68F0016XPE Calendarization in the Context of the Unified Enterprise Survey

Yanick Beaucage and Peter S. Tarassoff[1]

ABSTRACT

Many respondents to business surveys provide data for their respective fiscal years, which often differ both from one respondent to another and from the calendar year. Despite the mix of periods covered, business survey managers commonly aggregate all such data and publish so-called "annual" estimates. However, several problems ensue from the fact that these estimates and any related growth rates are calculated from data with different fiscal year-end dates. These results would have been different if all of the data had been based on true calendar years. Calendarization, the transformation of fiscal year data to a calendar year basis, is a remedy to this problem. Statistics Canada's new Unified Enterprise Survey (UES) represents an opportunity for calendarization's more widespread application. Under the aegis of the Project to Improve Provincial Economic Statistics (PIPES), the UES Pilot will target seven industries, and all respondent microdata are to be calendarized.

This document will show the importance of calendarization, explain how it is performed and what the necessary inputs are, and finally discuss certain aspects of the development of a calendarization system. These aspects include: the construction of seasonal profiles, the types of variable that are to be calendarized, the place of calendarization in a processing sequence, the constraints among variables and/or among units (enterprise/establishment) that calendarization ought to preserve, and finally the evaluation of calendarized values. This discussion will be made in an exploratory way since the UES calendarization system will not be ready until Fall 1998. Data from other surveys will be used to illustrate each point.

KEYWORDS: Calendarization, benchmarking, PIPES, annualization, quadratic minimization.

1. INTRODUCTION

The Unified Enterprise Survey (UES) is an annual survey that invites respondents to divulge certain financial information according to their fiscal years (corresponding to their financial statements) to facilitate the collection of that information. Clearly, each respondent could have a different fiscal period, according to its needs. When data are aggregated, one faces a multitude of overlapping periods and, by this very fact, a range of information covering a period of up to two years. The habitual procedure consists of not accounting for the exact reporting period and of simply aggregating the reported data and thus publishing estimates that supposedly pertain to a single and unique period, the calendar year, for example. However, despite the simplicity of this practice, is it rigourous? What can one do to produce estimates that truly pertain to the calendar year?

The purpose of this article is to show on one hand, the danger of simply aggregating data covering different periods. Secondly, we will discuss the use of two methods of calendarization which permit us to resolve this problem. Finally, we will tackle other aspects related to the calendarization of UES data.

2. CALENDARIZATION

In the context of the UES Pilot Survey of 1997, each questionnaire includes a "Reporting Period" section, which indicates the period that is targeted by the survey: "For the purpose of this survey, please report information for your **12 month fiscal period** for which the **FINAL DAY** occurs on or between January 1, 1997 and December 31, 1997". This practice is not uncommon, since most of the information that is provided by Canadian businesses is provided according to their own fiscal periods.

The respondent is required to indicate the period (start and end) for which its data are reported in terms of the day, month, and year. One can easily see that, in general, the responses can be staggered from the period

[1] Yanick Beaucage, Business Survey Methods Division, Statistics Canada, R.H.Coats Building 11-G, Ottawa (Ontario), Canada, K1A 0T6, E-Mail : beauyan@statcan.ca.
Peter S. Tarassoff, Enterprise Statistics Division, Statistics Canada, Jean-Talon Building GE-8, Ottawa (Ontario), Canada, K1A 0T6, E-Mail : tarapet@statcan.ca.

"January 2nd 1996 to January 1st 1997" to the period "January 1st 1997 to December 31st 1997". One thus has as set of responses that cover a period of two years, and that one would like to convert to a single, common period, i.e. January 1st 1997 to December 31st 1997, namely the calendar year.

To evaluate the extent of the periods for which Canadian businesses report their data, one must examine the distribution of their fiscal years (while supposing that their fiscal years do cover 12 months). There are several sources of administrative data that can yield such information. Table 1 presents the distribution of the number of establishments (in the form of percentages) by the month of the fiscal period end.

Table 1: Distribution of the Number of Establishments by Month of Fiscal Period End.

