CAI and Computational Statistic

By *P. Naeve*, Bielefeld[1])

The symposia on computational statistics are devoted to a systematic exploitation of modern computing techniques in mathematical statistics and applied probability. Or as is said in the preface of the proceedings of Compstat 74 they want statisticians to become aware of the great possibilities of modern computer facilities. Joining these efforts this paper would like to indicate the impact computers have or should have on teaching statistics.

As this is a wide field that can be attacked in quite different ways, it is necessary to make the authors position clear.

It will be that of a person involved in teaching statistics for more than ten years, that of a person used to work with computers from the very beginning. As it was at German universities, I gathered the experiences and facts my judgement is based upon I only can claim my critical statements to be true for teaching statistics in the Federal Republic of Germany — although I believe they will hold elsewhere with minor modifications.

To put it into one sentence: a German teacher of statistics will talk about teaching statistics at German universities in the computer age.

Let me guide you through this paper by placing three statements at the beginning.

Firstly : *Statistics is taught as if we were living in the precomputer age.*

Secondly : *Computers enter the field of teaching statistics twofold: as a tool and as a medium.*

Thirdly : *One need not be an expert in computer science to take computers into consideration when teaching statistics.*

Let us turn to the first statement. Stating that we teach statistics as if we were living in the precomputer age means that till now computers did not have much impact on the statistical curriculum or the way it is taught.

Three examples will make this more explicit.

Example 1: I took at random three German statistical textbooks from my bookshelf. In one way or the other all three authors are recommending the following procedure when one has to calculate the sample variance

$$s_x^2 \text{ of } x$$

[1]) Prof. Dr. *P. Naeve*, Universität Bielefeld, Fakultät für Wirtschaftswissenschaften, Postfach 8640, D—4800 Bielefeld.

i) Transform the observed values x according to the formula

$$z_\nu = (x_\nu - c_0)/c_1$$

{choose the constants so that the values of z_ν are easy to manipulate with}

ii) Calculate s_z^2, the sample variance of z

iii)
$$s_x^2 = c_1^2 \; s_z^2.$$

This formula gives the variance of x.

(This approach can be used for calculating the sample mean and other sample moments too as another helds it worthwhile to mention) Is there realy any need for this procedure when living among all sorts of computers and desk calculators with built-in statistical functions?

Example 2: This example will make the changed world of statistical calculations even clearer. Let us turn to the book on Time Series by *Kendall* [1973]. Scattered in the text and on 21 pages in the appendix you will find tables of coefficients for fitting a polynominal trend by moving averages. I myself work in the field of "Time series analysis" for many years. I have never met anybody who did this fitting business the old way: Time series: x_{t-3} x_{t-2} x_{t-1} x_t x_{t+1} x_{t+2} x_{t+3}
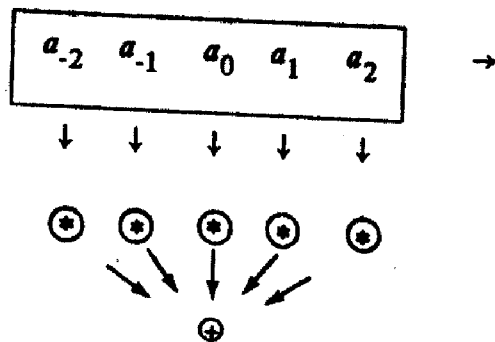
- choose the proper coefficients and write them down on a slip of paper
- move the slip along the recorded time series data step by step
- at each step multiply the data with the appropriate coefficient
- sum up the products
- store the result into the proper place.

No Sir, they all let a computer do this business for them.

Example 3: So far, as I know, there is no class-room equipped with a terminal to link a computer directly to a lecture if one wishes to do. More examples could be added easily. But for the moment I think these will be sufficient to strengthen the first statement.
     There has been a tremendous development in computer facilities at our universities.
- For instance at the Free University Berlin in 1966 an IBM 1130 (8K) was the biggest computer at hand. Today we have access to three CD machines (7200, 6500, 175) two German ones (TR 440, S4004/151) – and for sentimental people: the IBM 1130,

older and grown up, is still running — beloved by many a user for being so reliable (from a user's point of view its mean time between failures is to be measured in months compared with days or even hours when working with the bigger ones.)

But this huge expansion of computer equipment did not seem to have influenced the teaching of statistics.

Now what could be done to improve the situation?

First we must realize that computers have become a valuable tool in statistical applications. Just to give a short reminder.

— forecasting, especially in inventory control, cannot be thought of without mentioning computers
— econometric models will hardly be of any use if computers do not take over the burden of all the calculations necessary to estimate the coefficients of these models.
— to be somewhat more popular and present day — can you think of an election night without getting the latest forecast provided by statisticians with the aid of a computer?

So, one first step should be to adjust our statistical curriculum to the fact that computers have become an indispensable tool in the daily life of an applied statistician.

That calls for dropping those subjects mentioned before, i.e.
— how to calculate sample moments using transformed data
— fitting a polynominal trend the old way.

Instead room should be given for comments on simple numerical problems like those:
— Why may it happen that

$$\frac{1}{n} \Sigma (x_\nu - \bar{x})^2 \neq \Sigma \left(\frac{x_\nu - \bar{x}}{n} \right)^2 \neq \frac{1}{n} \Sigma x_\nu^2 - \left(\frac{1}{n} \Sigma x_\nu \right)^2 .$$

— How many digits are really reliable when one is confronted with computer output.

