Statistics Canada   Statistique Canada

Methodology Branch

Time Series Research and Analysis Division

Direction de la méthodologie

Division de la recherche et de l'analyse des chroniques

Canada

WORKING PAPER NO. TSRA-86-013E

TIME SERIES RESEARCH & ANALYSIS DIVISION
METHODOLOGY BRANCH

Views on

BENCHMARKING AND INTERPOLATION

by

Pierre A. Cholette and Rene Piche

WORKING PAPER NO. TSRA-86-018E
TIME SERIES RESEARCH & ANALYSIS DIVISION
METHODOLOGY BRANCH

Statistics Canada Statistique Canada

Ottawa, Canada

views on

BENCHMARKING AND INTERPOLATION

by

Pierre A. CHOLETTE

and

Rene PICHE

STATISTICS CANADA

presented at the

Statistics Canada / U.S. Bureau of the Census
Methodology Interchange

NOTE: The top each page corresponds to the content of the slides presented at the conference. The bottom of each page are possible comments made by the presentator.

Direct enquiries to:

Pierre A. CHOLETTE

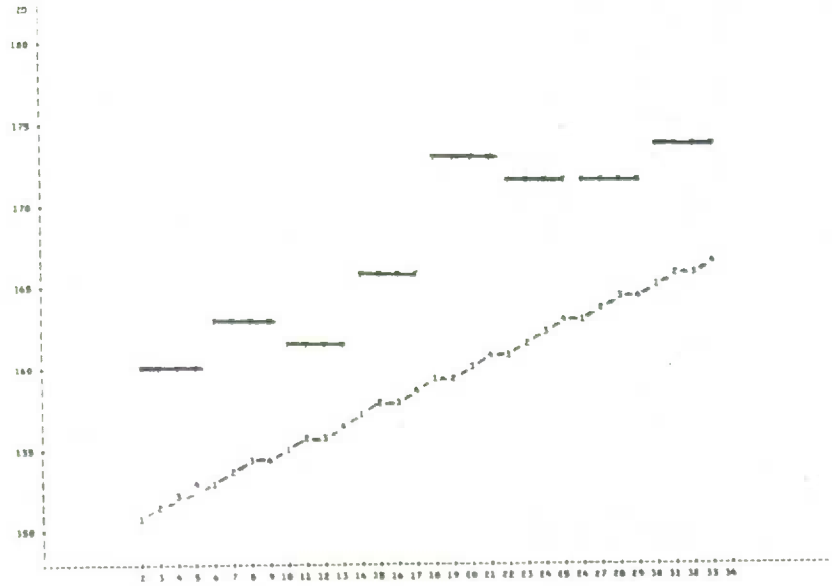
Time Series Research and Analysis
Statistics Canada
R.H. Coats Building, 13th Floor, "J"
OTTAWA, Canada
K1A 0T6

(613) 991-1600

BENCHMARKING SITUATIONS
COVERED IN THIS PRESENTATION

- annual benchmarks every year
- scatterred annual benchmarks (not available every year)
- scattered annual benchmarks and scattered sub-annual benchmarks
- discontinuities in movement
- seasonal patterns available for base years only
- interpolation by means of growth rates

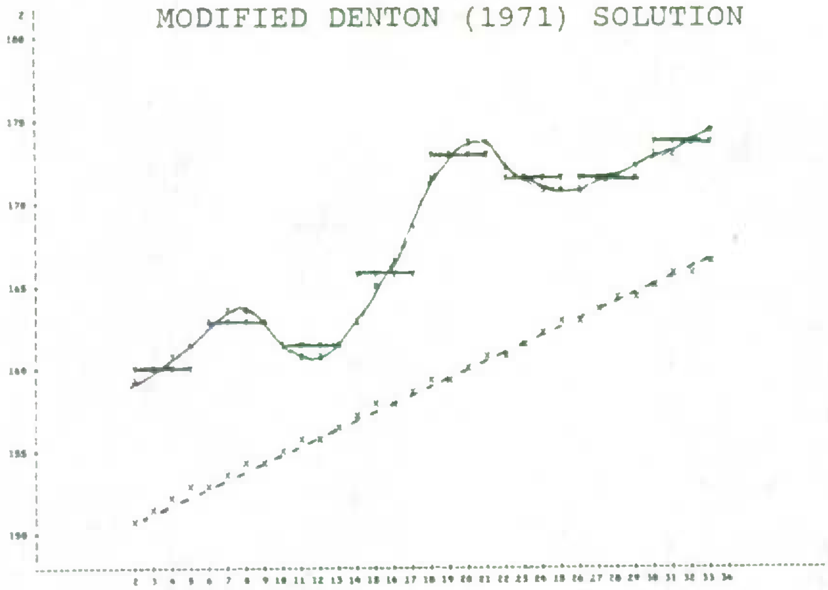
BENCHMARKS EVERY YEAR



The dotted curve stands for the original sub-annual unbenchmarked series.

The steps indicate the average annual level of the desired series dictated by the annual benchmarks. In the case of flow series, the values plotted are those of the annual benchmarks divided by the number of months or quarter in the year. In case of index series, the values plotted exactly correspond to the annual benchmarks.

The problem of benchmarking is one of correcting the original series so that its level complies with the annual benchmarks.



Cholette (1984), **Survey Methodology**, Statistics Canada, Vol. 10, pp 35-49.

The modified Denton solution (Cholette, 1978, 1984) consist of drawing through the average annual levels a curve which

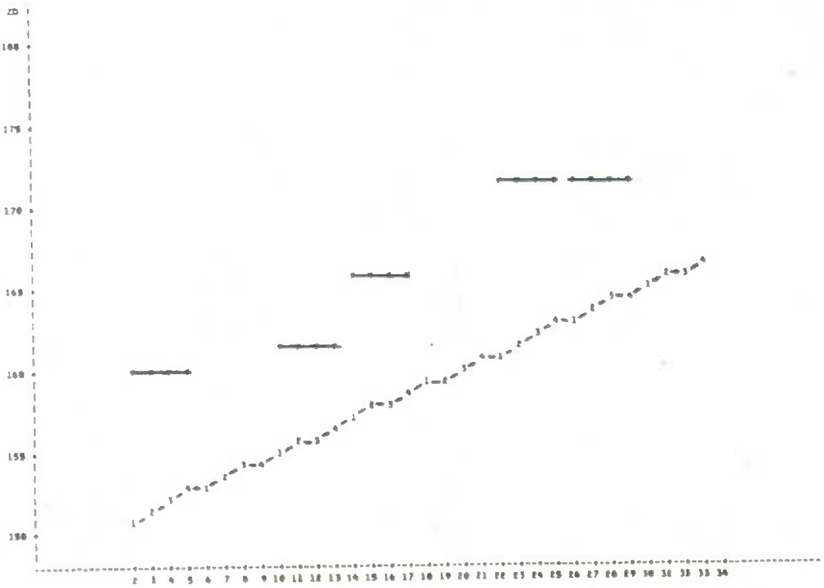
1) is as parallel as possible to the original, that is preserves the only available sub-annual movement,

2) covers the same annual surfaces as the average annual level.

