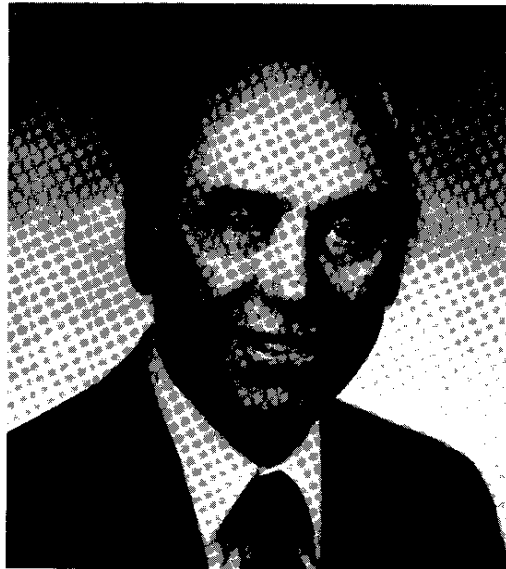


## THE ET INTERVIEW: PROFESSOR T. W. ANDERSON

*interviewed by Peter C. B. Phillips*



Theodore W. Anderson

As young scholars contemplate an academic career of teaching and research they are challenged and inspired by the work of senior scientists who have gone before them. Amongst econometricians, the Cowles Commission researchers of the 1940s occupy a special position of seniority and significance. Their contributions are seen by most of us as the mainspring of decades of subsequent research. They opened up new and promising fields of research, they set new standards of mathematical rigor in their work, and they forged new and productive areas of contact with sister disciplines like mathematical statistics. Ted Anderson was a young and energetic member of the Cowles

Commission research team in 1945–46. His individual and joint contributions were a vital element in the Cowles research endeavor during this productive period. His articles now stand as classic works in the history of econometrics, and they have helped to inspire many subsequent generations of econometricians.

In mathematical statistics, as well as econometrics, Ted Anderson is a scholar of immense stature. The scope and diversity of his research in statistical theory is almost a phenomenon in itself. Sometimes it seems that wherever one turns in the subject, Ted's mark and influence is already firmly established. His books on multivariate analysis and time series rapidly became accepted as major treatises and are now integral parts of the bookshelves of every statistician. His articles, like his books, have long established him as one of the best communicators in the statistics profession. His work is nothing short of exemplary, both in rigor and in exposition, and has been the fountainhead of entire fields of subsequent research.

Those of us who have had the privilege of knowing Ted as a person see him as a gentleman of humility and much personal charm. His students, his colleagues and his friends at other institutions have all been fortunate to have intellectual and personal contact with him. It is hoped that the following interview will bring Ted closer to a much wider audience, so that all of us will now have an opportunity to learn from his thinking about matters of research and teaching, his insights into the past, and the possibilities of the future.

Over the last forty years there has been great depth and diversity to your research, not only in mathematical statistics but also in neighboring fields like econometrics and psychometrics. At the beginning of your academic career in the early 1940s did you have any idea of the future course that your work would take?

I should go back a step to my undergraduate days. My major was in chemistry through the first three years of college, but I had a wide range of interests. I took as few courses as I could in chemistry and as many as I could in other subjects. When I came up to my senior year, I realized I really didn't like chemistry and I decided to change my major. At that time in mathematics I had taken only differential calculus; I had one course in economics, two semesters in psychology, and a variety of courses in other subjects. I realized that I liked mathematics best and could do it well. So I decided to major in mathematics and debated about carrying on my interest in economics, psychology or possibly some other subject. Finally I decided to minor in economics. In my senior year I had to fill all of the requirements in mathematics and all the requirements for a minor in economics. I was kept quite busy. During this last year I took a course in statistics and a course in econometrics. And my interest in statistics and econometrics began to develop. That was

towards the end of the Depression and opportunities for a professional career did not look very promising. At that time students of mathematics would often prepare themselves in actuarial mathematics and I thought about starting to take the actuarial exams. There were very few openings in industry. There were some in government as an alternative to university. But, when I decided to go into mathematics I already expected to do graduate work, hopefully getting a Ph.D., and making a career in teaching, though at what level, I wasn't sure. During my senior year, I was not well prepared in mathematics, so I stayed another year at Northwestern before looking for a more research-oriented mathematics department in which to do my graduate work. I went to Princeton for my Ph.D. because it had an excellent mathematics department and because Sam Wilks was known as one of the leading mathematical statisticians. That was perhaps the only place where one could get an outstanding training in mathematics and at the same time develop statistics. At that time, Harold Hotelling and Abraham Wald were at Columbia University, but not in a mathematics department. And Iowa State was more oriented towards applied statistics, though there was Cochran in statistics and Tintner in statistics and economics. At Northwestern in my first year of graduate work I also studied mathematical economics with Bill Jaffé. In the first semester, we read Hick's *Value and Capital* and Schultz's *Theory and Measurement of Demand*. In the second term, we studied Joan Robinson's *Economics of Imperfect Competition* and Chamberlain's *Theory of Monopolistic Competition*. I developed my interest in mathematical economics and also in statistics with Harold T. Davis. He taught the course in econometrics that I took in my senior year. A requirement was economic statistics which was taught, I think, by Horace Secrist, but it was known as a dull course. I took econometrics instead to fulfill that requirement.

I believe Harold Davis was at Indiana University for a number of years and was also associated with the Cowles Commission for some time.

That's right. He founded the Principia Press to publish some of his books and some of those were Cowles Commission Monographs. He had a considerable influence on my career because he was a very stimulating man. He had wide interests. He had written books on maybe a dozen different subjects: history, mathematics, econometrics, and so on.

When you selected Princeton as a place to do graduate work, were you already thinking about doing work in mathematical statistics?

Yes. In selecting a graduate school for my Ph.D. training I had that in mind. I think at the time, 1940, I was offered an assistantship at Iowa State. At Columbia, I did not have an offer of financial assistance, but at Princeton I had an offer of a "junior fellowship." That came about partly because there

was a connection between Princeton and Northwestern, in that Al Tucker, who eventually became chairman of the Princeton mathematics department, had married the daughter of Professor D. R. Curtis at Northwestern. So there was this personal tie and I wanted to go to Princeton because Wilks was there. I had already decided that I wanted to make statistics a major interest within mathematics.

So when you were at Princeton, no doubt Samuel Wilks was one of your major intellectual influences. Were there any others?

He was the major intellectual interest for me. I worked closely with him. First, I took courses with him and then he was my dissertation supervisor. I did have a contact with Oscar Morgenstern. Before the war got going very heavily, there was a seminar on mathematical economics that Morgenstern ran in which I participated. Von Neumann was developing the theory of games at the time. He gave lectures in Fine Hall Mathematics Department, either the first year I was at Princeton or the second year, but I must say, I really didn't understand it [the theory of games] very well when he lectured on it. I would see him with Morgenstern from time to time and Morgenstern would tell me about some of the things they were doing. I did get to hear about the theory of games early on. Maybe if I hadn't been so busy in war work at that time, I would have gotten into game theory much more.

Would you like to tell us about the war work that you got involved with?

Well, early in the war I continued my graduate work and then, in 1942, I got into full time teaching. There were many students at Princeton, not so many of the usual undergraduates, but special programs like the ASTP, the Army Special Training Program. There was a lot of teaching being done. But early in 1943 I got into full time war research work. This was directed by Wilks under the Applied Mathematics Panel of the NDRC, the National Defense Research Council. My first project was to evaluate long range weather forecasting. Long range, at that time, meant three or four days. There were competing organizations. There was a group in the Weather Bureau, there was a group in the Air Force part of the army, there was a private group [I think the meteorologist's name was Krick from Cal Tech] and one or two others. We amassed lots of data and punched them on cards. We used a tabulator and I had maybe half a dozen people using desk calculators to carry out these rather simple computations. Well, that went on for perhaps nine months. But by the time we got our report ready, the various wheels of politics and bureaucracy had moved and a decision had already been made as to which service would do the long range weather forecasting. But we found that even forecasting three or four days was not effective; the usefulness of a forecast petered out after one or two days. The biggest goal of long range weather forecasting was to forecast weather for

the invasion of Europe and our advice at that time, in late 1943, was you couldn't do better than use the climatological averages. As it was though, D-day took place in good weather. I suspect that was by chance.

The next project I was on, was together with Phil McCarthy and Bill Cochran. We were doing a project to advise the Navy on how battleships should fight gunfire battles [rather presumptuous]. We took into account the accuracy of gunfire by analyzing lots of target practice records of the battleships. I went out on a battleship to watch them carry out target practice and then we took into account the armor of the ship and the explosive power of the projectiles. This study went on for maybe two years. By that time there weren't any more gunfire battles among battleships, so we ended up with recommendations that were never put into use. The third project I got on was testing explosives and that was, I think, a useful task that we performed.

I remember reading that Harold Hotelling had developed the  $T_0^2$  statistic in the context of bomb sight testing, where you have natural multivariate dispersion around a target in the two-dimensional plane for field artillery and in three-dimensions for anti-aircraft fire. Did anything in the way of methodological developments such as this arise in the course of your own scientific war work?

Well, in the third project I was on, there were statistical methods being developed. The Army Ordinance tested the sensitivity of an explosive by dropping a weight on a small piece of it from a certain height, say, 40 inches and see whether it exploded. If it didn't explode they would take a greater height and so on, and if it did explode then they might take a smaller height. In effect, they had a sequential method of testing. Will Dixon and Alex Mood developed something called the up-and-down method for testing sensitivity and that's related to bio-assay, where you want to establish the median dose. That did develop into some new methodology. We related it to the recently developed sequential analysis of Wald, which of course, we knew about at that time.

A good deal of the statistics done during the war used known statistical methods. We did a kind of small scale crude Monte Carlo to get solutions to bombing problems that were too difficult or too complicated to obtain on a theoretical basis. It was during that time that I learned a good deal of applied statistics. My training at Princeton was primarily in mathematical statistics, but with the war work and with people like Cochran around I learned some of the problems about analyzing data. In that sense it was a good experience.

After the war, your first position was with the Cowles Commission for Research in Economics at Chicago in 1945-46. Would you tell us how this appointment came about?

Yes, at the end of the war we were winding down this war research. That was in the early fall of 1945; I had had my five years of monastic life at Princeton and I was ready to make a change. I wanted to get into a different kind of environment and I thought that applying statistics in economics would be a good field. I talked this over with Morgenstern and he suggested I go up to New York and talk to Arthur Burns, Director of the National Bureau of Economic Research. Having spent weekends there while at Princeton, I thought it would be fun to live in New York, so that was an acceptable idea. I went to have an interview and, when I told my qualifications to Burns, he said, "we have no use for anybody who is so well trained in statistics." [Two years ago, when I was invited to be the Wesley C. Mitchell Visiting Research Professor of Economics at Columbia, I felt that things had changed a great deal in that line, since Mitchell had been the leader of the National Bureau.] So the idea of a position with the National Bureau fell through. My backup plan was to do full time teaching at Princeton on a temporary basis. Because the war had changed academic schedules, the fall term was to start November fifth. Morgenstern had heard that Marschak was going to be in Rye, New York, for a conference on the peaceful uses of atomic energy and suggested that I talk to him. I suppose Morgenstern talked to him or wrote to him about me. I went up to see Marschak on Sunday, November 4th, and I was to start teaching the next day. But I had been told that if I wanted to change that commitment I could at any time until the morning of November 5th. I talked to Marschak and indicated my interest in economics and applying statistics to economics and he seemed impressed. But he was a cautious man and said, well, let me go back and talk it over with Tjalling Koopmans, my associate. I said, but I have to make my decision before tomorrow morning: if you can't offer me a position now then I can't do it. Well, finally he was ready to make a decision, probably he called up Koopmans and talked to him on the phone. So he did finally agree to give me a position. I only had a question about remuneration. I had applied for a Guggenheim Fellowship and he thought that, well, the Cowles Commission is a research institute and the proper stipend would be the same as you would get on a fellowship. I think we finally decided on an annual salary of \$2,500.00. Then, after I got to the Cowles Commission I pointed out that a Guggenheim Fellowship is tax-exempt and my salary was raised appropriately. So, as it was then November fifth, I did not teach but cleaned up my affairs at Princeton and went to the Cowles Commission later on in November.

