STATISTICAL DATA PROCESSING
IN THE CENTRAL BUREAU OF STATISTICS OF NORWAY

by Svein NORDBOTTEN

Central Bureau of Statistics, Norway

1. Introduction

I. Historical introduction

The Central Bureau of Statistics was founded in 1876, and the work was at that time satisfactorily performed by 15-20 employees. In connection with the Census of Population in 1900 it was decided to apply the first data processing equipment to make the processing more effective.

The equipment was Hollerith punched card machines, three counting machines with 70 counters and 21 mechanical punch machines, all working with 24 column cards.

Since that time the use of electro-mechanical equipment has increased continuously and has proved that these aids make the statistical processing more effective. The milestones are marked by acquisition of sorters in the 1910, tabulators in 1930, calculator and electronic statistical machine about 1950, and the first electronic data processing machine in 1958.

The last milestone, the purchase of a medium-size electronic computer in 1958, is of particular interest because of the extensive preparations and the role this machine plays in the present data processing in Norwegian Central Bureau of Statistics.

Encouraged by the information from the US Bureau of the Census about application of UNIVAC in the statistical data processing, our first studies of electronic data processing started about 1953-54. The systematic work in this field commenced, however, with preliminary studies and evaluation of systems in 1956 in order to select the most effective machines for our particular use. This evaluation concluded that for our purpose some medium-cost systems without tape equipment ought to be more closely studied.

The second part of our studies was embarked in the last months of 1956 with 3 people working on this project. Parts of the data processing of some representative statistics were programmed and partly tested on the selected equipment. The programming was a very time-consuming job, and much time was spent on machines which we never are going to use. This was not, however, regarded as lost time, because it gave the programmers an all round training which has later proved to be
very useful. Some of the findings of this detailed study were that several medium-
cost machines would be able to make the data processing more accurate at the same
time as the number of statistics worked out could be increased without any increase
in the cost of processing. It was pointed out that the Bureau itself at least needed
1,000 hours each year on a medium-cost machine. The cost calculation was based
on the assumption that the machine was rented out to other users for about the same
number of hours.

A special committee was appointed by the Ministry of Finance to evaluate our
conclusions and to search for other data processing jobs to finance the purchase of
a machine. This committee recommended the purchase of a DEUCE Mk II manu-
factured by the English Electric Company, and our order for this machine was signed
in December 1957.

The programming for DEUCE was commenced in the Bureau immediately
after the order was placed. As the machine delivered to the Bureau was one of the
two first Mark II produced, the programming also had to cover the necessary auxiliary
80 column input/output routines. When DEUCE was installed 12 months after
ordering, we were in the position to utilize the machine at once for about 5 hours
per day.

After using DEUCE for about one year, we found that it was a need also for
another medium-cost machine, and an IBM 1401 was ordered for delivery in May 1961.
These two rather unequal EDP-machines are assumed to supplement each other
in our statistical data processing.

2. Organization, resources and operating system

I. Organization

The Central Bureau of Statistics in Norway is not large enough to permit data
processing experts and equipment in each statistical division. Effective use of
modern data processing machines in the modest scale of the statistical activities of
a small country requires, however, a careful selection of suitable applications as well
as a centralization of uniform work. A central Data Processing Division has therefore
been established, and this division is responsible for all mechanical and automatic
data processing within the Bureau. The chief of the division reports directly to the
Director.

The division is subdivided into Section for system analysis, Section for program-
ing, Section for machine operation, Section for punching, and Section for desk
calculating.

The task for the Section for system analysis is to record and analyse the need
for data processing services, set up schedules, assist the statistical divisions in specifying
their processing jobs, record and analyse existing routines aiming at more effective
processing, and to evaluate technical developments of data processing equipment in view of the statistical data production.

The Section for programming is working in cooperation with the above mentioned
section, and plans the data flow through the machines. This includes programming
for EDP-machines, i.e. the development of detailed working instructions for programme-controlled machines, as well as deciding which type of auxiliary machines shall be used.

The detailed planning for the auxiliary punched card equipment, the operational scheduling, and the operation of the equipment are performed by the Section for machine operation. Within this section, there exists a further subdivision into working groups. The section is working according to a three-shifts system.

The Section for punching is preparing the punched cards for the machine operation section, while the Section for desk calculating is a service pool for pure desk calculation.