	Construction	Couriers	Real Estate
January	9%	6%	7%
February	7%	6%	4%
March	10%	8%	6%
April	8%	8%	6%
May	6%	6%	5%
June	6%	7%	7%
July	7%	8%	8%
August	6%	8%	7%
September	7%	8%	8%
October	7%	7%	7%
November	5%	4%	5%
December	22%	25%	31%
Total	100%	100%	100%

Note that the classification is approximate since the NAICS (North American Industrial Classification System) codes used for the [UES] Pilot Survey are not yet very widespread. A correspondence with the old Standard Industrial Classification (SIC-80) enabled us to construct the table. Clearly, even if the proportion of respondents having a December fiscal year end is about 25%, one can wonder about the distribution when one considers revenue. Table 2, constructed once again from fiscal data, shows this distribution in the form of percentages.

Table 2: Distribution of Annual Revenue by Month of Establishment's Fiscal Period End.

	Construction	Couriers	Real Estate
January	8%	1%	2%
February	7%	2%	1%
March	10%	3%	1%
April	7%	1%	2%
May	4%	17%	3%
June	4%	11%	2%
July	5%	4%	3%
August	5%	2%	2%
September	7%	7%	3%
October	8%	2%	3%
November	6%	26%	2%
December	30%	26%	78%
Total	100%	100%	100%

Clearly, the percent distributions of remitters and those of their revenues are quite different. One could say that in the real estate industry, remitters with large revenues generally have fiscal years ending in December, which minimizes the impact of not calendarizing. However, for the other industries, one observes that fiscal periods are distributed throughout the year, with the largest proportion in December.

For these cases, what damage is caused to the estimates when one does not calendarize? Actually, the answer to this question depends on the growth rate between Januray 1996 and December 1997 of the variables that are to be calendarized. Clearly, the more that this rate differs from 0, the greater is the importance of calendarization. One rough way of evaluating the error that one commits when one does not calendarize is to suppose that the data are reported for a period having a set lag (three months, for example) relative to the calendar year. By setting the

annual growth rates and calculating the equivalent compounded monthly rate, one can evaluate the error that is committed. Table 3 presents the error that is committed for lags of 3, 6, and 9 months, i.e. corresponding to fiscal years ending on March 31st, June 30th, and September 30th.

Table 3: Error Committed without Calendarization.

Annual Growth Rate	Lag in months (relative to December 31st)						
	-9	-6	-3	0	3	6	9
-10,0%	8,2%	5,4%	2,7%	0,0%	-2,6%	-5,1%	-7,6%
-5,0%	3,9%	2,6%	1,3%	0,0%	-1,3%	-2,5%	-3,8%
-2,5%	1,9%	1,3%	0,6%	0,0%	-0,6%	-1,3%	-1,9%
-1,0%	0,8%	0,5%	0,3%	0,0%	-0,3%	-0,5%	-0,8%
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
1,0%	-0,7%	-0,5%	-0,2%	0,0%	0,2%	0,5%	0,7%
2,5%	-1,8%	-1,2%	-0,6%	0,0%	0,6%	1,2%	1,9%
5,0%	-3,6%	-2,4%	-1,2%	0,0%	1,2%	2,5%	3,7%
10,0%	-6,9%	-4,7%	-2,4%	0,0%	2,4%	4,9%	7,4%

* This table applies to the case of flow variables

One can thus see that in the case of a negative lag of three months (i.e. fiscal period ending on September 30th of year *t*) and a growth rate of 5% (applied to the last three months), one commits an error of -1.2% in using reported data from the respondent to make a tabulation of whatever kind. One can see clearly that the further the lag and the growth rate are from zero, the greater is the error. These errors do not, unfortunately, cancel (if the growth rate goes in the same direction for each of them), since all of the survey's respondents will have negative lags relative to the end of the calendar year by the very nature of the survey's coverage period. One can thus easily conclude that the sum of such errors will cause an undesirable difference in the tabulations.

