Statistics Canada Statistique Canada

Methodology Branch

Direction de la méthodologie

Time Series Research and Analysis Division

Division de la recherche et de l'analyse des chroniques

Canada

[redacted]

Working Paper No. TSRA-89-007E
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]

May 1989

Paper submitted

for publication to

The Journal and Business and Economic Statistics

[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

Converting Fiscal Year Data
into Calendar Year Values

- abstract -

This document proposes a method to transform fiscal year data into "calendar" year estimates. The "fiscal year" data refer, for instance, to the months of April to March of the following year; and the calendar year estimates, refer to the month of January to December. The method, based on temporal disaggregation, may also be used to disaggregate 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 aggregates of weekly data into monthly values. The method is essentially an adaptation of the Denton (1971) benchmarking method and of the Cohen, Muller and Padberg (1971) interpolation methods. These methods are redeveloped in the framework of regression analysis.

KEYWORDS: Temporal disaggregation, Interpolation, Benchmarking, Calendarization, Quadratic Minimization, Regression, Generalized Least Squares

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

- résumé -

Ce document propose une méthode pour transformer des données d'années financières en estimations d'années civiles. Les chiffres annuels financiers couvrent, par exemple, les mois d'avril à mars de l'année suivante; et les estimations d'années civiles, les mois de janvier à décembre. 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 chiffres pluri-hebdomadaires en chiffres mensuels. La méthode constitue essentiellement une adaptation des méthodes d'étalonnage de Denton (1971) et d'interpolation de Cohen, Müller et Padberg (1971). Ces méthodes sont présentées à neuf dans le cadre l'analyse de régression.

1. INTRODUCTION

Most of the annual data published by statistical agencies reflect - or are perceived to reflect - the "calendar" year begining on January 1st and ending on December 31st. However, many companies and institutions supply the basic data in terms of their "fiscal years", which may 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 the early 1980's, 20% of Canadian retail trade sales were made by companies with fiscal years ending in January; 12%, by companies with fiscal years ending in March; and only 30%, by companies with fiscal years coinciding with the calendar year, i.e. ending in December. The same situation prevails for quarterly data, which often reflect the fiscal quarters of the respondents; and for monthly data, which are often fowarded to statistical agencies in aggregates of four or five weeks.

This situation creates the calendarization problem, which in the case considered in this paper consists of transforming fiscal year data into calendar year values. Two basic situations are distinguished. In the first and more favourable situation, there is sub-annual auxiliary information, i.e. an "indicator", for the variable of interest; but the fiscal sums of that indicator do not comply with the available fiscal year data. In this case, calendarization is viewed as a benchmarking problem. Classically, benchmarking consists of adjusting sub-annual measurements of a variable so that their annual sums conform with annual, more precise and separately obtained measurements of the same variable. In this paper, the benchmarking technique is applied to the case where the sub-annual values are an indicator - instead of measurements - of the target variable. The calendarized values then obtained by taking the calendar year sums of the "benchmarked" series. The benchmarking method by Denton (1971) is adapted in Section 3 for calendarization purposes. The benchmarked series obtained in such a context will be referred to as the interpolated series. We use the word interpolation both for flow and stock series.

In the second and less favourable situation, data are available only for the fiscal years, and there are no sub-annual data nor any auxiliary information on the variable of interest. The approach proposed is then to interpolate the sub-annual values between the fiscal year data and to aggregate the interpolations into calendar year values. That problem is closely related to that of interpolating monthly or quarterly series from annual series addressed by Boot, Feibes and Lisman (1967) and by Cohen, Müller and Padberg (1971). These approaches can be seen as special cases of the Denton method in which the sub-annual indicator is equal to zero.

Section 2 first present a unified approach to the calendarization problem, based on generalized least square regression. Chow and Lin (1971) also proposed an interpolation method based on regression; and Fernandez (1981), Litterman (1983), Silver (1986) and De Alba (1988) proposed variants of the Chow and Lin method. The main differences between these similar uses of the regression model and that proposed herein are the following. (1) In this paper the regressors are the deterministic columns of a design matrix; whereas in the methods of the Chow-Lin type, the regressors are related stochastic time series. (2) In this paper, the annual data may explicitely reflect fiscal years. (3) Finally in this paper, no initial conditions for the autocorrelation behaviour of the disturbances are specified. Unimportant in some applications, these initial conditions are crucial for calendarization purposes.

Section present a proportional variant of the calendarization method. Section 5 relates the proposed methods to existing empirical calendarization practices. Finally, Section 6 compares the performance of the former and the latter.

This paper assumes that calendarization takes place at a level of aggregation where the respondents to the survey or census have a common fiscal year - and common sub-annual indicator if applicable. That level is generally be lower than that of publication by the statistical agency. However, in some socio-economic sectors where all the respondents have the same fiscal years (e.g. the Canadian banking sector), the level of calendarization may coincide with that of publication. Despite the importance of the issue, we will refrain from discussing it further in this paper.

2 GENERAL MODEL FOR CALENDARIZATION

The approach to calendarization proposed in this paper consists of disaggregating the fiscal year values into **sub-annual**, monthly or quarterly, values; and, of temporally re-aggregating the **interpolated** sub-annual **values** into calendar year values. The interpolation of the sub-annual values is achieved by means of the following system of equations:

$$(2.1) \qquad \underline{s} = \underline{b} + \underline{e}, \qquad E(\underline{e})=\underline{0}, \quad E(\underline{e}\,\underline{e}')=\underline{V}_e$$

$$(2.2) \qquad \underline{f} = \underline{J}\,\underline{b} + \underline{\epsilon}, \qquad E(\underline{\epsilon})=\underline{0}, \quad E(\underline{\epsilon}\,\underline{\epsilon}')=\underline{V}_\epsilon$$

where $E(\underline{e}\,\underline{\epsilon}') = \underline{0}$.

The T-dimensional vector $\underline{b}$ denotes the "true" sub-annual values of the socio-economic variable to be interpolated by the system. Vector $\underline{s}$ denotes T sub-annual values of the indicator of the socio-economic variable to be calendarized. Vector $\underline{f}$ denotes the M fiscal year bencmark observations of the variable. Random vectors $\underline{e}$ and $\underline{\epsilon}$ are mutually independent. Their respective covariance matrices $\underline{V}_e$ and $\underline{V}_\epsilon$ are known and depend on the particular variants of the method presented in Sections 3 and 4. The model defined by (2.1) and (2.2) basically states that $\underline{b}$ is to be estimated on the basis of the fiscal year values $\underline{f}$ and of some sub-annual information $\underline{s}$.

More details are now given about some of the vectors and matrices. Vector $\underline{s}$ of (2.1) ideally denotes sub-annual measurements of the variable to be calendarized. In most calendarization situations however, sub-annual values of the target variable are not available. Vector $\underline{s}$ may then be an indicator of the variable. The indicator may consist of values of a variable observed sub-annually which is closely related to the target variable. The indicator may also be a seasonal and/or trading-day pattern (Young, 1965), borrowed from such a variable. Vector $\underline{s}$ may have an order of magnitude different from that of the fiscal benchmark data $\underline{f}$. Friedman (1962) provides a correlation criterion for selecting $\underline{s}$. When there is no sub-annual indicator, $\underline{s}$ will be set equal to zero.