There has recently been a good deal of interest in the early years of the Cowles Commission, particularly the 1940s which were so productive in research on econometric methods. Could you describe to us some of your own personal memories of the research environment at Cowles during those years?

Yes. The Cowles Commission was a very exciting place at the time that I arrived in late 1945. Besides Marschak and Koopmans, there was Leo Hurwicz and Lawrence Klein. Herman Rubin came back from his year of military service about that time and Roy Leipnik was a student in mathematics who was full time at the Cowles Commission. Marschak and Koopmans had almost a missionary zeal for this new approach to econometrics through simultaneous equation systems, and so everybody was gung ho on developing the methodology. It is very unusual that you get such a group of talented people that concentrate essentially on one subject. We had weekly seminars, some directed more towards mathematical economics or model building, and some towards the more statistical aspects. Koopmans was working particularly with Rubin and Leipnik on identification and full information maximum likelihood. When I got to the Cowles Commission I didn't know about the approach initiated by Haavelmo, so I read up on it. Then Koopmans said one of our problems was what to do with a single equation. We knew that we could use maximum likelihood on the whole system. Well this problem of a single equation was to become a special interest of mine. Now, it was about that time that a memorandum of Abe Girshick came through, late in 1945, that helped us on the problem. Haavelmo had an extended visit of several months, maybe six months, in the spring and summer of 1946.

One other aspect of the Cowles Commission was that, except for Marschak, we were all in one big room. The room had been partitioned into little cubicles. First, the partitions only went up three-quarters of the way to the ceiling. The architects said that the sound absorbing tile in the ceiling would keep it quiet, which wasn't the case with people like Hurwicz and Rubin. Later the partitions went up to the ceiling, but it still left a good deal of communication, both intentional and otherwise. That feature probably helped to make the group very cohesive. In the open part of the room where we held our seminars there were also some research assistants, like Sonya Adelman (Klein) and Marianne Ferber. Don Patinkin was also around often, though he wasn't part of the Cowles Commission. Then, there were others from the Economics Department. I found that the University of Chicago at that time was very stimulating, perhaps due to the fact that people were getting back to academic work after being taken away from it by war work.

Because of this intense and continual work, there was a lot of interaction. As soon as someone had developed some ideas, though not very far, he would write a memorandum and that would be discussed at a seminar. Of course, anything that was going to get to the point of publication would be reviewed by other members of the staff, particularly Koopmans. Koopmans was very well disciplined. He would not let anything go out that wasn't rigorous. One of the in-house jokes was, that whenever there was a seminar Koopmans would raise a question about the notation and, you know, the Cowles Commission got a reputation for its complicated notation.

Quite justifiably, I think, from Cowles Monograph No. 10. One senses, looking back on those years from afar, that there was a cohesive structure that really hasn't been attained since then in econometrics amongst workers with very complementary skills in mathematics, in economics and in statistics. Perhaps it had something to do with the genius of Marschak and Koopmans that they drew this extraordinary group of talented people together at a time when the subject was ripe for intellectual change. Did you have any sense at that stage that the work that you were doing would turn out to be as long-lived and as important as it has been?

Well, in a way we expected it would be even more important and long lived. I think that we expected that this new approach, first using probability in modern statistical inference in economics, but more precisely the simultaneous equations approach, would revolutionize economics; that it would give a scientific basis for economics to proceed by developing rigorous mathematical models and making inferences in modern scientific terms; and that this would be used virtually throughout economics. In one sense, I don't know how convinced each member of the group was that this would happen. I didn't have as big a stake in it as Marschak and Koopmans, as the more senior economists. I think we younger people were not as committed in the sense that our careers were starting and we would move onto other things if this didn't develop so much. It was expected that this would have a big impact, at least, in the long run. I don't think that we expected that the theoretical problems would continue so long: the distribution problems, the autoregressive model problems, and others that we are still struggling with. We had, I think, a pretty good theoretical grasp on the goals of the simultaneous equations models. We were unable to carry out the programs to a big extent because of the limitations of computational ability. That stood in the way, so it took maybe several decades before the methods could be really tried out. But, over the years, I think that we expected more in the way of accurate prediction and of advice to policy-makers, so simultaneous equations models are continuing to be developed and statistical econometric methods worked on. Still, the payoff seems somewhere in the distance. The big payoff.

To your memory was there ever any discussion amongst the Cowles Commission researchers about finding good alternatives to maximum likelihood methods in simultaneous equations? In the Cowles Commission archives there is frequent mention of a letter from Girshick about simultaneous equations estimation which was written in December 1945. Mary Morgan has unearthed it recently in her own researches in Marschak's personal papers and, to my reading, it seems to point to a "just identified" type of instrumental variable estimator, although that



name was not given. Now that you have examined the letter can you shed any further light on the history of this discussion? Do you yourself ever remember thinking about the possibilities of a least squares alternative to maximum likelihood in this context?

Yes. When I look at this copy, it looks vaguely familiar. I got to the Cowles Commission, as I recall, on November 19th preceding. That's about a month before this letter was written. When I came to the Cowles Commission I read some of the material like Haavelmo's probability approach to econometrics and some of the papers that were in progress at the Cowles Commission. There was a paper by Wald on incomplete systems that was published in Cowles Commission Monograph No. 10. I looked at that and was sort of feeling my way. Then this letter came in and we all looked at it. Marschak turned it over to Koopmans, Rubin and myself and there are minutes of a seminar meeting about January 15th or 20th, I think. We discussed this letter and we saw that the general idea was to estimate the reduced form by regression and then relate that to the simultaneous equations by taking the transformation back again. That is the first half of this letter—the method we now call indirect least squares and it works if the equation is just identified. I think quite quickly we saw that, as long as the equation is just identified, then you can solve back. The question then was what to do if you have over-identification. In that case, you had to estimate the reduced form so that the relevant part of the reduced form matrix has the proper rank, which would be one less than the number of endogenous variables in the equation. At that point I then drew on work in my doctoral dissertation where I estimated a matrix of means with arbitrary rank. It's an easy step to go from that to regression coefficients. So then Rubin and I worked on this, and developed it into the *Limited Information Maximum Likelihood* procedure.

The second part of Girschik's letter had to do with getting confidence intervals by using the  $t$  distribution and this is based on the idea that the coefficients of the reduced form have a normal distribution. So a linear combination of the coefficients of the reduced form has a normal distribution and if you take the corresponding quadratic form of the estimate of the covariance matrix of the reduced form you get something that is proportional to a chi-square and you build up the  $t$ -statistic from that. So Girschik gave this example and then Rubin and I generalized it. Once you see how to get the  $t$  statistic, then you can generalize this to an  $F$  statistic where you use a more general quadratic form in the numerator, and from that you then get exact small sample confidence regions. As I say, Rubin and I generalized it. The LIML procedure really came out of the question of how do you use the reduced form when you have over-identification. Then after we saw a way to do it, we derived it as the maximum likelihood estimator when you only use the identification conditions on the single equation. That became a single equation estimator. Now, Girschick came to the Cowles Commission

for a small meeting, I think a little later that winter. It was either later that winter or it was early in the spring of 1946 and we discussed this some more. By that time Rubin and I had worked out the LIML method pretty well. It says here in the letter that he was working on some black magic estimates for incomplete systems. But, as I recall, he didn't send more on that.

We thought that since we were using maximum likelihood, that this was the final answer to estimation in a single equation, especially if you assume normality. We thought that it was a desirable property that if you change the normalization then you just made that same transformation on the estimate. We didn't appreciate the fact that economists liked to write their equations in a particular way with a particular normalization. Another consideration in this was that, at that time, if you had an equation with two endogenous variables, there wasn't any problem computing. If you had very much more than that, it was impossible to compute, so there wasn't a great deal of incentive to find, you know, simpler computational procedures. It was really quite a few years before people were doing much practical work with this and one worried about the ability to compute. In 1946 to 1950 they were computing everything by desk calculators and even for solving linear equations you didn't go over three or four variables. It wasn't until another five or ten years that people thought they wanted to go on to larger systems. Even so, I think that probably the difficulty of computing LIML has been overestimated. The advantage of two stage least squares is that it's easy to compute. Of course now, in 1985, unless you have a tremendous system, you don't worry about computation. So from that point of view, computation isn't the critical aspect.

Now, in proving the asymptotic normal distribution of the estimators in the second of the two papers with Rubin, the first step in proving this was to show that the contribution of the smallest root times the estimate of the covariance matrix drops out. So in doing this we reduced the estimator to what's now called the two stage least squares estimator. So the two stage least squares estimator is actually in that second paper.

I recall that there was a meeting of the IMS in Ithaca in 1946 in which the Cowles Commission researchers presented their results to an audience that was composed of leading statisticians and probabilists: Wald, Feller, Hotelling, Tukey, Blackwell, Cochran and others who later became prominent like Bowker, Dwyer and Stein. I formed the impression from reading the transcript of this meeting in the Cowles archives that there was some disquiet amongst the community of mathematical statisticians about the methods that were being advanced. Do you have any memories of that conference and the reception that the papers received?

Yes, I remember the conference. In fact, Koopmans and I gave papers and wrote up the minutes of the meeting. At the Cowles Commission we, particularly Koopmans, had realized that computation was a big bottleneck

and so we used every opportunity to try to get some leads on how to tackle this problem. One person present at that meeting, I think, was John Mauchly, one of the inventors of ENIAC. Earlier we had had a discussion with Von Neumann when he came through Chicago. My impression was that the kind of problems we had were too special for standard procedures. For example, Hotelling was writing about methods of inverting matrices. But, we had special kinds of matrices, perhaps big ones. We also had nonlinear problems. Von Neumann hadn't gotten far along with developing electronic computers that could be programmed in a versatile way. It was about these problems that various statisticians gave us some suggestions. I don't think that it had a big impact on what we were doing. Later on, Herman Chernoff and Rubin and others did work on computational methods but we didn't really get very far until computers were much more advanced. The statistical community was sympathetic with the aim of bringing modern statistical inference into econometrics. Though some people were rather pessimistic about how useful it would be, because statisticians are accustomed to other kinds of data where perhaps you can gather as much data as you want, or you can have large sample surveys, or you can design experiments as in agriculture, or the measurements are more precise, or the structure is more permanent. I think that statisticians do not fully appreciate the kind of problems that simultaneous equation models put forward, such as the problem of identification and some of the problems of the time series autoregressive type. I don't think that any of the statisticians that attended that conference went home, sat down at their desks and started solving some of our problems.

Your doctoral thesis was on the noncentral Wishart distribution and some of your earliest published research in the 1940's is in that field. Looking back on that subject now with the benefit of hindsight, do you think that the main problems that concerned you then have now been solved?

Well, let me talk about a couple of aspects of that. My dissertation research really started out of reading a couple of papers by Fisher on discriminant analysis which I had difficulty understanding. Actually there were serious mistakes in them, which is one reason why I couldn't understand them. I got an impression of what Fisher had in mind, that is, problems of collinearity in multivariate analysis. This led me to consider multivariate statistical problems where one had a matrix of means of rank lower than the dimensionality of the number of vectors and then the statistical methods involved in the simplest case of two matrices, one being a central Wishart sample covariance matrix and the other being a noncentral sample covariance matrix where the rank of the noncentrality parameter would be equal to the rank of the matrices of means. So that led me to the noncentral Wishart distribution. It turned out when I got to the Cowles Commission and considered the single equation approach, that the use of the reduced form, the

limited information maximum likelihood method, amounts to estimating a part of the matrix of the reduced form with rank one lower than the dimensionality or the maximum possible rank. So this latter problem turned out to be very closely related to part of my dissertation work.