Technical maintenance is provided by the machine manufacturers. As to DEUCE, a Norwegian research institute takes care of the daily maintenance, while periodical overhauls is provided by the manufacturer.

II. Resources

The Data Processing Division is at present disposing of the following 69 positions by sections:

Section for system analysis ........ 3 positions;
Section for programming ............ 10 positions;
Section for machine operation ...... 26 positions;
Section for punching ................ 24 positions;
Section for desk calculating .......... 6 positions.

In addition the Bureau has a number of positions for the punching staff in connection with the 1960 Censuses of Population and Fisheries.

The Section for system analysis is recruited by people with experience from both statistical divisions and machine processing. The programmers need a general knowledge of all available equipment and are therefore recruited from the machine operating staff. It is assumed that the training of operators on all machines is compensated by their ability as programmers to find more flexible processing solutions within the available range of equipment.

The staff of the three operating sections is mainly recruited from undergraduates and is given full training by the Bureau.

In the spring of 1961, the Bureau disposes of the following data processing equipment:

a. EDP-machines:
   1 English Electric DEUCE Mk II;
   1 IBM 1401.

b. Auxiliary equipment:
   1 IBM Interpreter;
   2 IBM Electronic Statistical Machines;
   1 IBM Collator;
   6 IBM Sorters;
   1 Bull Electronic Sorter;
3 IBM Tabulators;
3 IBM Reproducers;
36 IBM Punch Machines and Verifiers.

c. Desk calculating machines.

In this context the English Electric EDP-machine DEUCE is of particular interest.

DEUCE is a general purpose, programme controlled, binary EDP-machine with a basic binary digit period of 1 microsecond and a word length of 32 bits which is equivalent to about nine decimal digits plus sign.

The internal operating time is 64 microseconds for addition and subtraction and approximately 2 millisecond for multiplication and division carried out in fixed point binary.

The storage is of two types, high speed store and backing store. The highspeed store is of the acoustic delay line type and has a capacity of 402 words. The backing store is a magnetic drum with a capacity of 8.192 words. The information is exchanged between the high speed store and the backing store and vice versa in blocks of 32 words. This transfer can be performed simultaneously with other operations.

Instructions are stored one per word. Each instruction specifies a Source and a Destination which may either be real locations in store or hypothetical locations representing special operations. The instruction also comprises a specification of the next instruction. All addresses in the machine are relative to the location of the instruction performed and are specified in special fields of the instruction. A special component for automatic instruction modifications is included in the machine.

The machine is equipped with an 80-column punched card input and output unit. The reading speed of this unit is 200 c.p.m. input and 100 c.p.m. output. A special part of the high speed store can be utilized as an input/output buffer which makes simultaneous computation possible. Up to 300 millisecond is therefore available for computations between the reading of two cards when the machine is reading at full speed. The machine has two different input/output modes, i.e. the binary and the decimal mode, which allows for direct processing of decimal cards as well as compact storage of programmes and intermediate results in binary form on cards.

The input/output data may be externally edited by a plug-board.

DEUCE is also equipped with a paper-tape reader. This reader works with a speed corresponding to 850 characters per second.

Our DEUCE is thus a very fast and powerful computer, but has rather modest input/output units and no possibility for direct printing. The IBM 1401 purchased is a model C-3 which makes it possible to add magnetic tape equipment to the machine if necessary. The main characteristics of our 1401-machine is the input and output speeds. The machine reads and punches 800 and 250 cards per minute, respectively, and prints 600 lines per minute. In contrary to DEUCE, the 1401-machine is thus fitted for jobs requiring much input/output operation combined with little computing.
III. Operating system

Extensive long-range planning for the data processing is necessary because of the budget rules for government agencies. The needs for data processing services are therefore recorded by the system analysis section each autumn. Based on the collected information a time schedule is set up for the next year with the month as a unit period. This main schedule has to be approved by the Director. During the year the chief of the Data Processing Division decides whether it is possible to undertake new jobs not recorded. He also gives priorities to different jobs if necessary and must be notified if change in the time schedule is wanted. The yearly schedule together with the operational statistics is also the basis for the man-power and machine budgets of the division.

The system analysts and the programmers are responsible for the planning, cost and time calculations. During this stage all communication with the statistical divisions passes through the system analysts and programmers. A main task for the former is to secure accurate and precise descriptions of the jobs.