3. CALENDARIZATION METHOD

Conceptually, calendarization can be divided into 3 steps:

(1) Interpolate the reported fiscal data into sub-annual data (such as monthly, weekly, or even daily) while ensuring that the sum is always equal to the annual total;
(2) Extrapolate, at the beginning and end of the covered periods, the behaviour of the fiscal data (always in a sub-annual form); and
(3) Sum the values (interpolated or extrapolated) that lie within the period of interest such as, for example, the calendar year.

There are several calendarization methods that can combine each of these steps in some manner to produce a calendarized value. Laniel and Poirier (1992) examine some of these methods by comparing them using a simulation. Likewise, Cholette (1998) also examines several methods.

Certain methods are simple and easy to apply; whereas others are more rigourous and are based on standard minimization criteria. The first method presented belongs to this first category; we will call it the "fractional method". This method divides each of the fiscal values in two parts according to the part of the [calendar] year that is covered by each of these parts. When one has divided two fiscal values thus (for the years *t* and *t+1*), one obtains the adjusted value for the year *t*:

$$X_t^c = \frac{kX_t^f + (12-k)X_{t+1}^f}{12}$$

where X_t^c denotes the calendarized value for year *t*, X_t^f denotes the fiscal value (collected) for year *t*, and *k* is the number of months elapsed since the beginning of calendar year *t*. If, for example, a respondent ends its fiscal year on May 31st 1997, the calendarized value for 97 is:

$$X_{97}^c = \tfrac{5}{12} X_{96-97}^f + \tfrac{7}{12} X_{97-98}^f$$

Note right away that this method requires the arrival of a second year of data (97-98 in this case) or that a means be found to forecast for 97-98. With a forecast for 97-98, one faces a preliminary estimation, followed by a final estimation once the fiscal data for 97-98 are finally available.

A second method from among the group of more rigourous methods is that presented in Cholette (1984). It uses the technique of benchmarking a sub-annual series (e.g. monthly) to calendarize the data over the desired period. Since this method is a modification of Denton's (1971) approach, let us call it the "Denton Method". This method benchmarks sub-annual data to fiscal data. One thus uses an objective function that is to be minimized subject to the constraint of the fiscal totals. The objective function is the following:

$$O(x) = \sum_{m=2}^{T} ((x_m - x_{m-1}) - (z_m - z_{m-1}))^2$$

where z_m corresponds to the monthly values that are to be benchmarked, and T represents the total number of months included in all of the fiscal years that are available for calendarization. When there is a single year of data, T=12. We are thus looking for the series of monthly values x_m that best reproduces the movement of the series z_m, subject to the constraint:

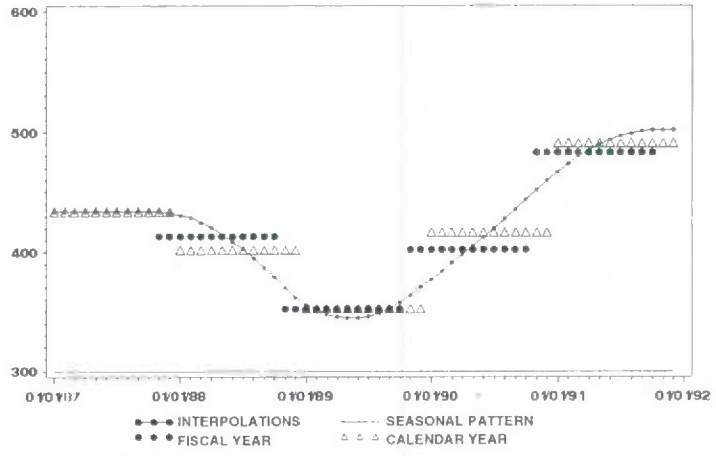
$$X_t^f = \sum_{m=1}^{12} x_m \quad \text{for each fiscal year } t \text{ involved.}$$

Note that one minimizes the first differences $(x_m - x_{m-1})$ of the series such that the two series are as parallel as possible. One could also minimize the second differences $(x_m - x_{m-2})$, which would yield a different objective function as well as different calendarized values. It can be shown that minimizing the second differences is equivalent to applying the fractional method (Cholette, 1998).