The benchmarked series is represented by the solid curve on the screen.

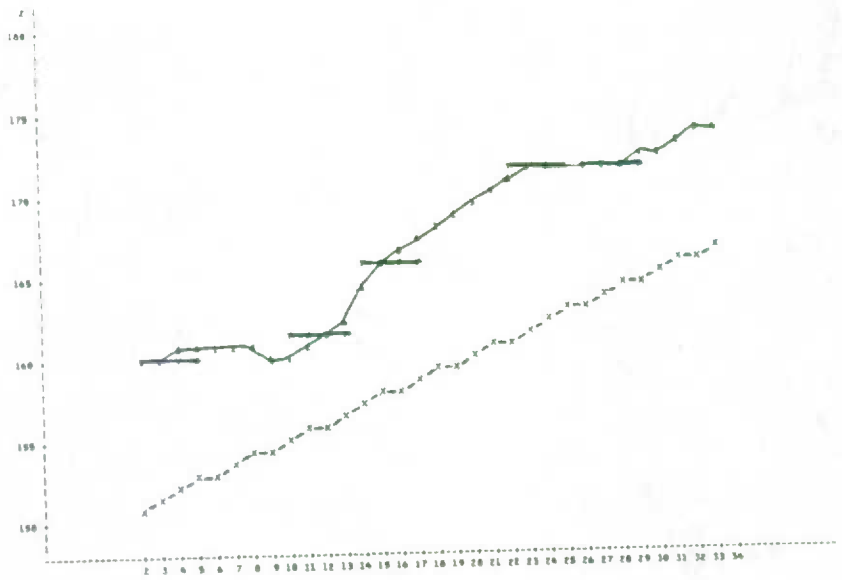
This illustration does not represent a realistic case: There is no seasonality and the movement of the orginal series is simplistic. However that simplicity will make easier to illustrate certain of the concepts in this presentation.

SCATTERED ANNUAL BENCHMARKS



In practical situations, it is quite common that annual benchmarks are not available every year.

SOLUTION

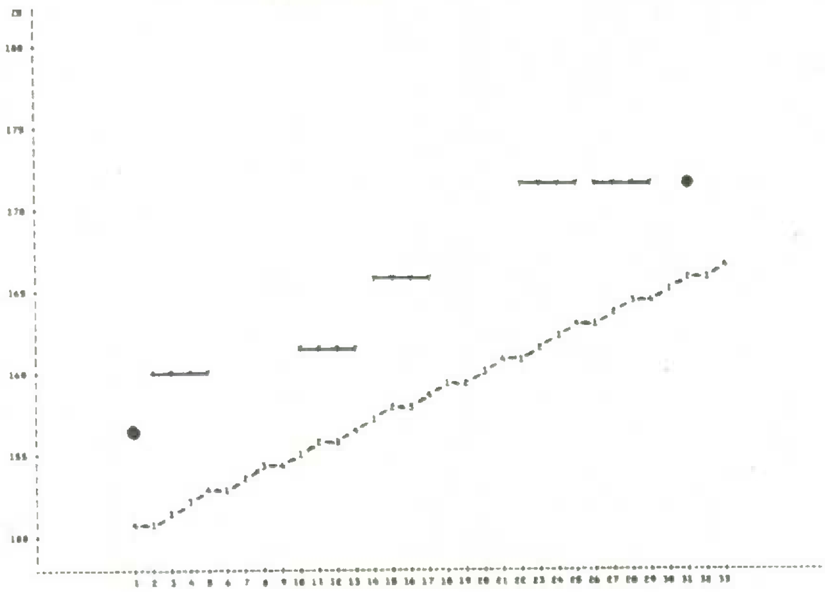


Again the benchmarked series (solid curve) covers the same annual surfaces as the average annual levels (steps) and is as parallel as possible to the original series (dotted) under the circumstances.

The solution is geometrically simpler than when benchmarks are available every year.

Super-impose previous slide. This illustrates the cost of not having annual benchmarks every year: cyclical movements are lost.

SCATTERED ANNUAL AND SUB-ANNUAL BENCHMARKS

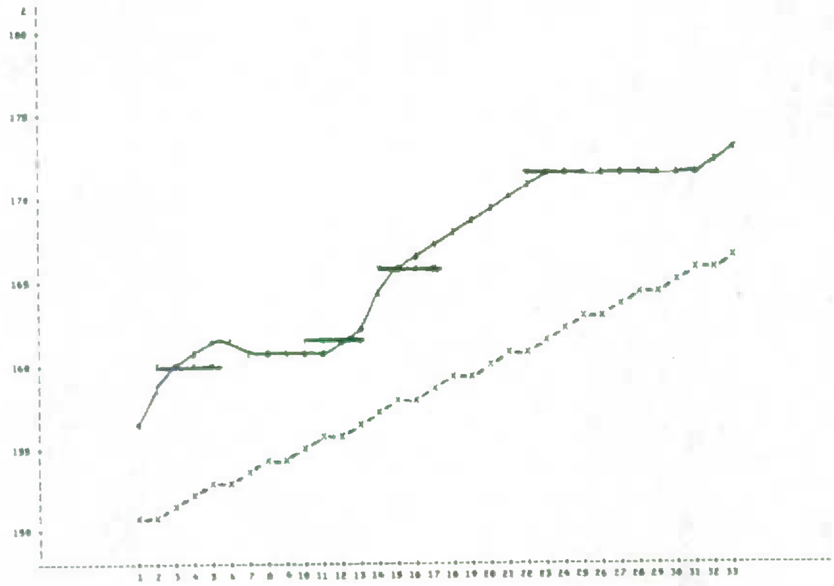


Apart from annual benchmarks, sub-annual benchmarks may be available. Those are pre-determined points through which the desired benchmarked series is required to pass. The sub-annual benchmarks are represented as big dots on the screen; the annual benchmarks, by the steps; and the original unbenchmarked series, by the dotted curve.

The first sub-annual benchmark may have to do with the fact that the point and those prior to it are "frozen", i.e. historical, benchmarked values.

