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

[redacted]

Working Paper No. TSRA-88-012E
Time Series Research and Analysis Division
Methodology Branch

CONVERTING FISCAL YEAR DATA INTO CALENDAR YEAR VALUES

by

Pierre A. CHOLETTE[1]

and

Andrew BALDWIN[2]

[1]Statistics Canada
Time Series Research
and Analysis Division
Coats Building, 13th floor "J"
OTTAWA, Canada
K1A 0T6

(613) 951-0050

[2]Statistics Canada
Prices Division
Jean-Talon Building
13th floor "D 2"
OTTAWA, Canada
K1A 0T6

(613) 951-3359

La transformation des chiffres d'années financières
en valeurs d'années civiles

- résumé -

Ce document présente une méthode pour rectifier les périodes de référence de données annuelles. Il s'agit par exemple de transformer des chiffres annuels financiers, couvrant les mois d'avril à mars de l'année suivante, en valeurs d'années civiles, se rapportant aux mois de janvier à décembre. Des variantes de la méthode conviennent aux séries de flux, de stock et d'indice. La méthode peut aussi s'utiliser pour désagréger des chiffres annuels (financiers ou pas) en valeurs trimestrielles ou mensuelles. L'approche accommode les situations où les périodes de référence des chiffres originaux varient d'une année à l'autre. Elle pourrait aussi s'adapter à la transformation de trimestres financiers en trimestres civils, à la transformation de paquets de chiffres hebdomadaires en chiffres mensuels.

- summary -

This document introduces a method to correct the reference periods of yearly data. It can be used to transform "fiscal year" data referring to the months of April to March of the following year, for instance, into calendar values referring to the month of January to December. Variants of the method are developed for flow, stock and index series. The method may also be used to dis-aggregate yearly data (fiscal or not) into monthly or quarterly values. The approach accomodates situations where the reference periods of the original data vary from occasion to occasion. It could also be adapted to transform fiscal quarter data into calendar quarters and bundles of weekly data into monthly values.

KEYWORDS: Interpolation, Benchmarking, Fiscal years, Calendarization, Quadratic minimization

INTRODUCTION

Much of the annual data published by statistical agencies reflect - or are perceived to reflect - the "calendar" year ranging from January to December. However, many companies and institutions supply the basic data in terms of "**fiscal years**", which differ from the calendar year. For example, the Canadian Federal and Provincial governments operate in terms of a fiscal year running from April 1st to March 31st. The U.S. Federal and most State goverments keep their accounts using a fiscal year running from July 1st to June 30th. In 1983, 20% of the Canadian retail sales were made by companies with fiscal years ending in January 1984; 12%, by companies with fiscal years ending in March 1984; and only 30%, by companies, with fiscal years ending in December 1983, that is with fiscal years coinciding with the calendar year).

This situation creates the fiscal year adjustment problem addressed by this paper, which consists of transforming fiscal year data into calendar year values. Two basic situations are distinguished. In the first and less favourable situation, the data are available only for fiscal years, and there are no sub-annual values for the variables of interest. The approach proposed is then to estimate the missing monthly values (from the fiscal year values) and to aggregate them into calendar year values. That problem is closely related to that of deriving monthly or quarterly series from annual series, addressed by Boot, Feibes and Lisman (1967). Their approach is generalized in Section 2 to accommodate fiscal years. It is shown that a simple and widespread fiscal year adjustment method is a particular case of the proposed method.

In the second and more favourable situation, there is a monthly or quarterly indicator - however approximate - for the variable of interest; but that indicator does not comply with the available fiscal year data. In this case, calendarization is viewed as a benchmarking problem. Benchmarking consists of adjusting sub-annual measurements of a variable so that they comply with annual and separately obtained measurements of the variable. The approach by Denton (1971) and Cholette (1984) is adapted in Section 3 to the problem addressed by this paper.

Section 4 shows the modification required to process stock and index series. Section 5 generalizes the method for situations where the reference periods of the fiscal year data change from year to year. Section 6 discusses implementational issues, like revisions and computational considerations. Section 7 concludes and suggests further extensions of the approach, such as the **calendarization** (which is the generic term) of fiscal quarter and of weekly data into quarterly and monthly values. The structure of the document is somewhat inspired by Ehrenberg (1982), who recommends results before methods.