I was doing my war research work eight hours a day and then doing the dissertation research in the evening. Wilks was busy with all kinds of war work and committee assignments and he didn't pay any attention to my dissertation. I just went plugging along by myself. I worked on the noncentral Wishart distribution and developed it for the noncentrality parameter of rank 1 and then rank 2. About the middle of 1944 I went to Washington for a meeting of the IMS and one of the contributed papers was by Girshick and it was on the noncentral Wishart distribution. He had done more or less the same thing as I had. I actually never saw anything written by him; I wrote up the paper that we published jointly on that topic. He had hoped to use this as a dissertation subject, but he eventually wrote a dissertation on sequential analysis because that was what he was excited about and because my dissertation already involved the noncentral Wishart distribution and the maximum likelihood estimates of matrices of lower rank and likelihood ratio tests.

During this time I was involved with various problems such as the distribution of roots of determinantal equations and a lot of these problems came down to integrating out some exponential of a quadratic form with respect to the Haar measure of orthogonal matrices. I remember going around Fine Hall asking how do you integrate with respect to the Haar measure and all these pure mathematicians could give me theorems about the existence of the Haar measure but nobody could tell me how to integrate with it. Well, I could see that something more was needed than the classical mathematics, Bessel functions, and so on, but I didn't really go further. I got more interested in these problems of inference and didn't push into this, but I could see that some new mathematical developments were needed. At Princeton, just a few years later, there was a student, Carl Herz. Herz ended up writing a big paper on hypergeometric functions with matrix arguments. His dissertation supervisor was Solomon Bochner, who was one of the mathematicians that I had talked to quite a bit about these problems of the noncentral Wishart distribution. Bochner had at least paid some attention to the expressions that I had in terms of Bessel functions and I think there's a connection there. Of course, the definitive work along this line was done by Alan James, who was also a graduate student at Princeton, a student of Wilks. My impression is that Alan James developed his work, which eventually led to zonal polynomials, independently of Herz's work. I think that Herz says that only late in the game did he know about Alan James' work. Alan did some of his work before he came to Princeton and after he went back to Australia. Alan developed the technique of integrating the exponential trace of a matrix product involving orthogonal matrices. He showed how to ex-

pand the result in terms of zonal polynomials and that's what the noncentral theory was waiting for, to have the proper kind of mathematical quantities for such expansions. I think that as far as noncentral distributions go, you can express them with zonal polynomials if you work hard enough and have some ingenuity. But the other side of it is that those expressions don't tell you a great deal about the properties of the distribution. It turns out that they haven't been so very useful in that respect and apparently they cannot very well be used for computation. I think it's an important part of the development of the subject, multivariate analysis, that we have those tools and those results. The other approaches to those problems are either numerical work or approximations, in particular, asymptotic expansions of various kinds, and those, of course, are carried out in several ways. Once one has developed the theory of zonal polynomials and general hypergeometric functions then one can consider approximations in terms of those functions or one can go back to the statistical problem and consider approximations or asymptotic expansions from that point of view.

At what stage did you meet Alan James and Carl Herz?

I can't remember that I ever met Carl Herz. Now, when did I meet Alan? Oh, I guess I must have met him when he was a graduate student at Princeton and then he was at Yale for a few years and I saw him every once in a while there.

Another focus of your early work was the theory of testing for serial correlation. Your article in *Skandinavisk Aktuarietidskrift* in 1948 gave some definitive results on the power of various tests of serial correlation and has recently been extensively used in econometrics, in particular by Sargan and Bhargava. What stimulated your own work on this topic?

My interest there goes back sometime to the early 40's when there was a good deal of interest in serial correlation. R. L. Anderson did his dissertation on the circular serial correlation. Tjalling Koopmans, at the same time, was actually living in Princeton, I think before the United States got into World War II, and he wrote an important article on serial correlation. Will Dixon was working on the topic for a dissertation and John von Neumann was writing on the subject. So at least four people that I was aware of were doing research on serial correlation. It therefore seemed to be an important topic in time series analysis and the application of statistics to economic time series in particular. Then, R. L. Anderson came to Princeton to do war work so we talked about it and we considered the circular serial correlation when you used deviations from fitted sine and cosine pairs, which you might want to use for seasonal adjustment. That worked out very nicely and it worked out because the trigonometric functions were characteristic vectors of the covariance matrix of the circular model. At the time I wrote the paper you're referring to, I was in Stockholm. I remember that R. L. Anderson and I were

having correspondence and revising our paper so I kept being reminded of this problem. Then I saw you would get more general results out of the same idea and I sat down and carried out the mathematical development. I remember that I enjoyed working on that paper because I realized that there was a kind of general problem with using the algebraic framework and putting serial correlation within a framework of the model using the Neyman-Pearson theory. The research went smoothly. I think that I spent a month developing the theory and writing the paper and had no hassles with it. Actually, there were several aspects of the paper that I didn't really push very much. One was that least squares and generalized least squares estimates give you the same results if the independent variables span the same space as an equal number of characteristic vectors of the covariance matrix and that was picked up quite a few years later and developed by Bill Kruskal and George Zyskind and others. Then there was another aspect in considering uniqueness of tests in that paper. I was on to the idea of completeness, but Erich Lehmann and Henry Scheffe were working on that at the same time and were publishing in *Sankhya*. Actually the paper, being published in the *Skandinavisk Aktuarietidskrift*, didn't really get attention very quickly. It was written in English but it wasn't read a great deal. I was in Sweden and was doing research at Harold Cramér's Institute for Insurance Mathematics and Mathematical Statistics and he was editor of the *Skandinavisk Aktuarietidskrift* so I thought it would be appropriate to publish there.

Can you tell us more about this year that you spent abroad in the late 1940's, visiting Scandinavia at the University of Stockholm and England at the University of Cambridge? Were these overseas visits important in helping to shape your academic and research interests? Did you form any impressions of the intellectual traditions in mathematical statistics in these countries and how they may have differed from that in the United States?

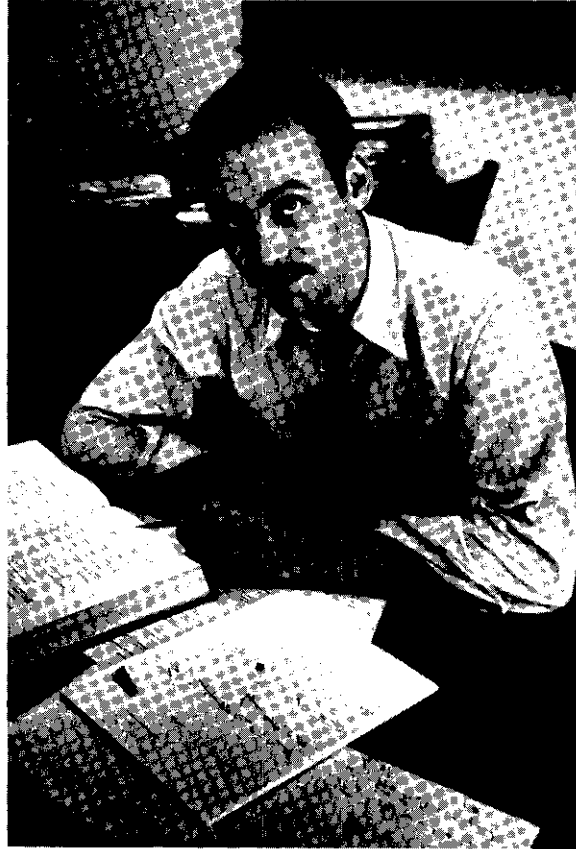
Yes, I had a Guggenheim Fellowship that year. I went to Sweden in August of 1947 and stayed until April of 1948 and then I went to England and returned to the United States in September. I wanted to go abroad, I had never been out of the United States and I thought that this would be an opportunity. In going to Sweden, I had been in touch with Cramér; he had just published his *Mathematical Methods of Statistics*. I had met him at Princeton. He was one reason for going to Sweden. Another reason was that I was interested in the land of my forebears. The third reason was that Sweden had not been hit by the war and it was more prosperous there than most of Europe. I wanted to go to England, in any case, because of the English development of statistics. After all, modern statistics pretty much comes from England with some help from the United States. I thought that was important. At the time I was in England I had an office in Richard Stone's Department of Applied Economics and we had a very good group there. There was D. Cochrane and Guy Orcutt, and Jim Durbin was a

student. Gerhard Tintner also came through. There were a number of economists that came around, like Joan Robinson and others. There were statisticians at Cambridge like John Wishart, Henry Daniels and Frank Anscombe. Of course, in England it wasn't hard to get to Manchester to visit Maurice Bartlett and London where I visited Maurice Kendall. R. A. Fisher was at Cambridge and since my research had been stimulated by work of his, I did visit him. But he was mostly interested in talking about his current work in genetics and showing me around his garden. For me, it was interesting but not very useful in developing my own research. C. R. Rao was at Cambridge also and I met him. Somehow we didn't really interact much in the way of statistics. But at Cambridge my contacts in statistics were Wishart, Daniels and Anscombe and in econometrics the ones that I mentioned. At that time Dick Stone was also more interested in econometrics. Actually, one of the things he was interested in then was factor analysis applied to economics.

Factor analysis and multivariate methods in psychometrics became another major area in which you worked particularly during the 1950's. Was your contact with Stone instrumental in getting that research going or were there other influences?

Well, factor analysis really started when I was at the Cowles Commission. I mentioned earlier that I had taken some courses in psychology and among the various interests I had in the social sciences, factor analysis appealed to me because it applied matrix algebra. One of the mathematical methods that I feel very comfortable with is matrix algebra, at least some aspects of it. In the Cowles Commission I worked a lot with Herman Rubin, and Thurstone was at University of Chicago. Herman also was interested in factor analysis. So Rubin and I did a lot of work together during that one year I was at the Cowles Commission. It also tied into econometrics because you have a model with errors in variables rather than errors in equations. If those errors are independent, then you have a version of the factor analysis model and Leo Hurwicz and I developed a number of memoranda on shocks versus errors in econometric models and one of the models was a renaming of the factor analysis model. Unfortunately, we never got to the point where we published that material. My work on factor analysis started while I was at the Cowles Commission and it was all part of the same package. Rubin and I went back and forth on it. We would develop some theorems and then Herman would get more general proofs, then a yet more general theorem. I finally wrote the work up in the 50's and presented some of the material at the Third Berkeley Symposium. But, unfortunately, that paper, being published in the *Proceedings of the Third Berkeley Symposium*, didn't get the audience it should have either.

When I came to Columbia in 1946, Wald had said that he would like to work with me on some econometric problems. But when I got there he was involved with statistical decision theory and wasn't interested in putting the time into econometrics. I found that nobody in the Economics Department



Professor T. W. Anderson in his office at Columbia University in 1949

was interested in this new direction of econometrics. So I continued some of the work that I had done at the Cowles Commission, on LIML, factor analysis and serial correlation. Paul Lazerfeld got me interested in latent structure analysis and then I saw the connection between this and factor analysis. I did quite a bit of work with Lazerfeld, so at the time I finished up this paper with Herman Rubin on factor analysis I was also looking at a wider range of problems. In 1961 I published a paper in the Harold Cramer Volume in which I set up a general model of this kind and got some results on statistical methods. This approach has recently been developed further by David Bartholomew. I was on the Board of Editors of *Psychometrika* for many years. In 1963 I had a paper on factor analysis and time series, where I combined my interests in factor analysis and time series.



Your research in multivariate analysis during the 1940's and 1950's seemed to culminate in the publication of your book *An Introduction to Multivariate Statistical Analysis*. By any standards this book rapidly assumed the status of a classic work. Was its extraordinary impact and lasting impression on statistical workers a surprise to you?