The next step is to introduce to the students statistical software at hand. Most people will use the computer as a tool via a canned statistical package. It is not so important which package one takes (SPSS, BMD to mention just two) but it must be available for students use. Students should learn that there are always the same three questions to answer.

i) How to get access to the computer and that special piece of software?
ii) How to link the data to the package?
iii) How to get the proper method to work on the data?

Don't loose yourself in explaining the difference between batch and dialog environment or in strongly advocating a certain kind of statistical package — not at hand or to be written in future. A statistician always hears *"Hic Rhodus hic salta"* and that means here is some software: use it.

But these three questions are basic and must be answered before any statistical work relying on software could be started. That is the thing a student should understand. — It prevents him from punching thousands of card, although the package expects the input on papertape.

Talking about software packages: did you ever realize how strong an influence those who produce and distribute software have on how and which statistical methods will be applied?

The non-statistician — but not only him — far too often just runs into the next available software package when looking for some statistical advice. There he finds some methods — though handy prepared — fitting by no means to his problem and data. There is a manual not only containing input description and operating advices but also worked out examples. For many an applicant this manual is the only source for backing up or even establishing statistical knowledge.

Without being insultant one may state only few software products provided by the computer manufactures reach the high standards of their hardware — naturally there are some who fail in both areas. For instance, when I had to give a lecture on experimental design, I used a BASIC-written package to back up my lesson. The manual contained worked out examples. That for analyzing a greco-latin design had just one little fault — it was no greco-latin design at all. This program product was provided by a German computer firm. But there was more than one hint that they copied all the material from someone in the United States — few of us have been so lucky in travelling around the world as that fault obviously was.

I see a twofold danger when using statistical software packages without some advice from a trained statistician.

— Wrong and misleading theoretical explanations together with wrong and misleading examples can produce a wrong understanding of what statistics is.
— To people not acquainted with implementing a statistical procedure on a computer, the software at hand develops a great power. Those methods are used because they are at hand and everybody else in the firm or institute uses them too. At last statistics boils down to those statistical methods the computer center has on its tapes.
— I think everybody doing some job as statistical consultant can add easily to a long list of examples.

Although computers were mentioned more than once in the last section you must not be an expert in computer science to incorporate these things into your lectures.

This will be true for my next point too.

Surely computational statistics stresses the importance of the algorithmic aspects of statistical methods. These aspects are far too often concealed by the hunt for streamlined elegance in the theoretical argumentation.

We must confront the student directly with the algorithm inherent in a statistical method. That means to present an algorithm to him whereever there is one. This can be done without the necessity of introducing a computer language with which one would formulate the algorithm. I advocate the way of representing an algorithm recommended by Knuth [1969].

Let me demonstrate this on an example:

Consider a truncated sequential test. The test is cited from the wellknown book by Wilks [1963, p. 492].

Truncation of Probability Ratio Sequential Test

If the probability ratio sequential test does not terminate for $n = 1, \ldots, N - 1$, suppose the following rule is adopted for terminating the test upon drawing x_N:

Accept H_0 if

$$(15.4.52) \quad \log k_0 < z_1 + \ldots + z_N \leqslant 0;$$

accept H_1 if

$$(15.4.53) \quad 0 < z_1 + \ldots + z_N < \log k_1.$$

Let this truncated sequential test be denoted by S_N and let α_N and β_N be Type I and Type II errors associated with S_N.

The inherent algorithm might be presented as follows:

Algorithm TST: Truncated Sequential Test

S is the test statistic, x_j the observation at the j-th draw, $f_0(x_j), f_1(x_j)$ are the p.d.f. (p.f) under H_0, H_1 respectively, k_0, k_1 are the boundary constants of the test, N is the drawing truncation point, j is the draw counter.

TST 1	*(initialisation)*	Set k_0, k_1 and N according to the desired probabilities for the Type I and Type II errors, $k_0 \leftarrow \log k_0$, $k_1 \leftarrow \log k_1$, $j \leftarrow 0, S \leftarrow 0$	
TST 2	*(next draw)*	$j \leftarrow j + 1$. Draw x_j and calculate $z = f_1(x_j)/f_0(x_j)$ $S \leftarrow S + z$	
TST 3	*(decision possible?)*	if not $k_0 < S < k_1$	$\rightarrow$ TST6
TST 4	*(N reached?)*	if $j < N$	$\rightarrow$ TST2
TST 5	*(decision after truncation)*	if $S \leqslant 0$ accept H_0, if $S > 0$ accept H_1	$\rightarrow$ TST7
TST 6	*(which hypotheses?)*	if $S \leqslant k_0$ accept H_0, if $S \geqslant k_1$ accept H_1	
TST 7	*(finis)*	end of algorithm TST	

Naturally this algorithm gives no information about that truncated sequential test that is not already contained in the text above. But the information is presented in a different way and this might ease understanding the procedure.

The essential features in representing an algorithm are:
— identifying characters
— the steps of the algorithm are identified by the same characters followed by a number
— a header introduces the variables and constants used within the algorithm
— each step begins with a brief description of the content of
— that step
— $\leftarrow$ is used to state assignment
— $\rightarrow$ is used to state a branch.

At last one could try to teach students how to implement statistical procedures on the computers. Certainly this goal demands knowledge of a higher computer language for both – students and teachers. At the moment there are strong institutional restrictions against this kind of teaching. Most of our examination rules don't give any credit for attending a programming course and gaining the necessary skill in that particular language. So I think this last point will be a vision for a long time.

But all the other things could be introduced at once as there are no institutional constrainsts preventing us from doing so.

So far we talked about the role computers as tooles should play in teaching statistics. Now let us comment on how a computer would be used as a medium when teaching statistics.