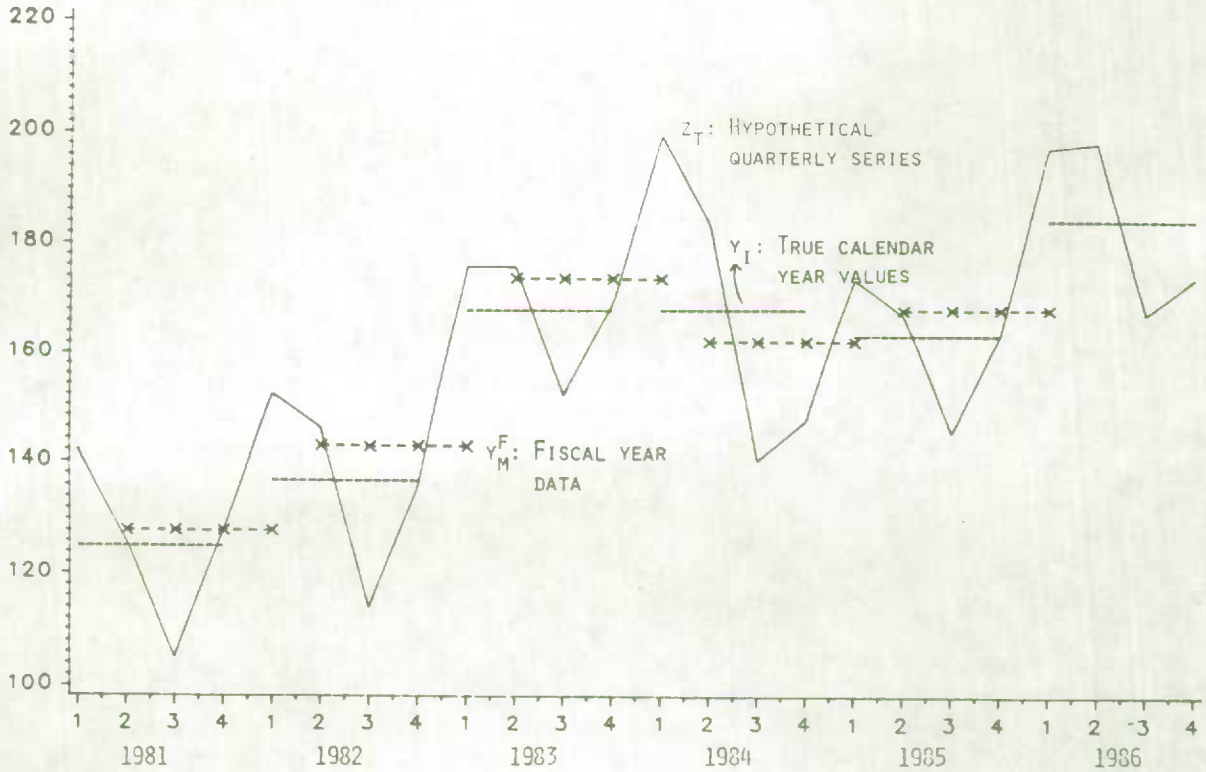

Figure 1: Quarterly series with corresponding fiscal year and calendar year values

2 CALENDARIZATION OF ANNUAL FLOW SERIES WITHOUT SUB-ANNUAL INDICATOR

For the purpose of calendarization, flow series are those whose annual values are the sum of their monthly or quarterly values. For instance, the number of cars manufactured in 1988 is the sum of the cars manufactured during each month of that year.

Figure 1 depicts a quarterly flow series along with its calendar year and fiscal year values represented by their average over each reference period. The fiscal year ends with the first quarter (March 31st). As illustrated, taking the fiscal year values for the calendar year values results in over-estimation for periods when the (trend-cycle of the) series rises; and, in under-estimation when the series declines. The quarterly series is unknown in practice and is displayed for the sake of illustration.

2.1 A Simple WideSpread Calendarization Method

Many statisticians are already acquainted with a simple method used in statistical agencies to transform fiscal year data into calendar year values. That method is a particular case of the method proposed in Section 2.2 and will be used as an introduction. For fiscal years ending with the first quarter (as in Figure 1), the ith calendar year estimate y^c_i is a weighted average of the two closest fiscal year values y^f_{i-1} and y^f_i:

(2.1') $$y^c_i = (1\ y^f_{i-1} + 3\ y^f_i)\ /\ 4, \quad i=2,3,...$$

Fiscal year i (i.e. the ith fiscal year) is given three times the weight the fiscal year i-1, because the last three quarters of calendar year i are in fiscal year i and one quarter of calendar year i is in fiscal year i-1. The more general formula is:

(2.1) $$y^c_i = (K\ y^f_{i-1} + (J-K)\ y^f_i)\ /\ J, \quad i=2,...,N-1,$$

where K is the number of periods (months or quarters) of calendar year i which are in fiscal year i-1. Parameter J is equal to 4 for quarterly series and to 12 for monthly series. Parameter N is the total number of calendar years in the series, that is the number of fiscal year plus one. For Figure 1, the parameter values are K=1, J=4 and N=6. For fiscal years running from September 1st 1981 to August 31st 1985, the parameters would be K=8, J=12 and N=5.

It can be shown that the first and last calendar year estimates consistent with the simple method are

$$\begin{aligned} y^c_1 &= ((J+K)\ y^f_1 - K\ y^f_2)\ /\ J, \\ y^c_N &= (-(J-K)\ y^f_{N-2} + (2J-K)\ y^f_{N-1})\ /\ J; \end{aligned} \qquad (2.2)$$

specifically in the situation of Figure 1,

$$y^c_1 = (5\ y^f_1 - 1\ y^f_2)\ /\ 4, \qquad y^c_6 = (-3\ y^f_4 + 7\ y^f_5)\ /\ 4. \qquad (2.2')$$

Those terminal estimates are extrapolations in the sense that some of their reference periods are outside any of the available fiscal years. Section 6 discusses the potential use of extrapolations.

The estimates corresponding to (2.1') and (2.2') were calculated for the fiscal year data of Figure 1. The resulting percentage errors with respect to the actual calendar year values displayed are -0.9%, 1.7%, -1.2%, -2.0%, 2.6% and 5.9% for years 1 to 6 respectively. (The error is larger for year 6, because its estimate is an extrapolation and obtained with second differences.) Considering the fiscal year data as calendar year values entails substantially larger errors of 2.1%, 4.4%, 3.5%, -3.8% and 3.7% for years 1 to 5 respectively (doing that yields no estimate for year 6.)

2.2 Proposed Calendarization Method - Second Differences Variant

The simple fiscal year calendarization method just described is in fact a special case of a more general one. The approach now proposed consists of estimating the unknown sub-annual, monthly or quarterly, values from the available fiscal year data; and, of temporally aggregating the resulting sub-annual estimates into calendar year values. The estimation of the sub-annual values is achieved by means of constrained mathematical optimization. Namely the following objective function is minimized

$$f(z) = \sum_{t=3}^{IJ} (z_t - 2z_{t-1} + z_{t-2})^2 \qquad (2.3)$$

subject to constraints

$$\sum_{j=1}^{J} (z_{(i-1)J+j+K}) = y^f_i, \quad i=1,..,I-1. \qquad (2.4)$$

Variable z_t is the desired sub-annual estimate for period t. Parameter I is the number of calendar years considered in the series interval ($3 \leq I \leq N$). Again J equals 4 for quarterly series and 12 for monthly series; and K is the number of periods (months or quarters) of calendar year i which are in fiscal year i-1.

Objective function (2.3) is a quadratic sum of second differences of the desired monthly or quarterly estimates z_t. These second differences specify that each desired sub-annual value z_t lies on the straight line running through z_{t-1} and z_{t-2}. The extent to which this local linearity is achievable is determined by the constraints. Constraints (2.4) specify that the desired estimates must sum to the fiscal year data y^f_i. Global linearity over the series interval is achieved only if the fiscal year data themselves behave in a perfectly linear manner. (The estimated sub-annual values correspond to the trend-cycle component of the series. This will deserve more comments later on.)