Well, needless to say, I was very pleased. I think I came along at the right time with that book, and yes, it did have more of an impact than I had anticipated. There is a history to that book too. When I was at Princeton during the war years, Wilks and Walter Shewhart started this series with Wiley. In 1944 Wilks and I signed a contract with Wiley for a book on multivariate analysis. Earlier, Wilks had an orange lithograph book called *Mathematical Statistics* and he wanted to update that and got Henry Scheffé and David Votaw and me to write a draft of a new version. I wrote a chapter on multivariate analysis and several other chapters. Actually, we were disappointed that we weren't listed as coauthors, rather than being thanked in the Preface. But writing that chapter was a predecessor to this contract. Then, after my year at the Cowles Commission, I went to Columbia and during the first year I was at Columbia I taught a year-long course on multivariate analysis. Since I had the contract, I thought, well, this is the chance to get started on writing my book. I was very fortunate in getting two students to write lecture notes. One was Charles Stein who wrote really a very concise and thorough report on my lectures in the first semester and then another student, Gobind Seth, who wrote in the second semester. It was around that time that I decided that I would work seriously on a book on multivariate analysis. I could see that Wilks wasn't going to be helping very much. We talked it over and decided that he was going to put his efforts into other books: an elementary book and then the big book of his that came out as *Mathematical Statistics*. So we parted ways and I got a new contract and continued on by myself. At Columbia we didn't have the facilities for reproduction that one has now. The students would put out lecture notes in dittoed form and sell them at cost. My manuscript, as it developed, would come out in that fashion and it grew. I think I had a pretty well finished manuscript by about 1954 or 1955 and then I remember in the summer of 1956 going over it intensively and putting it into shape. Somehow it took another 20 months before it came out. I remember being so happy with finishing the proofs on it and finally seeing the book arrive.

And now, of course, it's gone into its second edition.

Yes, it took me quite a while to get to the second edition. Well, as a result of the impact it had, many statisticians worked on the obvious problems that one could see by reading my book, lots of distributions under null hypotheses and under alternative hypotheses. There were other invariant procedures that I only just touched on, there were further distribution problems

there, comparisons of power functions, and other approximations to the distributions of test criteria. My book, among other things, I think, pointed out a lot of problems in what you would now call classical multivariate analysis. There are hundreds, maybe thousands, of papers written on subjects that came out of that book.

In a certain sense, the publication of the book helped to define a new subject area of statistics in itself, that of multivariate analysis.

Yes. I think I came at the right time for that book. The methods that I discussed have been fairly well accepted. There was a good theory to it and you could organize the theory pretty well. The methods that I discussed in the book are still used. In fact, they're being used more now because computers make it possible to carry out the procedures easily. So it went for many years without being revised because the material that was contained in it was pretty well accepted. Maybe there are some qualifications to make to that. For example, I didn't have any Bayesian methods and there are many statisticians who would object to that. Then, there is the whole development of Stein estimation. I have included some treatment of these subjects in the new edition.

In 1959 the *Annals of Mathematical Statistics* published your paper on the asymptotic theory of estimation in the first order autoregressive model. This paper brought the subject forward from the original Mann and Wald contribution of 1943. It has always been one of my personal favorites. Do you remember what stimulated you to write it?

Yes. I have already mentioned my interest in serial correlation and, along with that, autoregressive models and simultaneous equation models with lagged variables. It was natural that in the Cowles Commission work there were autoregressive models and in the year that I was at the Cowles Commission we had developed a lot of technique dealing with vector autoregressive models which didn't all get into print at that time. Now, one of the questions that came up then was the case when the stochastic difference equation did not model a stationary process. Rubin had a paper on consistency in the explosive case; and Hurwicz had a little note, too. So the paper of mine that you are referring to had two aspects to it. The more important one was to treat the usual autoregressive estimates when the process was not stationary. If the equation is written so that the condition for stationarity is that the roots are less than 1, then this paper treats the case of one or more roots being equal to 1 or several roots being greater than 1 and that's sometimes called the explosive case. My interest in the problem dates back to 1945–46 and the Cowles Commission days. I think what brought me to work on the problem was a paper by John White, who gave a talk at the 1955 ASA-IMS meetings. He later published a paper in the *Annals* the year before my paper. His presentation stimulated me to look more intensively at the problem. In the year 1957–58, I was at the Center for Advanced Study in the Behavioral

Sciences. I finished proofreading my multivariate book early in the fall so I had the time available for this work on autocorrelation and autoregression. It was one of several topics I worked on during that year. I remember that it was a paper that I enjoyed working on because there were some very clean results that came out of the paper. I treated a first order case for the scalar series and also a first order case for the vector process. And for that case I considered all of the roots to be greater than 1 and that led to a vector analogue of the scalar case. I got nice, neat theorems in that case. Now, I realize that there were other interesting cases where some roots were bigger than 1, some equal to 1, and some smaller than 1. Shortly after that, a man named M. M. Rao picked up the problem in his doctoral dissertation at the University of Minnesota. As I recall, he had one root larger than 1 and the rest smaller than 1. And other people have gone on from there. I realize that there were many interesting problems to come out of it, but I thought I had made my contribution and I had other problems I was interested in. There's one problem in that area that I kept coming back to. I had the feeling that one should be able to prove consistency of the maximum likelihood estimates in a way that didn't make use of the size of the roots as a condition. If the roots are bigger than 1 you get convergence faster than if the roots are less than 1. This suggests that if you have a proof of consistency for roots less than 1, it shouldn't be hard to extend it to roots bigger than 1. Although I came back to the problem from time to time I was never really happy with the method. The methods that I have seen published, I think, all depend on separating the roots into at least two classes, if not three. Then, of course, there is also the question of the limiting distribution of the estimates when properly normalized. That problem is part of a larger problem where you have a dependent variable that depends on some independent variables and these independent variables can be functions of the past. So the autoregressive model is one of these. Another model comes up in stochastic control where in each time period you adjust the independent variable to try to obtain some set value of the dependent variable. John Taylor and I wrote several papers on that topic. And Herbert Robbins and T. L. Lai and C. Z. Wei have written many papers in this broad area and I know of at least one paper by Lai and Wei on the autoregressive model. I think that they separate roots in order to treat the problem. I think there's still a general technique there that hasn't been discovered.

In this later work that you are referring to, there has been a serious search for minimal conditions for consistency and a move towards establishing almost sure convergence rather than convergence in probability. From my memory of your own paper one of its contributions was to introduce new methods of establishing the limiting distribution theory, methods that were quite different from those used by Mann and Wald in the original paper.

Yes. In the stationary case where the roots are less than 1 and you get asymptotic normality using the square root of the length of the series as the normalizing factor, Mann and Wald proved asymptotic normality but in a somewhat cumbersome way, which I might say parenthetically exemplified the way that Wald worked. He wanted to get a result and often he would just slug it out. He didn't like to take the time and effort to refine the methods. In studying the proof of Mann and Wald, it looked like you didn't really need to use the fourth order moments, that you could get along with only second order moments because what you average is the product of the current disturbance and a lagged value of the observed time series. The second order moment of that only depends on the second order moments of the disturbances. It therefore seemed to me that you didn't need to have anything more than the second order moments. I then set about proving this. It turned out that by rather simple machinery which is intuitively appealing you can prove a central theorem in this context. Just a little later in 1963, the IMS had an institute on time series analysis at Michigan State University. Morris Walker was a participant and we talked about this 1959 paper of mine. I said that I thought that this had much more general applicability, that you could have even an infinite autoregressive representation of a process and sample autocorrelations would have the same property. That is, for asymptotic normality all you need to assume is the second order of moment. He didn't believe it initially, but by the time we finished discussing it he saw that you could prove the result. Then we published a paper on it a little later. I think that this kind of idea is even more general. If you have a stationary autoregressive model with finite variance, then the sample autocorrelations have an asymptotic normal distribution and that only requires that the second order moment of the disturbance exists.

The 1960's proved to be a highly productive decade for you with 30 or more articles published in different fields like multivariate analysis, statistical inference and time series. Can you tell us how you came to maintain so many ongoing research projects during this period?

Well, since I saw this question, I went back to look at my bibliography and I found that in the next decade I also published 30 or more articles. What can I say? I have wide interests and lots of chances for stimulation. I probably take on more research projects than I should.

In 1968 you spent some time as a Visiting Professor of Mathematics at the University of Moscow. Can you tell us about this visit: how it came about, the scholars you met there and the Russian intellectual tradition in probability theory and mathematical statistics?

Yes, the background to that visit is that my multivariate book was translated into Russian in the early 60's. The International Congress of Mathematicians was held in Moscow in 1966 and I attended that Congress. I recall

that B. V. Gnedenko took me to the State Publishing House in Physics and Mathematics so I could receive my royalty check in rubles. I got this envelope with 500 rubles and a little change, and I said to him: "can I put some of this in a bank?" He said: "oh, no, you want to buy things." I said, "well, I don't know if I want to buy this much." And he said: "oh, my wife and I will help you spend it." In those days, to get the royalties you had to go there, collect it in rubles and spend it in the Soviet Union. Actually I did bank most of the royalties.

Well, Gnedenko was very friendly in that visit and I saw quite a bit of Yu. V. Linnik also. He had been at Berkeley the preceding year for the Berkeley Symposium. I also knew Akiva Yaglom and S. Kh. Sirazdinov from the Brown Symposium on Time Series in 1963. I had a number of Russian contacts and Gnedenko said that he would like to invite me to spend some time in Moscow. In 1967–68 I was on sabbatical leave at Imperial College, London, and I thought it would be convenient to go for a few weeks to Moscow. I went for about six weeks starting around the beginning of March. I stopped off briefly in Leningrad and saw Abram Kagan and I. A. Ibragimov and then went on to Moscow. I was at the University of Moscow. My wife and I had an apartment in the university and the Gnedenkos were very hospitable and took good care of us. Yu. V. Prohorov was at the university and he was another statistician that I spent time with. I gave a number of lectures, some in the Statistical Laboratory, one at the Academy of Sciences, one at the Moscow Mathematical Society and several others. Yaglom and I shared a particular interest in time series analysis and we had good discussions. Bolshev was at the Academy of Sciences and he had interests in some work I had done on sequential analysis which involved some calculations with the Wiener process. The impression I had in the Soviet Union was that statistics wasn't pursued in nearly as broad a fashion as in the United States and United Kingdom. Of course, probability was very highly developed. Time series analysis goes along with that because time series analysis can be based a good deal on stochastic processes. I had the impression that a good deal of the work in statistics was closely related to probability and these subjects could be developed by rigorous mathematics. On the other hand there were subjects that didn't seem to have much interest in the Soviet Union like simultaneous equation models, factor analysis, latent structural analysis, things of that kind. Also, at the time I was there there was a question about genetics, it was still the time of Lysenko and that's another field that seemed to be taboo. I think design of experiments was coming into some interest, but that's almost 20 years ago and things have changed a great deal.

Did you think that the work of R. A. Fisher on maximum likelihood and the Neyman-Pearson work on statistical inference had really penetrated the Russian intellectual tradition at the time that you were there?

I did give a lecture at the University of Leningrad and I think that, yes, they were pursuing these problems of statistical theory very vigorously and with a lot of competence. Now, that was the place that I saw in the Soviet Union which was rather like Berkeley, let's say, or Stanford. It was very up to date work. Of course, that school has suffered with Linnik dying. I suppose Leningrad would be the place where most of the work is in line with what we think of as modern statistical inference. I should think that by now the University of Moscow, which in some ways is the intellectual center of the Soviet Union, would have been developed a good deal more. Actually, the Statistical Laboratory seemed to have quite a few people and the Mathematics Department at the University of Moscow, which is what they call the Faculty of Mathematics, has a section on probability and statistics. So there must be a lot of activity and I just haven't kept up with what they are doing. We don't get very many Russian visitors. I did meet I. G. Zurbenko last spring. He was visiting here and he works on time series analysis and certainly in time series they are right up to date.