Matrix $\underline{J}$ of (2.2) is the following **generalized annual sum operator**:

$$(2.3) \quad \underset{M \text{ by } T}{\underline{J}} = \begin{array}{c} \text{columns: } \qquad \tau_1 \qquad\quad \kappa_1 \qquad\quad \tau_2 \qquad\quad \kappa_2 \qquad\qquad \\ \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 1 & \cdots & 1 & 0 & 0 & \cdots \\ \vdots & & & \vdots & & & & \vdots & & & & \vdots & & & \vdots & & \\ \vdots & & & \vdots & & & & \vdots & & & & \vdots & & & \vdots & & \\ \vdots & & & \vdots & & & & \vdots & & & & \vdots & & & \vdots & & \end{bmatrix} \end{array},$$

where τ_m and κ_m are the **reference** periods, i.e. the starting and ending periods, of the M fiscal year values f_m (τ_m and κ_m indicate the columns and m the rows to place the ones). For instance, fiscal year values of a monthly flow series have reference periods $\tau_1=4$, $\kappa_1=15$, $\tau_2=16$, $\kappa_2=27$, etc., if each fiscal year ends in March and if the series $\underline{b}$ is a monthly flow series. For the corresponding stock series (e.g. inventories), the reference periods are $\tau_1=\kappa_1=15$, $\tau_2=\kappa_2=27$, etc. Operator $\underline{J}$, which has never been explicitly proposed before, allows for fiscal years ending in any month; and also for fiscal years of different length, i.e. occasionally covering more or less than 12 months. Consequently, equation (2.2) states that the fiscal sums of the desired $\underline{b}$ are equal to the appropriate fiscal values $\underline{f}$ plus an error term $\underline{\epsilon}$. For calendarization purposes, however, $\underline{\epsilon}=\underline{0}$ and $\underline{V}_\epsilon=\underline{0}$, in which case the benchmarks f_m are **binding**, i.e. must be satisfied. For reason which will soon become apparent, $\underline{\epsilon}$ and $\underline{V}_\epsilon$ are kept in the model.

Throughout this paper and with no loss of generality, t=1 refers to a January or a first quarter; and t=KH=T, to a December or fourth quarter. Symbol H stands for the number of sub-annual periods per year, 12 or 4; and K, for the number of calendar years in the interpolated series. In other words, $\underline{b}$ and $\underline{s}$ ranges over K>M complete calendar years, and the M fiscal years are embedded within the latter. Some of the interpolations will therefore be **extrapolations**, in the sense that they are not embedded in the fiscal years.

The system of equations defined by (2.1) and (2.2) can be written as

$$(2.4) \qquad \underline{g} = \underline{X}\,\underline{b} + \underline{d}, \quad E(\underline{d}) = \underline{0},$$

$$E(\underline{d}\,\underline{d}') = \underline{V} = \begin{bmatrix} \underline{V}_e & \underline{0} \\ \underline{0} & \underline{V}_\epsilon \end{bmatrix} = \underline{C}^{-1} = \begin{bmatrix} \underline{C}_e^{-1} & \underline{0} \\ \underline{0} & \underline{C}_\epsilon^{-1} \end{bmatrix},$$

where $\underline{g}'=[\underline{s}'\ \underline{f}']$, $\underline{X}'=[\underline{I}'\ \underline{J}']$ and $\underline{d}'=[\underline{e}'\ \underline{\epsilon}']$. Equation (2.4) is a case of generalized least squares (GLS). The GLS estimator of $\underline{b}$ and its covariance matrix are then respectively $\underline{b}^* = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\,\underline{X}'\underline{V}^{-1}\,\underline{g}$ and var $\underline{b}^* = \underline{\Omega} = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}$. On noticing the content of $\underline{X}$, $\underline{g}$ and $\underline{V}^{-1}=\underline{C}$, the estimator and its covariance may respectively be written as

$$(2.5) \qquad \underline{b}^* = \underline{\Omega}\,[\underline{V}_e^{-1}\,\underline{s} + \underline{J}'\underline{V}_\epsilon^{-1}\,\underline{f}] = \underline{\Omega}\,[\underline{C}_e\,\underline{s} + \underline{J}'\underline{C}_\epsilon\,\underline{f}]$$

$$(2.6) \qquad \text{var } \underline{b}^* = \underline{\Omega} = [\underline{V}_e^{-1} + \underline{J}'\underline{V}_\epsilon^{-1}\underline{J}]^{-1} = [\underline{C}_e + \underline{J}'\underline{C}_\epsilon\,\underline{J}]^{-1}$$

where $\underline{b}^*$ is the interpolated sub-annual series and where $\underline{C}_e=\underline{V}_e^{-1}$ and $\underline{C}_\epsilon=\underline{V}_\epsilon^{-1}$ are what we call the **criteria matrices**, corresponding to the covariance matrices $\underline{V}_e$ and $\underline{V}_\epsilon$. As exemplified in Sections 3 and 4, the criteria matrices may be calculated directly, that is without inverting the covariance matrices.

Using matrix algebra identities (e.g. Hillmer and Trabelsi, 1987, p. 1067) and tedious substitutions, solution (2.5) may be written:

$$\underline{b}^* = \underline{s} + \underline{C}_e^{-1}\,\underline{J}'\,[\,\underline{J}\,\underline{C}_e^{-1}\,\underline{J}' + \underline{V}_\epsilon]^{-1}\ [\,\underline{f} - \underline{J}\,\underline{s}\,],$$

$$(2.7) \qquad = \underline{s} + \underline{V}_e\,\underline{J}'\,[\,\underline{J}\,\underline{V}_e\,\underline{J}' + \underline{V}_\epsilon]^{-1}\ [\,\underline{f} - \underline{J}\,\underline{s}\,],$$

$$= \underline{s} + \underline{W}\,\underline{r}.$$

This alternative solution is feasible only if $\underline{C}_e$ is non-singular, or if its inverse $\underline{V}_e$ is obtainable directly. In the latter case, solution (2.7) enormously reduces the calculations compared with (2.6). Contrary to (2.5), (2.7) also allows $\underline{V}_\epsilon=\underline{0}$. When this is the case however, the covariance matrix (2.6) must be calculated as a limit as $\underline{V}_\epsilon$ tends to zero, hence the reason for keeping $\underline{\epsilon}$ and $\underline{V}_\epsilon$ in the model.

Equation (2.7) establishes the interpolated values $\underline{b}^*$ as equal to the available sub-annual values $\underline{s}$ plus a linear combination $\underline{W} = \underline{V}_e\underline{J}'[\underline{J}\,\underline{V}_e\underline{J}' + \underline{V}_\epsilon]^{-1}$ of the fiscal year discrepancies, $\underline{r} = \underline{f} - \underline{J}\underline{s}$, between the fiscal values $\underline{f}$ and the corresponding fiscal sums of the sub-annual information $\underline{s}$. The desired calendar year values $\underline{y}^*$ are then the proper annual sums of $\underline{b}^*$:

$$(2.8) \quad \underline{y}^* = \underline{G}\,\underline{b}^* = \underline{G}\,[\,\underline{s} + \underline{W}\,\underline{r}\,] = \underline{G}\,\underline{s} + \underline{P}\,\underline{r},$$

where $\underline{G}$ is a K by T calendar year sum operator, which is a particular case of (2.3), e.g. with $\tau_k = 1, 13, 25, \ldots$ and $\kappa_k = 12, 24, 36$. From (2.6) the covariance matrix of the calendarized values $\underline{y}^*$ is

$$(2.9) \qquad \text{var}\, \underline{y}^* = \underline{G}\, \text{var}\, \underline{b}^*\, \underline{G}'.$$

Note that the method could be used as an opportunity to obtain calendar quarter as well as calendar year values, by designing an appropriate sum operator.

Depending on the contents of the vectors and matrices of this section, different variants of the proposed calendarization method obtain: namely the additive variant presented next and the proportional variant presented in Section 4.

3. THE ADDITIVE VARIANT OF THE CALENDARIZATION MODEL

The additive variant of the proposed calendarization method is a particular case of the regression model presented in Section 2. Section 3.1 presents the model; Section 3.2, discusses some of the issues arising from the model; and Section 3.3 illustrates the additive variant.