As to the programming the jobs are allocated to the programmers by the head of the section, who also is responsible for the training of programmers and control of the programming work. Previously the programmers were not specialized, and performed all kind of work from drawing flow diagrams to coding and programme testing. The staff is now specializing on programming for the DEUCE or for the IBM 1401 at the same time as we are training special coders.

Most of the programmes are worked out in the machine language for DEUCE and the standard programme library routines have rarely been used. The cause is that previous DEUCE users have been mainly mathematicians or engineers, and the automatic programming schemes and programme library have been mathematically oriented. But for some special jobs the automatic schemes and the very extensive programme library have been applied with great success. We have, however, established our own library for routines in the statistical data processing, which is at present containing about 75 special programmes and routines, representing about 10 programming years.

For programme testing the programmers are allowed up to half an hour each day in total, which seems to be more than really needed for rational testing. Apart from the testing, programmers are normally not permitted to operate machines, because they are apt to waste machine time by making doubtful corrections in their programmes on-line.

Tested programmes and overall processing plans are handed over to the machine operating section in a standard documented form including descriptions, operating instructions and packs of test cards. Usually not more than 10 jobs are planned in parallel and the time schedule causes little trouble at this stage. Progress reports are therefore only prepared once every month and examined by the chief of the division.

In the operating sections, however, more than 30 different jobs may be processed in parallel during a week and compete for man-power and certain machines. The
time schedule requires here current inspection. For each month it is broken up into week, day and shift schedules. The head of the operating sections therefore once every week sets up timetables for each day and shift and allocates operators and machines to the different jobs. Shift supervisors are responsible for the realization of this allocation, while the leaders of the working groups are responsible for the timetables for the jobs within their group.

All operators are trained to operate all available equipment including the EDP-machines. The shift supervisors are, however, responsible for testing the machines, and as to DEUCE, they particularly have to control that the test pack related to a programme is runned successfully before starting a new large processing.

Because of the large number of jobs performed and many interrupting factors, progress reports are prepared weekly and discussed with the chief of the division. They are based on the operators’ working reports, the supervisors’ machine reports and the machine log-books. All information is, of course, punched and processed to very useful operational statistics for the data processing.

The operational statistics indicated that the Data Processing Division in 1960 worked on about 90 different jobs of which some was repeated each month or quarter. In about 30 jobs DEUCE was used with one or more programmes. The statistical data processing required in total about 22,000 machine hours excluding punching and verifying. But according to cooperation schemes the equipment has also been utilized in other kinds of data processing.

For statistical purposes DEUCE was used about 1,750 hours in 1960. About 100 of these were used for programme testing, 1,400 hours were productive runs, and 250 hours were used for re-runs owing to inconsistent data and failures of the programmers, operators and the machine.

3. Applications of electronic data processing equipment

I. Survey of applications

The electronic data processing machine DEUCE has been extensively used in the Norwegian statistical data production. I shall not try to present a complete survey, but only mention some more important applications.

First, DEUCE is doing runs every month in processing the External Trade Statistics. These are typical applications to the data processing stage of data production, but one of the runs represents also an example of application within the stage of publication. These runs will be more completely described in the following section.

Another very successfull application was the 1958 Consumer Expenditure Survey. As in many other of our applications, DEUCE simulates the editing activity previously done by specially trained employees. After the editing, the data is aggregated. The 1958 survey did not include fishermen, farmers and old pensioners previously analysed. To make these results compatible with the 1958 Survey changes of population, income and prices were taken into account by regression analysis.
Each year detailed Tax Statistics are worked out on a stratified sample basis. This was one of the first annual statistics on which DEUCE was applied. DEUCE does a recomputing of different tax components controlling the data before weighting, inflation and aggregation is performed.

Other annual statistics for which the processing is based on DEUCE are the Industrial Production Statistics and Internal Trade Statistics. In contrast to the External Trade Statistics, these are characterized by a large mass of different information per unit and are therefore very difficult to control and edit consistently by manual methods.

DEUCE is also applied in the processing of our Wage Statistics, Transport and Communication Statistics and in the computation of different monthly, quarterly and annual indices. We are now preparing DEUCE programs for some parts of the 1960 Censuses of Population and Fisheries which, however, will mainly be processed by IBM 1401 and other equipment.