In 1971 your treatise on the *Statistical Analysis of Time Series* appeared. I remember Jim Durbin saying at the 1980 World Congress of the Econometric Society that he spent some time proving a result on partial autocorrelations and then found that, like everything else in time series, it was in your book! Did you set out to write a really exhaustive treatment at the time? Looking back now 15 years later, how do you see time series as having developed? What fields in time series do you think have really grown in significance since then?

Well, after publishing the multivariate book, I thought I would turn to the other interests that I had had over a long period of time and write a book on time series analysis. Since the multivariate book was so well received, I thought that wouldn't be such an effort, but it turned out to be a different experience. At first, I had in mind writing a very brief book. Vernon Johns had taken notes on my course in time series analysis at Columbia about 1955 or '56. I thought I would revise them, but then when I started getting into the subject, I was dissatisfied with the status of the book. I had in mind that I would include in the book the sort of old traditional time series analysis, like smoothing and the use of trigonometric functions to represent seasonal effects or the business cycle. When I went at these questions from the point of view of putting them within the framework of, let's say, Neyman-Pearson theory, I found that it really hadn't been done. I started serious work on writing this book during sabbatical leave, 1960–61. It happened that I was on the island of Madeira that year and was completely isolated. If I wanted to prove something I sat down and proved it myself. I didn't have the opportunity to ask other people. One of the first questions I had was; how do you determine the degree of a polynomial trend? That's in the older writings. The first thing you encounter, let's say, in economic time series is

to specify the trend, a simple linear trend or a polynomial trend and how do you determine the degree of polynomial? Well, nobody had really satisfactorily treated that question. So I sat down and worked out a theory, which was published in 1962. That approach also applied to determining the order of an autoregressive model. Of course, since that time there has been a tremendous amount of work on model selection, producing Akaike's amount of information criterion and other criteria. The point I was really making is that I tried to be thorough and what I was going to put into the book I wanted to have within this framework. Well, it took a lot of time and effort to do that and I went on to this question of having a trend represented by trigonometric functions which is one way of representing cycles and I found that that hadn't been treated from a theoretical point of view either and then I developed a theory there. Well, the result was that as I went along I was doing very thorough treatments. I remember that when I submitted the first four or five chapters to Wiley and it was reviewed that the reviewer remarked that it looked like I was doing an exhaustive treatment. It didn't look like it was going to be a short summary. About that time Hannan's short book on time series came out and purported to do what I had in mind. My own book had really gotten much more detailed than that, so I continued in that fashion. I think I got rather carried away by doing this and I was getting close to making an exhaustive treatment. Once you get started on that track, then it's hard to get out of it. Now, as I said before, I thought that I came out with the multivariate book at a good time because there were a lot of methods that were generally accepted and it was possible to organize the field. Time series just wasn't right for that kind of treatment, because it hadn't settled down. On the one hand, there was the time domain approach and on the other hand the frequency domain approach and they hadn't been worked together very thoroughly. As far as the time domain approach goes, the moving average aspect was not very popular and the methods that had been proposed for inference didn't seem very thorough. There were some original approaches by Durbin and Walker, but they did not seem very satisfying. Since the time I worked on the book, of course, Box and Jenkins have popularized an approach using moving average models, and there has been great deal of work on estimation of ARMA and ARIMA models. I have the feeling that the field still hasn't settled down. There is still a good deal of discussion on whether you need a moving average part and if you do, how do you determine the orders of autoregressive and moving average parts and what are appropriate ways of estimating the coefficients. When I teach time series analysis, I find that the computer programs generally use a procedure which was given in the book by Box and Jenkins in 1971 and it's a workable procedure. But there's no reason to believe that that's the best procedure either from a computational point of view or from a theoretical point of view. I have the feeling that there is a lot left to be done there. Of course, a lot is being done currently. Nevertheless, I am in the process

of starting to revise that 1971 time series book and it is a somewhat daunting prospect. In the next edition I do not intend to be exhaustive, but I would like to bring up to date the ARMA part of it and certain aspects of spectral analysis. But the field of time series has gotten too large to cover it all in one book. As I said, I don't think the field has really settled down. It's not yet time for definitive treatment.

There have been some major developments which have helped to ease the computational problem, particularly the Kalman filter and the construction of likelihood functions via that algorithm. But there still seem to be important practical problems of identification for vector ARMA models.

Yes. Vector autoregressive models involve a very large number of coefficients. Every additional order requires another matrix of coefficients and the number of components is the square of the size of the vector. Certainly in economics the number of observations that we have isn't going to permit a very large model. In the scalar case we have good ideas of how to balance the number of coefficients. We want this number to be large to reduce the bias, but we want it to be small to keep the variability down. How to balance one goal against the other in the vector case, it seems to me, is not clear as yet.

Have you looked at the work of Christopher Sims on the use of vector autoregressions in econometrics as an alternative modeling strategy?

Yes, I have looked at that some. I want to look at it much more closely because he has raised the question of whether in dealing with economic time series, vector autoregressive models are better to work with than simultaneous equations models. Simultaneous equations use economic theory whereas vector autoregressions really don't use theory to constrain the coefficients. Because of the problems that we just mentioned, I think that vector autoregressions are going to involve a lot of sampling variability, because there are so many coefficients and that's going to make it difficult to get useful results from such fitted models. Of course, the other side of it is that in simultaneous equations models, the structure of the economy keeps changing and the theory doesn't keep up with the changes in the structure. I think that neither one has been as useful as a statistician or an econometrician would hope, though. Simultaneous equations models have not been spectacularly successful in forecasting and the vector autoregressive models are not very useful for policy recommendations. There is certainly a lot more for econometricians to do in this area.

I remember Akaike told me once that he tried his information criterion with economic data and was most disappointed because every time it selected a first order vector autoregression. With data from other sciences, however, he found that there were much more interesting results.



Actually, his criterion theoretically tends to overestimate the order. So if the order 1 is an overestimate, then you are back to zero and it's independence. In teaching time series this last year, I was looking for examples of data that would give interesting, but simple, ARMA models. In particular, one of the things I teach is that if you have a second order autoregressive model and the roots of the associated polynomial equation are complex then you have a tendency to see cycles in the data. To illustrate this I wanted a set of data for which an estimated second order autoregression yields cycles, and I had a very hard time finding such a series. My TA looked at 15 or 20 series before he could find one.

Perhaps he should look at pre-Second World War data.

Oh, that's right. It makes a big difference. You could get that out of the 1919 to 1939–41 data. Also, I found in just generally applying ARMA models that you tend to get low orders. I don't think that's really because phenomenon are that simple, but I think that the mechanism is slowly changing and so the higher order terms are not well defined.

In 1973 *Econometrica* published your work with Takamitsu Sawa on the distribution of  $k$ -class estimators and their asymptotic expansions. This paper seemed to mark your return to active research in econometrics with a series of major articles during the 1970's and early 1980's. What prompted your renewed interest in econometric research and what made you select finite sample theory as the field to work in?

As I said earlier, when I went to Columbia, I didn't have colleagues who were particularly interested in simultaneous equations models. I found colleagues interested in other applications of statistics in the social sciences and I did work in other areas such as latent structural analysis and used Markov chains to represent panel studies in social research. When I moved to Stanford as a joint appointment in Economics and Statistics, I expected that there would be interest in simultaneous equations and econometrics generally and I had retained an interest in this but I hadn't been active. My actual re-involvement in this case came in 1968–69 when Bobby Mariano was a Ph. D student of mine in statistics. He was interested in econometrics and that was an opportunity to renew my active interest in the field. Going back again to my work with the Cowles Commission, I had the feeling that, for single equation methods, applying maximum likelihood would give the most efficient estimator. I had in the back of my mind the idea of investigating this question. Also, I had the idea that when you set up the single equation estimation problem, you have one matrix with a Wishart distribution, and another matrix with a noncentral Wishart distribution. I thought that the small sample distribution of estimators such as LIML or two stage least squares would involve these two matrices and that this would be the way to attack such a problem. So I had in mind this approach and Bobby Mariano

wanted a dissertation subject in econometrics, so I put him on to this problem. He did use the noncentral Wishart distribution and in the case of an equation with two endogenous variables the noncentrality parameter matrix is of rank 1 and just involved a Bessel function. It happened that Sawa had written a paper also on this subject and they had reached the results independently. Then they got together and published it. Because the resulting distribution came out in terms of, what was it, double infinite series, I could see that that wasn't going to tell us an awful lot, or at least, it wouldn't answer all the questions that I wanted to ask. So then I also suggested to Mariano that he do an asymptotic expansion. This was really the beginning of my work in small sample theory and asymptotic expansions. I was interested in trying to clear up the question of comparing two stage least squares and limited information maximum likelihood. I spent a lot of time and effort considering the simplest case, because I thought it was better to understand the simple case fairly thoroughly before going on to general cases where the mathematics would be much more difficult. Sawa came to Stanford in 1972–73. At that time I had an NSF grant in econometrics and had room for a research associate. I think it was Takeshi Amemiya who recommended him. We worked together on these problems, and we published the paper that you mentioned in 1973 and then the next year I published a paper on the asymptotic expansion of the LIML estimator and Sawa and Kimio Morimune and Naoto Kunitomo and I did a number of papers on distributions and asymptotic expansions. Then Sawa and later Morimune and later Kunitomo went back to Japan and with other colleagues there wrote many papers on these problems of distributions. I was impressed that Basman did his small sample theory work and yet never referred to the noncentral Wishart distribution. In effect, he worked out the noncentral distributions as he needed them.

Yes, he worked directly from the joint distribution of the reduced form coefficients and then by elementary mathematical transformations worked back to the structural parameter distributions via the transformations and the restrictions. It was D. G. Kabe, I think, who in 1963–64 formulated the problem so that the two stage least squares estimator was expressed as a function of the elements of a noncentral Wishart matrix in the  $2 \times 2$  case.

I always thought that the sufficient statistics were the error covariance matrix and reduced form coefficients. Methods like LIML and two stage least squares were based on the quadratic form that you get from the reduced form coefficients, so that gives you the noncentral Wishart matrix. I thought that these problems ought to be started at that point.

Your own recent work on LIML points to the (often markedly) superior performance of this procedure over 2SLS in finite samples. Do you feel that this helps to resolve a long standing issue over the choice of a single equation estimator in simultaneous systems?

In the case of two endogenous variables where you have just one coefficient estimate, Sawa and I published tables of the distributions for various values of the parameters. We did that for two stage least squares first and then with the help of Kunitomo, LIML. So in one sense you can make any comparison that you want because you have the numerical values there. We tabulated enough values so that with interpolation you can fill them in entirely. Then we have asymptotic expansions which give more structure. They show you what to look for. I think that with numerical comparisons, whether they are actual distributions or Monte Carlo results, in order to understand them, you have to put some structure into the problem. Asymptotic expansions tell you, or at least, suggest the kind of structure to look for. I think there is a very close relationship between exact numerical results and asymptotic expansions and what this work does is show that two stage least squares tends to be biased if there are many excluded exogenous variables, whereas LIML is only slightly biased. One way of putting it is that the LIML distribution is a little more spread out than that of two stage least squares, particularly if the number of excluded exogenous variables is large and if the standardized coefficient is different from zero, and in these cases two stage least squares has less variability but it's estimating the wrong thing. The graphs of densities in the Klein volume display these features. Our numerical results also show that for very special values of the parameters two stage least squares will do better, but that's sort of like saying that zero is an admissible estimator because if the parameter value happens to be zero, you've got a perfect estimator. Subsequent to that, of course, there is the higher order efficiency argument. This shows that if you consider centering in terms of the median instead of the mean then LIML is median unbiased and a  $k$ -class estimator will be median unbiased if you select the proper value of  $k$ . Then LIML as a median unbiased estimator dominates the  $k$ -class estimator. In theoretical terms, if you center LIML then it comes out to be best. I think that translates into nonrigorous terms as follows: if you want an estimator that really is sensitive to the data, follows the data, then LIML will be the best.