3.1 The Model

The first equation of system (2.1)-(2.2) states that the sub-annual indicator values $\underline{s}$ are equal to the "true" sub-annual values $\underline{b}$ plus an error $\underline{e}$. In the additive variant, the error $\underline{e}$ follows an autoregressive process of the form,

$$(3.1) \qquad \underline{D}\,\underline{e} = \underline{u}, \quad E(\underline{u}) = \underline{0},\ E(\underline{u}\,\underline{u}') = \sigma_u^2\,\underline{I},$$

where $\underline{u}$ is an independently distributed random variable with mean zero and known variance σ_u^2. As in Litterman (1983, eqs. (16) to (18)), given (3.1), $\underline{V}_e$ is equal to $\sigma_u^2\,[\underline{D}'\underline{D}]^{-1}$. The criterion matrix $\underline{C}_e = \underline{V}_e^{-1}$ may be calculated directly from $\underline{D}$ as $\sigma_u^{-2}\,\underline{D}'\underline{D}$. Matrix $\underline{D}$ is the first difference matrix operator

$$(3.2) \qquad \underset{(T-1)\text{ by }T}{\underline{D}} = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ . & . & . & . & & . & . & . \\ . & . & . & . & & . & . & . \\ . & . & . & . & & . & . & . \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}.$$

so that equation (3.1) may specifically be written as:

$$(3.1') \qquad \begin{aligned} e_t &= e_{t-1} + u_t, \quad t=2,\dots,T, \\ E(u_t) &= 0, \quad E(u_t\, u_{t-k}) = \begin{cases} \sigma_u^2 & \text{for } k = 0, \\ 0 & \text{for } k \neq 0. \end{cases} \end{aligned}$$

where σ_u is the standard deviation of the error in $\underline{s}$, reflecting the accuracy of the sub-annual indicator $\underline{s}$ of the socio-economic variable. This random walk model for $\underline{e}$ specifies the principle of **additive movement preservation** or **parallelism** widely used in benchmarking: The errors e_t change as little as possible from one period to the next, resulting into interpolations b^*_t as parallel as possible to s_t. Section 3.2 further discusses the selection of the difference operator.

The second equation of system (2.1)-(2.2) states that the annual sums of the desired $\underline{b}$ are equal to the appropriate fiscal benchmarks $\underline{f}$ plus an error term $\underline{\epsilon}$. The covariance matrix V_ϵ is $\sigma_\zeta^2\,\underline{I}$ where σ_ζ tends to zero. The criterion matrix is $\underline{C}_\epsilon = \underline{V}_\epsilon^{-1} = \sigma_\zeta^{-2}\,\underline{I}$. Equation (2.2) may then be written as

$$f_m = (\sum_{t=\tau_m}^{\kappa_m} b_t) + \epsilon_m\,, \quad m=1,\dots,M, \ (\tau_1 \geq 1\ \kappa_M \leq T),$$

where

$$E(\epsilon_t) = 0, \quad E(\epsilon_t\, \epsilon_{t-k}) = \begin{cases} \sigma_\zeta^2 & \text{for } k = 0, \\ 0 & \text{for } k \neq 0, \end{cases}$$

and where ϵ_m is randomly and independently distributed with variance σ_ζ^2.

Substituting $\underline{C}_e$, $\underline{V}_\epsilon$ and $\underline{r}=\underline{f}-\underline{J}\underline{s}$ into (2.7) yields the following interpolated values:

$$(3.3) \quad \begin{aligned} \underline{b}^* &= \underline{s} + [\sigma_u^{-2}\ \underline{D}'\underline{D}\ ]^{-1}\ \underline{J}'\ (\underline{J}\ [\sigma_u^{-2}\ \underline{D}'\underline{D}]^{-1}\underline{J}' + \sigma_\zeta^2\ \underline{I})^{-1}\ \underline{r}, \\ &= \underline{s} + \underline{W}\ \underline{r}. \end{aligned}$$

Consequently the corresponding calendar year estimates are

$$(3.4) \quad \underline{y}^* = \underline{G}\ \underline{b}^* = \underline{G}\ [\ \underline{s} + \underline{W}\ \ \underline{r}\ ] = \underline{G}\ \underline{s} + \underline{P}\ \underline{r}.$$

Substituting $\underline{C}_e$ and $\underline{C}_\epsilon$ into (2.5) and (2.9) yields the covariance of $\underline{b}^*$ and $\underline{y}^*$ respectively:

$$(3.5) \quad \text{var}\ \underline{b}^* = [\sigma_u^{-2}\ \ \underline{D}'\underline{D} + \underline{J}'\sigma_\zeta^{-2}\ \underline{I}\ \underline{J}]^{-1},$$

$$(3.6) \quad \text{var}\ \underline{y}^* = \underline{G}\ \text{var}\ \underline{b}^*\ \underline{G}'.$$

For calendarization purposes, $\underline{V}_\epsilon=\sigma_\zeta\underline{I}=\underline{0}$, in which case $\sigma_u 2$ cancels out in (3.3). The covariance matrices may then be calculated as the limit of (3.5) and (3.6) as σ_ζ tends to zero with $\sigma_u=1$. The resulting covariances are then interpreted as the covariance of the estimates per unit of variance assumed for the sub-annual indicator. Alternatively σ_u may be estimated from the residuals $\underline{\hat{e}}=\underline{s}-\underline{y}^*$ as $(\underline{\hat{e}}'\underline{C}_e\underline{\hat{e}}/T)^{1/2}$ and multiplied into (3.5).

3.2 Issues Related to the Additive Variant

In equation (3.2), we propose first differences. Second or third differences would reflect a more realistic behaviour of the discrepancies, and translate into much smaller residuals $\underline{\hat{e}}$ (better fit of the model) and into narrower confidence intervals. However, the interpolated values in the first and last calendar years become subject to heavy and even wild revisions. (Users may also try other operators in the general model of Section 2.)

It should also be stressed that the purpose of the first differences is to **specify** the criterion of movement preservation, that is of parallelism of the interpolations to the the sub-annual indicator $\underline{s}$. This principle is widely accepted and reasonable in the circumstances: $\underline{s}$ provides the sub-annual movement, e.g. the seasonal and trading-day components of the interpolated series; and the fiscal year data $\underline{f}$, the super-annual movement and the level, i.e. the trend-cycle component. The purpose of the first difference is not to reflect the behaviour of the "survey" errors, like in the Hillmer and Trabelsi (1987) benchmarking method.

Based on signal extraction, the Hillmer and Trabelsi (1987) benchmarking method specifies an ARIMA criterion to be maximized by the benchmarked series. The true sub-annual series, the signal $\underline{b}$, is assumed to follow an ARIMA model. In terms of the regression framework, the method could accurately be described by adding the following third equation to system (2.1)-(2.2): $\underline{\mu} = \underline{\Pi}\ \underline{b} + \underline{\eta}$, where the ARIMA parameters $\underline{\Pi}$ of the model are predetermined and μ is equal to zero for non-stationary series. In their method matrix $\underline{D}$ is an autoregressive operator to reflect the ARIMA model followed by the sampling errors $\underline{e}$.

Incorporating such ARIMA models in the calendarization method proposed in this paper would increase the number of degrees of freedom from T+M to a number closer to 2T+M. By making μ equal to $\underline{\Pi}$ times an intervention variable (Box and Tiao, 1975), the model could take additive trading-day (Bell and Hillmer, 1983) and other exogeneous variations not captured by a pure ARIMA model. One problem would be the level of time series expertise required for massive application in a statistical agency; another problem is that ARIMA modelling requires high signal-to-noise ratios. The method proposed in this paper on the other hand, preserves both the signal and the noise of the sub-annual series without need to partition them.