A very interesting application of DEUCE is the calculations of economic forecasts based on the large input/output model of the Bureau comprising several hundred relations. Special programmes for seasonal adjustments are the basis for a large scale automatic decomposition of monthly time series. DEUCE are also used for regression analysis comprising up to 25 factors, computation of variances and covariances besides more standard computations as matrix inversions, etc.

Most of the applications fall within the processing stage. We are, however, to a limited extent also applying DEUCE in planning and administration.

An extensive internal quarterly cost analysis giving the management a concentrated survey of the activities has been programmed and is performed by DEUCE. DEUCE has also been applied in resource allocation based on a linear programming model, and in simulation as part of research in methods.

II. Application to External Trade Statistics

The External Trade Statistics consists of the monthly import and export statistics, the weekly fish export statistics, special forms for publishing the results, and the annual statistics.

The data is sent to the Bureau from 64 ports of entry during the first week in the month after the one recorded. The data includes up to 140,000 imported and 30,000 exported consignments per month. The statistical results must give a classification in 6,000 commodity groups, 133 countries of production (consumption), 133 countries of purchase (sale), and 64 ports of entry, all in different combinations for the prescribed tables.

As early as 12 days after the end of the month, preliminary results are required. About 24 days after the end of the month, final totals are required for press release, and in 5 weeks the manuscript for the monthly bulletin must be complete. The annual statistics are prepared in the spring.

A very high degree of accuracy is required. The tables can be controlled against each other and an error may require the processing to be partly repeated.

Since January 1959 the External Trade Statistics have been processed by a new system in which DEUCE is the central processing component.
Previously, coding and control of the data was done by specialized clerks. The processing by machines was limited to such elementary operations as punching, sorting, listing and aggregating to tables by means of special 65-column punched card equipment reserved only for this job.

In the new system we use 80-column punching, allowing the application of our main machine equipment and DEUCE. DEUCE takes care of the control and aggregation. Simultaneously with the introduction of the new processing system, we also had to change to a new commodity nomenclature which greatly extended the processing job.

Before embarking on the new system, we had been working on it for some time. By means of punched card machinery, we also made test runs simulating the working of DEUCE in the system. This was of great importance for the setting of control parameters. After completing the programmes for DEUCE, we had the opportunity to test them before the machine was handed over. In the beginning of January 1959, we took the chance that everything would work as planned and dropped the very expensive double processing by the old and the new systems.

Below the discussion is limited to the monthly import statistics. The weekly and monthly export statistics procedures are not identical with the import statistics and require special programmes, but they are very similar to those discussed here.

A rough description of the whole machine processing precedes the DEUCE application in detail. The documents are first sent for punching. For each separate consignment a card including information about commodity, country of production, country of purchase, port of entry, quantity and value is punched. After punching, the cards are run through an IBM Electronic Statistical Machine giving the necessary results for the preliminary figures. The cards are then sorted into commodity groups, each preceded by a master card defining the special control parameters for the group. The general control procedure is, of course, included in the programme.

The cards are processed by DEUCE following a programme for control and aggregation of data cards. The main purpose of this process is to detect errors in the data cards and indicate these by punching error cards. Besides this error detection, DEUCE also aggregates the accepted data cards in basic tables, punching a preliminary summary card for each cell in the tables.

The error cards are separated from the preliminary summary cards, and returned to the Division for External Trade Statistics for a thorough inspection; they are later included in accepted form in the next DEUCE control run. The master cards and data cards are separated and filed in the master card file and the processed data in the datacard file.

The preliminary summary cards from all control runs are merged with the final summary cards from the preceding month's processing including cumulated figures for quantity and value so far this year; the merged pack is processed by means of the DEUCE aggregation programme.

DEUCE does here some further aggregation and produces the final summary cards. They are sorted and listed in several ways to give the required monthly tables.

The Bureau also offers a subscriber service. A private firm can subscribe on one or more table cells in which it has particular interest. For each subscriber, a register card has been prepared with name, address and subscriber's number. For each cell
on which the subscriber wishes information, a particular request card is established. These request cards contain both the subscriber's number and the identification for the particular cell. The request cards and the final summary cards are merged and processed in DEUCE by means of the subscriber programme. The results punched by DEUCE are the subscribers' information cards which are sorted by subscriber's number, merged with the subscribers' register cards and listed with name and address ready for cutting and posting in envelopes with windows.