The solution to this optimization problem is developed in the Appendix using matrix algebra. The resulting sub-annual estimates are weighted averages of the fiscal year data:

$$z_t = \sum_{m=1}^{I-1} w_{t,m}\, y^f_m \quad t=1,\ldots,IJ \qquad (2.5)$$

The required calendar year values are then simply the annual sums of the sub-annual estimates:

$$y^c_i = \sum_{j=1}^{J} z_{(i-1)J+j}, \quad i=1,\ldots,I. \qquad (2.6)$$

If the sub-annual estimates are of no interest in themselves, the desired calendar year values can be expressed directly in terms of the fiscal data, by substituting (2.5) into (2.6):

$$y^c_i = \sum_{m=1}^{I-1} v_{i,m}\, y^f_m, \quad i=1,\ldots,I \qquad (2.7)$$

The weights $v_{i,m}$ in (2.7) depend only on I, J and K. (The same also holds for weights $w_{t,m}$ in (2.5).) They do not depend on the observed fiscal data y^f_i. They can therefore be calculated once and for all and applied to any series with the same values of J and K. Precalculated weights are found in Cholette (1987a) for I=5, J=4 or J=12 and K=1,...,J.

For series interval I equal to 3, i.e. when the problem is specified on two fiscal years, the solution is given by equations (2.1) and (2.2). In other words, the simple fiscal year calendarization method described in Section 2.1 is a particular case of the one presented here. Estimates derived on 5-year series intervals (I=5) are probably more reliable, because they depend on four fiscal values instead of two.

The method of Boot, Feibes and Lisman (1967) for interpolating sub-annual estimates from annual data is also a particular case. Their method produces no extrapolated sub-annual values. The authors make no mention about the possible fiscal nature of the annual data and about the possible use of their method for calendarization purposes. The method presented in this section, on the other hand, may be applied to fiscal year data and produces extrapolated sub-annual values at the beginning and end of the series (as evidenced by (2.2) and (2.5)).

2.3 Proposed Calendarization Method - First Differences Variant

The calendarization method in Section 2.2 minimizes second differences of the estimated sub-annual values. For reasons which will become apparent, first differences may be more appropriate in most calendarization situations. The following objective function is then minimized

$$f(z) = \sum_{t=2}^{IJ} (z_t - z_{t-1})^2 \qquad (2.8)$$

subject to the same constraints (2.4). This new objective function specifies that the desired sub-annual values z_t are as constant as possible from one period to the next. The extent to which this local constant character is achievable is determined by the constraints. Constraints (2.4) specify that the desired estimates must sum to the fiscal year values y^f_i.

The solution developed in the Appendix still holds for first differences. The estimated sub-annual values z_t and the calendarized yearly values y^c_i are weighted averages of the available fiscal year values as in equations (2.5) and (2.7). The weights are now different but still independent of the fiscal year values. (However the first differences solution is no longer a generalization of the simple solution presented in Section 2.1.)

Figure 2 compares the sub-annual values obtained from first differences and from second differences. The two curves are almost identical implying very similar calendarized values (not displayed), except for the first and the last years of the series. In those years, the first difference sub-annual estimates level off, whereas the second difference estimates maintain their linear behaviour. This linear behaviour, associated with second differences, could be appropriate for series which behave monotonically (e.g. linearly, exponentially monotonically), in a very predictable manner. Grocery sales may be such a candidate. However, the first difference curve in the figure also implies that the next fiscal year value will be higher than the last one available. First differences may then be sufficient for most variables with predictable linear behaviour. Second differences on the other hand are clearly inappropriate for series

subject to the business cycle, like those in Figures 1 and 2: The resulting sub-annual and calendarized yearly values for the last year are subject to heavy revisions, if the next fiscal year value is lower than the last one available. In our opinion, first differences are preferable to second differences for the majority of socio-economic time series.

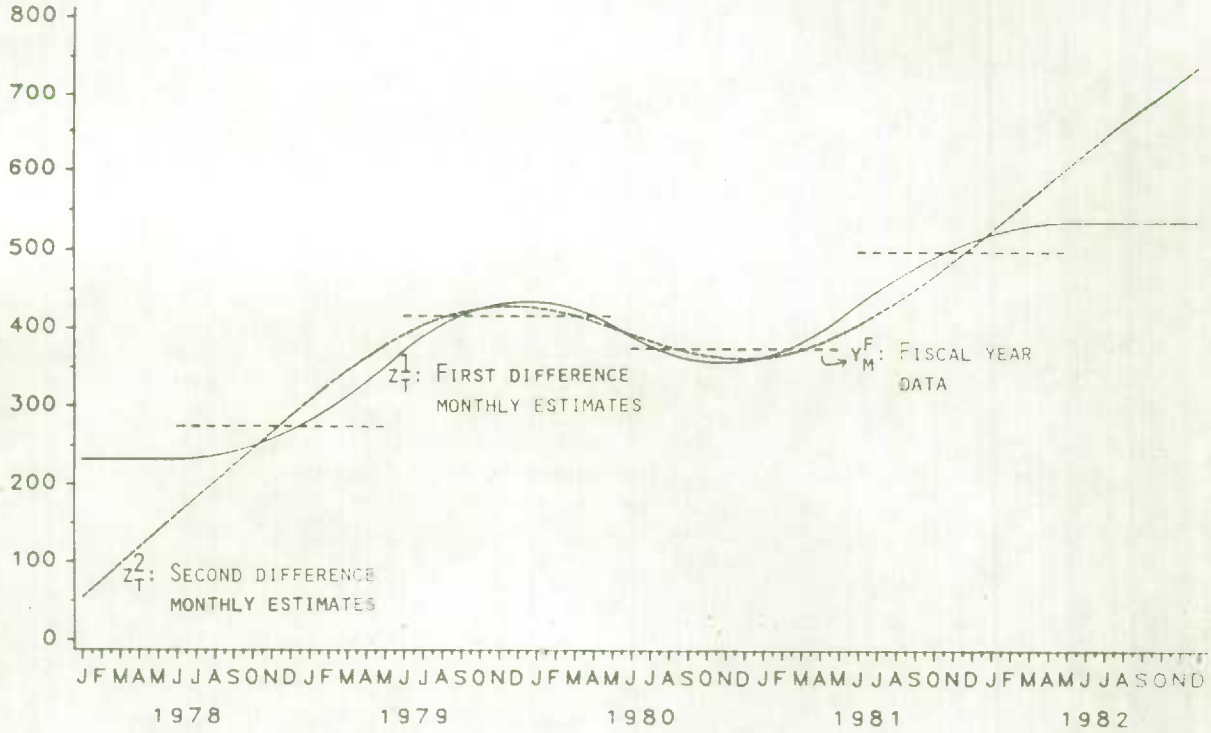

Figure 2: Interpolated sub-annual values obtained from minimizing first differences and minimizing second differences

