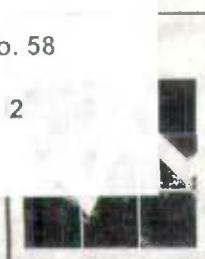


68N0003XPB

no. 58

c. 2



Project to Improve Provincial Economic
Statistics

Calendarizing Business Survey Data

Projet d'amélioration des statistiques
économiques provinciales

La calendrialisation des données des enquêtes- entreprises

Technical Series

Number 58

Série technique

Numéro 58



Internet: www.statcan.ca
Intranet: <http://pipes>



Statistics
Canada

Statistique
Canada

Canadä

Calendarizing Business Survey Data

Calendarization is the transformation of data from a fiscal period basis to a calendar period basis. Calendarization methodology is fairly well-documented, but its implementation is not. This paper reviews some of the practical considerations that arise when applying calendarization to annual business survey data.



Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

For further information on the materials covered in this paper, please contact Bonnie Bercik (613) 951-6790 or Diane Proulx (613) 951-7192
Fax: (613) 951-0411

Calendarizing Business Survey Data

Peter S. Tarassoff, Statistics Canada
Main Building 1005, Ottawa, Ontario, K1A 0T6, Canada
Peter.Tarassoff@statcan.ca

June 2000

Key Words: Annualization, Calendar Year, Fiscal Period, Irregularly-Spaced Data, Temporal Disaggregation, Benchmarking

1. Introduction

There is a fairly extensive theoretical literature about using the Denton (1971) or Boot, et al. (1967) method to benchmark time series and calendarize fiscal period data (e.g.: Cholette, 2000a; Cholette and Baldwin, 1989; Cholette and Chhab, 1991; and Cholette and Coster, 1998). However, except for Beauchage and Tarassoff (1998), little has yet been written about applying calendarization to the actual data of a business survey. Statistics Canada's Unified Enterprise Survey (UES) has represented the first major production application of Pierre Cholette's Denton-based calendarization method, and has acted as a rigorous proving ground for the calendarization method and the SAS module that embodies it. Without repeating the formal, theoretical basis for the calendarization methodology, this paper deals with the practical considerations that are entailed by the calendarization of business survey data.

Calendarization is the transformation of data from a fiscal period basis to a calendar period basis. Unlike the fiscal period, the calendar period is consistent over time as well as among different statistical units and industries. Calendarization takes explicit account of the following: first, that statistics are generally published with explicit reference to a certain calendar period; second, that respondents' fiscal periods often fail to coincide with that calendar period; and third, that the levels of the quantitative variables tend to change over time.

The alternative of leaving non-calendar period data as they are can give rise to biases, both in estimates of level (Cholette and Coster, 1998; Cholette and Higginson, 1987) and in estimates of calendar period over calendar period change. These biases can also be thought of as a *misplacement* of the level in time, e.g. such that turning points in the business cycle can be missed. Other problems could ensue if the mix of fiscal periods changes over time and/or differs among industries, sectors, and/or geographic regions. In the first case, a time series of non-calendar period estimates will be irregularly-spaced to the extent that the fiscal period mix will have changed. In the second case, non-calendar period estimates will not be fully comparable across industries, sectors, and/or regions. This latter case could give rise to difficulties in reconciling inter-regional and inter-industrial flows, e.g. between a business service industry and a manufacturing client industry, as well as to time-dependent biases in inter-governmental flows.

Finally, some jurisdictions (e.g. France) and agencies require that businesses' annual fiscal periods be on a calendar year basis; or require that respondents report on a calendar year basis, despite their non-calendar annual fiscal periods. These can be thought of as alternatives to calendarization, but the latter has several disadvantages. It imposes response burden, and it would yield data of dubious quality. There is no guarantee that respondents will bother to adjust their data, nor that they will do so consistently or correctly. Also, the fact, quality, and consistency of the transformations cannot readily be

verified. As might be gathered, such requirements for calendar year responses are uncommon in Canada. Nevertheless, Canadian income tax legislation requires newer, unincorporated businesses to have calendar year annual fiscal periods. Moreover, certain Canadian industries have uniform, yet non-calendar annual fiscal periods. For example, the banking industry has an annual fiscal period ending on October 31st, and the Canadian Government's annual fiscal period ends on March 31st.

2. An Overview of Calendarization

2.1 "Fiscality," "Calindexes," and Diagnosing the Need for Calendarization

"Fiscality" can be defined as the incidence of non-calendar fiscal periods among a set of data, i.e. as a lack of "calendarity." Refinements to this notion account not only for the incidence of non-calendar fiscal periods, but also for the extent to which fiscal periods deviate from calendar periods. However measured, a high degree of fiscality is not sufficient to indicate that calendarization is required, just as significant growth in pertinent key variables is not sufficient either. Instead, it is the combination of these two phenomena that indicates a need for calendarization.