At this point CAI *(computer assisted instruction)* enters the scene. May be you have heard instead of RGU, CUU or CMI – unavoidably one or the other of this abbreviations will be brought to ones attention when trying to use a computer as teaching medium. They all stand for a newly developed field of research and effort centering around the question: *How to use a computer in teaching?*

At first glance there seems to be no difference between all these abbreviations – at least to amateurs as we certainly are. But if one plunges deeper into this CAI business one will soon be struggling for safe ground to find ones way out again.

There are seldom two authors meaning the same when using the phrase *"computer assisted instruction"*. Much progress in this field seems to be gained by just throwing a new abbreviation on the market. Most writers brilliantly juggle *"computer assisted instruction"*, *"computer assisted learning"*, *"computer assisted teaching"* and some other balls of that kind. They never seem to loose the difference between tutorial oriented and training oriented instruction whilst blaming CAI to be convential compared with CMI.

It is a tremendous merry-go-round one believes to sit upon. But here is the magic spell. Just say: *"Here is my problem let us get started."* Then all that scientific dust suddenly disappears and very little is left that could be used to restructure our statistical lectures.

The German government must have gained the same impression too. There was a drastical reduction of funds provided for research in this area. Turning back to our attempt to give a precise meaning to CAI let us adopt a definition set out by the United Kingdoms National Development Programme in Computer Assisted learning (Quoted from a paper by *Miles* [1974] presented at a conference on RGU held in Hamburg).

Up to this definition the following subjects are included:

a) the computer as manager	– of classroom learning, time tabling and student records
b) the computer as tester	– student assessment, evaluation of learning system
c) the computer as tutor	– teaching a range of skills, concepts and facts, using various strategies
d) the computer as exerciser	– providing regular practice in skills, use of knowledge etc. in various ways

e) the computer as calculator — aid in learning numerically based subjects

f) the computer as laboratory — simulation of experiments, modelling, gaming and problem solving experiences

g) the computer as producer — generating learning materials in various media for use without computers.

This certainly is a very broad definition. Scarcely any teaching activity is excluded a computer might be involved in.

To make a long story short let me put my impression into three statements.

I. *Most work in the field of CAI either has a very theoretical touch without giving much thought to how one could apply it or at the other hand focusses on technical aspects like considering a new brand of courseware etc.*

II. *The whole CAI business seems to have started to get rid of the teachers.*

III. *Scarcely a paper deals the question of CAI from an economical point of view.*

I claim that CAI should concentrate on assisting the teacher and not to make him superfluous. The cooperation of the human teacher and the computer-teacher will overcome most difficulties which have been encountered when one was really practicing CAI up to now. At the same time one will reach economical feasible solutions. As far as I see there are three levels of approach

The black-box-level — *uses the computer to create a more realistic environment for statistical teaching and exercises*

The glass-box-level — *uses the computer to transform statistical methods into operational procedures*

The open-box-level — *uses the computer to demonstrate the dynamic properties of a statistical procedure.*

These rather abstract concepts will become clear in moment when we look at the following example.

Let us take an example from the field of "Time series analysis". If one has got a time series

$$x_1, x_2, \ldots, x_n$$

a natural thing to look at is the spectrum of the stochastic process generating this time series.

Looking for a spectral estimator, the first one will think of, is the periodogram

$$I_p = \frac{2}{n} \sum_{r=1}^{n} \sum_{u=1}^{n} x_r x_u \cos (r - u) \omega.$$

Unfortunately this estimator is not consistent.

Now *Bartlett* [1950] proposes the following procedure.

Break the time series into m consecutive strips of length k, i.e.

$$x_1^s, x_2^s, \ldots, x_k^s \qquad s = 1, \ldots, m.$$

Calculate the periodogram I_p^s for each strip.

Take the average (for frequencies) of all these periodograms

$$\bar{I}_p = \frac{1}{m} \sum_{s=1}^{m} I_p^s .$$

This is a consistent estimator.

Imagine we are teaching this subject to a class.

Black-box-approach: The emphasis of our teaching lies on interpreting spectra to see what can be learned from them about time series. The way the spectra are calculated is not considered in detail. There is a box called *"Bartlett's* estimator" where we input time series and get as output spectra.

Time series

$$\rightarrow \boxed{\begin{array}{c} Bartlett's \\ estimator \end{array}} \xrightarrow{\text{spectra}}$$

Purpose

Input: realizations of known processes to learn how the estimator will produce the shape of the spectra etc.

Input: empirical time series search for periodicities i.e. seasonality, business cycles etc.

Glass-box-approach: The emphasis of our teaching lies on the operational content of this statistical procedure. From a mathematical point of view, all what could be said is contained in the formula

$$\bar{I}_p = \frac{1}{m} \sum_{s=1}^{m} \frac{2}{k} \sum_{r=1}^{k} \sum_{u=1}^{k} x_v^s \, x_u^s \, \cos(r-u)\,\omega .$$

But much more insight is gained by transforming this procedure to a computer program (using a computer language to teach the computer what to do and at the same time let the computer control you by checking the logic and doing some calculations with test data). Below you see a program written in APL. The program is done in a style which clearly demonstrates the essential features of *Bartlett's* method. I am sure APL – fans will easily write a oneliner. But remember efficiency was not intended.

```
∇ Z ← L BARTLETT X
∩ CUT TIME SERIES X INTO L[1]PIECES OF LENGTH L[2]
∩ EACH ROW OF MATRIX XP WILL BE ONE SUCH PIECE
XP ← L ρ X
∩ CALCULATE THE PERIODOGRAM FOR EACH PIECE
∩ THE CORRESPONDING ROW OF MATRIX P WILL
∩ CONTAIN THE PERIODOGRAM
P ← L ρ I ← 1
LOOP : P [I;] ← PERIODOGRAM XP [I;]
→ LOOP IF L[1] ≥ I ← I + 1
∩ TAKE THE AVERAGE FOR FREQUENCIES
Z ← (+⌿ P) ÷ L[1]
```

The function PERIODOGRAM calculates the periodogram. One may take that given by *Grenander/Freiberger* [1971] — after correcting the obvious error in their program.