As illustrated in Figure 2, the estimated sub-annual values (under either variant) contain no seasonality, no trading-day and no irregular fluctuations. However the method does work for (sub-annual) series containing those fluctuations. Indeed by definition, constant seasonality cancels out over any consecutive 12 months (or 4 quarters); and moving seasonality, largely does. Both the fiscal year and the calendar year values are then exempt from seasonality. It is not necessary - or at least not crucial - to estimate seasonality since it would largely cancel out in the calendar year reaggregation (2.6). Or put differently, the two variants proposed to this point allow for any cancelling seasonal pattern in the underlying sub-annual series. The same can be argued about the other components of the series: the trading-day component cancels out on any consecutive 3 months and therefore on any 12 months; the irregular values tend to cancel each other in any sum. To the extent the seasonal, the trading-day and the irregular components cancel out yearly, it is useless to estimate them - for calendarization purposes at least.

The estimated sub-annual values - of Figure 2 for instance - consequently correspond to the trend-cycle component of the series. It should be emphasized that the trend-cycle values obtained are approximate and should not be used for business cycle and current economic analysis: The exact dates and the sharpness (more precisely the lack thereof) of the turning points are not significant; and no turning-point can occur in the last year or be predicted.

3 CALENDARIZATION OF ANNUAL FLOW SERIES WITH A SUB-ANNUAL INDICATOR

When one is not prepared to assume that the seasonal, the trading-day and the irregular components cancel out on 12 consecutive months, the methods proposed in Section 2 are not sufficient. The same applies if one needs seasonal (say) sub-annual values per se or precise trend-cycle values. A sub-annual indicator is then required.

The sub-annual indicator x_t may be a mere seasonal pattern borrowed from a related socio-economic variable observed sub-annually, or chosen by the subject-matter expert. The indicator may be seasonal-trading-day pattern; or ideally approximate sub-annual values for the variable of interest. For simplicity x_t is now a yearly seasonal pattern. A value equal to 1.5 means that the month is 50% higher than an average month; equal to 0.6, 40% lower. A value of 1.0 specifies the month to be average; and a value of 0.0001, practically nil. Only the relative values of x_t matter, and zeroes must be avoided.

An appropriate objective function to minimize is:

$$f(z) = \sum_{t=2}^{IJ} (z_t/x_t - z_{t-1}/x_{t-1})^2 \qquad (3.1)$$

This quadratic of first proportional differences specifies that the desired sub-annual values z_t are as proportional to the seasonal pattern x_t as possible (on the series interval considered). In other words the proportion z_t/x_t should change as gradually as possible from one period to the next. The extent to which this is achievable is determined by the constraints (2.4).

The objective function could also be based on second proportional differences. As explained in Section 2.3 however, first differences should be preferred for the majority of socio-economic time series. This is especially true if x_t is an indicator of all the components of a time series.

The solution developed in the Appendix remains applicable. The estimated sub-annual values z_t and the calendarized yearly values $y^c{}_i$ are weighted averages of the available fiscal year values as in equation (2.5) and (2.7). The weights however are now dependent on the values of the quarterly indicator chosen. It is therefore necessary to recalculate them for each set of value of the indicator, that is presumably for each series considered and probably for each series interval considered.

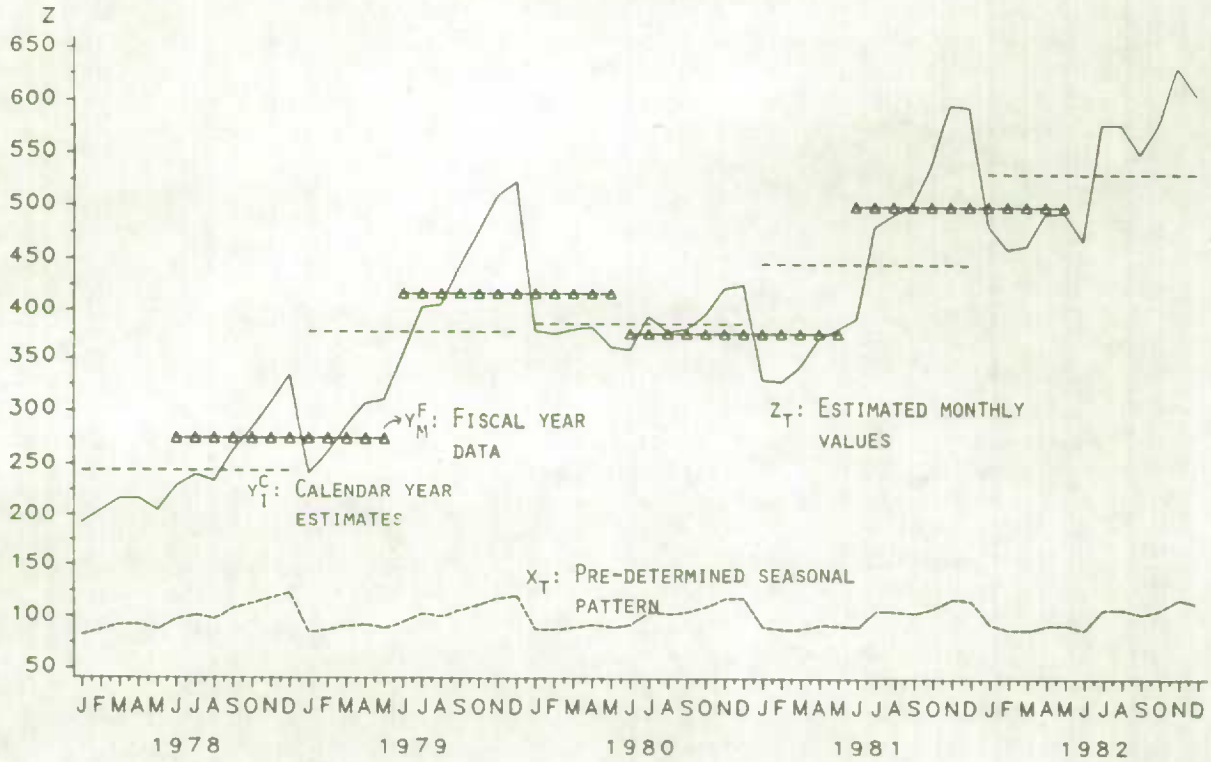

Figure 3: Seasonal sub-annual values estimated from fiscal year data and from a preselected moving seasonal pattern, and corresponding derived calendar year values.