The "Calindex" is one appropriate measure of "fiscality." Each Calindex variant depends on the assumed reference year concept, i.e. on the applicable mapping from a given fiscal year to the target calendar year, but a generalized Calindex can be computed as follows (Cholette, 2000b):

$$\text{Calindex} = (\sum_n |d_n| r_n) / M;$$

Where M is the maximum number of months in the target calendar period, i.e. 12; d_n is the displacement in months between the fiscal year end and the corresponding calendar year's end; and r_n is the proportion of units (or units' activity) in displacement class d_n . Among all reference year concepts, the calindex has a range from 0 (zero) to slightly over 1 (one), with 0 (zero) indicating no fiscality and 1 (one) indicating significant fiscality.

Note that d_n and r_n depend on the correspondence that is established between fiscal years and calendar years, i.e. on the notion of "reference year." At one extreme, and under a reference year concept like that of the UES, where a fiscal year is mapped to the calendar year in which it ends, the calindex and apparent fiscality are maximized. At the other extreme, a reference year which maps fiscal years to the calendar years with which they overlap the most would tend to minimize the calindex, hence apparent fiscality. Thus, some reference year concepts entail a closer correspondence between the fiscal year and the calendar year. Likewise, it is important to compute the Calindex according to the same reference year concept when comparing fiscality among industries and geographic regions. The Calindex's chief limitations are that it does not account for fiscal year length, and that it ignores movements in the trend-cycle.

Table 1. shows UES reference year concept, revenue-weighted Calindex values for certain 4-digit NAICS industries, for all of Canada, as well as for the provinces having minimum and maximum Calindex values. The table also shows revenue-weighted mean fiscal period start and end dates and the potential impact of calendarization, should the GDP at Market Price series be used as the calendarization indicator. Clearly, the degree of fiscality can vary by industry and by geographic region. Thus, if estimates at any level other than the national, all-industry level are contemplated, a "fiscality" assessment should be made at that more detailed level.

TABLE 1. Calindexes, Weighted Mean Fiscal Year Start and End Dates, and Expected Calendarization Impacts for the UES Samples of Certain 4-Digit NAICS Industries

4-Digit NAICS Industry		Provincial Calindex Extrema		All-Canada Calindex	Weighted Mean Fiscal Year Start and End Dates		Calendar to Fiscal % Difference
Code	Descriptor	Min.	Max.		Start	End	
1125	Animal Aquaculture	0.07	0.73	0.19	19/10/96	18/10/97	1.12%
2312	Building Construction	0.13	0.55	0.32	19/08/96	19/08/97	1.50%
2313	Engineering Construction	0.22	0.62	0.32	15/08/96	15/08/97	1.54%
2321	Site Preparation Work	0.03	0.57	0.36	25/08/96	25/08/97	1.43%
2325	Building Equipment Installation	0.30	0.68	0.36	06/08/96	04/08/97	2.24%
4853	Taxi & Limousine Service	0.00	0.69	0.33	13/09/96	11/09/97	1.82%
4921	Couriers	0.12	0.55	0.23	16/10/96	15/10/97	1.16%
5311	Lessors of Real Estate	0.11	0.76	0.26	06/10/96	04/10/97	1.56%
5312	Offices of Real Estate Agents & Brokers	0.00	0.50	0.34	07/09/96	06/09/97	1.59%
7221	Full-Service Restaurants	0.23	0.51	0.43	08/08/96	03/08/97	3.12%

Regardless of what the Calindex might show, the expected calendarization impact (EI) can be computed as follows:

$$EI = (EC - EF) 100 / EF;$$

Where EC, the expected calendar year value, is the sum of the sub-annual indicator's observations over the target calendar year, and EF, the expected fiscal year value, is the sum of the same sub-annual indicator's observations over the fiscal year that is delimited by the weighted mean fiscal year start and end dates. This computation presupposes that proportional rather than additive calendarization will be applied, and that the indicator series is appropriate. Note that EI would entail some error, since much information about fiscal periods is lost when the mean dates are computed. Finally, EF, hence EI, can be improved somewhat by using the Tarassoff modification (discussed below) such that the indicator sum includes the activity of partial months.

EI depends highly on the indicator series. In fact, the actual calendarization of the reference year 1997 Aquaculture data yielded an impact of from -21.53% to +3.80%, depending on the province. Although the impact of calendarization is often on the order of a few percent, the calendarization transformation is orthogonal to other sources of error. Thus, resources permitting, it is probably better to calendarize than not to calendarize.