In a certain sense the results are probably more definitive than could be reasonably hoped for given the complexity of the problem. Since the distribution of LIML is itself well located about the true value of the parameter there isn't really much adjustment that's needed in order to make it mean or median unbiased up to a certain order in  $1/\sqrt{T}$ . So is your conclusion that the best an investigator can hope to do, on average, is just to apply the LIML procedure itself?

There is another modification of LIML we haven't mentioned, that was suggested by Fuller, where the root of the determinantal equation is made a little bit smaller and then the estimator will have moments and that is asymptotically equivalent to correcting for the bias of the LIML. So I would

say: if the sample size is large, use LIML, if you want to adjust for the bias use Fuller's modification or correct for the asymptotic bias another way. There is another aspect about this that I want to mention. These results that we are talking about are in terms of the normalization of an equation with one term having the coefficient unity. Now, the LIML estimation procedure doesn't depend on what coefficient you set equal to unity, in the sense that if you change the normalization, then you just change the estimators in the corresponding fashion. Other methods of estimation do depend upon that normalization. Now, originally, when Rubin and I were approaching this problem, we had in mind estimating a relationship and thought that it was an advantage to have a procedure which transformed the same way as you transform the normalization because the meaning of the equation is the same no matter what normalization you use. We didn't really fully appreciate that economists liked to set a particular coefficient equal to 1. I have been doing some historical work on R. A. Fisher's contributions to multivariate analysis and Fisher was very strong on having an estimation method that was invariant with regard to transformations or maybe to put it better, estimates that transform as the parameters, and that was perhaps a major reason why he put forward maximum likelihood as an estimation procedure. For example, in the case of the product moment correlation coefficient, he considered various transformations and the accepted Fisher z transformation was one of these. He pointed out that maximum likelihood was invariant with regard to such transformations. This general property of maximum likelihood occurs in the case of single equation estimators where it refers to transformations of the parameters or changes of the normalization.

If you go back to your original approach and normalize in a different way, perhaps using a quadratic form to define a manifold which contains the parameter vector, then you avoid the problem of the infinite moments of the LIML estimator that occur with the conventional normalization.

Yes. Infinite moments come up because you have a particular normalization and one of the normalizations that we considered was normalizing so the error variance was 1. That is symmetric in the coordinates of all the observed endogenous variables.

The field of finite sample theory has excited some negative reactions in the econometrics profession such as that of William Taylor in his article for *Econometric Reviews* in 1983, which you discussed. By contrast, the statistics profession seems to have maintained a very supportive view of the field with entire subject areas like that of multivariate analysis being deeply embedded in an intellectual tradition of investigations of this type. Do you see a reason for these differences in professional reaction?

I think that from the point of view of science that you investigate problems of interest in order to understand them and that you don't ask, at every

turn of the game, what use is it going to be, or whether you are going to be able to use it tomorrow? I think that this criticism of small sample theory is unscientific. Taylor wanted to examine the applicability of it all immediately instead of waiting for 5, 10 or 25 years, when we may understand the methodology better. I think that in statistics finite sample theory has been fundamental, that a large part of our current knowledge in statistics is based on this theory. I don't see why it should be essentially different in econometrics than in other areas in which statistics is applied. It reminds me of earlier criticisms that economists didn't see any point in the simultaneous equations approach and now, even if not totally accepted, it certainly is used in a large part of econometrics. I believe that time will show that his criticisms are irrelevant.

In the past 10 or 15 years you have had a steady output of first rate doctoral students in econometrics. Many of them have been young Japanese scholars. Can you tell us about your doctoral education program at Stanford: what mathematical background do your students typically possess; do they all follow a similar course structure; how do you help them to get started in their research; do you have a similar number of students in econometrics and in mathematical statistics?

Let me begin to answer this question by saying that the attitude in the economics profession towards mathematics and statistics, as you might say, has been revolutionized during the course of my career. When I was an undergraduate student, it was unusual for an economist to know any calculus and statistics was quite rudimentary. Economists typically would not understand regression, even simple regression. I would say that things have come a long way since then, and now many economists are very competent and well trained in statistics as well as mathematics. The graduate students that we get in economics generally at Stanford would have some mathematics and a good grounding in statistics, though ones that come into our program in econometrics are usually highly trained in mathematics and statistics. Most of the students that have worked with me and with Takeshi Amemiya take Master's Degrees in statistics; that means they take the equivalent of a full year of courses in the statistics department. Many of those students take basic courses that Ph.D.'s in statistics would take. I am thinking of probability at the level of Chung or statistics at the level of Ferguson, Lehmann and DeGroot. So they take very rigorous, sophisticated courses in statistics. Many of them will also take an advanced course in multivariate analysis or the advanced course in time series analysis. Of course, they take the basic requirements in economics, but the research that they do would be of a caliber that for most cases would be acceptable in the Department of Statistics. In the case of some students, like Asad Zaman, his dissertation could just as well have been in statistics as in economics. In other cases there are students such as Bobby Mariano whose dissertation was in statistics but it could just as

well have been in economics. It goes both ways. In some cases like Kunitomo, the motivation is so strongly from economics that it is more appropriate for it to be in the Economics Department. But now the university bureaucracy is such that to take some kind of an interdepartmental field for a Ph.D. is a hassle and students just don't do that. So a student more or less has to be in one department or the other. I think that the caliber of the students that I've gotten as Ph.D. students has been high in both economics and statistics. I wouldn't really want to make any comparison of them. I've had possibly more Ph.D. students in statistics than in economics since I have been at Stanford, but it's roughly the same. Of course, there are students in economics who do their dissertations with Amemiya but they have worked with me and eventually collaborate with me, such as Cheng Hsiao and Kimio Morimune. We've had a number of students from Japan, not only in economics but also in statistics. That's perhaps partly due to the fact that Amemiya is on the faculty, partly that we are on the West Coast of the United States and partly due to econometricians and statisticians in Japan recommending Stanford. In statistics, I recently had Akamichi Takamura; and both he and Kunitomo were recommended by Kei Takeuchi. I think that Sawa has recommended students to come to Stanford. So we have certainly done well with our Japanese connection.

Your own articles are famous for their rigor, their attention to detail and a clarity of exposition which makes difficult things seem easy. Are these strengths that you encourage your students to work towards? Or is it more a standard which perhaps they try to follow by way of example?

Well, I believe these aspects of rigor, detail and clarity of exposition are important. I think that an author of a research paper has a responsibility to communicate what he has done to other readers. I believe that science progresses more if the communication is made easier. It's unfair to the reader, as well as the editor, to put out papers which are difficult to read, not because of the difficulty of the material, but because of the sloppiness of the work and the carelessness in exposition. I do encourage students to consider these aspects and I am probably a tough supervisor. That discourages some students from working with me. When I read a draft of a dissertation, I try to insist on having the details right and for them to have at least a reasonable exposition. I was recently on a Ph.D. exam in which I went through the dissertation and got more and more upset and quite angry. I went to the exam after reading about one quarter of it because it was so hard to read, because of the lack of rigor and the lack of detail. I had gotten the draft I was reading the day before the exam and then during the exam when I started to ask my questions I was told by the candidate: "Well I've changed that and some things work better in this version," which was on the day of the exam. When that happened about the third time I just blew up and said, "look, I can't examine this if you are going to tell me everything in yesterday's version

was wrong and that you corrected them today.” That was the last straw on that exam.

Taking an overview of ongoing research in econometric theory and in mathematical statistics, would you single out any particular areas or developments as being of special significance for further research?

Well, we have already referred to some of these problems such as simultaneous equations models versus vector autoregressive models and other time series methods. There are problems of model selection within those areas and more generally. I think that one of the major aspects of current research in any statistical field is how to exploit the use of the computer. That is, not only in effective computation of methods that have been developed on the basis of some theory, but methods which use the computer in a more creative fashion. I am thinking of such things as the bootstrap method of evaluating variability, for example. I think econometricians do use some simulations as ways of trying out forecasting. Related to this are data analytic methods such as methods of visualizing multidimensional data, and nonparametric methods such as projection pursuit. These exploit the computer a lot. Data analytic methods, methods which are not within the context of significance levels and confidence levels are a kind of data mining. I think there’s more theory that is needed for that, especially in econometrics where the data may be limited.

Another ongoing field in statistics and econometrics is robust estimation which, of course, is getting a lot more attention now. We also haven’t talked about use of cross-section data. That may be a way of increasing the amount of statistical information that can be applied to economic problems. I think that the tying together of cross-section data and macrodata is important and there’s still much to be done along those lines. There are also a host of theoretical problems, such as these small sample problems, which are yet to be tackled. You, yourself, have been using a lot of this noncentral theory we have been talking about to get exact distributions. There is certainly a lot left to do there.

If you were a Ph.D. student, a beginning researcher, what topic would you now select for research?

It’s hard to answer that question because, for example, a Ph.D. student now would have a lot more background in computation than I would. So he’d be in a much better position to go into that area; my ideas are constrained by my own background. So it’s really hard for me to respond to that question. What I can say is what approaches I would tend to suggest to students. The students that work with me, on the one hand, have their own backgrounds and training, but on the other hand, they are going to draw on my expertise and use fields that I’ve worked in. If I myself were starting out again, I would be going in a somewhat different direction. There’s no doubt

that, in general terms, this area involving econometrics and statistics is a very fruitful one and has lots of research problems in it. There are many problems in econometrics like limited dependent variable models that I haven't made direct contributions to and these are other areas that ought to be mentioned. These days econometricians are very highly trained in mathematics and statistics; much more so than statisticians are trained in economics: and I think that there will be more cross-fertilization, more joint activity.

You are certainly a person who has crossed traditional academic boundaries by working in a variety of subjects, going from statistics to economics and to psychology. Looking back on your career of professional activity in these various fields, can you think of any particularly valuable instances of such cross-fertilization?

Well, there is a lot of work that goes on in theoretical statistics. I think that the eventual motivation in statistics comes from the applications, from what sense you can make of data in various substantive fields. In that sense the statistician, maybe not directly, but certainly indirectly, needs the contact with the problems that require statistical analysis. In my own case, I was stimulated first by wartime research and then by my contact with economics and simultaneous equations and also time series analysis, and as you mentioned, psychometrics and also sociology. I think it's important that statisticians do take up the problems of fields. I was going to say that at Stanford in the Statistics Department there are quite a few joint appointments, so we do make a big point of having this contact with fields of application: Brad Efron, Lin Moses and Rupert Miller are in the medical school; Paul Switzer is in geology; Tom Cover is in electrical engineering; and Jerry Friedman is at Stanford Linear Accelerator Center. So we have a lot of that contact going on and think it is important. Econometrics also benefits by having this close contact with statisticians.

Looking back over your career are there any major problems in mathematical statistics you would have liked to have tackled and solved?

There are a lot of problems that I had in mind working on at some time and other people have tackled them and carried them out. For instance, there was the problem of the selection of the order of an autoregression. Since I published a paper on the topic in 1963, there's been an awful lot done and I would like to come back to that problem but it's quite a jump to get into it again. We talked about autoregressive models and more general models with stochastic control and so on. That's gone a long way. And noncentral distributions, I had in the back of my mind to do something more there. Well, that's been pretty well taken care of. One opportunity, perhaps that I missed, was getting in on the theory of games at the inception. The theory of games was very closely related to statistical decision theory,



at least in the earlier days. And I didn't really get into that. If things had gone differently, if I had stayed at the Cowles Commission, then I would have done a lot more with simultaneous equations. And maybe I would have finished that paper with Hurwicz. The paper with Rubin on factor analysis would have been done a lot sooner. Shortly after I went to the Cowles Commission, I got an offer from Abraham Wald to go to Columbia where he was setting up a new department of mathematical statistics, and since he was the major figure in statistics at that time, that was an offer I couldn't refuse. Well, I went to Columbia, as I said earlier, and Wald had said that he wanted to work on some of these problems of econometrics, but then he got into decision theory. I could have really gotten into decision theory while that field was developing, but then Wald got killed in a plane crash in 1950 and that changed the whole setting. It is hard to say what would have happened if I had gone in other directions.

