the things that I have always felt is knowledge is the edge, and statistics is the way to discover knowledge. We should preach to the public: "Knowledge is the edge, and statistics is the way to discover knowledge." But, this means that we have to take the body of knowledge that is statistical theory and somehow digest it and package it so that the gap between what we know and what is being applied is reduced. This is what I call statistical methods learning: reducing the gap between what we know in the theory of statistics and what is being applied. I have a plan for how to write a concise encyclopedic book about statistical methods learning (unification).

There are many people, as in data mining and statistical learning, who are happy to rob statistics of its methods, and make very successful careers, and say they are so happy to be able to do this without the competition of statisticians.

On the other hand, why should a person choose a career in statistical science? Let's face it; I can't give the universal answer. Statistics itself has to give the answer. What can we do to show young people who are considering entering the discipline of statistics that it is a rewarding enough and exciting enough career? I think if you go to the statistics meetings this summer, you will see that people are doing a lot of things that aren't inference. The key is to find the connecting link between all these different activities, and once you learn the connecting link, the question is "why does this connecting link exist?"

Newton: That's a philosophy that many in statistics would agree with. It also seems to be a good way to end our conversation. Thanks very much Manny for sharing your thoughts about your career and about statistical science with us.

Parzen: Thank you, Joe, for your friendship for 30 years.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* **AC-19** 716–723.

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

FINDLEY, D. F. and PARZEN, E. (1995). A conversation with Hirotugu Akaike. *Statist. Sci.* **10** 104–117.

GRENANDER, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195–277.

HIRSCHFIELD, H. O. (1935). A connection between correlation and contingency. *Proc. Cambridge Philos. Soc.* **31** 520.

JENKINS, G. M. (1961). General considerations in the analysis of spectra. *Technometrics* **3** 133–166.

LANNING, J. H. and BATTIN, R. H. (1956). *Random Processes in Automatic Control*. McGraw-Hill, New York.

MAKRIDAKIS, E., NEWTON, H. J., PARZEN, E. et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecasting* **1** 111–153.

NEWTON, H. J. (1988). *TIMESLAB*: *A Time Series Analysis Laboratory*. Wadsworth, Pacific Grove, CA.

OGDEN, T. and PARZEN, E. (1996). Data dependent wavelet thresholding in nonparametric regression with change point applications. *Comput. Statist. Data Anal.* **22** 53–70.

PARZEN, E. (1958). On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. *J. Roy. Statist. Soc. Ser. B* **20** 303–322.

PARZEN, E. (1961a). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

PARZEN, E. (1961b). Mathematical considerations in the estimation of spectra. *Technometrics* **3** 167–190.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

PARZEN, E. (1969). Multiple time series modeling. In *Multivariate Analysis—II* (P. Krishnaiah, ed.) 389–409. Academic Press, New York.

PARZEN, E. (1974). Some recent advances in time series modeling. *IEEE Trans. Automatic Control* **AC-19** 723–730.

PARZEN, E. (1977). Multiple time series: Determining the order of approximating autoregressive schemes. In *Multivariate Analysis—IV* (P. Krishnaiah, ed.) 283–295. North-Holland, Amsterdam.

PARZEN, E. (1979). Nonparametric statistical data modeling (with discussion). *J. Amer. Statist. Assoc.* **74** 105–131.

PARZEN, E. (1982). ARARMA models for time series analysis and forecasting. *J. Forecasting* **1** 67–82.

PARZEN, E. (1992). Comparison change analysis. In *Nonparametric Statistics and Related Topics* (A. K. Saleh, ed.) 3–15. North-Holland, Amsterdam.

PARZEN, E. (1992). Time series, statistics and information. In *New Directions in Time Series Analysis* (D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu, eds.) 265–286. Springer, New York.

RAMSAY, J. O. and SILVERMAN, B. W. (1997). *Functional Data Analysis*. Springer, New York.

SACKS, J. and YLVISAKER, D. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Statist.* **39** 40–69.

TUKEY, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics* **3** 191–219.

BIBLIOGRAPHY

The above list of references is a minimal list to show direct links between research by Parzen and seminal papers by others.