Open-box-approach: The emphasis of ones teaching lies on giving insight how that particular method works on the data, how they are transformed, which information is ignored, which data are involved at the various steps of the procedure.

Below is shown a brief outline how this could be achieved.
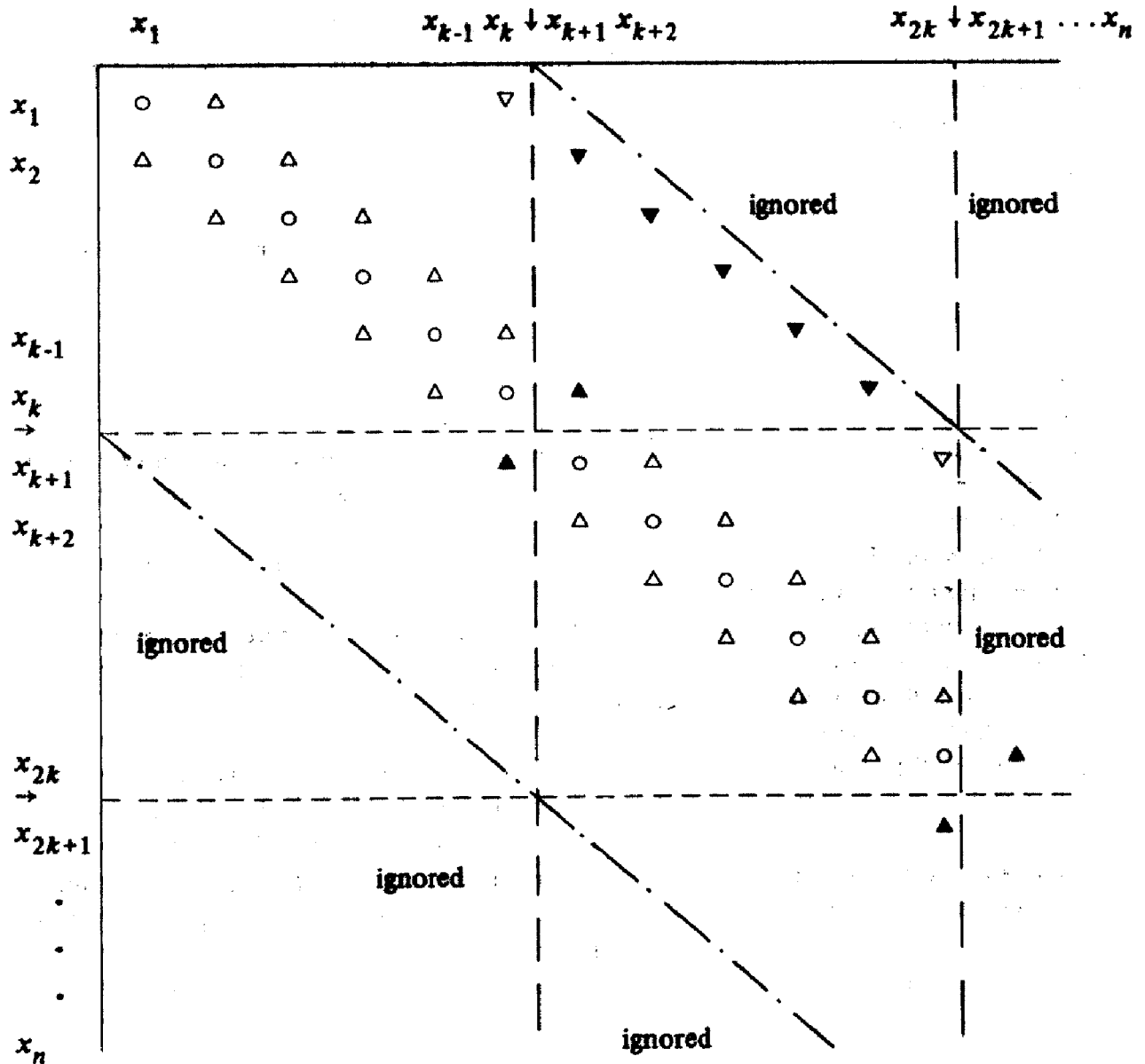
The arrows mark the points where the original time series is broken up to get the m pieces. This partions the matrix of crossproducts into the square matrices denoted by

the broken lines. Calculating the periodogram for each strip means that only the matrices on the main diagonal enter the calculations. For instance when calculating the periodogram at frequency ω using *Bartlett's* estimator the marked crossproducts are ignored compared to the periodogram calculated by using the original time series directly. The crossproducts inside the two dotted lines are used when one adopts a modification of *Bartlett's* estimator proposed by himself. All that is shown in this picture could be produced by a computer with the aid of a graphical display step by step. Thus showing the dynamic properties of *Bartlett's* procedure.

Here the three levels were listed according to the requirements in data processing and computer science capabilities these levels ask for by the teacher. If one is more concerned with the student's learning process then one should revise the order.