With D equal to the strict first difference operator (3.2) however, $[\sigma_u^{-2}\ \underline{D}'\underline{D}\ ]$ is singular, so that the expectation cannot be calculated with (3.3). Redefining $\underline{D}$ as the quasi first difference operator (e.g. Harvey, 1981, p. 190)

$$(3.7)\qquad \underset{T\text{ by }T}{\underline{D}} = \begin{bmatrix} (1-\rho^2)^{1/2} & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -\rho & 1 & 0 & \ldots & 0 & 0 & 0 \\ . & . & . & . & & . & . & . \\ . & . & . & . & & . & . & . \\ . & . & . & . & & . & . & . \\ 0 & 0 & 0 & 0 & \ldots & 0 & -\rho & 1 \end{bmatrix},$$

where ρ is lower than but very close to 1.0, (e.g. 0.999999), entails an invertible matrix. Furthermore, the inverse is now known algebraically:

$$(3.8)\quad [\sigma_u^{-2}\ \underline{D}'\underline{D}\ ]^{-1} = \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{T-3} \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \end{bmatrix} \sigma_u^2\ /\ (1-\rho^2).$$

With ρ very close to 1.0, the quasi first difference operator (3.7) causes an infinitesimal sacrifice to the movement preservation principle. Indeed it is equivalent to adding the term $(1-\rho^2)^{1/2}\ e_1/s_1$ to (3.1'), which is negligible. Furthermore, Bournay and Laroque (1979) show that as ρ tends to 1.0, the solution reached tends to that obtained with strict first differences. The denominator of (3.8) does tend to zero, but that zero cancels out in solution (3.3). One insisting on the strict first difference operator (3.2) could calculate the expectation by means of (2.5) (provided $\underline{V}_\epsilon > \underline{0}$). Alternatively, one could use numerical optimization techniques (instead of regression) adopted by Denton (1971), by Boot, Feibes and Lisman (1967) and by Cohen, Müller and Padberg (1971).

Incidently, for binding benchmarks, i.e. $\underline{V}_\epsilon = \underline{0}$, and with $\underline{J}$ as the calendar year sum operator, solution (3.3) coincides with Denton's additive variant, except for the first difference operator $\underline{D}$. Denton's $\underline{D}$ is given by (3.7) with ρ=1, but with the first element equal to 1. Cholette (1979 and 1984) observed that this operator entails important movement distortion if the absolute size of the first discrepancy r_1 in $\underline{r}=\underline{f}-\underline{J}\underline{s}$ is large, which is likely to be the case for calendarization. Under the same two conditions on $\underline{V}_\epsilon$ and $\underline{J}$, and in the absence of any sub-annual information, i.e. $\underline{s}=\underline{0}$, (3.3) coincides the method proposed by Boot, Feibes and Lisman (1967) and by Cohen, Müller and Padberg (1971), to interpolate quarterly values from yearly data and interpolated any more frequent values from less frequent data. The algorithm of Boot et al. and Cohen et al. is more elaborate than that proposed here, but does not rely on the quasi first difference operator (3.7). Users reluctant to use (3.7) are referred to those authors. Baldwin (1980) developed the proportional variant corresponding to the Cohen et al. algorithm.

3.3 Illustration of the Additive Variant

As observed by Boot et al and Cohen et al., sub-annual information about the target variable may be totally unavailable, i.e. $\underline{s}=\underline{0}$. The method proposed may still be very useful. As illustrated in Figure 1, the movement preservation principle sets the interpolations b^*_t as parallel to zero, i.e. as flat, as possible. These smooth interpolations are an approximation of the trend-cycle component of the series. That trend-cycle has annual sums equal to the fiscal data, that i.e. is consistent with the latter, and levels off in the first and last years. Consistency is illustrated by the fact that $f_m/4$ covers the same surfaces as b^*_t over each fiscal year.

Such interpolations are simplistic but sufficient, if

(1) each fiscal year comprises the same 12 consecutive months or 4 consecutive quarters and

(2) the seasonal and the trading-day components can be assumed to cancel on such consecutive periods (which is usually not a strong assumption).

Under these two conditions, the seasonal and trading-day components have cancelled out in the fiscal values and will also cancel out in the calendar year values. It is therefore useless to estimate them. If some fiscal years cover more or less than 12 months, the fiscal values contain seasonality and possibly trading-day variations. One then needs a non-trivial $\underline{s}$.

Whether the sub-annual indicator $\underline{s}$ is equal to zero or not, the weights $\underline{W}$ and $\underline{P}$ in the additive variant are independent of $\underline{s}$ and $\underline{f}$. They depend only on the number K of calendar years considered and on the reference periods of the M fiscal year (i.e. on matrix $\underline{J}$). The weights $\underline{W}$ and $\underline{P}$ of (3.4) can therefore be calculated once and for all and applied to any series whose fiscal years have same reference periods. Cholette (1988) tabulated $\underline{P}$ weights for fiscal years referring to any consecutive 12 month or 4 quarters with $\underline{V}_\epsilon=\underline{0}$. The weights are to be applied in a 2- (K=2), 3-, 4- and 5-year moving average manner. The calendarized estimates $\underline{y}^*$ are then simply $\underline{G}\ \underline{s} + \underline{P}\ (\underline{f}-\underline{J}\underline{s})$. If $\underline{s}=\underline{0}$, then $\underline{y}^* = \underline{P}\ \underline{f}$. Table 1 A and B give examples of weights $\underline{P}$ for fiscal years ranging from April to March and from July to June. For fiscal years ranging from October to September, the weights would be those in Table 1 A in the reverse order. The weights of Table 1B were used for Figure 1.