Denton also proposes a proportional objective function where the differences are divided by the corresponding observations z_m. Cholette (1984) succeeded in approximating numerically the proportional Denton Method for application to daily rather than monthly interpolations. Extrapolations for incompletely covered calendar years are a by-product of this technique.

Here is a first illustration in which there are four fiscal values (from Nov. 87 to Oct. 88, ..., and from Nov. 90 to Oct. 91), represented by solid circles in Figure 1.

Figure 1: Simple Application of the Denton Method.



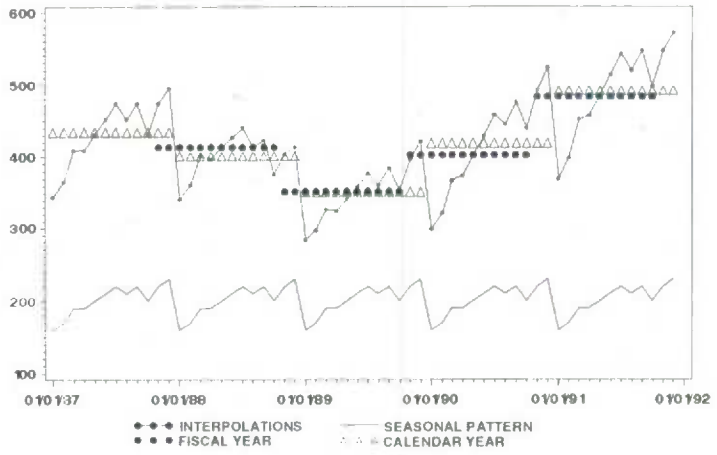
Note that an annual value is first of all divided into 12 equal parts before being represented in the figure. Below, there is the series z_m (representing seasonality) that is equal, for the purposes of the example, to the constant 300. Thus, the method seeks the series of interpolations x_m that minimizes the objective function while ensuring that the four fiscal totals are respected. The series of x_m (interpolations) is represented by the smooth curve.

Once the interpolations (and the extrapolations for the period of Jan. 87 to Oct. 87 and that from Nov. 91 to Dec. 91) are calculated, all that remains is to sum over the 12 months of interest, i.e. those of the calendar year. One

thus obtains the calendarized values (5 in this case), once again divided into 12 equal parts represented by the hollow triangles. Note that the extrapolations are flat and that they simply repeat the last interpolation.
When one calendarizes a single observation on a twelve-month fiscal period using the method presented thus far, one obtains the same result as the fiscal value that one seeks to calendarize. As in the case of the fractional method, one must obtain more than one annual fiscal observation to produce a "true" calendarized value. This implies that the calendarized data will be revised and that there could be more than one revision. Nevertheless, Cholette (1984) mentions that after the second revision, the changes become very small.

The beauty of this method is its adaptation to the contribution of auxiliary information. Actually, instead of using a series of constant z_m, one can improve the method by the aid of a series of monthly indicators. On one hand, this precludes obtaining flat extrapolations at the beginning and end of a series, and on the other, it permits improved calendarization in the case of fiscal data from a non-12 month period. If the fiscal period exceeds 12 months, seasonality enables a partition within the fiscal period such that one can extract the value that corresponds uniquely to the calendar year. If the fiscal period is less than 12 months, knowing the seasonal pattern enables an improved inference on the rest of the year (as the case may be). Figure 2 illustrates the same example, but in which the series of z_m (seasonality) is no longer constant, but obeys a given pattern.

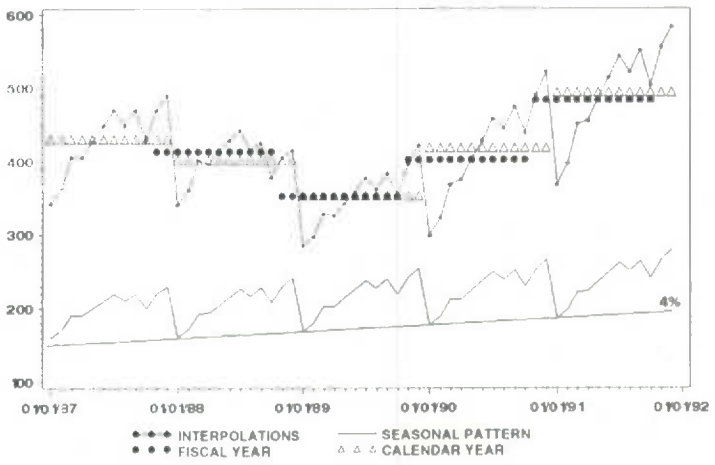
Figure 2: Addition of a Seasonal Pattern.