I will comment on these approaches by giving some more detailed examples.

Black-box-approach:

Having introduced a new statistical method and shown its theoretical qualities inevitably there comes the point where one has to demonstrate at an example how to use it.

As nobody wants to do lengthy calculations the classroom examples will be boiled down to just a few data. One just wants to demonstrate the necessary steps. What is good for the teacher is good for the students too. So their exercises are handy pieces far from statistical reality. There is no evil in avoiding lengthy calculations. But there is more to learn about a statistical method than just to know the steps. For instance: is it robust against violations of its premises, what happens under these circumstances and so on. This can only be figured out by trial and error – so one cannot avoid doing lengthy calculations.

At this point the computer should enter the stage. Let him do the calculations and concentrate on the interpretations.

I will demonstrate this by an example. When I had to give a lecture on econometrics, I decided to use the Econometric Program System – EPS (This package was presented by *Erber/Haas* [1974] at the first Compstat Symposium).

All I had to do was:

Introduce EPS to the students	– this was easily done because the necessary commands and symbols had a close resemblance to econometric terminology.
Provide access to a computer and the system EPS for the students	– a standardized set of cards were handed over to the students (we used a CD-computer as a batch machine).
Create the data	– this was done by us. The connection of data files to the EPS-System was easy to realize within our standard decks of cards.

With these aids we could proceed like this

Introduce a theoretical concept	– ordinary least squares – multiple regression theory.
Introduce a problem	– which variables should be included into the regression (the danger of a one sided statistical argumentation).

Create a data base	— confer *Rao/Miller* [1971]
	— confer *Koutsoyiannis* [1973].
Let the students find out	— along a discussed line of approach (confer
what happens	*Koutsoyiannis, Rao/Miller*) by trial and error.

The next page shows a short example of input and output for the first problem.

Although EPS was written for a dialog machine it worked fine at this low level of computer support.

The students and I found it a pleasure to work in this way. Some students soon were stimulated to investigate economic problems of their own with the tools provided by us.

Glass-box-approach:

As has been said before computational statistics emphazises the algorithmic aspects of statistics. Therefore the necessity and importance of formulating a statistical procedure as an algorithm has been stressed when talking about the computer as a tool. The glass-box approach offers insight and deeper understanding of a statistical method by transforming the appropriate algorithm into a routine or program ready to run on a computer. It is not intended that the result should be a very efficient piece of software (minimizing core load, CPU-time or what else you want). The goal is not the resulting software but the process of transformation. One may even call it a process of teaching.

The student being the teacher, the computer playing the role of the student (a very stupid one indeed), the subject to be taught: that particular algorithm. I think you can learn a statistical method in this way more easily than by doing many examples. An algorithm is a description of the steps one has to take to get the desired statistic or what ever is in the core of that statistical method. When we practice the method on many examples, we are just trying to filter out and learn these steps.

So why not jump directly into the algorithm.

Clearly the glass-box-approach demands knowledge of a higher computer language. Without going into any details I will declare that in my opinion APL is a computer language well suited for this kind of business. Many concepts developed in the field of data processing will prove usefull too — for instance structured programming.

On this and the next pages there are two more examples for the glass-box-approach — both not completely worked out.

I.   In econometrics one introduces the method of two stage least squares when dealing with dependent systems. Below you find the formula given in every textbook. I think, more insight will be gained when we write this down as on page 20 using some principles of structured programming and APL.

II.   The theorem on page 21 can be used to construct a random vector with normal probability distribution function. I wonder if every student will know how to proceed when we ask him to generate a random vector according to this theorem. The flow chart on page 22 surely will help.

```
COMMENT        RAD MILLER  P 38 F.
FORMAT         (10F8.0)
   FORMAT IS
   (10F8.0)
LENGTH              22
READ TS        0 0 6
   DATA INPUT FINISHED

LABEL          1,TEA
               2,CEYLON
               3,COFFEE
               4,INCOME
               5,FOOD
               6,INDIAN
               0

LIST DATA


NO. OF STEQ    3
READ MODEL     1 0 0 2 3 3 0 0 2 4 0 0 2 2 0 0 2
               1 0 0 2 2 3 0 0 2 4 0 0 2 2 0 0 2 6 0 0 2
               1 0 0 2 4 3 0 0 2 4 0 0 2 2 0 0 2 6 0 0 2
```

```
================================================================================

   EQUATION  NO.   1
   ===================

   DEP.      1 (    0     0     2 )
   INDEP.    3 (    0     0     2 )    COFFE              LN
             4 (    0     0     2 )    INCOM              LN
             2 (    0     0     2 )    CEYLO              LN
       ESTIMATION METHOD  - - - O L S
================================================================================
```

```
********************************************************************************
*                                                                              *
*    BETA        STD.-ERROR     STUDENT-T    SIGN. LEVEL   VARIABLE             *
*                                                                              *
********************************************************************************
*                                                                              *
*   1.09994        .60555        1.81643       .09390      ABSOLUT    1         *
*    .09917        .13946         .71113       .49769      3 (  0   0   2 )     *
*    .51653        .11548        4.47275       .00122      4 (  0   0   2 )     *
*    .28195        .52349         .53860       .39403      2 (  0   0   2 )     *
********************************************************************************
*                                                                              *
*  C.I. P(  -.17767   BETA    2.37755) = 0.95      ABSOLUT    1                 *
*  C.I. P(  -.19506   BETA     .39340) = 0.95      3 (  0   0   2 )             *
*  C.I. P(   .27288   BETA     .76018) = 0.95      4 (  0   0   2 )             *
*  C.I. P(  -.82253   BETA    1.38644) = 0.95      2 (  0   0   2 )             *
********************************************************************************
* COEFF. OF DETERMINATION R**2 =  .575028   CORRECTED R**2    = .475034        *
* TEST OF HOMOSC.  F( 11, 11 ) =  3.713     LEVEL OF SIGN. =    .05276          *
* DURBIN-WATSON            = 2.37189                                           *
* COEFF. OF AUTOCORR.    =  -.3627                                            *
*                                          SUM OF SQUARES   =      .5138       *
* F-VALUE =    8.1186                       LEVEL OF SIGN.   = .000878         *
********************************************************************************
  MINIMUM CM    0470008
```