#### THE PUBLICATIONS OF THEODORE W. ANDERSON

##### BOOKS

- An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc., New York, 1958, xii + 374 pp. [Russian translation: *Vvedenie v Mnogomerni Statisticheski Analiz*, Gosudarstvennoe Izdatel'stvo Fizikomatematicheskoi Literatury, Moscow, 1963, 500 pp.] [Reprint: Wiley Eastern Private Limited, New Delhi, 1972.] Second Edition, John Wiley & Sons, Inc., New York, 1984, xvii + 675 pp.
- The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., New York, 1971, xiv + 704 pp. [Russian translation: *Statisticheski Analiz Vremennykh Rjadov*, Izdatelstvo "MIR," Moscow, 1976, 755 pp.]
- A Bibliography of Multivariate Statistical Analysis* (with Somesh Das Gupta and George P. H. Styan), Oliver & Boyd, Edinburgh, and Halsted Press, 1972, x + 642 pp. [Reprinted: Robert E. Krieger Publishing Co., Malibar, Florida, 1977.]
- Introductory Statistical Analysis* (with Stanley L. Sclove), Houghton Mifflin Co., Boston, 1974, xv + 499 pp.]
- An Introduction to the Statistical Analysis of Data* (with Stanley L. Sclove), Houghton Mifflin Co., Boston, 1978, xvi + 704 pp. Second Edition, The Scientific Press, Palo Alto, 1986.
- MINITAB Guide to the Statistical Analysis of Data* (with Barrett P. Eynon), The Scientific Press, Palo Alto, 1986, viii + 203 pp.

##### ARTICLES

###### 1943

- "Some significance tests for normal bivariate distributions" (with D. S. Villars), *Annals of Mathematical Statistics*, Vol. 14, pp. 141-148.
- "On card matching," *Annals of Mathematical Statistics*, Vol. 14, pp. 426-435.

###### 1944

- "Some extensions of the Wishart distribution" (with M. A. Girshick), *Annals of Mathematical Statistics*, Vol. 15, pp. 345-357. (Correction, Vol. 35, 1964, pp. 923-924.)

## 1946

"The non-central Wishart distribution and certain problems of multivariate statistics," *Annals of Mathematical Statistics*, Vol. 17, pp. 409–431. (Correction, Vol. 35, 1964, pp. 923–924.)

## 1947

"A note on a maximum-likelihood estimate," *Econometrica*, Vol. 15, pp. 241–244.

## 1948

"On the theory of testing serial correlation," *Skandinavisk Aktuarietidskrift*, Vol. 31, pp. 88–116.  
 "The asymptotic distributions of the roots of certain determinantal equations," *Journal of the Royal Statistical Society, Series B*, Vol. 10, pp. 132–139.

## 1949

"Estimation of the parameters of a single equation in a complete system of stochastic equations" (with Herman Rubin), *Annals of Mathematical Statistics*, Vol. 20, pp. 46–63. [Reprinted in *Readings in Econometric Theory* (J. Malcolm Dowling and Fred R. Glahe, eds.), Colorado Associated University Press, 1970, pp. 358–375.]

## 1950

"Distribution of the circular serial correlation coefficient for residuals from a fitted Fourier series" (with R. L. Anderson), *Annals of Mathematical Statistics*, Vol. 21, pp. 59–81.  
 "Estimation of the parameters of a single equation by the limited-information maximum-likelihood method," *Statistical Inference in Dynamic Economic Models* (Tjalling C. Koopmans, ed.), John Wiley & Sons, Inc., New York, pp. 311–322.  
 "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations" (with Herman Rubin), *Annals of Mathematical Statistics*, Vol. 21, pp. 570–582. [Reprinted in *Readings in Econometric Theory* (J. Malcolm Dowling and Fred R. Glahe, eds.), Colorado Associated University Press, 1970, pp. 376–388.]

## 1951

"Classification by multivariate analysis," *Psychometrika*, Vol. 16, pp. 31–50.  
 "The asymptotic distribution of certain characteristic roots and vectors," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), University of California Press, Berkeley, 1951, pp. 103–130.  
 "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *Annals of Mathematical Statistics*, Vol. 22, pp. 327–351. [Correction, *Annals of Statistics*, Vol. 8 (1980), p. 1400.]

## 1952

"Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes" (with D. A. Darling), *Annals of Mathematical Statistics*, Vol. 23, pp. 193–212.

## 1954

"Probability models for analyzing time changes in attitudes," *Mathematical Thinking in the Social Sciences* (Paul F. Lazarsfeld, ed.), The Free Press, Glencoe, Illinois, pp. 17–66, 418–419.  
 "On estimation of parameters in latent structure analysis," *Psychometrika*, Vol. 19, pp. 1–10.  
 "A test of goodness of fit" (with D. A. Darling), *Journal of the American Statistical Association*, Vol. 49, pp. 765–769.

## 1955

- "Some statistical problems in relating experimental data to predicting performance of a production process," *Journal of the American Statistical Association*, Vol. 50, pp. 163–177.
- "The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities," *Proceedings of the American Mathematical Society*, Vol. 6, pp. 170–176.
- "The Department of Mathematical Statistics," *A History of the Faculty of Political Science, Columbia University*, Columbia University Press, New York, pp. 250–255.
- "Some recent results in latent structure analysis," *Proceedings of the Invitational Conference on Testing Problems, October 30, 1954*, Educational Testing Service, Princeton, New Jersey, pp. 49–53.

## 1956

- "Statistical inference in factor analysis" (with Herman Rubin), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), University of California Press, Berkeley, Vol. 5, pp. 111–150.

## 1957

- "Statistical inference about Markov chains" (with Leo A. Goodman), *Annals of Mathematical Statistics*, Vol. 28, pp. 89–110. [Reprinted in *Readings in Mathematical Psychology* (R. Duncan Luce, Robert R. Bush, and Eugene Galanter, eds.), John Wiley & Sons, Inc., New York, Vol. 1, 1963, pp. 241–262.]
- "Maximum likelihood estimates for a multivariate normal distribution when some observations are missing," *Journal of the American Statistical Association*, Vol. 52, pp. 200–203.

## 1959

- "On asymptotic distributions of estimates of parameters of stochastic difference equations," *Annals of Mathematical Statistics*, Vol. 30, pp. 676–687.
- "Some scaling models and estimation procedures in the latent class model," *Probability and Statistics: The Harold Cramér Volume* (Ulf Grenander, ed.), Almqvist and Wiksell, Stockholm, pp. 9–38.

## 1960

- "A modification of the sequential probability ratio test to reduce the sample size," *Annals of Mathematical Statistics*, Vol. 31, pp. 165–197.
- "Harold Hotelling's research in statistics," *The American Statistician*, Vol. 14, pp. 17–21.
- "A limitation of the optimum property of the sequential probability ratio test" (with Milton Friedman), *Contributions to Probability and Statistics; Essays in Honor of Harold Hotelling* (Ingram Olkin, Sudhish G. Ghurye, Wassily Hoeffding, William G. Madow, and Henry B. Mann, eds.), Stanford University Press, Stanford, California, pp. 57–69.
- "Some stochastic process models for intelligence test scores," *Mathematical Methods in the Social Sciences* (Kenneth J. Arrow, Samuel Karlin, and Patrick Suppes, eds.), Stanford University Press, Stanford, California, pp. 205–220.

## 1962

- "The choice of the degree of a polynomial regression as a multiple decision problem," *Annals of Mathematical Statistics*, Vol. 33, pp. 255–265.
- "Least squares and best unbiased estimates," *Annals of Mathematical Statistics*, Vol. 33, pp. 266–272.
- "Classification into two multivariate normal distributions with different covariance matrices" (with R. R. Bahadur), *Annals of Mathematical Statistics*, Vol. 33, pp. 420–431.

"On the distribution of the two-sample Cramér-von Mises criterion," *Annals of Mathematical Statistics*, Vol. 33, pp. 1148–1159.

## 1963

"The use of factor analysis in the statistical analysis of multiple time series," *Psychometrika*, Vol. 28, pp. 1–25.

"Asymptotic theory for principal component analysis," *Annals of Mathematical Statistics*, Vol. 34, pp. 122–148.

"A test for equality of means when covariance matrices are unequal," *Annals of Mathematical Statistics*, Vol. 34, pp. 671–672.

"Determination of the order of dependence in normally distributed time series," *Proceedings of the Symposium on Time Series Analysis* (M. Rosenblatt, ed.), John Wiley & Sons, Inc., New York, pp. 425–446.

"Some inequalities on characteristic roots of matrices" (with S. Das Gupta), *Biometrika*, Vol. 50, pp. 522–524. [Correction: Vol. 52 (1965), p. 669.]

## 1964

"Monotonicity of the power functions of some tests of the multivariate linear hypothesis" (with S. Das Gupta and G. S. Mudholkar), *Annals of Mathematical Statistics*, Vol. 35, pp. 200–205.

"Monotonicity of the power functions of some tests of independence between two sets of variates" (with S. Das Gupta), *Annals of Mathematical Statistics*, Vol. 35, pp. 206–208.

"Some approaches to the statistical analysis of time series," *Australian Journal of Statistics*, Vol. 6, pp. 1–11.

"A monotonicity property of the power functions of some tests of the equality of two covariance matrices" (with S. Das Gupta), *Annals of Mathematical Statistics*, Vol. 35, pp. 1059–1063.

"On Bayes procedures for a problem with choice of observations," *Annals of Mathematical Statistics*, Vol. 35, pp. 1128–1135.

"On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process" (with A. M. Walker), *Annals of Mathematical Statistics*, Vol. 35, pp. 1296–1303.

"Sequential analysis with delayed observations," *Journal of the American Statistical Association*, Vol. 59, pp. 1006–1015.

## 1965

"Samuel Stanley Wilks, 1906–1964," *Annals of Mathematical Statistics*, Vol. 36, pp. 1–27. [Reprinted in S. S. Wilks: *Collected Papers, Contributions to Mathematical Statistics* (T. W. Anderson, ed.), John Wiley & Sons, Inc., New York, 1967, pp. vii–xxvi.]

"Some optimum confidence bounds for roots of determinantal equations," *Annals of Mathematical Statistics*, Vol. 36, pp. 468–488.

"Some properties of confidence regions and tests of parameters in multivariate distributions" (with discussion), *Proceedings of the IBM Scientific Computing Symposium in Statistics, October 21–23, 1963*, IBM Data Processing Division, White Plains, New York, pp. 15–28.

## 1966

"Some nonparametric multivariate procedures based on statistically equivalent blocks," *Multivariate Analysis* (Paruchuri R. Krishnaiah, ed.), Academic Press, Inc., New York, pp. 5–27.

## 1967

"Some inequalities among binomial and Poisson probabilities" (with Stephen M. Samuels), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Lucien LeCam and Jerzy Neyman, eds.), Vol. 1, University of California Press, pp. 1–12.

"Approximating the upper binomial confidence limit" (with Herman Burstein), *Journal of the American Statistical Association*, Vol. 62, pp. 857-861.

### 1968

"Multivariate analysis: classification and discrimination," *International Encyclopedia of the Social Sciences* (David Sills, ed.), The Macmillan Co. and the Free Press, Vol. 10, pp. 553-559. Updated: *International Encyclopedia of Statistics* (William H. Kruskal and Judith M. Tanur, eds.), The Free Press, New York, 1978, Vol. 1, pp. 628-635.