In the first mentioned DEUCE run, the data cards and master cards are first checked as to correct sorting. By means of the master cards and a general control procedure included in the programme, DEUCE checks each data card according to the same principles as the previous manual editing, but much more safely and systematically. DEUCE computes, for instance, the unit price of the consignment from the data card and compares this with a price variation interval worked out a priori for this particular kind of import. The bounds of this interval are punched in the master card and can easily be justified. Other checks are also performed and DEUCE either accepts a data card or indicates one or more of the following error types:

1. Unusually high unit price
2. Unusually low unit price
3. Unusually large quantity or value
4. Non-existent commodity code
5. Non-existent code for country of production
6. Unusual combination between commodity and country of purchase
7. Non-existent code for country of purchase
8. Non-existent code for port of entry

When DEUCE detects one or more errors in a data card, a copy of the data card is punched out including also some extra punching defining the type of error.

There are two main types of controls. The first is a control for logical errors which by definition must be rejected, for example the control of the commodity code. The second type is control against information which we assume from previous experience to be rare and therefore probably an error, for example the unit price being outside a price variation interval based on previous experience.

In other words, some error cards indicated by an error of the second type may be accepted by the specialists as correct. The programme is therefore constructed so that an error card can be read into DEUCE again, and the machine then drops any control of the second type which is already indicated in the card as an 'extra punching.' Thus, if only errors of the second type are indicated in the card and it is read again, it will be accepted and aggregated. About 12 per cent of the data cards are rejected.

Under control of the same programme DEUCE also simultaneously aggregates the accepted data cards and punches the above mentioned preliminary summary cards for five different tables. Four out of these five have quantity and value in the heading. The rows of the first table give a specification for each commodity group within which specifications for both country of production and purchase are given. The second table gives figures for each port of entry within each commodity, with separate numbers for
The fourth table gives figures for some special commodities with specification for country of production. The fifth table, produced mainly for control reasons, gives figures for larger commodity groups. The last table indicates how many errors of the different types have been detected within specified commodity groups.

DEUCE punches each set of table rows as preliminary summary cards as soon as the rows are completed. The preliminary summary cards are identified by row and table numbers.

The programme for control and aggregation contains about 800 instructions. The working speed corresponds to about 6,000 data cards per hour.

The second DEUCE application in the import statistics processing is based on the aggregation programme. The main job is aggregation based on the preliminary summary cards and the final summary cards from the preceding months' processing. The programme also performs a technical routine check of the summary cards against double punching, non-existent codes, etc. Three tables are built up. The first gives final figures for quantity and value both in this month and so far this year for each commodity by port of entry. The second gives final figures for the same characteristics, specified for each commodity both by country of production and country of purchase, while the third really consists of marginal columns of the two previous tables.

This programme is relatively simple, consisting of about 200 instructions. The work could have been performed by modern tabulators. It is mentioned here, however, because we would need at least three tabulators to do the same work in the same time. At present this would be more expensive than to use DEUCE.

In the DEUCE processing for the subscriber service, the request cards are sorted by commodity and table number together with the relevant final summary cards. The summary cards are again internally controlled against each other to ensure that the reading has been correct. DEUCE then reproduces the information from each summary card according to the request cards, transferring also the identifications from the latter to the information cards punched out. At the same time, totals for the information given to each subscriber are built up in the machine and punched out as the last results in order to have full control over the punching.

This programme consists of about 700 instructions. The speed corresponds to about 4,000-5,000 punched information cards per hour. In my opinion this is a very interesting application of DEUCE in the information giving activity, and the cost calculations also indicate that it is a very economical application.

The processing of the External Trade Statistics by DEUCE is now based on 6 programmes representing several man-years of programming work. In this connection, some of the experience gained might be reported here.

As most others, we have found that the programmers themselves should not be allowed to run the machine. This leads to the problem of writing unambiguous operator's instructions. We have found it worth while to increase the programming effort when it means simplified operating. Our aim is to make programmes with standard operator's instructions with, for example, failure indications each requiring a standard action from the operator.
An important problem for the programmers seems to be how to check the DEUCE input and output. It should here again be stressed that the popular opinion that statistical processing does not need to be as exact as for instance book-keeping, must be rejected. A set of interdependent statistical tables is a well defined system comprising a large number of definitional relations which cannot be violated.

