

Perspectives of Statistics as a Scientific Subject at a University

Helmut Strasser
Vienna University of Economics and Business Administration

Lecture, given on Nov. 5th, 2003, at the evaluation workshop of the Department of Statistics, Vienna University of Economics and Business Administration

1 Introduction

Dear Mrs. Vicerector, dear colleagues!

I would like to thank you on behalf of the department of statistics for taking part in the presentation of our research. For us this evaluation is a welcome opportunity to stop a little in the everyday business and to think about us and our role. It will be a large profit for us if you accompany and support us.

My task at this talk superficially consists in introducing the department of statistics to you and in giving you a first summary of the research performance and the research strategies of the institute. But much more detailed information will be presented to you within the next two days, both by personal conversation and by a small presentation programme. Therefore in my talk I would like to take a slightly more general point of view.

If we want to describe the role and the condition of a university department, then several points are of importance. The department supports a scientific subject in two different ways namely in research and education. Furthermore the department is integrated into an environment which permanently changes. One part of the environment is the scientific community in which the interplay of cooperation and competition is in the foreground, and the other part is the public which provides the financial resources of the research and consumes its results. For us the university is also a part of the public, since it consists of potential customers of our research.

Between these cornerstones there are complicated relations, each of which giving sufficient material for a study in sociology of science. But since today our topic is research, I will confine myself to the relations between research and the remaining cornerstones.

I will briefly talk about the difficulties of research conducted teaching. Then I will say some words about the public perception of statistics as a science, particularly in the view of our customers. The main part of my talk will be devoted to remarks about my own view of the role and function of statistics as a scientific subject.

All these questions could be discussed both on a general level and on a special level. On the one hand the present position of the department depends on the general role and reputation of statistics as a whole. On the other side there are very special historical and institutional framework conditions which have an effect on a single university department. Both have to be taken into account if one wants to evaluate the present situation of a department. Only if we understand both the general and the special framework conditions, then we are able to develop convincing strategic options for the future.

2 Local Framework Conditions

I will now start with a brief description of the historical and institutional framework conditions. Today the department is relatively large, it consists of about 14 scientists, having teaching authority and doing scientific research in their own responsibility. This was not always the case.

About 35 years ago, the former university for world trade put the mathematics and statistics education, which wasn't really established until then, into the hands of Prof. Roppert. Prof. Roppert is the founder and long-standing head of the department and I am particularly pleased that he stays with us today. After and after, with the growth of the university, more academic positions were added to the department. At the moment the department has four positions of full professors one of which, however, is vacant.

For a long time the department has been organized in the traditional way of time-honoured European universities, namely according to the chair principle. The chair, later called division, formed the smallest operative unit with a pyramid-shaped organisational structure and with a peak consisting of one single person, the chair owner. Several political attempts to change this system failed because of the realities in Austria.

At the moment the traditional university landscape in Austria is changing in a very radical way, but this opens up innovative possibilities for designing the department's future. The institute for statistics has the opportunity, to convert into a modern research and service department, where scientists being responsible for themselves but with strongly related interests, formulate common aims and attempt to realized them with common strategies. From the present evaluation and thus from our colleagues acting as experts we expect advice and help with the execution of this redesign process.

It is very important for me to stress a point which is perhaps a little strange for colleagues from the international community. It is the fact that at our university there is no degree in statistics. To explain this fact, both the causes and the consequences, I must go a little way back in my story. This is necessary because the complete lack of such a degree in statistics is on the one hand very unusual in the international comparison, and on the other hand it has serious consequences for the research strategies of the department. In particular, the research conducted teaching comes off very badly.

It is widely known that Great Britain and the USA are leading in the field of statistics. Although in the US this leading role has been built up by numerous emigrants from Europe, it is the valuable contribution of the US to have given to statistics that recognized status in the international science business which statistics till then had merely in Great Britain, but which was refused to statistics more or less in the rest of Europe.

In the German area for a long time statistics has been considered exclusively as a support science of the social and economic sciences. However, at the beginning of the seventies Germany made an attempt to establish statistics in universities more strongly. On the one hand the mathematical departments were forced by the political authorities to write out chairs in statistics on a large scale. On the other hand a few fields of studies were founded with graduation in statistics. In English-speaking countries, especially in the USA, such studies have been quite usual for a long time. In Germany up to now there are no more than two flourishing studies of this type, namely in Munich and in Dortmund. We are very happy that one of our experts comes from one of those departments.