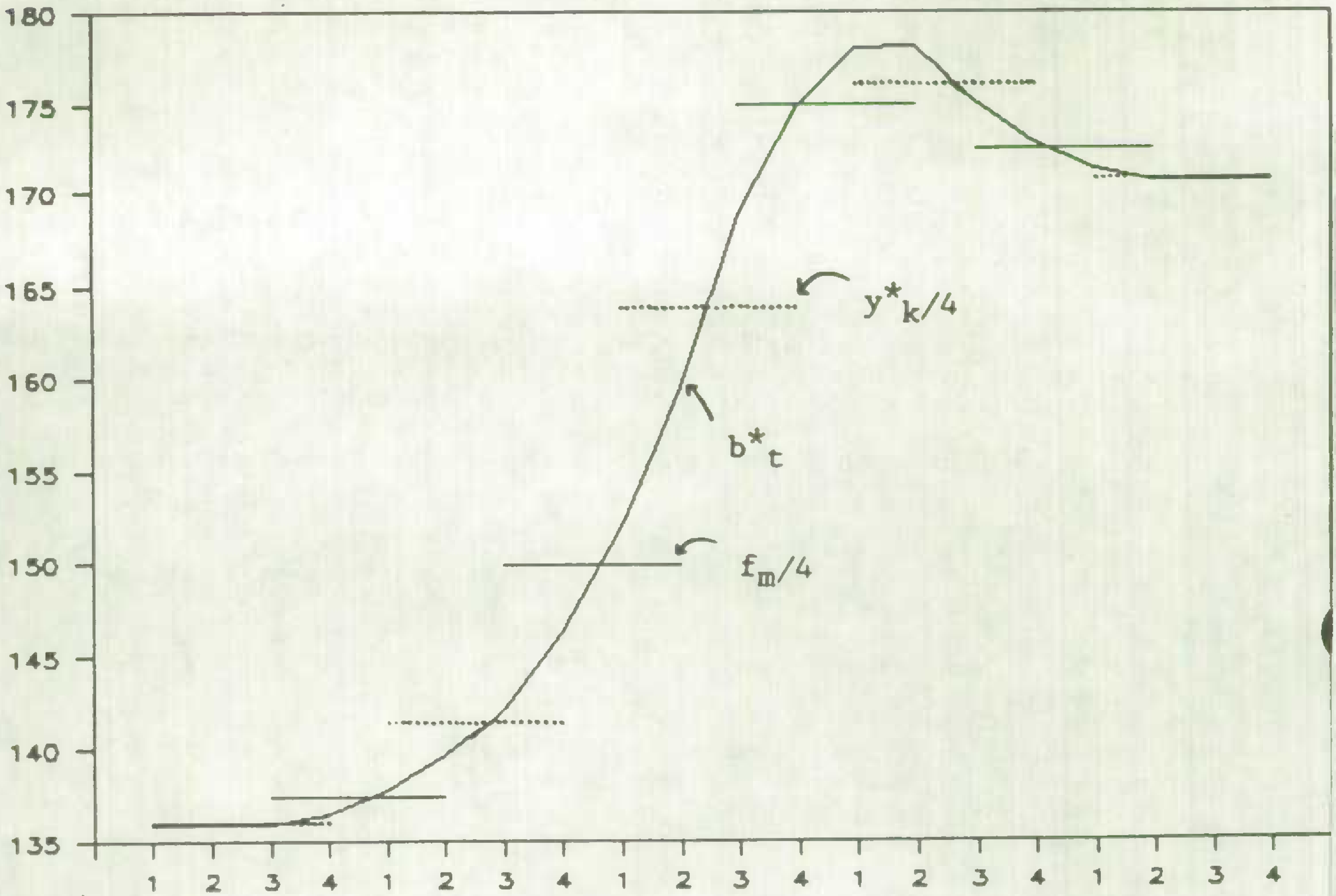

Figure 1: Values b^*_t interpolated between the fiscal data f_m and corresponding calendar year estimates y^*_k, obtained by the additive variant of the proposed calendarization method, in the total absence of sub-annual information (s_t=0)

Table 1: 5-year weights $\underline{P}$ obtained by the additive variant of the proposed calendarization method, when K=5, M=4, $\underline{V}_\epsilon=\underline{0}$, $\underline{D}$ is the quasi first difference operator (3.7) and each fiscal year covers 12 consecutive months

A: when the fiscal years cover from April to March

to estimate	weights applied to fiscal year 1	2	3	4	no. of implicit sub-annual interpolations
calendar year 1	1.154	-0.194	0.051	-0.010	3
" " 2	0.202	0.898	-0.124	0.025	0
" " 3	-0.050	0.267	0.873	-0.090	0
" " 4	0.015	-0.075	0.301	0.759	0
" " 5	-0.017	0.086	-0.330	1.261	9

B: when the fiscal years cover from July to June

to estimate	weights applied to fiscal year 1	2	3	4	no. of implicit sub-annual extrapolations
calendar year 1	1.232	-0.293	0.076	-0.015	6
" " 2	0.474	0.633	-0.133	0.026	0
" " 3	-0.093	0.593	0.593	-0.093	0
" " 4	0.026	-0.133	0.633	0.474	0
" " 5	-0.015	0.076	-0.293	1.232	6

4. THE PROPORTIONAL VARIANT OF THE CALENDARIZATION MODEL

This section presents the proportional variant of the proposed calendarization method and discusses some issues related to that variant.

4.1 The Model

Equation (2.1) of the system (2.1)-(2.2) states that the sub-annual indicator values $\underline{s}$ are equal to the "true" sub-annual series $\underline{b}$ plus an error $\underline{e}$. In the proportional variant, the error $\underline{e}$ follows an autoregressive and heteroscedastic process of the form,

$$(4.1)\quad \underline{D}\,[\mathrm{diag}(\underline{s})^{-1}]\,\underline{e} = \underline{u}, \qquad E(\underline{u}) = \underline{0},\ E(\underline{u}\,\underline{u}') = \sigma_u^2\,\underline{I},$$

where $\underline{u}$ is an independently distributed random variable with mean zero and known variance σ_u^2. Given (4.1), the criterion matrix $\underline{C}_e$ is σ_u^{-2} $[\mathrm{diag}(\underline{s})^{-1}\underline{D}'\underline{D}\ \mathrm{diag}(\underline{s})^{-1}]$. With $\underline{D}$ as in (3.2), equation (4.1) may specifically be written as:

$$(4.1')\quad \begin{aligned} &e_t/s_t = e_{t-1}/s_{t-1} + u_t, \quad t=2,\ldots,T, \\ &E(u_t) = 0,\ \ E(u_t\,u_{t-k}) = \begin{cases} \sigma_u^2 & \text{for } k=0, \\ 0 & \text{for } k \neq 0, \end{cases} \end{aligned}$$

where σ_u is the proportional standard deviation of the error in $\underline{s}$, reflecting the accuracy of the sub-annual indicator. Model (4.1') specifies the **proportional movement preservation principle** encountered in the widely used benchmarking methods of the Denton type. The ratio of each error e_t to the corresponding sub-annual observation s_t changes as little as possible from one period to the next, resulting in interpolated values b^*_t as proportional as possible to s_t. The proportional variant requires that all values of $\underline{s}$ be positive.

The matrix $\underline{V}_\epsilon$ chosen is $\sigma_\zeta^2\ \mathrm{diag}(\underline{f})$, $(\underline{C}_\epsilon = \underline{V}_\epsilon^{-1} = \sigma_\zeta^{-2}\ \mathrm{diag}(\underline{f})^{-1})$, so that equation (2.2) may be written as

$$f_m = (\sum_{t=\tau_m}^{\kappa_m} b_t) + \epsilon_m\,, \quad m=1,\ldots,M,\ (\tau_1 \geq 1\ \kappa_M \leq T),$$

where

$$E(\epsilon_t) = 0,\ \ E(\epsilon_t\,\epsilon_{t-k}) = \begin{cases} \sigma_\zeta^2\, f_m & \text{for } k = 0, \\ 0 & \text{for } k \neq 0, \end{cases}$$

where σ_ζ is the proportional standard error of $\underline{f}$. Again, for calendarization, $\underline{V}_\epsilon = \underline{0}$ implying $\sigma_\zeta = 0$.

Substituting $\underline{C}_e$, $\underline{V}_\epsilon$ and $\underline{r} = \underline{f} - \underline{J}\underline{s}$ into (2.7) yields the following interpolated values:

$$(4.4)\quad \begin{aligned} \underline{b}^* &= \underline{s} + [\sigma_u^{-2}\ \mathrm{diag}(\underline{s})^{-1}\ \underline{D}'\underline{D}\ \mathrm{diag}(\underline{s})^{-1}\ ]^{-1}\ \underline{J}' \\ &\quad \{\underline{J}\ [\sigma_u^{-2}\ \mathrm{diag}(\underline{s})^{-1}\ \underline{D}'\underline{D}\ \mathrm{diag}(\underline{s})^{-1}\ ]^{-1}\underline{J}' + \sigma_\zeta^2\ \mathrm{diag}(\underline{f})\ \}^{-1}\ \underline{r}, \\ &= \underline{s} + \underline{W}\ \underline{r}. \end{aligned}$$

In the proportional variant, it can be shown that (4.4) reduces to $\underline{b}^* = \underline{W}\ \underline{f}$. Consequently the calendar year estimates are $\underline{y}^* = \underline{G}\ \underline{b}^* = \underline{G}\ \underline{W}\ \underline{f} = \underline{P}\ \underline{f}$.

Note that $[\sigma_u^{-2}\ \mathrm{diag}(\underline{s})^{-1}\ \underline{D}'\underline{D}\ \mathrm{diag}(\underline{s})^{-1}]^{-1}$ in (4.4) is equal to $\mathrm{diag}(\underline{s})(\sigma_u^{-2}\ \underline{D}'\underline{D})^{-1}\ \mathrm{diag}(\underline{s})$, so that (3.8) can be used, if $\underline{D}$ is specified as the quasi-first difference operator (3.7).