Two Stage Least Squares

Consider the first equation of the simultaneous-equation system

$$
\begin{pmatrix} \hat{\gamma}_* \\ \hat{\beta}_* \end{pmatrix} = \begin{pmatrix} Y_*'X\,(X'X)^{-1}\,X'Y_* & Y_*'X_1 \\ X_1'Y_* & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_*'X\,(X'X)^{-1}\,X'y_1 \\ X_1'y_1 \end{pmatrix} .
$$

Notation: variables

$\hat{\gamma}_*$ vector of coefficients of the dependent variables

$\hat{\beta}_*$ vector of coefficients of the predetermined variables

X matrix of all predetermined variables

X_1 matrix of predetermined variables of the first equation

Y matrix of all endogenous variables

Y_* matrix of those endogenous variables encountered in the first equation

y_1 vector of the first endogenous variable

Notation: operators

$+ . \times$ matrix multiplication

⊟   inversion

φ   transpose

, [i]   catenation along the i-th coordinate

S operating on X, (Y) results in X_1, (Y_*) respectively

Program Tree for
'Two Stage Least Squares'

$$\begin{pmatrix} \hat{\gamma}_* \\ \hat{\beta}_* \end{pmatrix} = \begin{pmatrix} Y'_* X (X'X)^{-1} X' Y_* & Y'_* X_1 \\ X'_1 Y_* & X'_1 X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'_* X (X'X)^{-1} X y_1 \\ X'_1 y_1 \end{pmatrix}$$

Program Tree of
Matrix
$$\begin{pmatrix} Y'_* X (X'X)^{-1} X' y_1 \\ X'_1 y_1 \end{pmatrix}$$
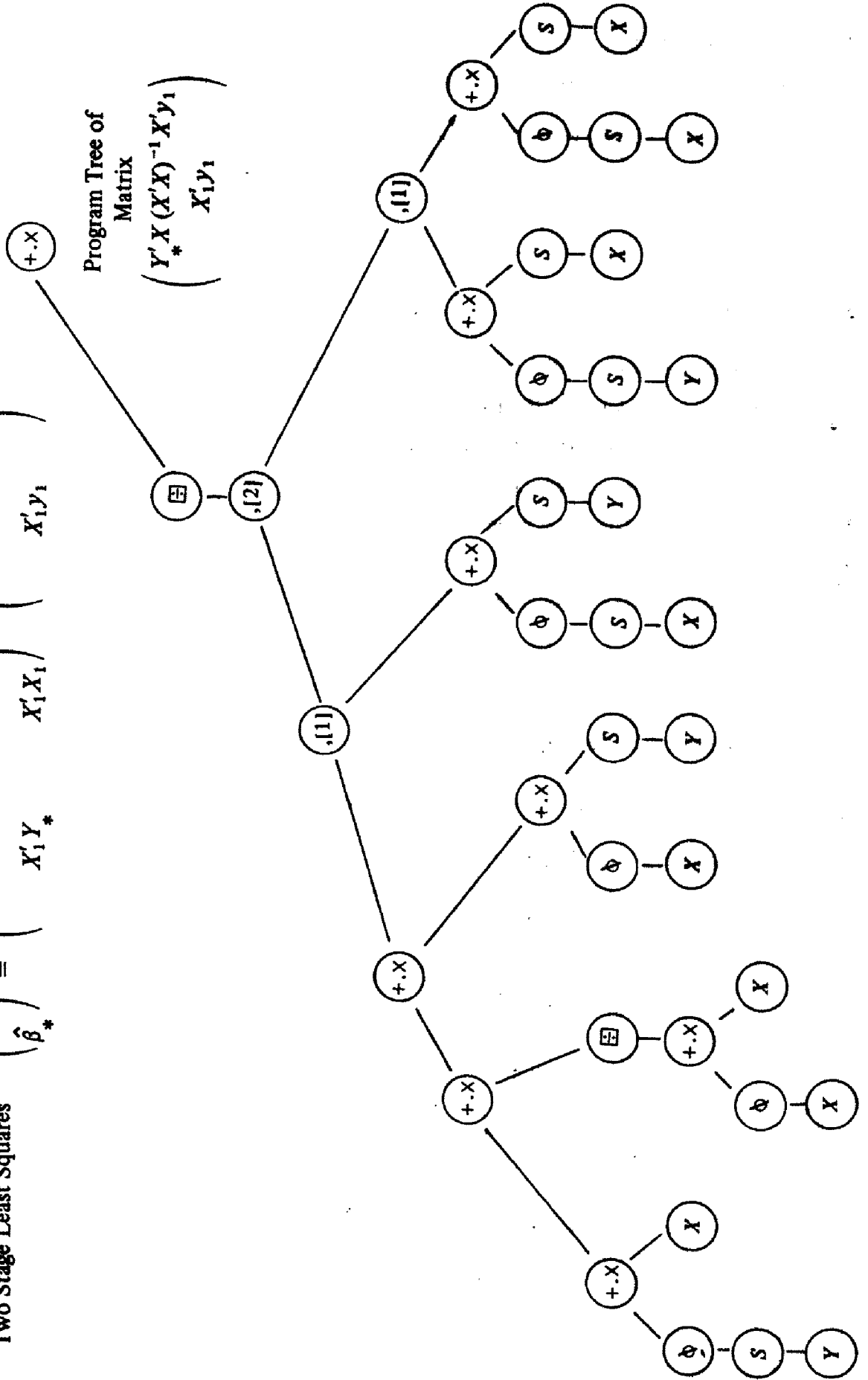
Theorem