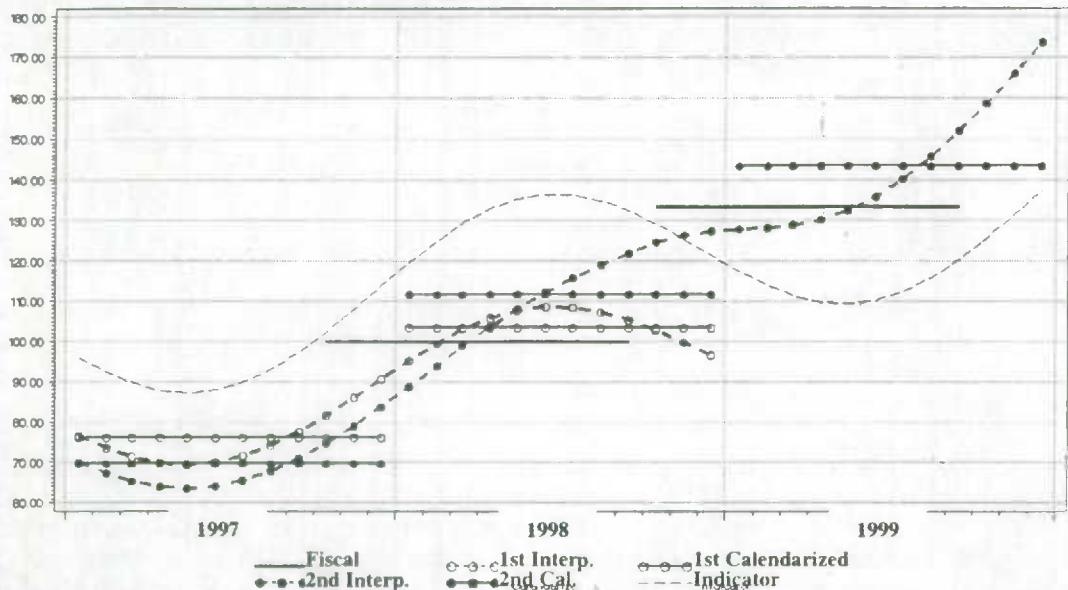
2.2 Cholette-Denton Calendarization

In non-mathematical terms, Cholette-Denton calendarization consists of fitting a series of sub-annual interpolations to fiscal period benchmarks, and then aggregating the interpolated series over the target calendar period(s). The sub-annual series may either be generated directly from the benchmarks themselves or by constraining a sub-annual indicator series to fit the benchmarks. The method is such that the sub-annual interpolations follow the indicator's movement, especially at the ends. It can treat both stock and flow variables. Given the Tarassoff modification of using trading day weight-derived fractions to split calendar period values, it is now possible to handle fiscal periods that do not start or end at the ends of a calendar period, i.e. irregularly-spaced data (Cholette, 2000).

In a survey processing context, there are two notional calendarization phases for each reference year's data, both of which are illustrated in Figure 1. In the first phase, the current reference year is also the most recent reference year, so that the fiscal benchmark would not necessarily cover the entire target calendar year. The series of sub-annual interpolations must thus be extrapolated according to the movements of the sub-annual indicator series, or according to a set of convergence parameters. The

second phase occurs when the next fiscal benchmark that completes the coverage of the calendar year becomes available. In this phase, calendarization depends much less on the indicator series. In summary, whereas the first phase enables one to produce relatively timely preliminary calendarized data, the second phase entails a less timely revision. However, the timeliness of first-phase calendarization is gained at the expense of relying upon what may only be proxy indicator series.

FIGURE 1. Interpolations and Calendarized Values Obtained In the First Phase (Circles) and the Second Phase (Dots).



Calendarization indicators provide the movement that is reflected in the sub-annual interpolations, and are especially critical in first-phase calendarization. Six desirable indicator characteristics are: spanning the union of the fiscal and calendar periods of interest; including trend-cycle and seasonal components; being of the same content as the variables that are to be calendarized; being at a fine temporal resolution; and pertaining as closely as possible to the businesses that are to be calendarized, or at least to the industrial-geographic combinations for which estimates are to be generated.

Unfortunately, sub-annual indicator series are of limited availability, and proxies often have to be used. Thus, some indicator series may be of unknown or questionable appropriateness. One means of evaluating sub-annual indicator series would be to regress their series of annual aggregations against the corresponding annual series for the units that are to be calendarized, although the best solution would be to ensure that second-phase calendarization and its corresponding revision take place.

Some variables may be deemed relatively constant over time, unlike the pertinent indicator series, and may not then be transformed by calendarization. Likewise, different indicators could be used for different variables. Both of these facts imply that a post-calendarization balancing step is essential, to ensure that the variables add up properly.

Before any target period constraints are applied, Cholette's SAS implementation of the Cholette-Denton method generates one record for each calendar period that is overlapped by the fiscal period, even if the unit in question is not in sample for both reference periods. Although these "bonus" records are not currently used in sampling and estimation, they do represent an opportunity for augmenting adjacent reference periods' samples.