4.2 Issues Related to the Proportional Variant

The discussion of Section 3.2 about the difference operator remains entirely applicable here. There is a more fundamental issue however. Numerically, solution (4.4) works. Indeed the wide-spread proportional variant of the Denton benchmarking method coincides with (4.4), if $\underline{V}_\epsilon = 0$ and if $\underline{J}$ is equal to the calendar year sum matrix operator (and if $\underline{D}$ is the same). The problem is that strictly speaking the inverse of the criterion matrix $\underline{C}_e$ cannot be interpreted as the covariance matrix of $\underline{e}$, because the weights $\underline{L} = [\underline{D}\ \mathrm{diag}(\underline{s})^{-1}]$ in (4.1) are not fixed. A similar situation prevails for $\underline{V}_\epsilon$. It is thus statistically illegitimate to view proportional benchmarking as a particular case of the regression model of Section.

The case at hand is one known as dependent variable heteroscedasticity. The variance of the observations s_t (ignoring f_m and assuming $\underline{D}=\underline{I}$ for the moment) is assumed to be proportional to $\sigma^2 s_t^2$. The GLS estimator (4.4) is then inefficient. Harvey (1981, Chapt. 3.4) describes an slightly different specification, where the variance of the dependent variable is proportional to the square of its expectation, $\sigma^2 E(b_t)^2$, and for which GLS is usable. The technique consists of two steps. First, apply ordinary least squares ignoring heteroscedasticity (i.e. apply the additive variant) and calculate the OLS estimates $\hat{b}_t$ of b_t. Second, apply GLS using $\sigma^2 \hat{b}_t^2$ as the variance of the observations, i.e. weighting the observations by $1/\hat{b}_t$. Under a Gamma distribution the resulting estimator is asymtotically efficient. For calendarization, that distribution would in fact be more appropriate than the normal, because it excludes negative observations. In order to avoide negative values in the first step, the technique further requires that the scale of $\underline{s}$ is first changed to that of $\underline{f}$, by means of a suitable factor, e.g. $[\min(f_m/(\kappa_m-\tau_m))]/[s_{.}/T]$.

That 2-step estimation yields interpolations $\underline{b}^*$ which are neither proportional nor parallel to $\underline{s}$ but mixed. For that reason and for simplicity, we recommend solution (4.4). This estimator $\underline{b}^*$ is a numerical estimator of the Denton (1971) type, with undetermined covariance matrix. It is nevertheless pedagogically useful to view proportional calendarization through the familiar regression framework.

4.3 Use of the Proportional Variant

The usefullness of the proportional variant of the proposed calendarization method are best described in opposition to that of the additive. The additive variant keeps the interpolated values as parallel as possible to the the original sub-annual values s_t. That parallelism implicitly requires that s_t - if different from zero - has the same order of magnitude as the fiscal year values. Consider the following simple quarterly example, where s_t is distributed around 100 with an amplitude of 20 and $f_m/4$ is a constant equal to 1 million. The resulting interpolations b^*_t are parallel to s_t and distributed around 1 million with an amplitude of 20, i.e. basically display no movement. With the proportional variant on the other hand, b^*_t are distributed around 1 million with an amplitude of 200,000. The proportional variant transform the scale of $\underline{s}$, whereas the additive does not. The proportional variant is especially be usefull when the fiscal years have different lengths. A non-trivial seasonal pattern is then required, and the scaling problems encountered with the additive variant are avoided. It is also a lot easier to think of seasonality in terms of percentages than in terms of units of the series.

5. EXISTING CALENDARIZATION METHODS

This section relates the proposed calendarization method to existing procedures, which were all developped on an empirical basis.

For stock series ($\tau_m=\kappa_m, m=1,\ldots,M$) with binding benchmarks ($\underline{V}_\epsilon=\underline{0}$), both variants of the proposed calendarization method reduce to simple algebraic formulae. The corrections c_t made to s_t to obtain b^*_t are linear interpolations between the sub-annual discrepancies d_m and constant extrapolations of the first and last discrepancies:

$$c_t = d_m + (t-\tau_m) * [(d_{m+1}- d_m)/(\tau_{m+1}- \tau_m)], \quad \tau_m \le t \le \tau_{m+1}, \; m=1,\ldots,M-1$$

(5.1)

$$c_t = d_1, \; t \le t_1, \quad c_t = d_M, \; t \ge t_M$$

where $d_m=f_m-s_m$ in the additive variant and $d_m=f_m/s_m$ in the proportional. The proportional variant of (5.1) generalizes for fiscal years one of the empirical formulae described by Friedman (1962, eq. (6)).

A particular case of (5.1) occurs when only one year is considered (M=1)

(5.2) $$c_t = d_m = f_m/s_m \quad \rightarrow \quad b^*_t = s_t\,(f_m/s_m) = (s_t/s_m)\, f_m$$

where s_m and s_t respectively refer to the March (say) and to the previous December sub-annual values and f_m is the fiscal year benchmark referring to March. Method (5.2) applies the (inverse) growth in the sub-annual values to the benchmark. This common technique generates steps between years, if f_m/s_m changes from year to year. Unfortunately in some operational circumstances, it is the only one feasible.

For flow series, many statisticians are acquainted with another empirical method:

(5.3) $$y^*_2 = (N/H)\, f_1 + ((H-N)/H)\, f_2),$$

where N is the number of periods (months or quarters) of calendar year 2 which are in fiscal year 1 and where H is equal to 4 or 12. For fiscal years running from April 1st to March 31st, the parameters of (5.3) are N=3 and H=12; the fractions are then 1/4 for f_1 and 3/4 for f_2. For fiscal years running from July to June (N=6), the fractions are 1/2 and 1/2.

Any such convex weights reduce the amplitude of business cycles, because the estimates always lie between two neighbouring fiscal year values. This is illustrated in Figure 2, in the July to June case. Each calendar year estimates lies exactly in the middle of the fiscal year values on each side. The figure also shows the flattest possible trend-cycle consistent with such estimates and with the fiscal data (in calendar and fiscal year sums respectively). That trend-cycle displays a complete cycle, i.e. two turning-points, during the two last fiscal years. This event is very unlikely: the fiscal year data provide evidence of one downwards turning point but no evidence of a second - upwards - turning-point. An alternative trend-cycle with only one downward turning point would have to bend backwards during the second fiscal year to satisfy consistency. The additive variant of Section 3, on the other hand, yields a much more likely trend-cycle with one downwards turning-point, which is the one displayed in Figure 1. Also note that the fourth calendar year estimate lies outside the neighbouring fiscal data. It can be shown that method (5.3) thus implies an unlikely trend-cycle whenever the fiscal data change direction or level off.

Method (5.3), is in fact a particular case of the general model of Section 2, under the following conditions:

(1) The interpolation is specified over two fiscal years, i.e. $M=2$ and $K=3$.

(2) The fiscal years all refer to 12 consecutive months.

(3) The sub-annual information is totally absent, i.e. $\underline{s}=\underline{0}$.

(4) Matrix $\underline{D}$ in $\underline{C}_e=\underline{V}_e^{-1}=\underline{D}'\underline{D}$ is the second difference operator.

(5) The variance $\underline{V}_\epsilon$ is zero.

This produces an additive variant like that of Section 3, with second instead of first differences. The resulting interpolations lie on a straight line running through the middle of the two fiscal values divided by 12 (or 4). Such a model works (a) if the underlying seasonality is rather stable and

(b) if the fiscal years comprise 12 months.

(c) and if the fiscal values - and therefore the underlying trend-cycle - behave linearly without levelling off.

In Section 6, method (5.3) will be referred to as the traditional method and used to assess the performance of the proposed additive and proportional methods.