Figure 3 displays seasonal monthly values z_t obtained from the same fiscal year values as in Figure 2 and from an evolving seasonal pattern x_t. For each year the surface under the fiscal year data (average) is equal to that under the interpolated sub-annual values. This illustrates that constraints (2.4) are satisfied. The estimated values z_t now correspond to the aggregate of the trend-cycle and of the seasonal components of the series. The former is supplied by the fiscal year data; and the latter, by the seasonal sub-annual indicator x_t. The calendarized values are substantially different from the original fiscal values. In many cases like the one illustrated, the end of the fiscal year does not coincide with the end of a calendar quarter. It is then necessary to disaggregate the fiscal year data into monthly (as opposed to quarterly) values, in order to operate the calendarization.

Whatever the level of disaggregation required, calendarization provides an opportunity to generate monthly or quarterly values, or both: Monthly values are obtained first; and then, the quarterly values, by taking the appropriate sums of the monthly values. The calendarization operation is not transitive: Generating quarterly values from fiscal year data, and then generating monthly values from the resulting quarterly values, does not yield the same monthly values as generating monthly values from the fiscal year data. This strongly suggests that such generation of data at

the various frequencies be integrated into one operation performed by the same series expert, using one single sub-annual indicator. This insures the consistency of the data at the various frequencies.

Monthly flow series generally contain trading-day variations. In real applications - especially when interest focuses on the sub-annual interpolations -, the sub-annual indicator should (at least) be the product of a seasonal pattern s_t and of a trading-day pattern, for instance:

$$x_t = s_t * \sum_{k=1}^{7} n_{tk} D_k / n_{t.}, \quad t=1,\ldots,IJ \qquad (3.2)$$

where $D_1, D_2,\ldots,D_7$ represent a constant weekly trading pattern. (One could also specify moving trading pattern.) The daily weights D_k are preselected in much the same manner as seasonal factors and have a similar interpretation. A weight $D_1=0.1$ means that Monday is only 10% as important as an average day; a weight $D_5=2.0$, that Friday is twice as important as an average day. A daily weight equal to zero, e.g. $D_7=0.0$, specify no activity for that day. Parameter n_{tk} is the number of each day in month t; and $n_{t.}$ is the total number of days in month t. In the absence of predetermined seasonality ($s_t=1$, $t=1,\ldots,IJ$), $n_{t.}$ must be set equal to 30.4375 (=365.25/12). The length of month effect, normally captured by the seasonal factors, is then accounted for by the trading-day component.

The approach to calendarization presented in this section is in fact an application of the proportional variant of the Denton (1971) benchmarking method generalized by Cholette (1987b). In a benchmarking context, the values of the sub-annual indicator x_t are in fact measurements of the socio-economic variable of interest. That variable is also measured yearly by y_m, in a more reliable and independent manner. Benchmarking consists of adjusting x_t, in such a way that it sums to y_m over the reference periods of y_m (whether calendar or not) and that the movements of x_t are preserved. When possible, the calendarization operation should in fact be carried out as a by-product of benchmarking. The calendarized fiscal year values are simply the appropriate sums of the benchmarked sub-annual series. This is indeed feasible in cases where whole socio-economic sectors (or sub-sectors) have a common fiscal year, e.g. school board, by using the method presented in this and the next two sections.

4. CALENDARIZATION OF ANNUAL STOCK AND INDEX SERIES

This section generalizes the calendarization method of Section 3 to stock and index series. The developments are still based on objective function (3.1):

$$f(z) = \sum_{t=2}^{IJ} (z_t/x_t - z_{t-1}/x_{t-1})^2 \qquad (4.1)$$

where x_t is an appropriate sub-annual indicator. Note that if x_t set equal to 1 (actually to any non-zero constant), (4.1) reduces to (2.8). In other words, the objective function required in the absence of sub-annual indicator is a special case of that required in the presence of indicator.

(The same holds for second differences.) Depending on the indicator selected, variable z_t represents the trend-cycle of the sub-annual series, the aggregate of the trend-cycle and of the seasonal components, etc.

4.1 Stock Series

For the purpose of this paper, stock series are those whose annual values correspond to one of the sub-annual values. For instance, the annual values of inventories correspond to the December values; the annual values of the population of Canada, to the June values. The fiscal year or the calendar year values of stock series therefore contain seasonal, trading-day and irregular fluctuations (unless the underlying sub-annual series contains none of those fluctuations). Indeed, if the month assigned to the calendar year or to the fiscal year is seasonally high, the corresponding yearly values is higher than if the month assigned is seasonally low; and vice versa. For stock series, it is then crucial to specify a sub-annual seasonal indicator x_t in objective function (4.1). The appropriate constraints for stocks series are:

$$z_{tm} = y^f_m, \quad m=1,\ldots,M \quad \text{(read subscript tm as } t_m\text{)} \qquad (4.2)$$

They specify that the desired sub-annual estimates z_t are equal to the fiscal year data for the time periods t_1, $t_2,\ldots,t_M$ referred to by the fiscal years (e.g. all months of March).

The sub-annual and the calendar year estimates may be expressed as weighted averages of the fiscal year data, as in equation (2.5) and (2.7). However, the solution exactly reduces to:

$$z_t = x_t * p_t \qquad (4.3)$$

where

$$p_t = d_m + (t-t_m)*[(d_{m+1}-d_m)/(t_{m+1}-t_m)], \quad t_m \le t \le t_{m+1}, \; m=1,\ldots,M-1 \qquad (4.4)$$

$$p_t = d_1, \; t \le t_1; \quad p_t = d_M, \; t \ge t_M \qquad (4.5)$$

The desired sub-annual estimates z_t are equal to the corresponding sub-annual indicator value x_t times a correction p_t. These corrections are linear interpolations between the proportional discrepancies d_m on each side of a time period considered. These discrepancies are the ratios of the fiscal year data to the applicable sub-annual value of the indicator, $d_m = y_m/x_{tm}$ (read subscript tm as t_m). For the periods preceeding the first and following the last discrepancies, the corrections repeat the values of first and last discrepancies. The calendarized values are simply equal to the applicable sub-annual values of z_t (e.g. all the months of December).