In Austria any attempts to offer studies with statistics degrees have never achieved a chance of long-term survival. In Austria statistics at universities is dependent on the possibility that in courses of existing studies divisions are allowed with low student numbers, where emphasis can be given to specializing in statistics. However, our recent university reforms go into a completely different direction. They reduce the flexibility of university studies instead of increasing it. This goes at the expense of very specialized subjects whose contents cannot be taught in mass studies.

It should be undisputed that also such subjects are indispensable for the science landscape which aren't studied by masses of students. I would like to explain why statistics is part of these indispensable scientific disciplines.

3 The Perception of the Subject Statistics

Why are such justifying arguments necessary at all? The reason is that scientific statistics has a problem: That is the perception of the subject from the view of the public, and from the view of the scientific customers.

The public view again is a German speciality. In German, a bureau of census is called a statistics office. Obviously, this is only an amusing misuse of language, but it leads the public view of scientific statistics into a completely wrong direction.

However, much more disturbing for us is the perception of statistics in the minds of our scientific customers. These customers are the scientists that are working empirically. Their view of statistics is distorted and partly completely inappropriate.

Many scientists working with empirical data have acquired as autodidacts a narrow selection of so-called statistical methods. The curricular structure at universities still promotes this way. The idea of how statistical instruments work, mostly consists in a very doubtful user illusion. Disappointments arise as a necessary consequence, and are ascribed to the subject statistics. The failure of empirical research by unsatisfactory and incompetent dealing with statistical instruments is finally avoided by doubtful publication strategies, such as cheating with data and interpretation.

One of the deeper reasons for this situation is probably the deplorable condition of the mathematical and statistical education. As a result, a young scientist faces a problem if he wants to achieve an economically sound position within an appropriate time. He has to put priorities. Shall he make up in years of effort those methodological foundations which would enable him to understand really what he uses as a technical instrument in his empirical research? He won't do it, he cannot do it, he may not do it, if he doesn't want to risk his career and with that his economic existence.

Of course, inadequate dealing with statistical methods does not imply immediate dangers. No bridges collapse and no aeroplanes crash. It may happen that an important medicine comes onto the market too late, it can be that wrong economic or business decisions are taken. However, the immediate causal connection between bad statistics and wrong decisions with fatal consequences is usually invisible.

In case you think that my remarks reflect a one-sided pessimistic philosophy of life I would like to tell you an anecdote. Already 15 or 20 years ago at a special conference in a famous research institute the most powerful man in the field of statistics at that time has

called a special session at a nightly hour. It was Prof. van Zwet, being both in charge of the *Annals of Statistics*, the undisputed flagship of the statistics journals, and also chairman of the IMS, which is the international union of the theoretically working statisticians. All of us were very curious to see what might be the topic of this special session. When we were assembled, he went silently to the blackboard and wrote up the following sentence with chalk: Why are we not popular?

So, at that time, the problem was already known. Today, the question is more relevant than ever.

4 Scientific Statistics in the Spectrum of Sciences

I would like to explain now, how I see the subject statistics and its role in the spectrum of sciences.

The relations between scientific subjects are governed by competition and cooperation. This can take place among single scientists or among departments, and then remains mostly within the limits of the subject. However, the relation can also be discussed at the more general level of the subjects themselves: Which scientific approach provides to mankind the most promising concepts for solving the burning questions?

In the following I would like to talk about the problem at this more general level. I am deeply convinced that in the political discussion there is no sufficient attention devoted to the general position of statistics. My claim will be: Scientific statistics plays a pivotal role in the spectrum of sciences. It is an indispensable component of the scientific progress in all fields of empirical research.

4.1 Statistics and the Cognition Problem

I would like to start with two propositions. The first is:

The object of statistics is the cognition problem, i.e. the formation of scientific theories based on empirical data.

The second proposition is:

The process of scientific cognition has great analogies to the process of the everyday perception by sensual impressions, i.e. how sensual impressions lead to knowledge about things.

Now, I am going to explain these propositions in somewhat greater detail.

It was Galileo Galilei, who suggested that cognition in the natural sciences should not be based on wishes and ideologies but rather on empirical data and conscientious dealing with those. About the same time the philosophical cognition theory started to discuss the question whether empirical data alone can lead to perceptual cognition, or whether certain ideas about the reality, which are existing from the start as a prejudice or a model, play the major part in the cognition process. Soon, there was a struggle between extreme philosophical positions for this perceptual cognition problem.

On the one hand there were the so-called empirists, represented by Hume in the purest form. Roughly speaking, they thought that in principle a machine would be conceivable

which is fed on one side with an empirical data stream and which produces cognition on the other side as output.

On the other hand the so-called rationalists fought for the opinion that cognition is produced essentially by mental ideas.