"Large-sample distribution theory for estimates of the parameters of a latent class model," *Latent Structure Analysis* (by P. F. Lazarsfeld and Neil Henry), Houghton Mifflin, Appendix B, pp. 273-287.

"Approximating the lower binomial confidence limit" (with Herman Burstein) *Journal of the American Statistical Association*, Vol. 63, pp. 1413-1415. [Editor's correction, Vol. 64 (1969), p. 669.]

### 1969

"Statistical inference for covariance matrices with linear structure," *Multivariate Analysis II* (P. R. Krishnaiah, ed.), Academic Press, Inc., New York, pp. 55-66.

"Confidence limits for the expected value of an arbitrary bounded random variable with a continuous distribution function," *Bulletin of the International Statistical Institute*, Vol. 43, Book 2, pp. 249-251.

"Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices," *Essays in Probability and Statistics* (R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao, and K.J.C. Smith, eds.), Statistical Publishing Society, Calcutta, and University of North Carolina Press, Chapel Hill, 1970, pp. 1-24.

### 1970

"Some notes on doing research in mathematical statistics," *Scientists at Work: Festschrift in Honour of Herman Wold* (T. Dalenius, G. Karlsson, and S. Malmquist, eds.), Almqvist and Wiksell Book Co., Uppsala, pp. 21-27.

### 1972

"Efficient estimation of regression coefficients in time series," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Lucien M. LeCam, Jerzy Neyman, and Elizabeth L. Scott, eds.), Vol. 1, University of California Press, pp. 471-482.

"Tests for randomness of directions against equatorial and bimodal alternatives" (with M. A. Stephens), *Biometrika*, Vol. 59, pp. 613-621.

### 1973

"Asymptotically efficient estimation of covariance matrices with linear structure," *Annals of Statistics*, Vol. 1, pp. 135-141.

"Asymptotic evaluation of the probabilities of misclassification by linear discriminant functions," *Discriminant Analysis and Applications* (T. Cacoullos, ed.), Academic Press, Inc., New York, pp. 17-35.

"An asymptotic expansion of the distribution of the Studentized classification statistic  $W$ ," *Annals of Statistics*, Vol. 1, pp. 964-972.

"Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions" (with Takamitsu Sawa), *Econometrica*, Vol. 41, pp. 683-714.

### 1974

"An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system," *Journal of the American Statistical Association*, Vol. 69, pp. 565-573. [Correction: Vol. 71 (1976), p. 1010.]

## 1975

"Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure," *Annals of Statistics*, Vol. 3, pp. 1283–1304.

## 1976

"Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics" (with discussion), *Journal of the Royal Statistical Society, Series B*, Vol. 38, pp. 1–36.

"Strong consistency of least squares estimates in normal linear regressions" (with John B. Taylor), *Annals of Statistics*, Vol. 4, pp. 788–790.

"Some experimental results on the statistical properties of least squares estimates in control problems" (with John B. Taylor), *Econometrica*, Vol. 44, pp. 1289–1302.

## 1977

"Two-stage least squares: In which direction should the residuals be minimized?" (with Takamitsu Sawa), *Journal of the American Statistical Association*, Vol. 72, pp. 187–191.

"Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences," *Econometrica*, Vol. 45, pp. 509–518.

"A comment on the test of overidentifying restrictions" (with Joseph B. Kadane), *Econometrica*, Vol. 45, pp. 1027–1031.

"Estimation for autoregressive moving average models in the time and frequency domains," *Annals of Statistics*, Vol. 5, pp. 842–865.

"Identification of parameters by the distribution of a maximum random variable" (with S. G. Ghurye), *Journal of the Royal Statistical Society, Series B*, Vol. 39, pp. 337–342.

"On maximum likelihood estimation of parameters of autoregressive moving average processes," *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians*, Academia, Publishing House of the Czechoslovak Academy of Sciences, Prague, and D. Reidel, Dordrecht, pp. 37–45.

"The generalized variance of a stationary autoregressive process" (with Raul P. Mentz), *Journal of Multivariate Analysis*, Vol. 7, pp. 584–588.

## 1978

"A new development in multivariate statistical analysis," *Journal of the Japan Statistical Society*, Vol. 8, pp. 9–13.

"Unique factorization of products of bivariate normal cumulative distribution functions" (with S. G. Ghurye), *Annals of the Institute of Statistical Mathematics*, Vol. 30, pp. 63–69.

"Repeated measurements on autoregressive processes," *Journal of the American Statistical Association*, Vol. 73, pp. 371–378.

"An extremal problem for positive definite matrices" (with I. Olkin), *Linear and Multilinear Algebra*, Vol. 6, pp. 257–262.

## 1979

"Evaluation of the distribution function of the two-stage least squares estimate" (with Takamitsu Sawa), *Econometrica*, Vol. 47, pp. 163–182.

"Pao-Lu Hsu 1909–1970" (with K. L. Chung and E. L. Lehmann), *Annals of Statistics*, Vol. 7, pp. 467–470. [Updated: *Pao-Lu Hsu Collected Papers* (Kai-Lai Chung, ed.), Springer Verlag, New York, 1983, pp. 1–4.] [Chinese translation: *Knowledge and Practice of Mathematics*, No. 3 (1980), pp. 3–5.]

"Hsu's work in multivariate analysis," *Annals of Statistics*, Vol. 7, pp. 474–478. [Reprinted in *Pao-Lu Hsu Collected Papers* (Kai-lai Chung, ed.), Springer-Verlag, New York, 1983, pp.

8–12.] [Chinese translation: *Knowledge and Practice of Mathematics*, No. 3 (1980), pp. 8–12.]

- “Strong consistency of least squares estimates in dynamic models” (with John B. Taylor), *Annals of Statistics*, Vol. 7, pp. 484–489.
- “Panels and time series analysis: Markov chains and autoregressive processes,” *Qualitative and Quantitative Social Research* (J. S. Coleman, R. K. Merton, and Peter H. Rossi, eds.), The Free Press, New York, pp. 82–97.
- “Some relations between Markov chains and vector autoregressive processes,” *International Statistical Institute: Contributed Papers, 42nd Session, December 4–14, 1979*, Manila, pp. 25–28.

### 1980

- “An inequality for a sum of quadratic forms with applications to probability theory” (with John B. Taylor), *Linear Algebra and Its Applications*, Vol. 30, pp. 93–99.
- “Recent results on the estimation of a linear functional relationship,” *Multivariate Analysis V* (P. R. Krishnaiah, ed.), North-Holland Publishing Co., Amsterdam, pp. 23–34.
- “Maximum likelihood estimation for vector autoregressive moving average models,” *Directions in Time Series* (D. R. Brillinger and G. C. Tiao, eds.), Institute of Mathematical Statistics, Hayward, California, pp. 49–59.
- “Finite-state Markov chains and vector autoregressive process,” *Proceedings of the Conference on Recent Developments in Statistical Methods and Applications*, Directorate-General of Budget, Accounting and Statistics, Executive Yuan, Taipei, Taiwan, Republic of China, pp. 1–12.
- “On the structure of the likelihood function of autoregressive and moving average models” (with Raul P. Mentz), *Journal of Time Series Analysis*, Vol. 1, pp. 83–94.

### 1981

- “Estimation of dynamic models with error components” (with Cheng Hsiao), *Journal of the American Statistical Association*, Vol. 76, pp. 598–606.

### 1982

- “Maximum likelihood estimation in autoregressive and moving average models” (with Raul P. Mentz), *Time Series Analysis: Theory and Practice 1* (O. D. Anderson, ed.), North-Holland Publishing Co., Amsterdam, pp. 23–29.
- “Cochran’s theorem, rank additivity, and tripotent matrices” (with George P. H. Styan), *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, eds.), North-Holland Publishing Co., Amsterdam, pp. 1–23.
- “Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient” (with Takamitsu Sawa), *Journal of the Royal Statistical Society Series B*, Vol. 44, pp. 52–62.
- “Formulation and estimation of dynamic models using panel data” (with Cheng Hsiao), *Journal of Econometrics*, Vol. 18, pp. 47–82.
- “Sampling permutations for nonparametric methods,” *Statistics in Theory and Practice: Essays in Honour of Bertil Matern* (Bo Ranneby, ed.), Swedish University of Agricultural Sciences, Umeå, pp. 43–52.
- “Evaluation of the distribution function of the limited information maximum likelihood estimator” (with Naoto Kunitomo and Takamitsu Sawa), *Econometrica*, Vol. 50, pp. 1009–1027.
- “Some recent developments on the distributions of single-equation estimators,” *Advances in Econometrics* (Werner Hildenbrand, ed.), Cambridge University Press, Cambridge, pp. 109–122.
- “A new proof of admissibility of tests in the multivariate analysis of variance” (with Akimichi Takemura), *Journal of Multivariate Analysis*, Vol. 12, pp. 457–468.

## 1983

- "Notes on the estimation of parameters in vector autoregressive models" (with Raul P. Mentz), *A Festschrift for Erich L. Lehmann* (Peter J. Bickel, Kjell A. Doksum, and J. L. Hodges, Jr., eds.), Wadsworth, Inc., Belmont, California, pp. 1–13.
- "The numerical values of some key parameters in econometric models" (with Kimio Morimune and Takamitsu Sawa), *Journal of Econometrics*, Vol. 21, pp. 229–243.
- "Comparison of the densities of the TOLS and LIMLK estimators" (with Naoto Kunitomo and Takamitsu Sawa), *Global Econometrics, Essays in Honor of Lawrence R. Klein* (F. Gerard Adams and Bert G. Hickman, eds.), MIT Press, Cambridge, Massachusetts, pp. 103–124.

## 1984

- "Estimating linear statistical relationships" (1982 Abraham Wald Memorial Lectures), *Annals of Statistics*, Vol. 12, pp. 1–45.

## 1985

- "Best invariant estimation of a direction parameter" (with Charles Stein and Asad Zaman), *Annals of Statistics*, Vol. 13, pp. 526–533.
- "Components of variance in MANOVA," *Multivariate Analysis-VI* (P.R. Krishnaiah, ed.), North-Holland, Amsterdam, pp. 1–8.
- "Maximum likelihood estimation of the parameters of a multivariate normal distribution" (with I. Olkin), *Linear Algebra and Its Applications*, Vol. 70, pp. 147–170.

## 1986

- "Comparing single equation estimators in a simultaneous equation system" (with Naoto Kunitomo and Kimio Morimune), *Econometric Theory*, Vol. 2, pp. 1–32.
- "Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance" (with Blair M. Anderson and Ingram Olkin), *Annals of Statistics*, Vol. 14, pp. 405–417.
- "Maximum likelihood estimators and likelihood ratio criteria for multivariate elliptically contoured distributions" (with Kai-Tai Fang and Huang Hsu), *Canadian Journal of Statistics*, Vol. 14, pp. 55–59.
- "Why do noninvertible estimated moving averages occur?" (with Akimichi Takemura), *Journal of Time Series Analysis*, Vol. 7, in press.

## MISCELLANEOUS PUBLICATIONS AND PAPERS

This is a short list of additional papers and commentaries relevant to econometrics and to the Interview (Ed.).

- "Statistical models with disturbances in equations and/or disturbances in variables" (with Leonid Hurwicz), Cowles Commission for Research in Economics, June 1946. [Resume: "Errors and shocks in economic relationships," *Econometrica*, Vol. 17, Supplement (1949), pp. 23–25, also *Bulletin of the International Statistical Institute*, Vol. 31, Part 5, pp. 23–25.]
- Comments on "On the relevance of finite sample distribution theory," by William E. Taylor, *Econometric Reviews, Communications in Statistics* (Series D), 1983.
- Discussion of "Statistical inference in relation to the size of the model," by G. S. Maddala, *Large-Scale Macro-Econometric Models* (J. Kmenta and J. B. Ramsey, eds.), North-Holland Publishing Co., 1981.