Note that for stocks series where annual values correspond to one of the sub-annual observations (e.g. inventories), the annual benchmarks would actually take the form of sub-annual benchmarks. (There would be no steps in the figure.)

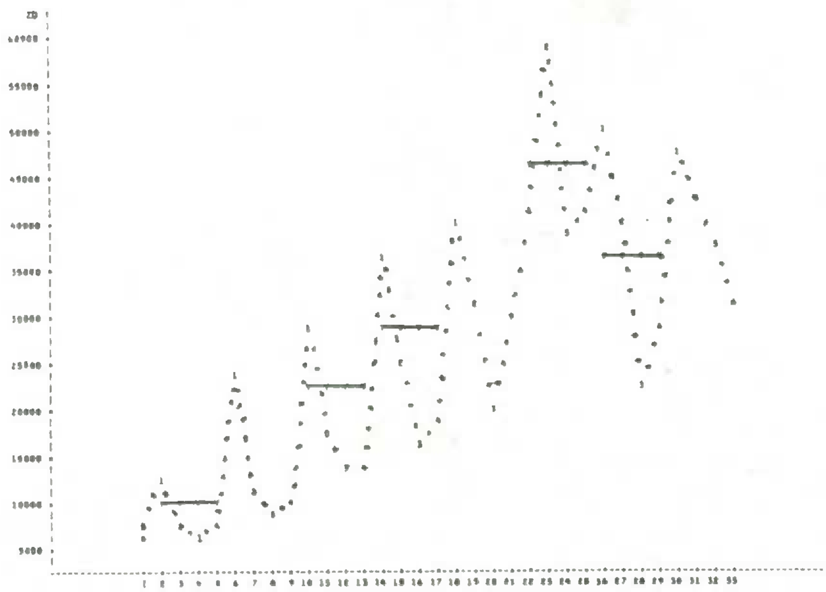
SOLUTION



Again, the benchmarked series (solid curve) is as parallel as possible to the original (dotted) under the prevailaing circumstances. The curve does pass through the sub-annual benchmarks (big dots) and covers the same annual surfaces as the average annual levels (steps).

Super-impose previous slide.

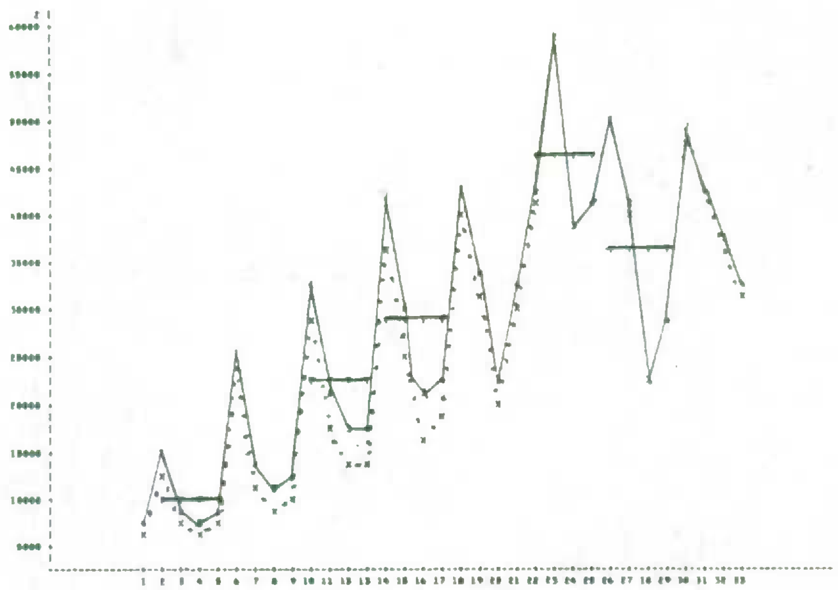
SCATTERED ANNUAL AND SUB-ANNUAL BENCHMARKS
IN A MORE REALISTIC SITUATION



The series contains seasonality.

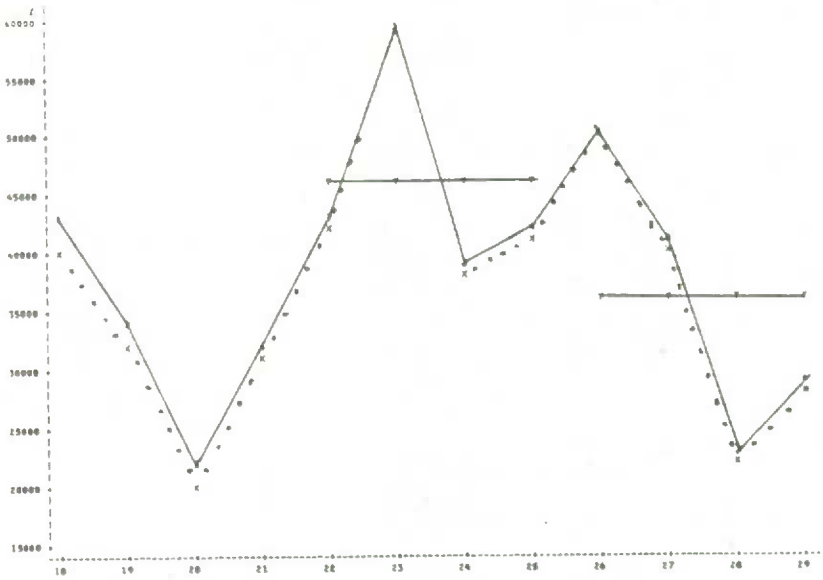
There is no annual benchmark available for the last year which is typical. Indeed the annual benchmark often becomes available quite a few months after the year is completed.

SOLUTION



Again, the benchmarked series (solid curve) is as parallel as possible to the original (dotted) ... even where this is not so desirable, namely in year 6.

CLOSE-UP OF YEARS 5, 6 AND 7



The seasonal pattern of year six is not at all similar to that of neighbouring years five and seven. Data usually recorded in the first quarter were possibly captured in the second.

We will soon see how the preservation of undesired movements can be prevented; in other words, how benchmarking can be used as an opportunity to correct the series for unjustified fluctuations (and not only for level discrepancies).

First the mathematical objective function underlying the previous illustrations will be presented.