One less known - but perhaps not less widespread - practice with respect to fiscal year data is the following. Depending on the month ending their fiscal year, the respondents to a survey can have 12 possible fiscal years; the data of respondents with any of the twelve fiscal years ending between April (say) of year 1 and March of year 2 are classified in the calendar year 1. For the variable considered, the annual estimate y^*_1 is simply the sum of all the fiscal data classified in calendar year 1. Note that this scheme produces an estimate y^*_1 supposedly referring to year 1, even if all respondents have a common fiscal year ending in March (say). More generally however, each calendar year value derived in that manner implicitly involves 23 months of the underlying unknown monthly data, i.e. May of year 0 to March of year 2. Cholette and Higginson (1987) conclude that for flow series such a scheme produces biased annual values and that the bias changes with the phase of the business cycle. Unfortunately, in some circumstances this practice is the only one feasible for the statistical agency.

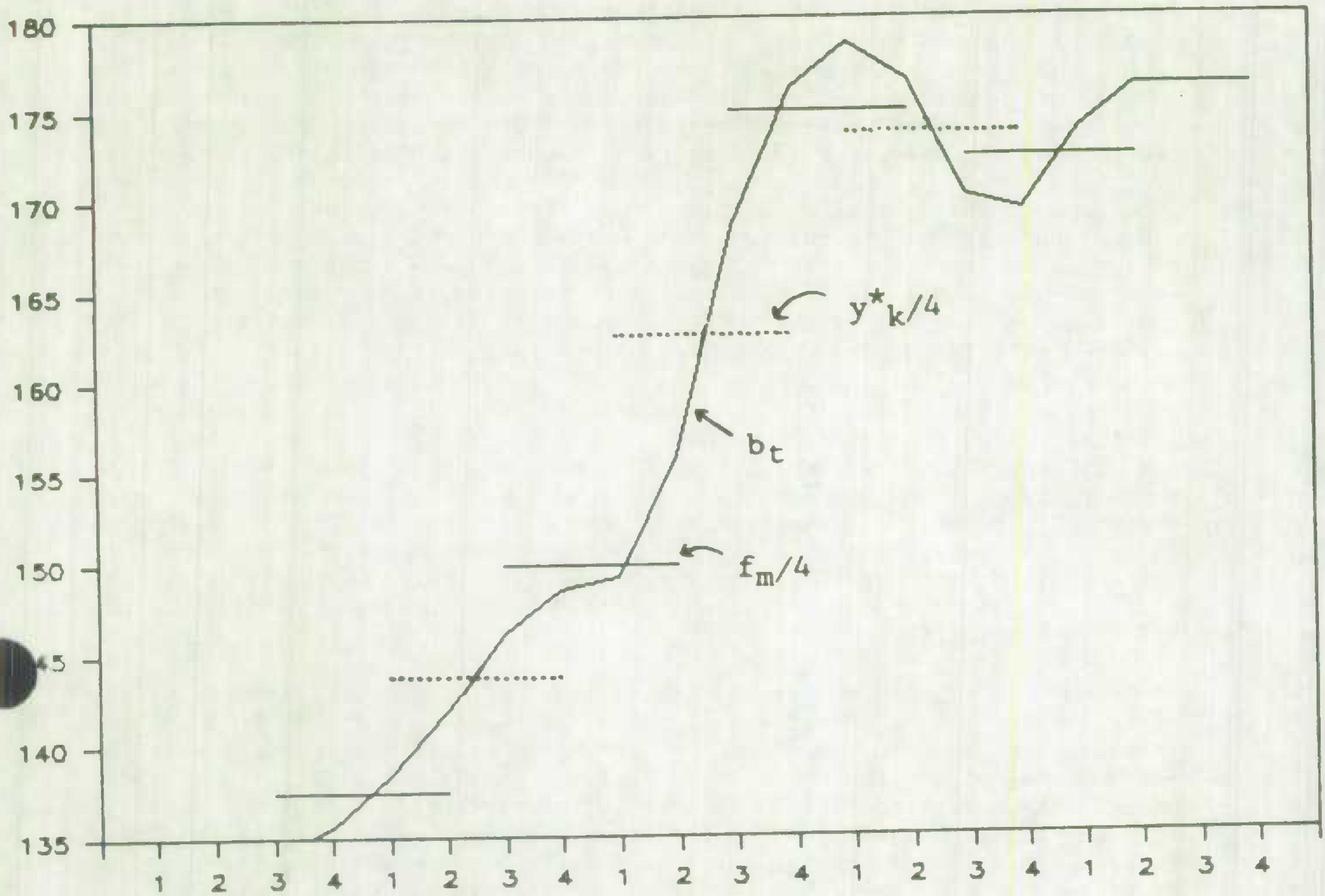

Figure 2: Calendar year estimates y^*_k obtained by the traditional method of equation (5.3) and flattest (simplest) trend-cycle b_t compatible with both the fiscal data f_m and y^*_k

6. NUMERICAL EXAMPLE

This section briefly presents examples of calendarization involving four retail trade series. Normally, calendarization would take place at a level of aggregation (over respondents), at which the respondents have common fiscal years and - in the case of the proportional method - at which they also share common sub-annual indicator $\underline{s}$. That level would usually be lower than that of publication by the statistical agency. The examples herein use published series because the "true" monthly series $\underline{b}$ and therefore the "true" calendar year values $\underline{G}\ \underline{b}$ are then known exactly and can be used to assess the interpolation and the calendarization procedures.

Column (3) of Table 2 A displays the true calendar year values $\underline{y}=\underline{G}\ \underline{b}$; and column (4), the fiscal year values $\underline{f}=\underline{J}\ \underline{b}$ for years ending in the month N indicated. Column (5) displays the calendar year estimates, $\underline{y}^* = \underline{G}\ \underline{b}^*$, obtained with the proportional variant of Section 4. The sub-annual indicator $\underline{s}$ required by this variant consists of the seasonal-trading-day pattern calculated for each series by the X-11-ARIMA seasonal adjustment method (Dagum, 1980). Using such an indicator simulates the unavailability of the monthly data. Column (6), displays the calendar year estimates obtained with the additive variant of Section 3, in the total absence of any sub-annual information, $\underline{s}=\underline{0}$. In both variants, the interpolation process was applied to the five calendar years (K=5) 1980 to 1984. Column (7) displays the estimates obtained with the traditional method of equation (5.3) applied in a 2-year moving average manner: the method is first applied to the 80-81 and the 81-82 fiscal values, to yield the 1981 estimate; second to the 81-82 and to the 82-83 fiscal values to yield the 1982 estimate; etc.

Columns (3), (4) and (5) of Table 2 B display the percentage errors corresponding to the proportional, the additive and the traditional methods respectively; and columns (6), (7) and (8), the differences between the absolute values of those errors. Column (6) shows little difference between the proportional and the additive variants. This would indicate that the additive variant be preferred: There is then no need of auxiliary information $\underline{s}$, and - as explained in Section 4 - the weights applied $\underline{P}$ are known in advance. Note however that if the fiscal years have unequal length (which is not the case here) a $\underline{s}$ different from zero must be supplied.

The predominance of negative entries in columns (7) and (8) of Table 2 B points to a superior performance of the proportional and of the additive variants, over the traditional method. The few positive entries are much smaller than the negative ones, except for years 1981 and 1983 for Used cars. Note however that for year 1982 of the same series, the traditional method is substantially out-performed. This result could be due to the pronounced year-to-year (cyclical) movement of that particular series, observable in Table 2 A. In other words that series is harder to calendarize, and all methods produce large errors, as shown in columns (3) and (4) of Table 2 B. However this case does suggest the following: The type of series for which the traditional method is superior or equivalent should be determined, so that the simplest method could be used; perhaps the traditional method should be used at the end of series and the proportional or additive variants in the center; research is needed (e.g. Monte Carlo simulation).