For the second difference variant the solution is the same, except for the first and last corrections. These become linear extrapolations of the central corrections:

$$p_t = p_{t1} + (t-t_1)*[p_{t1+1} - p_{t1}] \qquad t \le t_1$$

(read subscripts t1 and tM as t_1 and t_M) (4.5')

$$p_t = p_{tM} + (t-t_M)*[p_{tM} - p_{tM-1}] \qquad t \ge t_M$$

4.2 Index Series

For the purpose of this paper, index series are those whose annual values correspond to the average of the sub-annual values. Examples are the Index of Industrial production, the Consumer Price Index; but also Unemployment. Indeed by convention, the annual level of unemployement in Canada in 1986 is average of unemployement figures in each of the month of 1986.

The appropriate objective function remains (4.1). The constraints become:

$$\sum_{j=1}^{J} (z_{(i-1)J+j+K})/J = y^f_i, \quad i=1,..,I-1. \qquad (4.6)$$

or more conveniently

$$\sum_{j=1}^{J} (z_{(i-1)J+j+K}) = J*y^f_i, \quad i=1,..,I-1. \qquad (4.6')$$

where J is equal to 4 for quarterly and 12 for montly series. The latter equation transform index series into flow series by rescaling the fiscal year data. All the methods presented and comments made earlier for flow series therefore become exactly applicable.

5. FISCAL YEARS WITH IRREGULAR REFERENCE PERIODS

All the calendarization variants presented up to this point assumed that all the fiscal years of the series considered had the same reference periods and covered 12 months, for instance from April to March on every occasion. In practice, fiscal year periods may vary: The first fiscal year of a new company is likely to comprise more than 12 months; a company taken over by another one is also likely to change its fiscal year. In cases of irregular fiscal periods, the above methods cannot work for flow and index series, because the resulting fiscal year data contain seasonality: Seasonality cancels over 12 consecutive months (or 4 quarters), but not on 18 or 11 months.

That new situation requires that the fiscal year data be disaggregated into (at least) seasonal sub-annual values (a sub-annual indicator x_t is essential); and that the constraints be parametrized in a more general manner:

$$\sum_{t=\tau_m}^{\rho_m} z_t = y^f_m, \quad \rho_m \geq \tau_m, \quad m=1,\ldots,M; \qquad (5.1)$$

where τ_m and ρ_m are the reference periods of each fiscal year value y^f_m. Thus the first fiscal year may cover (say) from periods 1 to 18 (τ_1=1, ρ_1=18); the second, from periods 19 to 30 (τ_2=19, ρ_2=30); the third, from 31 to 42. These modified constraints also allow for missing fiscal year data for some of the years, i.e. M may be less than I-1. They are also valid for stock series by setting τ_m equal to ρ_m.

Objective function (4.1) is minimized subject to constraints (5.1). The resulting sub-annual values z_t are then aggregated into calendar year values.

Because of objective function (4.1) and of the generalized constraints (5.1), the calendarization variant presented in this section actually includes all the other variants proposed as particular cases. This unified variant is therefore the one most suitable for implementation purposes. It is the solution to this variant which is developed in the Appendix. (For stock series solution (4.3) to (4.5) was already valid for irregular fiscal periods.)

6. IMPLEMENTATION

This section discusses several implementational issues of calendarization, namely the level of aggregation at which calendarization should be performed, the (partial) extrapolation of calendar year values, the revision of estimates and computational issues.

In principle, calendarization should be carried out at the highest level of aggregation for which the respondents (e.g. companies) have the same sub-annual indicator (e.g. seasonal pattern) and same fiscal year periods. A neat example is that of the education sector, with fiscal year extending from September to August. All the annual data of that sector should be calendarized jointly. However if no seasonal pattern is specified (e.g. x_t=1 in (4.1)) the respondents do not need to have same seasonal pattern.

As mentioned, all variants of calendarization discussed produce estimates for the calendar years overlapped by the fiscal years. The terminal estimates for the first and last years are thus extrapolations, in the sense that some of the months in those years are not included in any available fiscal year. These less reliable terminal estimates do not have to be used. Consider the case however, where the last fiscal year ends in September 1987. The terminal estimate for 1987 involves only 3 sub-annual extrapolated values, those of October, November and December. If the series behaves in a fairly monotonic manner, the estimate is probably worth using - at least as preliminary. The alternative is to wait until the next fiscal year data becomes available (i.e. fall of 1988 or the winter of 1989), before calculating the 1987 calendar year as non-terminal.

This example raises the issue of revisions. The 1987 terminal calendar year estimate should be recalculated and revised on the availability of the October 1987 - September 1988 fiscal year data. This would improve its reliability. The 1986 estimate should also be revised, because the reliability of the estimates improves as they be come more central in the series interval. That revision is less imperative however: as years are added to the series, the revisions become negligible and may be ignored.

The Appendix shows the matrix operations required to carry out calendarization according to the method proposed in Section 5. Cohen et al. (1971) have shown that the matrix to be inverted becomes computationally ill-conditioned if too many years are adjusted together,

reducing the reliability of the estimates. Also, considering (i.e. minimizing (4.1) on) all the years of a series - say 15 years - produces practically the same results for years 3 to 13 as considering 5 years at the time: years 1 to 5, then 2 to 6, 3 to 7, etc. That 5-year moving average-type implementation saves on the amount of calculations and produces final results, i.e. not subject to revision, after 2 years. These points argue in favour of implementing calendarization over moving series intervals, for instance on 5-year moving intervals. A case could be argued for a 7-year moving interval.

As explained in the Appendix, the scale of the calculations required by calendarization may be reduced to the point that the operation can be carried out on micro-computers, at the individual respondent level. Calendarization could thus be integrated to Computer Assisted Telephone Interview systems.

Some of the methods presented in this paper are currently being tested on real business data by Laniel and Poirier (1988).

7. DISCUSSION AND CONCLUSION

Methods for the transformation of fiscal year data into calendar year estimates were presented. Calendarization is viewed as interpolation or benchmarking problems. The generalized variant of Section 5 can handle all the calendarization situations encountered. All previous variants are particular cases of the latter. A simple method often used in statiscal agencies is also a particular case. That method works when the fiscal year data have regular reference periods, when the seasonal, trading-day components of the series can be assumed to cancel yearly and when the series behaves in a fairly linear or exponential manner. When those conditions are not met, it is desirable - and in some cases imperative - to use other methods, like the general variant proposed.