Note once again that in the case of a single fiscal datum with a period of 12 months, the addition of a seasonal profile to the method will not change the calendarized value. The latter will be equal to the fiscal value since seasonality cancels out over a twelve month period.

One could improve the process further by applying a trend parameter to the data. In Figure 3, an annual growth parameter of 4% has been applied by computing the compounded monthly equivalent. Also, the same growth rate has been used from one year to the next. The line below the series z_m shows this growth.

Figure 3: Addition of a Trend Parameter.



Once again, one uses auxiliary information (the trend) to improve the method. This time, if one were to calendarize a single fiscal observation that is twelve months in length, the resultant value would be different.

Note finally that a fourth variant of the method consists of specifying only the trend parameter, without using a series of monthly indices. The series of z_m would then be represented by the line (of 4%) that appears in Figure 3. In fact, this variant corresponds to calendarizing at a macro level (at the level of an aggregate or group) and then applying it at the level of the microdata. A major disadvantage of macro-calendarization is the need to consider a common period for all units of the group (which is rarely the case in practice). In such cases, however, knowing the annual growth rate for a given group (e.g. industry by province), one can apply it (by means of calendarization) severally to each of the observations in the group. One thus retains the advantages of macro-calendarization, while accounting for the individual period at the moment of calendarization.

Another advantage of micro-calendarization is the ability to use calendarized microdata to construct any tabulation required, without having to re-calendarize the new desired total. Clearly, the ideal would be to know the exact trend at the level of each unit, but this may be rather difficult to obtain. Finding the trend at the level of the group turns out to be a good compromise while being easy to apply.

The Denton Method thus has the following advantages:

1) It is a rigourous method based on an objective function that is to be minimized subject to certain constraints;
2) It can benefit from the use of various kinds of auxiliary information (trends, profiles, ...);
3) It adapts easily to fiscal years of a length other than 12 months; and
4) It can be used with a single observation.

In addition, the method can estimate missing values when there is one (or more) gaps in the data; and it can emulate the fractional method when the objective function is adjusted. This method thus seems promising, and we hope that we can manage to implement it to generate calendarized data in a routine way.

4. OTHER ASPECTS

Clearly, choosing one method over another does not eliminate all of the problems related to calendarization. Among these problems, we immediately note that of choosing when to calendarize. Laniel and Poirier (1992) suggest that calendarization be applied directly to captured data, before proceeding to editing and imputation. In the context of the UES, the collected data enter into an automated editing process that permits the evaluation of the quality of the collected information as well as dealing with outlier data. Rejected records are next submitted to manual review before being reintegrated into the editing and imputation process. One cannot thus exit the system after automated/manual editing and before imputation in order to calendarize. The latter thus takes place after, at the end of the processing steps, and just before estimation.

Clearly, if one could calendarize immediately after editing, one could then impute for non-respondents using

already-calendarized values. This would reduce the number of records to be calendarized while obviating the need to evaluate non-respondents' reference periods. In calendarizing after imputation, it is necessary first to find a reliable source to establish the fiscal period of the non-respondent, and then to take this period into account when choosing a donor before proceeding to imputation. Once imputation is completed, the reported data (respondents) and the imputed data (non-respondents) are calendarized. From an operational standpoint, the task was seen to be more efficiently treated by having the calendarization module at the end, even if steps are thereby added to the process.

The UES collects over 500 different variables that are to be calendarized. Treating all of these variables separately would entail considerable time and development costs. Moreover, most of the variables are in fact "details" of global variables: the revenues are divided among categories, as are the expenses, etc... The variables that correspond to "totals" are thus the variables that are going to be calendarized and the detail variables will be pro-rated.