The traditional method produces no estimates for the first and last calendar years 1980 and 1984. (This would require availability of the 1979-80 and 1984-85 fiscal data). As obvious from columns (3) and (4) of Table 1 B, the 1980 and 1984 end estimates obtained with the proportional and additive variants of the proposed method are less accurate than the estimates for 1981 to 1983. As explained in Section 2, the end estimates involve sub-annual extrapolations outside the span covered by the fiscal years. The number of backwards extrapolations is equal to the ending month N of the fiscal year; and the number of forward extrapolations, to 12-N+1. These end estimates should then be ignored or used with caution, e.g. when N is small and when the socio-economic variable is not subject to turning-points. For variables subject to turning-points, the end estimates would greatly benefit from some forecasted fiscal year value; again research is needed.

As more (fiscal) years are added to the series the end estimates become more central; and therefore, more accurate. (This may explain why the 1982 estimate of Used Cars is more accurate than the other estimates). Empirical work with benchmarking suggest a 5-year moving average implementation of the method (Cholette, 1984) provides acceptable estimates: A calendar year estimate then becomes central and final (ceteris paribus) on the third estimation, and the final estimate is based on two fiscal year values on each side (i.e. is symmetrically embedded in four fiscal years).

The sub-annual interpolations $\underline{b}^*$ produced with the proportional variant had the following mean absolute errors: 2.14% for Department Stores, 7.78% for Used Cars, 2.54% for Family Clothing and 3.19% for Men Clothing.

Table 2: Estimating the Calendar Year Values for a Few Selected Retail Trade Series

A: True Annual Values With Their Proportional, Additive and Traditional Estimates

(1) series ident.	(2) year	(3) true value	(4) fiscal value	ending month N	(5) prop. estim.	(6) add. estim.	(7) trad. estim.
	80	9367			9607	9578	.
	80-81		9675	4			
Department	81	10218			10081	10042	10015
Stores	81-82		10184				
	82	10208			10249	10249	10286
	82-83		10336				
	83	10930			10928	10847	10846
	83-84		11100				
	84	11385			11248	11293	.
	80	439			525	523	.
	80-81		508	6			
Used	81	495			475	475	480
Cars	81-82		452				
	82	462			461	464	473
	82-83		493				
	83	501			514	516	511
	83-84		529				
	84	559			535	534	.
	80	1188			1198	1192	.
	80-81		1213	2			
Family	81	1367			1361	1358	1341
Clothing	81-82		1366				
	82	1275			1285	1285	1298
	82-83		1285				
	83	1487			1479	1467	1468
	83-84		1505				
	84	1575			1560	1571	.
	80	973			1005	1002	.
	80-81		1016	5			
Men	81	1073			1065	1060	1060
Clothing	81-82		1091				
	82	1114			1124	1119	1123
	82-83		1146				
	83	1250			1229	1215	1213
	83-84		1261				
	84	1323			1282	1287	.

Table 2 - continuation

B: Corresponding Percentage Errors

(1) series ident.	(2) year	(3) prop. estim.	(4) add. estim.	(5) trad. estim.	(6) abs(3) -abs(4)	(7) abs(3) -abs(5)	(8) abs(4) -abs(5)
	80	2.57%	2.25%	.	0.32%	.	.
Department	81	-1.34%	-1.72%	-1.99%	-0.38%	-0.65%	-0.27%
Stores	82	0.40%	0.40%	0.76%	0.00%	-0.36%	-0.36%
	83	-0.02%	-0.77%	-0.78%	-0.74%	-0.75%	-0.01%
	84	-1.21%	-0.80%	.	0.40%	.	.
	80	19.54%	19.07%	.	0.46%	.	.
Used	81	-4.16%	-4.04%	-3.07%	0.13%	1.09%	0.97%
Cars	82	-0.21%	0.53%	2.39%	-0.31%	-2.17%	-1.86%
	83	2.52%	2.90%	1.94%	-0.38%	0.58%	0.96%
	84	-4.37%	-4.43%	.	-0.07%	.	.
	80	0.83%	0.34%	.	0.49%	.	.
Family	81	-0.43%	-0.67%	-1.92%	-0.24%	-1.50%	-1.26%
Clothing	82	0.74%	0.77%	1.83%	-0.04%	-1.09%	-1.06%
	83	-0.52%	-1.38%	-1.26%	-0.86%	-0.74%	0.12%
	84	-0.97%	-0.26%	.	0.71%	.	.
	80	3.29%	2.98%	.	0.31%	.	.
Men	81	-0.74%	-1.26%	-1.25%	-0.52%	-0.50%	0.02%
Clothing	82	0.89%	0.44%	0.84%	0.45%	0.04%	-0.40%
	83	-1.63%	-2.78%	-2.94%	-1.15%	-1.31%	-0.16%
	84	-3.14%	-2.77%	.	0.36%	.	.

7. CONCLUSION

This paper proposed methods to transform fiscal year data into calendar year values. The calendarization problem is viewed either as a benchmarking or as an interpolation problem. In the first case, the method is a straightforward adaptation of the proportional variant of the Denton (1971) benchmarking method; in the second, an adaption of the interpolation methods of Boot, Feibes and Lisman (1967) and Cohen, Müller and Padberg (1971). The calendar year estimates are then the proper annual sums of the benchmarked or interpolated values.

The same strategy could be applied to calendarize fiscal quarter data and to transform data covering four or five weeks into monthly values. The inherently seasonal nature of such data however implies that the end estimates could be subject to very substantial revisions.

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: 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 topic certainly deserves more attention.

REFERENCES

Alba, E. de (1979), "Temporal Disaggregation of Time Series: a Unified Approach", **Proceedings of the Business and Economic Statistics Section, American Statistical Association**, pp. 359-370.

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.

Bell, W.R., Hillmer, S.C. (1983), "Modelling Time Series with Calendar Variations, **Journal of the American Statistical Association**, Vol. 78, pp. 526-534.

Bournay, J., Laroque, G. (1979), Réflexions sur la méthode d'élaboration des comptes trimestriels", **Annales de l'I.N.S.É.É.**, Vol. 36, pp. 3-30.

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

Box, G.E.P., Tiao, G.C. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems", **Journal of the American Statistical Association**, Vol. 70, p. 70-79.

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., and 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.

Cholette, P.A. (1988), "Weights to Calendarize Fiscal year Data Referring to Any Consecutive 12 Months or 4 Quarters" **Statistics Canada**, Time Series Research and Analysis Division, Research Paper No. TRSA-88-023E.

Chow, G,C., and 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., Müller, W., and Padberg, M.W. (1971) "Autoregressive approaches to the Disaggregation of Time Series Data", Applied Statistics, Vol. 20, pp 119-129.

Dagum, E. (1980), **The X-11-ARIMA Seasonal Adjustment Programme**, Statistics Canada, Cat. 12-564E.

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.

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

Friedman, M. (1962), "The Interpolation of Time Series by Related Series", **Journal of the American Statistical Association**, Vol. 57, No. 300, pp. 729-757.

Ginsburg, V.A. (1973), "A Further Note on the Derivation of Quarterly Figures Consistent with Annual Data", **Applied Statistics**, Vol. 22, No. 3, pp. 368-374.

Harvey, A.C. (1981), **The Econometric Analysis of Time Series**, Philip Allan.

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

Litterman, R.B. (1983), "A Random Walk, Markov Model for the Distribution of Time Series", Journal of Business and Economic Statistics, Vol. 1 No. 2, pp., 169-173.

Nasse, P. (1973), "Le système des comptes nationaux trimestriels", Annales de l'I.N.S.E.E., Vol 14, pp. 119-165.

Silver, J.L. (1986) "Two Results Useful for Implementing Litterman's Procedure for Interpolating a Time Series", **Journal of Business and Economic Statistics**, Vol. 4, No. 1, pp. 129-130.

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

c. 2