3. Calendarization Pre-requisites, Survey Data Problems, and Solutions

3.1 Microdata vs. Macrodata

The Cholette-Denton calendarization methodology is best applied to *microdata*. Despite the all too common practice of benchmarking *macrodata*, and despite the apparent efficiency of dealing with fewer series at a higher level of aggregation, this practice can result in biases. These biases would stem from the fact that since most *macrodata* represent aggregations of diverse fiscal periods, *macrodata* are only very rarely delimitable by a single pair of dates. Even if the *macrodata* were to be delimited using the central tendencies of the fiscal period start and end date distributions, the same errors as those mentioned above with respect to expected calendarization impacts would occur. Once again, the cause would be the loss of information about the individual fiscal periods that occurs when the mean dates are computed.

3.2 Fiscal Period Delimitation

The fiscal periods of the data that are to be calendarized ought to be well-delimited in the sense that each fiscal period benchmark should have both start and end dates, that these dates should be precisely measured down to the day, and that they should be error-free. Where they differ from the fiscal periods, periods of seasonal activity should also be well-delimited, as should any significant periods of temporary closure.

Fiscal period dates are the most important calendarization parameter, as they mark when the activity of interest took place in the case of a flow variable, or when a given level was attained in the case of a stock variable. By over- or under-measuring the time period during which activity took place, missing or erroneous dates can have a significant impact on calendarization. Thus, when calendarization is an objective, dates become more important than they have traditionally been as mere indicators of whether or not a given survey response fell within the reference period.

Consequently, date fields ought to be edited and imputed more stringently. Besides missing values, errors to look for include: originally out-of-reference-period dates that have been “corrected” to fit the reference period; ambiguous date digits on questionnaires; ambiguous date corrections that are made by respondents; seasonal period dates that are reported as annual fiscal period dates; and day and month elements that are inverted during data capture or by the respondents themselves. Clearly, business-demographic data can also be useful as a source of explanatory information regarding events that give rise to changes in date patterns.

Sometimes, the sampling units can differ from the statistical units, e.g. a single questionnaire might cover a group of establishments. When this occurs, the fiscal period dates on the questionnaire could be presumed to apply to most of the establishments in the group, but it is quite possible that some of the establishments would have different, unmeasured fiscal periods. One option would be to include fiscal period date fields on any supplementary worksheets that are sent for the purpose of allocating data from the collection entity to establishment level.

Usually, event dates such as dates of sale ought to be used to constrain the target calendar period, so that the inter- and/or extra-polations do not exceed the period(s) in which the unit is deemed to have existed and been active. However, this also relates to calendarization-estimation interactions. There are some events that cannot be adequately handled by post-stratification, e.g. the sale of a take-all enterprise's establishments partway through their fiscal year. Given the typical delay in updating a frame, or even despite it, those establishments that do not change their identity when sold will continue to exist, but will no longer be subject to sampling and will not affect post-stratification counts. Without

calendarization, their truncated fiscal periods would result in under-estimation of business activity. With calendarization, however, their activity could be extended beyond the date of sale.

3.3 Time Series Length and Parallel Calendarization

The Cholette-Denton calendarization methodology is best applied to time series that consist of at least two consecutive observations. Ideally, each target calendar period should be fully spanned by adjacent fiscal period observations. However, although they are ultimately intended to provide macro-level time series, many business surveys are, in fact, designed and processed from a more cross-sectional perspective. Non-longitudinal samples mean that most units will have single-observation time series. The levels of variables for those parts of the target calendar period that are not covered by available fiscal periods must then be extrapolated using sub-annual indicator series, in a kind of perpetual first-phase calendarization.

Unfortunately, the revision which can eventually be obtained in the second-phase calendarization of a multiple-observation series is not directly available for a single-observation series. However, parallel calendarization using multiple-observation administrative series may provide a solution. This would entail linking from the survey unit to its administrative counterparts, obtaining and calendarizing the administrative series of the latter units, and then applying the resultant matrix of calendarization weights back to the single observation. A strategy for carrying out such parallel calendarization using administrative series is being developed; and another strategy for the second-phase calendarization of extinct variables using related variables' movements is being contemplated.

3.4 Constant Business Structure

The Cholette-Denton calendarization methodology is best applied to business structures that are constant over time, e.g. where the same number of establishments is present in each successive observation of a given enterprise. Nevertheless, because growth by acquisition can be deemed another dimension of growth, besides inherent growth, this pre-requisite may permit some flexibility. However, there may be some methodological implications, including a requirement that acquisitions be reflected with an abrupt instead of smooth change in the interpolated series' level.