Moreover, the UES collects data at the level of the enterprise as well as at that of the establishments belonging to the enterprise. Therefore, just as one would like to ensure that the sum of the non-calendarized establishment values equals the value reported at the enterprise level, the same goes for the calendarized values. This constraint is not simple to formulate in the context of the proposed methodology and we hope that we can gradually find a solution to this problem.

The figures presented thus far only consider flow variables such as revenues and expenses, of which the value represents an accumulation over the fiscal period. However, the UES also collects stock variables (such as inventories). Their values, unlike those of flow variables, represent a level as of the last day of the fiscal period. Fortunately, Cholette's algorithm can treat flow and stock variables in a similar way, but a seasonal pattern is necessary for the latter to be calendarized properly. Without a pattern, the calendarized [stock] value is simply a linear interpolation between two stock values.

Throughout the previous section, one has been able to see that the method can function in four different modes: no auxiliary information (basic mode), seasonal patterns (first improved mode), trend (second improved mode), and seasonal patterns with trend (ultimate mode). It is thus important to use as much auxiliary information as possible and to attempt to apply the ultimate mode. This is especially true when there is a single datum, as the basic mode produces a calendarized value equal to the original fiscal value in the case of flow variables and the [calendar] value is not credible in the case of stock data. This has an impact on the importance of revisions, which will increase in extent once the survey attains its second year if calendarization will have been minimal during the first.

One would thus like to try to find the seasonal pattern as well as the trend in each group (industry/province) that is deemed to be different, and this for each of the variables that is to be calendarized. This kind of information can be very painstaking to obtain. Auxiliary data will be the main source used to create these patterns and to compute the trends. As well, one will have to use "proxy" variables and hope that these cover each type of variable that is to be calendarized, whether flow or stock. These patterns and trend will be stored for inspection and revision.

Note that when the data for the 1997 reference year are going to be calendarized, 1998 will already be fairly advanced. The availability of 1997 and 1998 series at that time will be very useful for evaluating the trends as well as the patterns.

Finally, as for all good informatics systems, one must proceed to verify the calendarization process. One would thus like to implement the methodology necessary for following a record throughout the calendarization process. One could thus ensure calendarization quality and modify the value where necessary. Secondly, one would like to implement a process that permits the verification of the calendarized value. According to the trend observed in the calendarized variable's group (industry/province), one would like to use limits deduced in a similar way to those of Table 3. One could thus determine a neighbourhood within which the calendarized value should lie. Values beyond the neighbourhood would then be sent to a manual review process wherein human judgment would decide which treatment to adopt.

5. CONCLUSION

A good number of operational problems remain to be dealt with before the complete (and even partial) running of the calendarization process in the UES context. We know, however, that calendarization has its place in the survey. It obviates a too-often ignored bias while giving managers increased confidence in the representativity of their data on a calendar year basis. The choice of the method is fairly clear but, to do a proper job of calendarization, we must be sure to have available the auxiliary data that are necessary for functioning at least in the first improved mode in the case of stocks, and at least in the second improved mode in the case of flows. All of this is to preclude the major revisions that would arise if only the basic mode were used for the first year. Most other problems related to calendarization have been resolved in a simple way or using promising ideas that are still waiting to be put to the test. Programming these is going to be a crucial element in the implementation of the process but, budget and resources permitting, the UES will have the most sophisticated calendarization system in the world.

6. REFERENCES

Cholette, P.A. (1984). "Adjusting Sub-Annual Series to Yearly Benchmarks." *Survey Methodology*, Vol. 10, No.1, 35-49.

Cholette, P.A. (1998). Comparison of Methods to Calendarize Fiscal Year Data (draft). Internal document, Statistics Canada.

Denton, F.T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. Journal of the American Statistical Association, Vol. 66, No. 333, 99-102.

Poirier, C., & Laniel, N. (1992). "Ajustement des données d'enquêtes annuelles à l'année civile[:] comparaison de méthodes." Internal document, Business Survey Methods Division, Statistics Canada, March 1992.