Appropriate calendarization - or the lack thereof - obviously impacts on the quality of time series produced by statistical agencies. Calendarization conditions all the other statistical processes applied thereafter: benchmarking, seasonal adjustment, integration into accounting frameworks (e.g. the National Accounts), econometric modelling, forecasting, etc. Despite that, we failed to encounter any specific reference on the subject.

The approach to calendarization presented can easily be adapted for the calendarization of fiscal quarter data and for the transformation of bundles of weekly data (covering 4 or 5 weeks for instance) into monthly values. This will be the subject of our next papers. An approach has yet to be designed to handle fiscal years or quarters which do not end at the end of a month.

Developments are needed with respect to the assesment and the statistical properties of calendarized values. As Hillmer and Trabelsi (1987) point out in the context of benchmarking, one difficulty with the "numerical" approach adopted herein is that it does not provide reliability statistics as a by-product, as a "statistical" model-based

approach would. It is questionnable however whether the benchmarking approach advocated by the authors would be appropriate. The method relies on ARIMA modelling. These are known to smooth the irregular and trading-day components of a series. ARIMA modelling also requires much time series expertise and a sizable signal-to-noise ratio in the series, which prevents large scale application.

Numerical methods, on the other hand, require relatively less time series expertise (to fit models) and are massively applicable - even at the low levels of aggregation contemplated for calendarization. At those levels, the low signal-to-noise ratio and the ignorance of the stochastic properties often preclude the use of ARIMA methods. Numerical methods are then one way to incorporate subject matter expertise, that is the intimate knowledge by series builders of the socio-economic processes and variables involved. That expertise is sometimes the only information available apart from the fiscal year values. The numerical methods proposed in this document are all designed to rationally incorporate subject matter knowledge, namely in the form of seasonal and weekly trading patterns in the sub-annual indicator x_t. Criteria for selecting indicator series - maybe along the lines of Friedman (1962) - should be developed.

APPENDIX: Solution of the Generalized Calendarization Problem

In matrix algebra, objective function (4.1) is written:

$$F(Z) = Z'X^{-1}{}'D'D\,X^{-1}\,Z \qquad \text{(A.1)}$$

Vector Z contains the T=IJ desired sub-annual values z_t. Matrix D is the first difference operator D_1 or the second difference operator D_2:

$$\underset{\substack{(T-1)\\ \text{by } T}}{D_1} = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ . & . & . & . & & . & . \\ . & . & . & . & & . & . \\ . & . & . & . & & . & . \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}, \quad \underset{\substack{(T-2)\\ \text{by } T}}{D_2} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 \\ . & . & . & . & & . & . \\ . & . & . & . & & . & . \\ . & . & . & . & & . & . \\ 0 & 0 & 0 & 0 & \dots & -2 & 1 \end{bmatrix} \qquad \text{(A.2)}$$

Vector X^{-1} is a diagonal matrix. Its diagonal elements are the inverse of the sub-annual indicator values:

$$\underset{T \text{ by } T}{X^{-1}} = \begin{bmatrix} 1/x_1 & 0 & 0 & \dots & 0 \\ 0 & 1/x_2 & 0 & \dots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & . & & 1/x_T \end{bmatrix} \qquad \text{(A.3)}$$

(For computational reasons, the diagonal may have to be standardized, by multiplying it by the average of the x_t's for instance. Only the relative values of this matrix matter.) In the absence of sub-annual indicator X^{-1} is the identity matrix.

Constraints (5.1) write:

$$B\,Z = Y \tag{A.4}$$

Vector Y contains the fiscal year values:

$$Y' = [\, y^f_1 \;\; y^f_2 \;\dots\; y^f_M \,] \tag{A.5}$$

For index series, the y^f_m stand for the fiscal values multiplied by the number of months in the year (4 or 12). Matrix B is an annual sum operator. Its values are arranged in such a way that when B is multiplied by vector Z, the fiscal year sums of Z are obtained:

$$\underset{M \text{ by } T}{B} = \begin{array}{ccccccccccccccccc}
 & \text{columns:} & & & \tau_1 & & & \rho_1 & & & \tau_2 & & & \rho_2 & & & \\
[& 0 & \dots & 0 & 1 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots &] \\
[& 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 & 1 & \dots & 1 & 0 & \dots &] \\
[& . & & . & . & . & & . & . & & . & . & & . & . & &] \\
[& . & & . & . & . & & . & . & & . & . & & . & . & &] \\
[& . & & . & . & . & & . & . & & . & . & & . & . & &] \\
[& & & & & & & & & & & & & & & &]
\end{array} \tag{A.6}$$

Parameters τ_m and ρ_m stand for the reference periods of each fiscal year. For stock series, $\tau_m = \rho_m$.

Objective function (A.1) is minimized subject to the constraints (A.4). This is accomplished by the Langrangian augmented objective function:

$$F^\lambda(Z) = Z'X^{-1\prime}D'D\,X^{-1}\,Z - 2\Lambda'(B\,Z - Y) \tag{A.7}$$

where Λ is a M by 1 vector containing the Lagrange multipliers.

The values of Z which minimize this hyper-parabola are required. At the minimum, the derivative with respect to the unknowns Z and Λ are equal to zero. This leads to the normal equations:

$$\begin{aligned} dF/dZ &= 2\,X^{-1}\,D'D\,X^{-1}\,Z - 2\,B'\,\Lambda = 0 \\ dF/d\Lambda &= -2\,B\,Z + 2\,Y = 0 \end{aligned} \tag{A.8}$$

The solution to this system of equation is

$$\underset{(T+M)\text{ by }1}{\begin{bmatrix} Z \\ \\ \Lambda \end{bmatrix}} = \underset{(T+M)\text{ by }(T+M)}{\begin{bmatrix} X^{-1\prime}D'D\,X^{-1} & -B' \\ & \\ B & 0 \end{bmatrix}^{-1}} \underset{(T+M)\text{ by }1}{\begin{bmatrix} 0 \\ \\ Y \end{bmatrix}} = \begin{bmatrix} W_0 & W \\ & \\ W^\lambda_0 & W^\lambda_y \end{bmatrix} \begin{bmatrix} 0 \\ \\ Y \end{bmatrix} \tag{A.9}$$