Let $u' = (u_1, \ldots, u_g)'$ be a g-dimensional random vector with normal distribution and parameters

$$E(u) = 0 \qquad E(uu') = \Sigma .$$

Partition u and E according to

$$u' = (u_1, u_2)' \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Then for the p.d.f. function $f(u)$ we have

$$f(u) = f(u_1 | u_2) f(u_2)$$

where $f(u_1 | u_2)$ is p.d.f. of a normal distribution with parameters

$$E(u_1 | u_2) = \Sigma_{12} \Sigma_{22}^{-1} u_2$$

$$E(u_1 u_1' | u_2) = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

and $f(u_2)$ is p.d.f. of a normal distribution with parameters

$$E(u_2) = 0 \qquad E(u_2 u_2') = \Sigma_{22} .$$

Open-box-approach:

Psychologists claim the importance of the connection between practical and theoretical understanding for the process of learning. I think the open-box-approach is in line with these findings.
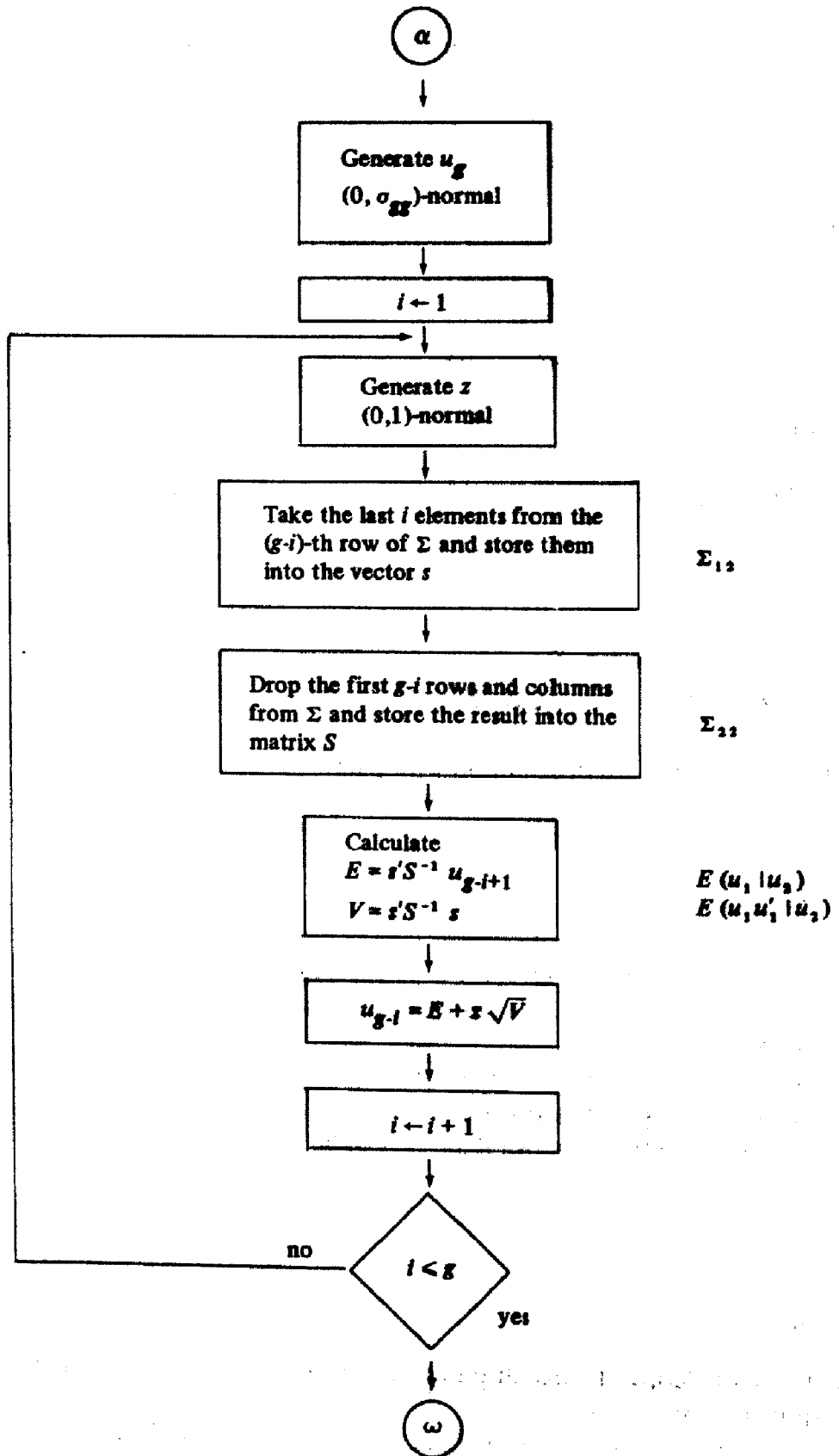
Here one has the opportunity to demonstrate a statistical method in action. The student can see how the appropriate algorithm works on the data. For example consider a lesson on exponential smoothing. All that it is can be written down in just two lines.

$$\hat{x}_t = \hat{x}_{t-1} + \alpha (x_{t-1} - \hat{x}_{t-1})$$

$$\hat{x}_t = \sum_{j=0}^{\infty} \alpha (1 - \alpha)^j x_{t-j} .$$

The next page is a hardcopy of three displays from a worked out open-box-approach of "teaching exponential smoothing".