To minimize the risk of processing errors, we have tried to estimate the optimum length between re-run points, i.e. how often special summary cards should be punched out. Another problem is how to reduce the subsequent control by tabulators. The solution must be to build up and check the tables internally in DEUCE as far as possible. This is, however, a question of storage capacity.

We have very good experience of the time required for programme-testing and test runs on DEUCE. The time required is less than estimated and reported from many other EDP installations.

The statistical product increased considerably at the same time as the External Trade Statistics was transferred to DEUCE. This was due to larger tables and increased data card volume. Thanks to DEUCE we managed the increase at the same time as improving the quality and increasing the service to the customers of statistics.

III. Application to Consumer Expenditure Survey

One of the first applications of DEUCE was the processing of the 1958 Consumer Expenditure Survey. Quite early in the planning stage it was decided that DEUCE was going to be used. The processing routine relied on 3 main DEUCE programmes: one combined control and aggregation programme, one control and correction programme and finally an aggregation programme. In addition to these main programmes, it was necessary to write about 15 auxiliary programmes. Also standard DEUCE programmes were used in order to do regression analysis on the material.

The 1958 Consumer Expenditure Survey included nearly 4,800 households. The information about the households and their expenditures were punched in 4,800 master cards and 300,000 detail cards. The master cards held common information, regarding the household, while each detail card contained in addition to household number, up to 5 items each of which consisted of commodity code, quantity and value.

First, the detail cards were sorted by household number and were run through DEUCE under control of the combined control and aggregation programme. As the title indicates the main purpose of this first DEUCE run was editing of the data. Each item had to pass the following tests in order to be accepted:

1. Commodity code should be valid code.
2. The commodities for which both quantity and value were required, should have both specified.
3. The commodities for which only quantity or only value was required, should have only quantity or value specified.
4. The price of a commodity item for which both quantity and value were required should be between bounds specified for each commodity.

The author is grateful to Mr. Thor Aastorp of the Central Bureau of Statistics for his assistance during the preparation of this section.
5. The value of the commodity item for which only value was required, should be within bounds specified for each commodity.

In order to carry out above tests, the programme included a table that contained the necessary information about all valid commodities, as commodity code, bounds for price or value, etc. This table was stored on the drum.

In addition to the tests mentioned above, all cards were checked against incorrect sorting and punching errors as blank columns and double punching.

Items which did not pass all the tests satisfactorily were rejected, and the machine punched a copy of all cards that contained one or more rejected items. In these copies, punched markers identified which item or items were rejected.

The aim of the programme was also preliminary aggregation. About 30 commodities were specially processed in DEUCE. As to these commodities, estimates for quantity and value were wanted according to two different principles: a. Quantity and value purchased, and b. Quantity and value consumed. For every household DEUCE punched one card for each of these 30 commodities containing quantity and value after both principles. Rejected items were excluded from the sums which therefore were incomplete.

During the same run through DEUCE the value of all accepted items were aggregated in a table of 30 commodity groups. For each household this table was punched out, 5 groups on a card. These cards were likewise incomplete summary cards because only accepted data were included.

In addition, it was punched two control cards per household. The first card contained total income and total expenditure, which should be equal because all accounts should balance. The other card contained information about number of rejected items, number of accepted items, how many cards that contained only accepted items, and how many that had one or more rejected items. This second control card was very useful for evaluating the editing and aggregation process.

The rejected cards were inspected and corrected manually and later processed by the same programme giving at last a complete set of summary cards. This programme consisted of 1,200 instructions including constant tables.

Based on a study of the 30 specially processed commodities it was decided to use the principle of purchase for all commodities except 3 which had to be corrected.

The second programme consisting of about 500 instructions was made to correct information in the commodity groups containing these three commodities. During this run DEUCE produced a new, corrected set of cards for commodity groups for each family. The cards were now also supplied by codes for the size and income of the family, and the information was again controlled for consistency. These cards gave the complete basis for tabulating all tables of the first statistical report of the Consumer Expenditure Survey.

A main purpose of the Consumer Expenditure Survey was to provide a weight basis for a new cost of consumption index. For this purpose a much more detailed specification than the above mentioned 30 groups was wanted.

All detail cards were therefore sorted in groups by the code of the position of the family head, by family size and composition and by income. The third main run through DEUCE based on a 2,000 instruction programme resulted in a set of summary
cards each containing quantity and value for each commodity code within each group. This processing gave about 60,000 cards which were ready for tabulation.