If a business is included in both sub-annual and annual surveys, the data from the former could be used to calendarize the latter. However, the business structure would then have to be accounted for explicitly. This is since the annual fiscal observation would be as per the structure at the end of that fiscal period, whereas the sub-annual series would record the activity of a series of business structures. The problem is akin to that of constructing longitudinal enterprise panels. Lavallée and Struijs (1998) may provide the basis of a solution, although it only deals with a single structural transition.

An important unresolved question is that of which structure should form the standard basis. Should it be the structure as of the end of the target calendar year, that as of the end of the fiscal period, the structure that lasted the longest in the fiscal period or the calendar period, or the structure which spanned the longest duration in the overlap between the fiscal period and the calendar year? The need to have a longitudinally constant business structure, and the fact that allocating data from a higher-level structure to an elemental level has a temporal aspect, together imply that there are some commonalities between allocation and calendarization, and that an integrated approach may be fruitful.

3.5 Stable Variable Content

The Cholette-Denton calendarization methodology is best applied to variables that are measured throughout the time span of interest, i.e. to business data for which the record layout does not change over time. However, questionnaires and the sets of metadata that they cover may change over time. For example, three detail-level revenue variables in one year may correspond to two in the next, and four in

the year after that. For calendarization purposes, a series should refer to the same content throughout. Thus, and in this example, one would have to consider temporarily aggregating each year's variables to a common basis, calendarizing this continuous, but ad hoc series, and then balancing each year's uncalendarized detail against the corresponding calendarized value of the ad hoc series. An algorithm is being developed for determining lowest common denominator continuous content from variable mappings.

4. Conclusions and Intended Future Developments

Calendarization can improve the quality of business survey data by introducing consistency both over time and across industries, while making the data become a better representation of the calendar period concept that they are usually purported to represent. Calendarization technology is being adopted at Statistics Canada, and has already been applied to the 1997 Canadian aquaculture and Canadian capital expenditures data.

However, implementing calendarization in a business survey context does pose certain challenges; for which the following solutions are under development: improved date editing and imputation; parallel calendarization using administrative data; structural reconciliation; and determining lowest common denominator continuous content. Additional investigation is needed in: evaluating and selecting calendarization indicator series; and in measuring the impact of calendarization indicator series on cross-sectional coefficients of variation at the survey estimate level. The integrated calendarization of annual fiscal data and benchmarking of sub-annual data is another goal for the very long term. These two processes have traditionally been separate, even though they share essentially the same data and methodology. Other work for the long term includes the retrospective insertion of "bonus" calendar year observations from one year's sample into the prior year's sample; and integrating calendarization and allocation.

5. Acknowledgements

The author would like to acknowledge the ideas and contributions of the following: George Beelen, Zhao-Guo Chen, Pierre Cholette, Michel Hidiroglou, Stuart Pursey, and Philip Smith. Page constraints preclude further details.

6. References

- Beaucage, Y., and P.S. Tarassoff (1998), "La calendrialisation dans le cadre de l'enquête unifiée sur les entreprises," *1998 Proceedings of the Survey Methods Section, Statistical Society of Canada*, pp.95-101.
- Boot, J.C.G., W. Feibes, and J.H.C. Lisman (1967), "Further Methods of Derivation of Quarterly Figures from Annual Data," *Applied Statistics*, **16**, pp.65-75.
- Cholette, P.A. (2000a), "Interpolating Monthly Values from Data Covering Various Fiscal Periods," Draft paper, Ottawa, Canada: Business Survey Methods Division, Statistics Canada.

Cholette, P.A. (2000b), "Formulae for the Cal-Index [sic] Statistics," Internal note, Ottawa, Canada: Business Survey Methods Division, Statistics Canada.

Cholette, P.A., and Baldwin, A. (1989), "Converting Fiscal Year Data into Calendar Year Values," *Working Paper No. TSRA-89-007E*, Ottawa, Canada: Time Series Research and Analysis Division, Methodology Branch, Statistics Canada.

Cholette, P.A., and Chhab, N.B. (1991), "Converting Aggregates of Weekly Data into Monthly Values," *Applied Statistics*, **40**, pp.411-422.

Cholette, P.A., and J.-L. Coster (1998), "Interpolating Time Series from Administrative Data with Diverse Reference Periods," *Symposium 1997, New Directions in Surveys and Censuses: Proceedings*, Ottawa, Canada: Statistics Canada, pp.233-238.

Cholette, P.A., and J. Higginson (1987), "A Note On Aggregating Data Referring to Different Financial Years," *Working Paper No. TRSA-87-015E*, Ottawa, Canada: Time Series Research and Analysis Division, Methodology Branch, 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*, **66**, pp.99-102.

Lavallée, P., and P. Struijs (1998), "Treatment of Enterprise Restructuring in Business Panels," *1998 Proceedings of the Survey Methods Section, Statistical Society of Canada*, pp.205-212.