It was Immanuel Kant who realized that these two extreme viewpoints were philosophical reductions. He recognized that the perceptual cognition process is characterized by an interaction between mental models, coming from biological evolution, and sensual impressions. On the top, human reason is acting as a referee between mental models and sensual impressions, modifying models, supporting them or questioning them.

Now, what does such a philosophical discourse about sensual perceptions have to do with the practical reality in the everyday science business? Why do I tell all that to you?

Indeed, statistical methods do exactly the same for the scientific cognition process (of course I mean here the empirical sciences) that the philosophical cognition theory ascribes to human reason within the process of perceptual cognition.

Of course, statistics does not deal with sensual impressions and their everyday interpretation, but rather with such empirical data which cannot be interpreted by senses. Instead of sensual organs coming from biological evolution statistics applies technical instruments, mainly from information technologies. Such technical instruments were not or only very little available in the beginnings of statistics. Therefore the early information technologies consisted in counting data and making summarizing diagrams with paper and pencil. However, for about 30 years we have undergone an explosive expansion of the technical information processing. Of course, this expansion left its marks on statistics.

For the interpretation of the data, statistics applies theories and models which aren't available from biological evolution as a priori models, but other models which have been produced by creative human reason in the history of civilization of mankind, in particular models in terms of mathematics. Indeed, mathematics is that science, which is concerned with everything what can be constructed and built by human reason, purely based on the laws of logic.

Statistics plays the role of a link between the intellectual models of abstract mathematics and the empirical data which we obtain from the instruments of information processing. Using the picture of sensual perception as an analogy, statistics plays the role of human reason which in the end takes the decision, which mental model of the reality should be assumed as a valid explanation of the empirical data.

4.2 The Historic Formation of Scientific Statistics

However, statistics as an independent science didn't already arise with the emergence of the natural sciences. In classical natural sciences the characteristic logical problems of the cognition process don't get visible completely. And this leads to my next proposition:

It was the cognition problem in the social and economic sciences, which led to the formation of a new scientific discipline being concerned with the cognition process itself, namely the field of statistics.

Statistics arose together with the emergence of the social and economic sciences in the 19th century. An early claim was to develop something like physics of the society. However, soon it was realized, that the data situation was different from that in the natural

sciences. Primarily, there are two important differences.

First, in the social and economic sciences, stochastic noise is much louder than in the natural sciences. And something else is also different from natural sciences: There is a considerable shortage of data, especially relative to the uncertainty of the mathematical models which shall be calibrated and selected. Thus, sample size is a cost factor of considerably higher importance in the social and economic sciences.

These two differences are the reason why a new methodological basis for empirical research had to be worked out.

It was Karl Popper in his *Logik der Forschung*, who described convincingly the interplay between the hypothetical formation of models on the one hand, and on the other hand the validation of these models at the reality of empirical data. Popper made clear that it isn't possible to verify hypothetical theories by empirical data. On the other hand he realized that falsification of theories is possible and that falsification is the corner stone for the progress of the natural scientific theories. However, in the presence of stochastic noise the situation is different. In the presence of stochastic noise also falsification of theories must be questioned. The statistical testing theory by Neyman and Pearson gives the necessary and final answers to this issue.

Nowadays it is generally accepted, that statistical decision methods have to make sure that erroneous falsifications are practically impossible in the process of statistical modelling. The label is the concept of significance.

The second difference, lacking of data, led to the issue of the efficiency of statistical methods, a point which hadn't been taken into account by the natural sciences till there either.

It was formulated as a primary goal of statistical research to quantify exactly the relation between information gain and sample size, and to develop methods which extract as much information as possible from a given data sample. Here the label is the concept of efficiency.

This scientific program of statistics was defined more or less in the thirties of the 20th century. In Great Britain, Ronald Fisher extended the scope of applications to the natural sciences, in particular to the biological disciplines. In the forties Abraham Wald succeeded in establishing the at the time sensational theory of games by Neumann and Morgenstern as a frame for the statistical decision theory.

All this was reason enough for the following remark of the famous Indian statistician Mahalanobis:

The subject statistics will be the leading information technology of the future.

So far the good news.

4.3 Problems with Identity

There are also less good news, however.

Although now more than 150 years have now passed since the beginning of scientific statistics, the family of sciences still has not got accustomed to this outsider, which takes such a strange intermediate position. The family does not really know where to put this child, and thus is trying again and again to reduce statistics to a special case of a more legitimate family member.

This leads to my next proposition.

The problem of cognition by sensual perception was always subject of controversial philosophical discussions. The typical philosophical reductions in these discussions are quite the same as those in the current discussion about the role of the subject statistics.