OBJECTIVE FUNCTION

$$q(z) = \sum_{t=2}^{T} [(z_t - z_{t-1}) - (x_t - x_{t-1})]^2$$

constraints for the sub-annual benchmarks:

$$z_t = z^d_t, \quad t=t_1,t_2,\ldots,t_K, \quad K<<T$$

constraints for the annual benchmarks:

$$\sum_{j=1}^{J} z_{Jo+(i-2)*J+j} = y_i, \quad i=i_1,i_2,\ldots,i_M, \quad M<=I$$

T: number of observations in series
I: number of years in series (incl. imcomplete)
K: number of sub-annual benchmarks (to impose)
M: number of annual constraints
J: number of months per year
Jo: number of months in first year

The objective function minimizes the quadratic slope discrepancies between the original and the desired benchmarked series. In other words, it specifies that the two series should run as parallel as permitted by the constraints.

The sub-annual constraints simply specify that the values of the benchmarked series are already determined for some periods of time.

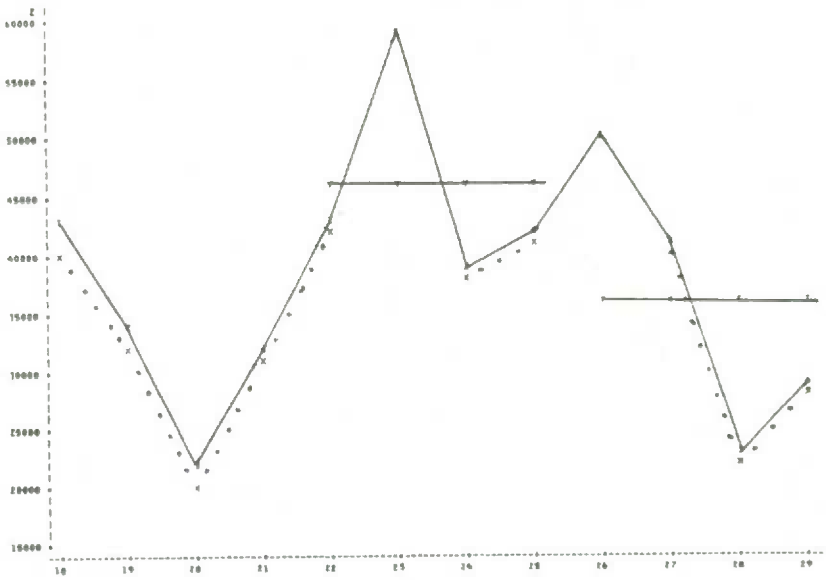
The annual constraints on the screen specify that the annual sums of the desired benchmarked series should be equal to annual benchmarks for some of the years.

For index series, multiply the annual benchmarks by the number of months per year J, since the annual benchmark correspond to the average of the year (e.g. prices indexes).

Do the same for stock series whose annual values correspond to the average of the year (e.g. unemployment).

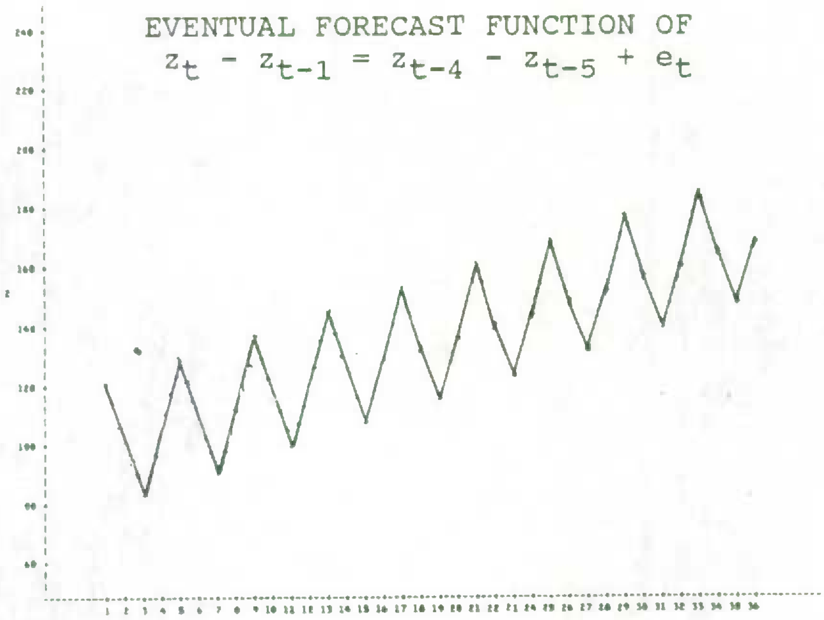
For stock series whose annual values correspond to one of the sub-annual values (e.g. inventories), the annual benchmarks can actually treated as sub-annual.

CLOSE-UP OF YEARS 5, 6 AND 7



The objective function specifies that the two series should run as parallel as permitted by the constraints. This is why all the movements of the original series, including the faulty seasonal pattern of year 6, are reproduced in the benchmarked series.

Avoiding this requires a modified objective function.



$$z^f_t = z_{t-1} + z_{t-4} - z_{t-5}$$

Before modifying the objective function let us look at some properties of simple seasonal autoregressive models. May-be these models are what is required in the objective function.

The simplest autoregressive model one can think of for a seasonal time series is displayed on the screen. The change between the first and the second quarter of this year (say) is basically the same as the change which occurred between the first and the second quarter of last year. Put differently, movements between two adjacent months or quarters tend to repeat themselves from year to year. In fact, such a simplistic model explains more than 85% of the variance of most socio-economic time series.

When the corresponding forecasting eauation is applied to the five initial values displayed on the screen and then to the forecasts themselves, a linear trend with a constant seasonal amplitude is generated.

FAMILY OF SEASONAL AUTOREGRESSIVE MODELS

$$z_t - z_{t-1} = z_{t-J} - z_{t-J-1} + e_t$$

is a special case of

$$z_t - f_1 z_{t-1} =$$
$$F_1(z_{t-J} - f_1 z_{t-J-1}) + e_t$$

where f_1 and F_1 were both equal to 1.0

Corresponding forecasting equation:

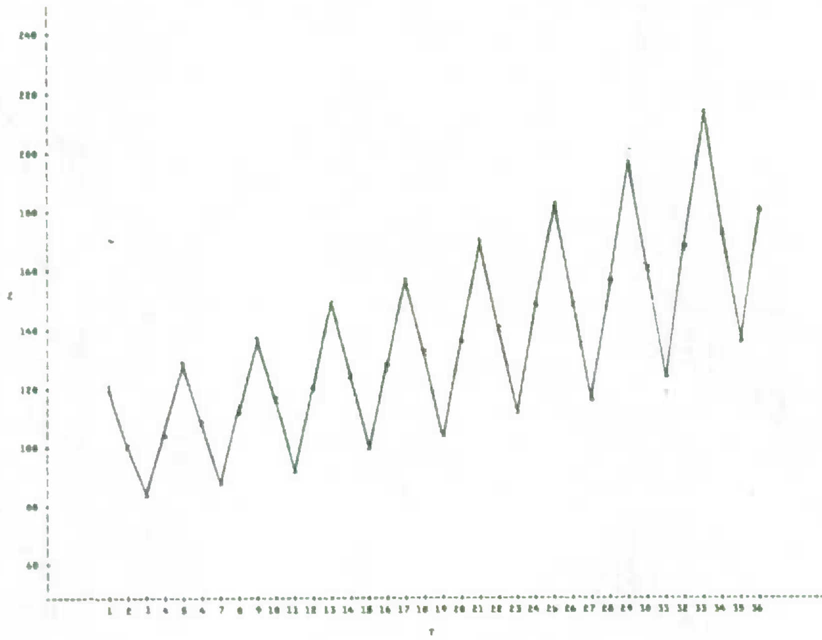
$$z^f{}_t = f_1 z_{t-1} + F_1(z_{t-J} - f_1 z_{t-J-1})$$

J: number of months per year

We will now explore which extra possibilities this more gereral model offers.

EVENTUAL FORECAST FUNCTION OF

$$z_t - 1.00z_{t-1} = 1.10(z_{t-4} - 1.00z_{t-5}) + e_t$$

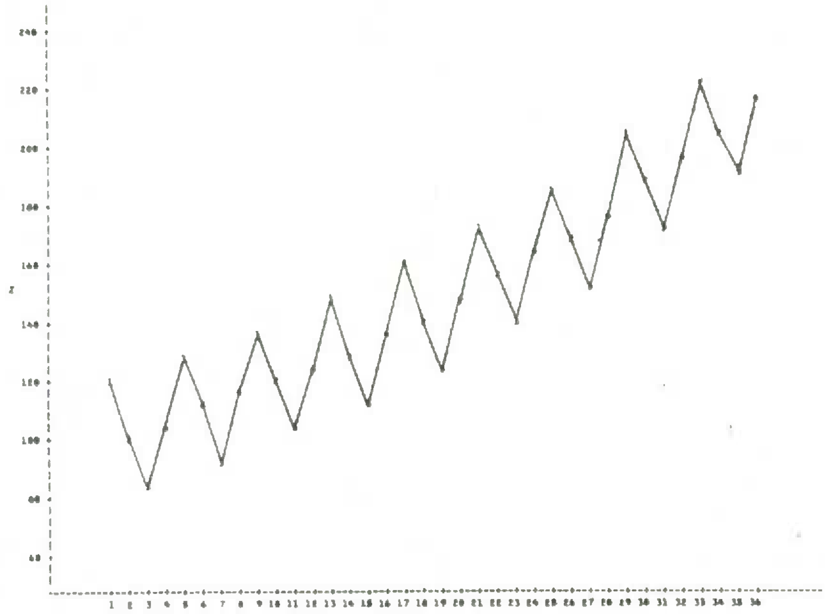


Movements between two adjacent months or quarters tend to increase by 10% from year to year.

The corresponding forecasting equation generates a linear trend with an increasing seasonal amplitude.

EVENTUAL FORECAST FUNCTION OF

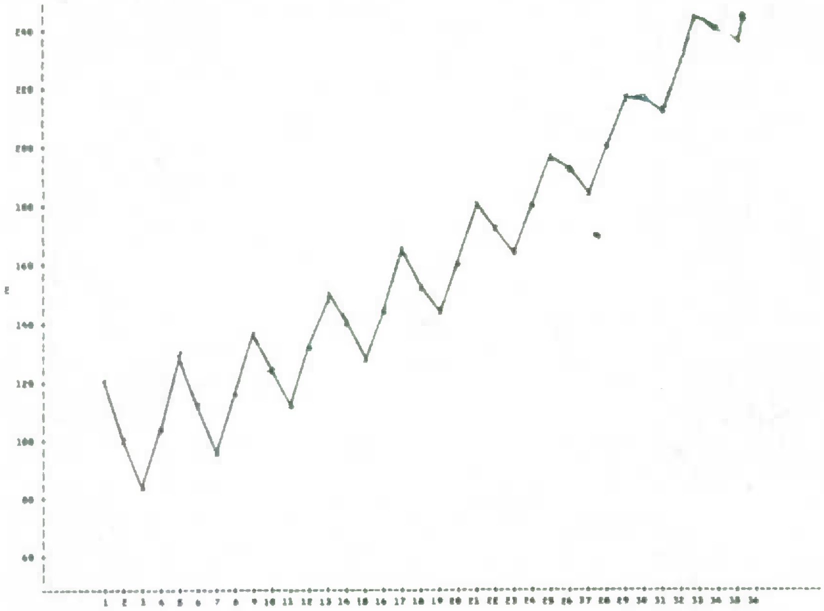
$z_t - 1.03z_{t-1} = 1.00(z_{t-4} - 1.03z_{t-5}) + e_t$



The corresponding forecasting equation generates an exponential trend with a constant seasonal amplitude.

EVENTUAL FORECAST FUNCTION OF

$$z_t - 1.03z_{t-1} = 0.90(z_{t-4} - 1.03z_{t-5}) + e_t$$



The corresponding forecasting equation generates an exponential trend with a shrinking seasonal amplitude.

As you begin to see the autoregressive models offer a rich variety of behaviours for our desired benchmarked series. They are now introduced in the objective function in the following manner.

MODIFIED OBJECTIVE FUNCTION

$q(z) =$

(parallelism criterion)

$$w \sum_{t=t_1}^{t_n} [(z_t - z_{t-1}) - (x_t - x_{t-1})]^2$$

$$+ (1-w) \sum_{t=J+2}^{T} [(z_t - f_1 z_{t-1}) - (F_1 z_{t-J} - f_1 F_1 x_{t-J-1})]^2$$

(seasonal autoregressive criterion)

w: weight of parallelism criterion, e.g. 0.95;
(1-w): weight of autoregressive structure

f_1, F_1: pre-determined by subject matter expertise

The constraints are the same as before.