Technical Series - Index

September 11, 2000

PIPES has a series of technical paper reprints primarily for internal purposes. A list of the reprints currently available is presented below. For copies, contact Bonnie Bercik at (613) 951-6790 or Diane Proulx at (613) 951-7192, fax number (613) 951-0411 or write to Statistics Canada, 13th Floor, Jean Talon Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

1. Unified Enterprise Statistics Program – Project to Improve Provincial Economic Statistics – May 5, 1997 – PIPES Project Managers.
2. PIPES Evaluation Framework – September 15, 1997 – Philip Smith.
3. Report on the Unified Enterprise Survey & Reporting Arrangements Business Consultations – August 1997 – Guy Gellatly, Larry Murphy and Junior Smith.
4. Update on PIPES Progress: Notes for a Briefing for Federal and Provincial Finance Officials, Halifax, Nova Scotia, March 12, 1997 – Philip Smith.
5. An Overview of The Project to Improve Provincial Economic Statistics – November 1997 – George Beelen, Francine Hardy and Don Royce.
6. Using Databases to Design, Generate and Store Business Questionnaires at Statistics Canada – November 5, 1997 – Alana M. Boltwood.
7. The How and Why of Business Statistics – January 1999 – Elise Mennie. (*Not for external dissemination*)
8. An update on PIPES Fifteen Months into the Project – April 24, 1998 – Philip Smith.
9. Key Provider Manager (KPM) – 1997-98 Annual Report – May 1998 – Vicki Crompton.
10. A Framework for Planning Unified Enterprise Survey Data Collection – October 28, 1998 – Alana Boltwood.
11. Impact of the PIPES Funding on the Services Division Programme and Achievements in 1997-98 – April 1998 – Gordon Baldwin. (*Not for external dissemination*)
12. PIPES Organization and Decision-Making Structure – August 17, 1998 – Philip Smith. (*Not for external dissemination*)
13. The Central Goal of PIPES – November 17, 1997 – Philip Smith.
14. The Terminology and Framework of the Unified Enterprise Questionnaire – Revised March 1999 – Philip Smith.
15. Realizing and Measuring Quality Improvements in Provincial Economic Accounts – August 1998 – Philip Smith.
16. Annual Report 1997-98 – Ombudsman for Small Business Response Burden – July 1998, – Michael Issa. (*Not for external dissemination*)
17. Decision Making in PIPES – October 1, 1998 – Philip Smith.
18. Task Force on Electronic Data Reporting – April, 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. PIPES Information Package – October 1998 – Philip Smith.
20. UES and the Non-Business Sectors – September 17, 1997 – Art Ridgeway.
21. CATS User Guide – April 1998 – Janet Howatson. (*Not for external dissemination*) Not available

22. Report on Collection and Data Capture Operation OID for UES 1997 – September 3, 1998 – Anne Ladouceur. *(Not for external dissemination)*
23. SDD Contribution to PIPES 1998-1999 – September 1998 – Shirley Dolan.
24. The Harmonized Sales Tax Revenue Allocation Formula – August 1998 – Karen Hall. *(Not for external dissemination)*
25. Task Group on Data Acquisition Strategy Report – July 22, 1998 – François Maranda and Don Royce.
26. Roles and Responsibilities in the Unified Enterprise Statistics Program – December 15, 1998 – George Andrusiak, Richard Barnabé, Albert Meguerditchian, Ray Ryan and Philip Smith. *(Not for external dissemination)*
27. Paper on the Project to Improve Provincial Economic Statistics from the Joint IASS/IAOS Conference – July 22, 1998 – Don Royce.
28. Respondent Relations Task Force – March 5, 1999 – Wayne Smith.
29. Response Analysis Follow-up Survey – March 1999 – Kristen Underwood.
30. Data Sharing Information Package – March 1999 – John Crysdale. *(Not for external dissemination)*
31. Coherence Analysis – Case Study from the Key Provider Manager Program – April 23, 1999 – Rachel Bernier and Julie Mandeville.
32. Evaluation of Collection Support Material used during the 1997 Unified Enterprise Survey – November 16, 1998 – Yvèle Paquette.
33. Waiver Information Package – May 1999 – John Crysdale. *(Not for external dissemination)*
34. The PIPES Plan for 1999-00 – June 14, 1999 – Philip Smith. *(Not for external dissemination)*
35. BTS + Forum Post-conference Actions – April 1999 – Cornwall Conference Participants.
36. Report of the Task Force on Sources of Business Information – March 1999 – Vicki Crompton and Mark Marcogliese.
37. Field 5 Task Force Report on Improving Generic Boards – August 1999 – Mel Jones.
38. Study of Business Survey Questionnaires – June 1999 – Jason Gilmore.
39. Complexity Scale for Business Questionnaires – June 1999 – Jason Gilmore.
40. Update on PIPES – September 1999 – Philip Smith.
41. Exclusion Thresholds & Sampling Practices for Business Surveys – Implementation Strategy – September 1999 – Implementation Strategy Team.
42. Use of Tax Data in the Production of Provincial Economic Statistics – October 1999 – Peter Bissett.
43. Data Quality Survey 1996 – March 1999 – Ed Bunko. *(Not for external dissemination)*
44. Estimates of Information Cost to Business Respondents, 1998 – September 16, 1999 – Linda Grant and Michael Issa. *(Not for external dissemination)*
45. Data Security Task Force – January 1999 – Mark Steski. *(Not for external dissemination)*
46. Key Provider Manager Program – 1998-99 Program Report – June 1999 – Monique Gaudreau.
47. Treatment of Head offices and other support units in the UES – October 1999 – Charles Delorme.