There are essentially two possibilities of reducing statistics in a scientifically inadmissible way and letting it be occupied by another discipline. These possibilities correspond exactly to those two components between which statistics should provide a connection.

On the one hand one can consider statistics as a part of applied mathematics. On the other side one can think that statistics is an information technology and therefore it is a branch of computer science.

Both possibilities of reduction and absorption actually have been carried out by the university policy in the recent past. Thus, if we want to develop strategies for the future of statistics we first have to ask what are the reasons for such undesirable developments and how can we avoid them. In this way, we soon get into our main topic, namely the future of this institute of statistics.

4.4 Statistics and Mathematics

Let me say at first a few words about the incorporation of statistics into mathematics. This is an episode which could be observed mainly in continental Europe, especially in Germany. Great Britain and the USA by contrast were more or less immune to this kind of temptation.

To be honest, the emphasis on the mathematical aspects of statistics is absolutely justified. Mathematical statistics is an important component of basic research like mathematical physics. And there was a gigantic progress just in this area at a time when by chance also much money was available for setting up new chairs. Within a short period the number of chairs in statistics at German mathematics departments was multiplied.

However, this for statistics very pleasant expansion unfortunately was carried out with a certain one-sidedness in favour of statistical basic research. And how every one-sidedness leads to contrary movements with the aim of returning to the balance this was also here the case.

And as a result of such a countermovement the subject statistics is presently faced with a threat from the opposite direction.

4.5 Statistics and Information Technology

The announcement of Mahalanobis according to which statistics would be the leading information technology of the future has turned out to be not completely correct. In fact, still in the sixties the beginnings of computer technology were incorporated into the statistics departments of the universities, as far as such departments were available at all. However, the extent of the following expansion process of computer technology was for all of us and of course also for Mahalanobis unforeseeable and inconceivable. Computer science became very fast an independent scientific subject whose role goes far beyond of being merely a supplier for the technology side of statistics.

Of course, computer technology opens up previously unimagined possibilities for statistics. Several famous statistical theorists have early pointed out these chances and have also used them.

However, these chances are accompanied by dangers. Actually, some prophets of computer science think that with such machines almost all problems of mankind could be solved. And of course also this were true for the cognition problem which were only a small problem at the edge of the road leading directly into the computer paradise.

Engineers tried to prove their abilities of future prediction by forecasting share prices with neural nets. They put forward the strong AI hypothesis, this is essentially the possibility of constructing a homunculus computer. And finally they also have announced the end of statistics. Some years ago statistics was actually concerned with technological concepts, where some experts hoped that they could convert a largely arbitrary data stream into useable scientific theories without any further contribution of a rational being.

This third announcement is philosophically identical with the empiristic reduction which is overcome by the philosophical cognition theory, at least since Kant. And since the clarification of the foundation problem of statistics by Popper, Neyman-Pearson and Abraham Wald such ideas have only a nostalgic attraction.

In the meantime it has gotten more quiet around the universal claim of computer science. In the daily routine of universities, however, we are still feeling the direction change which is triggered by the siren songs of computer specialists. Of course this happens at a much more trivial level than in the philosophical discourse described before. However, all the greater is the danger connected with it.

Computer technology undoubtedly opens up fascinating possibilities for statistics. Statistics gets so to speak new sense organs, or in another analogy new immensely efficient microscopes, telescopes and like that.

However, and unfortunately, many statistical theorists spare themselves the effort which is connected with learning to apply computers in a scientific way. This conservative inflexibility is the reason why on one side whole statistics departments are infiltrated by pure computer scientists simply because nobody else is able to produce those coloured little pictures which nowadays are necessary to get one's attention. On the other side, and we in Vienna are at the moment in an alarming way confronted with such a situation, it happens that statisticians who are flexible and open for the new technologies, move to computer science departments in crowds, because they feel no more understood in their traditional community.

5 Perspectives

How can we counter these challenges successfully?

The way forward is simple: A statistics department has to meet the challenge resulting from by the rapid expansion of computer technologies, but it must not give itself up. Statistics departments mustn't lose their identity: They are neither mathematics departments nor computer science departments. A deviation in any direction can be a fatal slip into a one-sidedness.

From this we obtain quite obvious guidelines for the research policy of an institute for

statistics.

It is necessary to keep the competence both in the range of mathematical models and in technological data analysis. Both the mathematical exploration of models and the information technological processing of empirical data are indispensable components of statistical research. Statistics cannot be reduced to or subsumed under a single one of these two components. The aim must consist in guaranteeing and maintaining a high scientific competence in both components of the subject at the same time.

Our department of statistics gets close to this ideal. However, the rapid development on the information technological side requires further structural improvements. This will be our task for the nearer future.