The 1958 Consumer Expenditure Survey did not, however, comprise family groups the heads of which were farmers, fishermen or old age pensioners. These had been analysed previously, and to make the results compatible with results of the 1958 Survey, a regression analysis programme was utilized. The regression processing took into account changes in the average family size and income as well as in the prices from observation periods to 1958.

An intensive regression analysis was also performed on the 1958 data which gave very useful knowledge about details in the consumption pattern of families.

All processing plans for the 1958 Consumer Expenditure Survey have not been completed. Valuable experience can, however, be reported. Only by an extensive application of automatic editing it has been possible to obtain such a high degree of statistical accuracy and detailed information. On the other hand, we also learned that large processings of this kind causes queueing conflicts which may delay the processing and require revised time-tables.

So far, the processing of the 1958 Survey has required about 18 months of programming and 500 DEUCE-hours. A similar processing by means of conventional punched card machines would have required punching of 4 times as many cards and a very large number of calculating and tabulating hours as well as large amount of clerical work.

IV. Application to time series decomposition

A recent application of DEUCE is automatic decomposition of monthly time series. In order to test and evaluate different methods, several special decomposition programmes have been made of which two are described here. The methods are now applied to large number of monthly series.

Method A is based on the widely used techniques of ratios between original series and moving averages. The procedure assumes a multiplicative composition of three components, a trend-cycle, a seasonal and an irregular component. DEUCE produces a set of 13 tables for each monthly series. The tables are:

**Table 1:** Identification of the months, i.e., the heading of the following tables.
**Table 2:** Original series.
**Table 3:** 12-month moving averages of original series.
**Table 4:** 13-month weighted, moving averages of original series.
**Table 5:** Preliminary uncentered ratios between original series and 13-month averages.
**Table 6:** Preliminary centered ratios between original series and 13-month averages.
**Table 7:** Preliminary seasonal factors computed as a 5-month weighted averages within each column of table 6.
TABLE 8 : Preliminary seasonally adjusted series as ratios of original series and preliminary seasonal factors.

TABLE 9 : 5-month averages of preliminary seasonally adjusted series.

TABLE 10 : Final uncentered ratios between original series and above 5-month averages.

TABLE 11 : Final centered ratios between original series and 5-month averages.

TABLE 12 : Final seasonal factors computed as table 7 on final centered ratios.

TABLE 13 : Final seasonally adjusted series.

The programme includes an extrapolation technique which gives substitutes for the lost values. This loss of information at the end of the series is a serious drawback because it is just this information which is of highest value for the current analysis of the economic development. From a theoretical point of view, the basis of this method is obviously not satisfactory.

Method B is more satisfactory. It is based on logarithmic transformation of the original series and application of a modified version of the method of A. Wald on the additive logarithmic components and finally a retransformation of the results from logarithmic to absolute basis. Under control of programme B, DEUCE processes the original series and produces 4 tables for each series :

TABLE 1 : Identification of months.

TABLE 2 : Original series.

TABLE 3 : Trend-cycle factors.

TABLE 4 : Seasonal factors.

Subject to a few general assumption, programme B gives values of both trend-cycle factors and seasonal factors also for the very last months of the series. This feature makes the method superior to method A for practical applications in current time series analysis.

Both programmes comprise about 1,500 machine instructions. As DEUCE is a two level storage machine and the programmes are made for both short and long series, they are not very fast compared with the speed of similar processing on more advanced machines. The processing time varies from 8 to 15 minutes.

So far, a large number of series has been analysed, but it is yet too early to report any conclusions. It may also be necessary to develop special simulation programmes to evaluate this type of automatic time series analysis.

V. Evaluation of electronic data processing

After two years of electronic data processing, our general experience is that a medium-size computer is a very powerful instrument in the statistical data production. Our EDP-machine substitutes a large amount of clerical work as well as tabulating and calculating machines and contributes to a larger statistical product without corresponding increase in the cost of processing. The statistical product is increased by a
larger number of statistical results and a higher quality is obtained because of a more consistent processing.

The clerical work saved is due partly to the transfer of the editing operations to automatic processing and partly to the integration of smaller processes into larger avoiding manual data handling. Here a gain in processing time has been obtained, but from an overall point of view a faster processing cannot yet be reported because there are still too many manual operations interrupting the automatic. Looking upon special and more complex jobs as computation of variances and covariances, regression estimates, time series components, the experience, however, indicates also a great reduction in processing time.