The dotted lines separate the three displays. The second display shows the necessary calculations to get in period 6 the forecast for period 7. The two values which enter the calculation are marked by stars the result is marked by equals sign

α

Generate u_g
$(0, \sigma_{gg})$-normal

$i \leftarrow 1$

Generate z
$(0,1)$-normal

Take the last i elements from the $(g\text{-}i)$-th row of Σ and store them into the vector s Σ_{12}

Drop the first $g\text{-}i$ rows and columns from Σ and store the result into the matrix S Σ_{22}

Calculate
$E = s'S^{-1} u_{g\text{-}i+1}$
$V = s'S^{-1} s$ $E(u_1 | u_2)$
    $E(u_1 u_1' | u_2)$

$u_{g\text{-}i} = E + z\sqrt{V}$

$i \leftarrow i + 1$

$i < g$

no

yes

ω

— the observed value in period 6: that is 6.8
— the forecast for period 6: that is 6.1
— the forecast for period 7: that is 6.3

In this way you may forecast the series step by step. The displays 1 and 3 demonstrate how the weighting scheme moves along the time series as time goes on. Display 1 shows the position of the weighting scheme in period 5 as display 3 does it for period 6.

The system that is behind these display is far to complex to be demonstrated without a terminal at hand.

It allows for

— going to final result that is the smoothed series — the series of all forecasts
— hard copy feature etc.

```
NR  WERT          GEWICHTUNG (PROZENT)

1   4.1  XXXXXXX  6.59
2   5.2  XXXXXXXXXX  9.88
3   5.9  XXXXXXXXXXXXXXX 14.81
4   6.3  XXXXXXXXXXXXXXXXXXXXXXX 22.21
5   7.1  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 33.30
6   6.8
7   7.0
8   5.8
9   5.1

SUMME DER GEWICHTUNGSFAKTOREN VORHERIGER, HIER
NICHT AUFGEFUEHRTER DATEN IN PROZENT  =  13.20
```

```
SCHRITT: 6
                              *
   4.1  5.2  5.9  6.3  7.1  6.8  7.0  5.8  5.1

   4.7  4.9  5.2  5.6  6.1  6.3
                       *    =

EXP. GEGLAETTETER WERT DER VORPERIODE =   6.1
ZU GLAETTENDER WERT IN PERIODE 6      =   6.8

BERECHNUNG:
0.333 * 6.8  +  0.667 * 6.1  =   6.3
```

```
NR  WERT          GEWICHTUNG (PROZENT)

1   4.1  XXXX  4.40
2   5.2  XXXXXXX  6.59
3   5.9  XXXXXXXXXX  9.88
4   6.3  XXXXXXXXXXXXXXX 14.81
5   7.1  XXXXXXXXXXXXXXXXXXXXXXX 22.21
6   6.8  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 33.30
7   7.0
8   5.8
9   5.1

SUMME DER GEWICHTUNGSFAKTOREN VORHERIGER, HIER
NICHT AUFGEFUEHRTER DATEN IN PROZENT  =  8.81
```

This lecture did not intend to give a report on the state of the art in the field of CAI nor in computational statistics.

I just tried to make clear that there are many reasons why computers should have impact on what is taught in statistical courses and how teaching is done.

Here are the main points again:

Firstly : *Statistics is taught as if we were living in the precomputer age.*

Secondly : *Computers enter the field of teaching statistics twofold: as a tool and as a medium.*

Thirdly : *One need not be an expert in computer science to take computers into consideration when teaching statistics.*

When preparing this paper I had many helpful discussions with members and students of our statistical institute. So when I succeeded in convincing you that it might be a challenging job to adjust our teaching to the needs and possibilities of computer age cheer them, otherwise blame me.

References

Bartlett, M.S.: Periodogram Analysis And Continuous Spectra. Biometrika 37, 1950, 1—16.

Grenander, U., and *W. Freiberger*: A course in computational probability and statistics. New York 1971.

Haas, H.J., and *G. Erber*: A Econometric Program System — EPS. Compstat 74 — Proceedings in computational statistics. Ed. by G. Bruckmann, F. Ferschl and L. Schmetterer. Wien 1974.

Kendall, M.G.: Time Series. London 1973.

Knuth, D.E.: The Art of Programming 1. Reading 1969.

Koutsoyiannis, A.: Theory of Econometrics. London 1973.

Kreyszig, E.: Statistische Methoden und ihre Anwendungen. Göttingen 1975.

Miles, R.: National Development of Computer Assisted Learning: Progress and Prospects. Rechner-Gestützer Unterricht. Ed. by K. Brunnstein, K. Haefner and W. Händler. Berlin 1974.

Rao, P., and *R.L. Miller*: Applied Econometrics. Belmont 1971.

Schaich, E., u.a.: Statistik I. München 1974.

Wetzel, W.: Statistische Grundausbildung für Wirtschaftswissenschaftler. I Beschreibende Statistik. Berlin 1971.

Wilks, S.S.: Mathematical Statistics. New York 1963.