However, this development of the information technological competence must not be misunderstood. It is exclusively devoted to maintaining the leading role of statistics in the production of cognition instruments for empirical sciences.

Of course our institute must prove itself in the competition and in the cooperation with other comparable institutions. I think that our institute is well on its way in this regard. We have given many details in our research report.

We are particularly proud of the high share of A-rated journals in which our papers are published and of the recognition by the scientific community which results in functions of the members of the institute in national and international institutions and committees. I think we will stand an assessment of the institute based on such perceptible signs successfully.

However, I wouldn't like to conceal that there are points in which we must get even considerably better. To express it in the language of our own university: We must become more customer-friendly and we must actually sell our products to our customers.

Our products are instruments for empirical research, and our customers are researchers from sciences working empirically. We must prove ourselves in this market also, and not only in the market of the innovative journal publications.

In order that a market actually works, supply and demand must be balanced.

We are responsible for the supply. Our main weakness in the past was that we often haven't developed our products to that point where the customer is waiting. Although this strategy has improved our performance parameters on the publication sector, it weakened our market opportunities with our customers. We want to improve this aspect considerably in future.

However, in order that a market works, the demand must be intact, too.

We produce nothing with which one can achieve a financial profit. The profit obtainable by statistics is exclusively the quality of the empirical research and the validity of the achieved scientific knowledge. Any demand for the products of methodological research can therefore arise only in a highly developed knowledge society which evaluates the results of scientific research according to objective quality. Only in such an environment scientific statistics will be able to survive in the long run.

Whether there is such an environment, is beyond our responsibility.

6 Conclusion

The evaluation of the institute for statistics which starts here today can have consequences in university policy. As a result of the present economic situation the university landscape in Austria will be restructured. It will be considered very carefully which subjects and institutes should grow and which one should shrink. These decisions will at first be taken at the university level and therefore the universities and their managers need information and advice which help them to take the right decisions.

So all about, in this evaluation it is not so important for us to keep or get a front place in an olympic competition. Competing institutes are rare in Austria and badly equipped. A comparison with international conditions is also only very conditionally possible due to completely incomparable framework conditions.

Any emphasis of the rivalry between statistical departments would be water on the mills of those which are ready to make money at the expense of statistics resources. It is much more important to define the position of statistics in the collection of scientific subjects at a university. This position is endangered and I have put a large portion of the time in my talk into describing the way of endangering.

It is therefore our wish to the colleagues who are acting as experts here, to confirm to the management of the university and thus also to the university as a whole, the value and the role of statistics as a scientific subject. We think to have made contributions with our research which support and justify such a confirmation.

Fortunately there is a recent event with which the position of scientific statistics is improved in the international public view. This year there was a very pleasant decision for statistics by the Nobel price committee. Let me conclude my talk with some historical remarks to put this event in the appropriate light.

The subject statistics did not exist as an independent subject at times of Alfred Nobel. But also the very venerable subject mathematics has been passed over at the foundation of the Nobel prize. There is a rumor that Alfred Nobel hasn't included mathematics into the list of the prize subjects from private feelings of resentment of a certain mathematician. Anyway, scientific achievements which are related to mathematics have been excluded from Nobel prizes for a long time. Only later, in the course of the Nobel prize for economics, created afterwards, sometimes and in recent time more frequently, there were prizes for mathematical achievements. E.g., think of Leontieff and the input output analysis, or of Markowitz and the portfolio theory in financial mathematics. Much public attention has got the awarding of the Nobel prize to Merton and Scholes, a few years ago. For the public the interesting point was that the economic topic is there connected to finance capital and stock exchanges. However, the scientific achievements of the laureates were primarily based on a mathematical model which comes from mathematical probability theory. This was the first case where that mathematical discipline was honored which also forms the mathematical basis of statistics. This year, and this is the pleasant event I mentioned before, finally the Nobel prize for economics was awarded for scientific achievements which are part of the subject statistics in the narrow meaning.

It is extremely gratifying for statisticians that at the level of the Nobel prize committee the scientific subject economics offers public recognition of the subject statistics. Our university is the largest European university for economics and business administration,

and it houses the at the moment largest institute for statistics in Austria. It would appropriate for this university to offer to statistics a similar amount of recognition and support, as it receives internationally.

Author's address:

Univ.Prof. Dr. Helmut Strasser
Department of Statistics
Vienna University of Economics and Business Administration
UZA II, 5th floor
Augasse 2-6
A-1090 Wien

Tel. +43 1 31336 5051
Fax +43 1 31336 734
E-mail: Helmut.Strasser@wu-wien.ac.at