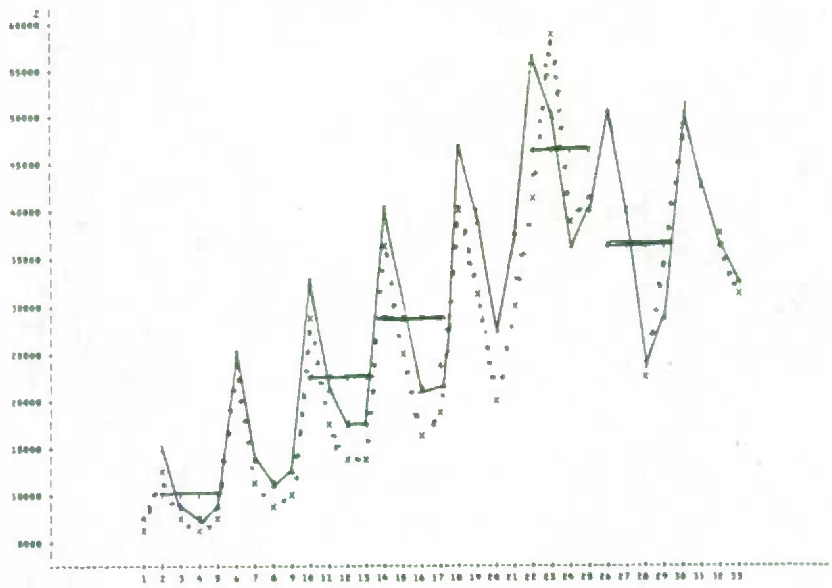
The second term of the objective function specifies that the desired benchmarked series should behave as much as possible as a pre-selected seasonal autoregressive process.

The index of the first summation controls the time periods for which the parallelism criterion is active. For those periods the autoregressive criterion is also active. However, if w is chosen high enough (e.g. 0.95), the autoregressive will be dominated by the parallelism criterion. In other words with high values of w, the autoregressive criterion is inactive in practice when the parallelism criterion is active.

In the periods for which the parallelism criterion is not active only the autoregressive criterion is active and therefore fully operational.

It would be possible to have the autoregressive criterion active only for the periods required. However such precision did not seem necessary in practice.

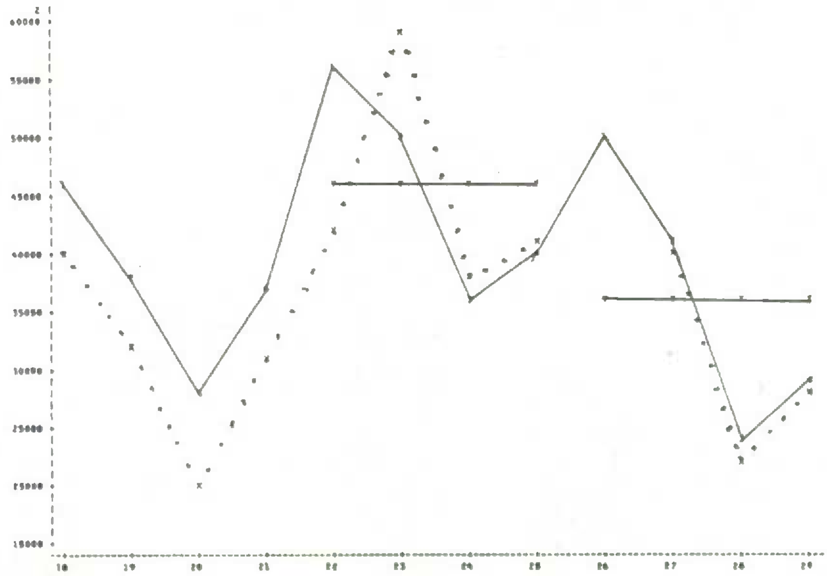
ILLUSTRATION OF
AUTOREGRESSIVE BENCHMARKING



The parallelism criteria was specified to be active only for periods 2 to 21 and 25 to 33. For these periods the benchmarked series remains as parallel as before to the unbenchmarked series.

For periods 22 and 23, only the autoregressive criterion was active. The corresponding benchmarked values fall in line with the same-quarter values of the other years.

CLOSE-UP OF YEARS 5, 6 AND 7

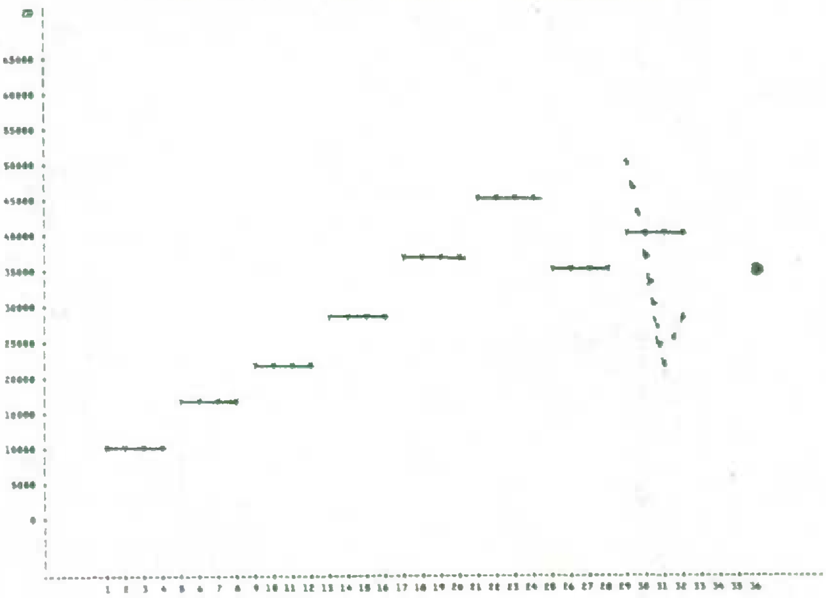


The seasonal pattern of year 6 is very similar to that of years 5 and 7.

The values of f_1 and F_1 were set to 1.0 and 1.05 respectively.

The value 1.05 was chosen for F_1, because the seasonal amplitude was increasing for the other years.

RETROPOLATING SEASONAL PATTERNS
WITH AUTOREGRESSIVE BENCHMARKING

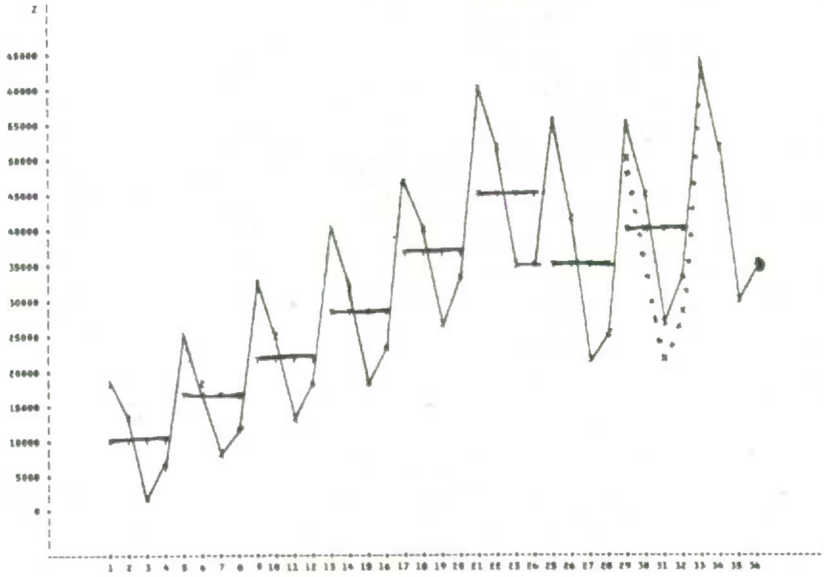


The situation depicted on the screen is the kind of benchmarking situation which initiated this idea of adding the autoregressive criterion to the benchmarking objective function.