We have not so far been disappointed, but some new problems have arisen of which the flexibility and queueing problems shall be mentioned. The use of programme controlled EDP-equipment in an integrated process requires a far more thorough preparation than conventional processing. The objectives of a processing have therefore to be specified completely at an earlier time, the consequence of which is a less flexible processing schedule.

The second problem is the queueing problem. Previously, the work was performed in parallel by several tabulators, etc. In a processing based on electronic systems everything has to be processed by one or a few units giving rise to queueing problems. This does not necessarily mean that the overall processing time is prolonged, but we do get a new class of priority problems.

4. *Future application of EDP-equipment in the statistical data production*

I shall finish this paper with some considerations about the future. The need for statistical information about the society will increase with the growth and complexity of the society itself. In Norway, the statistical data product indicated by the number of estimates produced has increased roughly 75 per cent during the last decade.

I think the solution of future requirements will be a statistical system based on files of collected and partly processed information. The collection and editing of data must be carried out continuously and quite independently of the subsequent processing, while the processing will have to be completely based on the modern estimation theory and methods. The collected and edited information has to be quickly transferred and stored in files, the units of which correspond to the statistical units. Within a unit record, information is recorded chronologically. One main feature of the recording is cross references to other files.

Compared with the present system, this statistical file system will help to give us a more steadily data production avoiding the up and downs owing to the periods and points of observation as month, quarter, year and decade now used, more effective utilization of the machines, the possibility of utilizing historical information independently of conventional statistical grouping, and of studies of the trends on unit basis rather than on aggregated basis as at present.

The statistical file system consists, from a processing point of view, of two main
processes of rather different nature. The first is recording, file maintenance and information retrieval combined with relative simple arithmetical and logical operations to maintain the necessary control of the collected data. The second process is a computational process to utilize retrieved data in statistical analysis. To avoid serious queueing trouble it must be possible to carry out these processes simultaneously.

Electronic data processing machines will be necessary in order to establish such a system. The requirements of a statistical file system to the technical development seem not to be peculiar. Manufacturers of data processing equipment are to-day working with and solving problems as automatic character recognition, updating of files, information retrieval and simultaneous processing.

The preparation for operating a statistical file system by means of electronic data processing equipment seems also to be of two main types, preparing programmes for the filing and retrieval process and preparing programmes for the statistical-analysis on basis of retrieved information. The present draw-back of such a system is that it will not be flexible enough because even a small change requires a time-consuming programming preparation. At this point, I think that in the future our work can be based on time-saving automatic programme generators and programming languages in which the present extensive routine parts of the programming is left to the machine itself.

Programming for filing and retrieval processes can be reduced very much by preparation of a standard generator programme for this part of the work by which the machine itself only by a few parameters produces its own filing-retrieval programmes. As to the statistical analysis, I think many simple processing programmes also may be generated by a similar report or statistical tabulating generator programme. There will, however, always be a need for special processing programmes which I hope can be worked out in the automatic programming languages now under development.
Depuis 1900, le Bureau central de Statistique en Norvège a traité sa documentation statistique au moyen de machines à cartes perforées. Au commencement de 1959, le Bureau a aussi employé un calculateur électronique du type DEUCE Mk II et, en mai 1961, un calculateur du type IBM 1401 sera installé.

Tout le travail fait par les machines sur les documents statistiques est centralisé dans le département mécanographique. Cette section organise le travail des programmes et dispose de l’équipement mécanographique.

L’équipement électronique est utilisé pour beaucoup de travaux de statistiques différents. On a discuté les programmes pour la Statistique du commerce extérieur, pour la recherche de la consommation 1958 et pour les statistiques mensuelles. Les avantages du travail électronique se sont manifestés principalement en donnant de meilleurs résultats, ainsi que des résultats statistiques plus détaillés pour les mêmes coûts.

Le besoin de l’information statistique continuera à croître rapidement et à présenter de grandes exigences sur les méthodes statistiques. Un système fondé sur la construction systématique des archives statistiques qui permettent de choisir facilement l’information et de procurer par le seul travail des machines les documents de base nécessaires sera une condition absolue. C’est inconcevable qu’un tel système puisse s’établir sans s’appuyer sur les méthodes très précises d’utilisation électronique de la documentation statistique.