For more details see Boot et al (1967). The solution implies $Z = W\,Y$. In other words the interpolated sub-annual values are weighted averages of the available fiscal year values

$$z_t = \sum_{m=1}^{M} w_{t,m}\, y_m \tag{A.10}$$

The matrix to be inverted in (A.9) may be large. For 5-year monthly intervals with 4 fiscal years, the dimensions of the matrix is 64 by 64. With the approach of Baldwin (1980), which adapts that of Cohen et al. (1971) for the proportional variant of benchmarking, the necessary inversion may be reduced to that of a (M+1) by (M+1) matrix, for first differences; and, to that of (M+2) by (M+2) matrix, for second differences. For first differences, an almost exact approximation is inspired from Chow and Lin (1971):

$$\underset{T \text{ by } 1}{Z} = V \; B'\underset{M \text{ by } M}{(B \, V \, B')^{-1}} \, (Y - B \, X) \qquad \text{(A.11)}$$

where B is given by (A.6), Y - B X contains the annual discrepancies and V is as follows

$$\underset{T \text{ by } T}{V} = \begin{bmatrix} 1 & x_1x_1 & \alpha\, x_1x_2 & \alpha^2\, x_1x_3 & \dots & \alpha^{T-1}x_1x_T \\ \alpha & x_2x_1 & 1\; x_2x_2 & \alpha^1\, x_2x_3 & \dots & \alpha^{T-2}x_2x_T \\ \alpha^2 & x_3x_1 & \alpha\, x_3x_2 & 1 \;\; x_3x_3 & \dots & \alpha^{T-3}x_3x_T \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & . & . & . & & . \end{bmatrix} \qquad \text{(A.12)}$$

with α lower but as close to 1.0 as possible (e.g. 0.9999999). Matrix V is such that V^{-1} is an approximation to $X^{-1}{}'D_1{}'D_1\, X^{-1}$ of (A.2) and (A.3). The inversion required in (A.11) is now that of a M by M matrix instead of a T+M by T+M in (A.9) (M being the number of fiscal years considered). The elements of V B' can easily be expressed algebraically.

Calendarization may also be performed without resorting to any matrix algebra. Section 4.1 provided the exact algebraic formulae for stock series (which is also valid for irregular financial years). A variant of that algebraic approach was developed (Cholette, 1987b) for flow and index series. The method basically consists of iteratively fitting corrections to the discrepancies observed between fiscal year data and the corresponding annual values of the sub-annual indicator and of smoothing the corrections by means of simple moving averages. That approximated solution yields almost the same results as (A.9), except for the terminal years. In those years, the approximation may actually be preferable. The approximation works for all the situations and variants examined in this paper.

REFERENCES

- Alba, E. de (1988), "Temporal Disaggregation of Time Series: a Bayesian Analysis", **Journal of Business and Economic Statistics**, Vol. 6, No. 2, pp 197-206.

- Baldwin, A. (1980), "Extensions of Quadratic Minimization Methods", Gross National Product Division, Statistics Canada, Research Paper.

- Boot, J.C.G., Feibes, W., Lisman, J.H.C. (1967), "Further Methods of Derivation of Quarterly Figures from Annual Data", **Applied Statistics**, Vol. 16, no. 1, pp. 65-75

- Cholette, P.A. (1979), "Adjustment Methods of Sub-Annual Series to Yearly Benchmarks", in J.F. Gentleman ed.: **Proceedings of the Computer Science and Statistics, 12th Annual Sumposium on the Interface**, University of Waterloo, pp. 358-366.

- Cholette, P.A. (1984), "Adjusting Sub-Annual Series to Yearly Benchmarks, **Survey Methodology**, Vol. 10, pp. 35-49.

- Cholette, P.A. (1987a), "Correcting the Reference Periods of Annual and Quarterly data", **Statistics Canada**, Time Series Research and Analysis Division, Working Paper No. TRSRA-87-010E.

- Cholette, P.A. (1987b), "Benchmarking and Interpolation of Time Series", **Statistics Canada**, Time Series Research and Analysis Division, Working Paper No. TRSRA-87-014E..

- Cholette, P.A., Baldwin A. (1988), "Converting Bundles of Weekly Data into Seasonal Monthly Values", **Statistics Canada**, Time Series Research and Analysis Division, Research Paper forthcoming.

- Cholette, P.A., Baldwin A. (1988), "Converting Fiscal Quarter Data into Calendar Quarter Values", **Statistics Canada**, Time Series Research and Analysis Division, Research Paper forthcoming.

- Cholette, P.A., Higginson, J. (1987), "A Note on Aggregating Data Referring to Different Financial Years", **Statistics Canada**, Time Series Research and Analaysis Division, Research Paper No. TRSA-87-015E.

- Chow, G,C., An-loh, L. (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series", **The Review of Economics and Statistics**, Vol. 53, pp. 372-375.

- Cohen, K.J., Padberg, M.W. (1971) "Autoregressive Approaches to the Disaggregation of Time Series Data", **Applied Statistics**, Vol. 20, pp 119-129.

- Denton, F.T. (1971), "Adjustment on Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization", **Journal of the American Statistical Association**, Vol. 66, No. 333, pp. 99-102.

- Ehrenberg, A.S.C. (1982), "Writing Technical Papers", **The American Statistician**, Vol. 36, pp. 326-329.

- Fernandez, R.B. (1981), "A Methodological Note on the Estimation of Time Series", **Review of Economic and Statistics**, Vol. 63, pp. 471-476.

- Hillmer, S. C., Trabelsi, A. (1987), "Benchmarking of Economic Time Series", **Journal of the American Statistical Association**, Vol. 82, pp. 1064-1071.

- Laniel, N., Poirier, C. (1988), "Adjustment of Annual Business Survey Data to the Calendar Year", **Statistics Canada**, Business Survey Method Division, Research Paper No. , Forthcoming.

- Leser, C.E.V. (1961), "A Simple Method of Trend Construction", **Journal of the Royal Statistical Society, Series B**, pp. 97-107

- Young, A.H. (1965), "Estimating Trading-Day Variation in Monthly Economic Time Series", **U.S. Bureau of the Census**, Technical Paper No. 12.

c. 2