A colleague of Statistics Canada came to me with a benchmarking problem essentially like that on the screen. From 1975 to 1981, he had no sub-annual unbenchmarked values at all. Sub-annual values started in 1982. His job was to create a quarterly series from the values he had, from his expertise of the socio-economic variables in question and from any direct or indirect information he could get.

Normally, the practise would have been to do the job by trial and error on a desk calculator.

SOLUTION
with f_1=1.0 and F_1=1.075

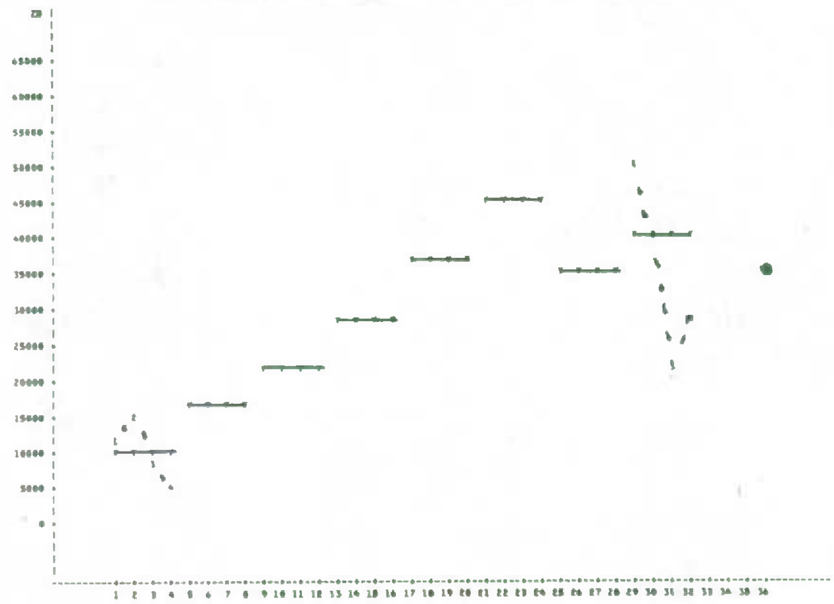


With the proposed autoregressive benchmarking method many series can be handled in a short period of time. The series experts can devote their time to what they are most competent in doing, namely to determining the inputs to the benchmarking method: the sub-annual and annual benchmarks, the most appropriate seasonal pattern, etc.

As shown on the screen, the seasonal pattern is successfully backasted.

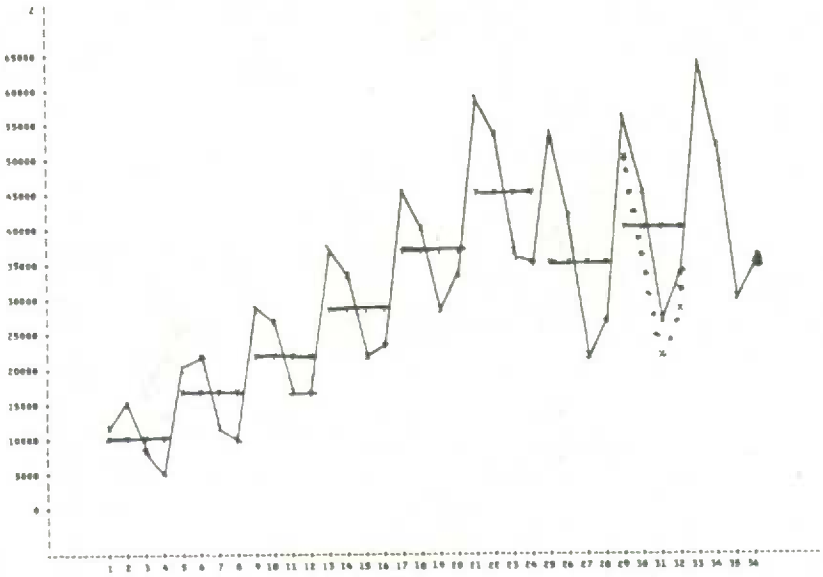
A potentially usefull by-product of the autorecressive criterion is the optinal availability of forecasts for year 9. For the sake of illustrating the possibilities of the method, the forecasts depicted here are required to pass through a point pre-determined by the expert, i.e. a sub-annual benchmark.

INTERPOLATING BETWEEN SEASONAL PATTERNS
BY AUTOREGRESSIVE BENCHMARKING



In the situation on the screen years 1 and 8 are "base years" and are the only ones with available seasonal patterns (big dots). The job is now to interpolate a seasonal curve between those two years consistent with the annual benchmarks (steps).

SOLUTION



The seasonal pattern of the benchmarked or interpolated series (solid curve) gradually changes from that in the year 1 to that in year 8.