48. UES Walkthrough – December 6, 1999 – Philip Smith.
49. The UES Strategy to Combine Enterprise, Establishment and Legal-Entity Data – November 29, 1999 – Philip Smith.
50. The Data Quality of the 1997 Unified Enterprise Survey (UES) Pilot – January 10, 2000 – Stuart Pursey.
51. Options for the Survey of Household Spending – January 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. (*Not for external dissemination*) AWAITING APPROVAL BY MIKE SHERIDAN
52. PIPES Budget Targets for 2001-02 and Beyond – March 28, 2000 – Project Management Team. (*Not for external dissemination*)
53. Unincorporated Business Statistics from the Tax Estimates Program – March 8, 2000 – Philip Smith.
54. PIPES and the Use of Statistics for Administrative Purposes – April 12, 2000 – Philip Smith.
55. Information Manual for Interviewers on the 1999 Survey of the Construction Industry – February 2000 – Claude Grenier. (*Not for external dissemination*)
56. The Conceptual and Practical Evolution of the Unified Enterprise Survey Integrated Questionnaire – May 2000 – Mario A. Vella and Annette Laurent.
57. Statistics Canada's Broad Strategy for Business Statistics – May 25, 2000 – Philip Smith.
58. Calendarizing Business Survey Data – June 2000 – Peter S. Tarassoff.

Other documents

PIPES Project Structure – Revised July 1999
Confidentiality of Statistical Data – April 1995 – D. Binder and L. Desramaux
The Business Register Hellerman Autocoder Project – November 1999 – Business Register Division
(English version only)
The Why of Business Surveys – March 2000

La calendrialisation des données des enquêtes-entreprises

La calendrialisation est la transformation de données se rapportant à une année fiscale en données correspondant à l'année civile. La méthodologie de la calendrialisation est assez bien documentée, mais sa mise en œuvre ne l'est pas. Le présent exposé passe en revue certains aspects pratiques de l'application de la calendrialisation aux données des enquêtes-entreprises annuelles.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Pour plus de renseignements sur ce document, veuillez communiquer
avec Bonnie Bercik (613) 951-6790 ou Diane Proulx (613) 951-7192
Télécopieur : (613) 951-0411

La calendrialisation des données des enquêtes-entreprises

Peter S. Tarassoff, Statistique Canada
Immeuble Principal 1005, Ottawa (Ontario) Canada K1A 0T6
Peter.Tarassoff@statcan.ca

Juin 2000

Mots clés : Annualisation, année civile, exercice financier, données réparties de façon irrégulière, désagrégation temporelle, étalonnage

1. Introduction

Il existe une documentation théorique relativement abondante sur l'utilisation de la méthode proposée par Denton (1971) ou par Boot et coll. (1967) pour étalonner les séries chronologiques et calendrialiser les données liées à un exercice financier (p. ex. Cholette, 2000a; Cholette et Baldwin, 1989; Cholette et Chhab, 1991; Cholette et Coster, 1998). Toutefois, exception faite pour Beaucage et Tarassoff (1998), on a peu traité de l'application de la calendrialisation aux données concrètes d'une enquête-entreprise. L'Enquête unifiée auprès des entreprises (EUE) de Statistique Canada a représenté la première grande application productive de la méthode de calendrialisation de Pierre Cholette, fondée sur Denton, et elle a servi à la mise à l'essai rigoureuse de la méthode de calendrialisation et du module SAS qui la met en oeuvre. Sans vouloir répéter l'exposition formelle et théorique de la méthodologie, l'auteur traite des aspects pratiques de la calendrialisation des données des enquêtes-entreprises.