GROWTH RATE CRITERION

$$q(z) = c \sum_{t=t_1}^{t_n} [(z_t - z_{t-1}) - (x_t - x_{t-1})]^2 \quad \text{(parallelism criterion)}$$

$$+ (1-c) \sum_{t=J+2}^{T} [(z_t - F_{t1} z_{t-J})^2 \quad \text{(growth rate criterion)}$$

c: weight of parallelism criterion e.g. 0.95;
(1-c): weight of growth rate criterion

F_{t1}: yearly growth rates pre-determined by subject matter expertise

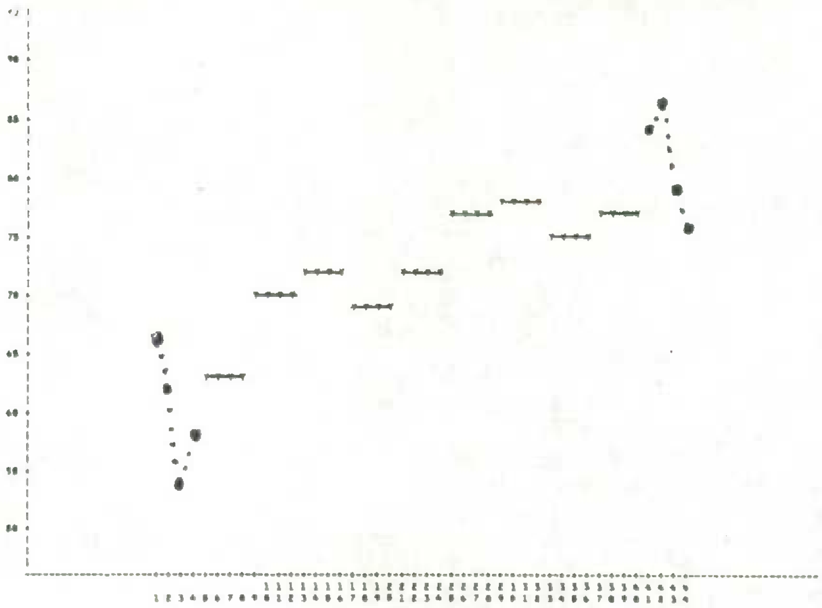
Cholette (1985), **Statistics Canada**, Methodology Branch, Working Paper TSRA-85-01EF.

Growth rates are determined from a combination of related series and relevant information by the series experts. That combination can be changed suddently according to the circumstances prevailing.

In the autogressive criterion the seasonal autoregressive parameter F_1 was not allows to change through time.

The constraints are the same as before.

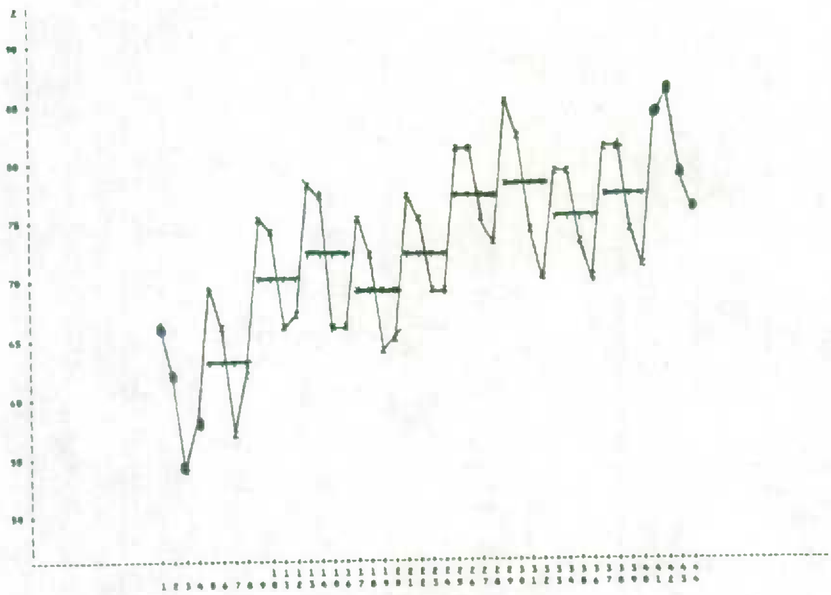
ILLUSTRATION OF GROWTH RATE BENCHMARKING



The growth rate situation typically arises when the current observations are generated by applying yearly growth rates to the observations of the previous year. This implicitly amounts to applying a counpounded rate to an initial base year displayed on the screen. In fact there are no real observations or unbenchmarked series to benchmark.

On the screen, years 1 and 11 are base years (big dots) and contain the only available seasonal patterns in the form of sub-annual benchmark. There are no unbenchmarked values, so that the parallelism is actually absent from the objective function. The question is: how can the growth rates and the annual benchmarks (steps) be combined with the base years to produce a satisfactory sub-annual (benchmarked) series?

SOLUTION



The solution is depicted on the screen. The benchmarked series (solid curve) obtained gradually departs from the seasonal pattern in the initial base year to adopt that of the terminal base year. Again the benchmarked series covers the same annual surfaces as the average annual values.

Note that the presence of the terminal base year automatically "freezes" the preceeding benchmarked values. Adding new years (of annual benchmarks and growth rates) after that terminal base year will have no impact on years 1 to 11, since the curve has to pass through the sub-annual benchmark values of the terminal base year.

POSSIBLE EXTENTIONS OF THE METHOD (conclusion)

- incorporating the ralated information explicitely in the objective function as Chow-Lin (1971), Fernandez (1981) and Litterman (1983)

- simultaneously benchmarking several series which are components of an aggregate

- indicative benchmarks

Litterman (1983), J.B.E.S., Vol. 1, pp. 169-173

Simultaneous benchmarking situations arise when each component series of an aggregate have to sum (or satisfy some identity with respect) to an aggregate series for each sub-annual period of time on top of its own annual and sub-annual benchmarks. Many such cases occur in financial statistics, international trade data, national accounts, etc.

The Canadian export series for instance are broken down by products and by country of destination. For each period of time the sum of the exported products should be equal to the total aggregate exports for the period. Furthermore, for each period of time, the exports totaled over products should be equal to the exports totaled over country of destination.

Mathematically the problem can be easily specified in a very satisfactory manner. Computationally, however, the size of the problem soon becomes unmanagable as the number of series and their length increase.

OBJECTIVE FUNCTION
WITH INDICATIVE BENCHMARKS

$$q(z) = \quad \ldots + \quad \text{(as before)}$$

$$+ \sum_{k=1}^{K} d_k \, (z_{tk} - z^d_{tk})^2$$

$$+ \sum_{m=1}^{M} g_m \left(\sum_{j=1}^{J} z_{Jo+(im-2)*J+j} - y_{im} \right)^2$$

d_k: weights of each sub-annual benchmark
g_m: weights of each annual benchmark

no constraints

(read t_k and i_m instead of tk and im)

In many situations, the annual and sub-annual benchmarks are not values which are fully and absolutely reliable. They may actually be less reliable than the unbenchmarked series. However they constitute information, and it is felt that they should be taken into account in deriving the benchmarked series.

The appropriate variant of the benchmarking methods is presented on the screen. The benchmarks are entered into the objective function as weighted terms instead of into constraints. In other words, the benchmarks are indicative as opposed to binding. Their weights is pre-determined by subject matter expertise.

This works and yields exactly the same results as constrained benchmarking if the weights are high enough. Actually this variant is computationally easier to implement than constrained benchmarking, because the matrices involved have smaller dimensions.