La calendrialisation est la transformation de données se rapportant à une année fiscale en données correspondant à l'année civile. Contrairement à l'exercice financier, l'année civile est uniforme au fil des ans aussi bien que pour différentes unités statistiques et différentes branches d'activité. La calendrialisation tient compte explicitement de ce qui suit : premièrement, de ce que les données statistiques sont généralement publiées en fonction d'une certaine période de l'année civile, deuxièmement de ce que l'exercice financier des répondants diffère souvent de l'année civile et, troisièmement, de ce que le niveau des variables quantitatives a tendance à changer au fil des ans.

L'autre possibilité, qui consiste à laisser inchangées les données liées à une période autre que l'année civile, peut être une source de biais, tant pour ce qui est des estimations de niveau (Cholette et Coster, 1998; Cholette et Higginson, 1987) que pour ce qui est des estimations des changements d'une période de l'année civile à l'autre. Ce genre de biais peut également être conçu comme le fait de mal situer le niveau dans le temps, de sorte que l'on risque, par exemple, de manquer les tournants d'un cycle économique. D'autres difficultés peuvent survenir si la combinaison d'exercices change au fil des ans ou diffère d'une branche d'activité, d'un secteur ou d'une région géographique à l'autre. Dans le premier cas, une série chronologique d'estimations fondées sur une période autre que l'année civile sera répartie de façon irrégulière dans la mesure où la combinaison d'exercices aura changé. Dans le deuxième cas, les estimations liées à une période autre que l'année civile ne seront pas tout à fait comparables d'une branche d'activité, d'un secteur ou d'une région à l'autre. Ce dernier cas risque d'entraîner des difficultés de rapprochement des flux interrégionaux et intersectoriels (p. ex. entre un secteur de services aux entreprises et un secteur de clients manufacturiers) ou encore des erreurs systématiques et diachroniques dans les flux intergouvernementaux.

Enfin, certaines instances (p. ex. la France) et certains organismes exigent que l'exercice financier des entreprises se fonde sur l'année civile, ou que les répondants présentent leurs données en fonction de l'année civile, même si leur exercice se fonde sur une période autre que l'année civile. On peut considérer ces possibilités comme une solution de rechange à la calendrialisation, mais la dernière possibilité comporte plusieurs inconvénients. Elle impose un fardeau de réponse et la qualité des données résultantes serait douteuse. On ne peut pas garantir que les répondants se donneront la peine de rajuster leurs données, ni que cela se fera de façon cohérente ou correcte. De plus, on ne peut pas facilement vérifier la réalité, la qualité ni l'uniformité des transformations. Tel qu'on pourrait le deviner, il n'est pas courant au Canada d'exiger que les réponses se fondent sur l'année civile. Néanmoins, les lois canadiennes de l'impôt sur le revenu exigent que les nouvelles entreprises non constituées en société adoptent un exercice financier fondé sur l'année civile. De plus, certaines branches d'activité canadiennes comportent des exercices annuels uniformes mais non fondés sur l'année civile. Par exemple, le secteur bancaire comporte un exercice se terminant le 31 octobre, tandis que l'administration fédérale canadienne comporte une année financière se terminant le 31 mars.

2. Aperçu de la calendrialisation

2.1 La « fiscalité », le « Calindex » et le diagnostic du besoin de calendrialisation

La « fiscalité » peut se définir comme la présence de périodes comptables autres que l'année civile dans un ensemble de données, c'est-à-dire comme le manque de « calendrialité ». Les raffinements d'une telle définition tiennent compte non seulement de la présence de périodes comptables autres que l'année civile, mais également de la mesure dans laquelle ces périodes diffèrent de l'année civile. Peu importe comment elle est mesurée, une fiscalité élevée n'est pas suffisante pour indiquer qu'une calendrialisation est nécessaire, pas plus d'ailleurs qu'un accroissement appréciable des variables clés pertinentes. C'est plutôt la combinaison de ces deux phénomènes qui indique le besoin de calendrialisation.

Le « Calindex » représente une mesure appropriée de la « fiscalité ». Chaque variante de Calindex dépend du concept adopté pour l'année de référence, c'est-à-dire de la correspondance applicable entre un exercice financier donné et l'année civile cible, mais il est possible de calculer un Calindex généralisé comme suit (Cholette, 2000b) :

$$\text{Calindex} = (\sum_n |d_n| r_n) / M;$$

où M est le nombre maximal de mois de la période civile cible, soit 12, d_n est le déplacement en mois entre la fin de l'exercice et la fin de l'année civile correspondante et r_n est la proportion d'unités (ou de l'activité des unités) dans la classe de déplacement d_n . Parmi tous les concepts d'année de référence, le Calindex comporte une gamme allant de 0 (zéro) à un peu plus de 1 (un), 0 (zéro) indiquant l'absence de fiscalité et 1 (un) indiquant une fiscalité appréciable.

À noter que d_n et r_n dépendent de la correspondance établie entre les exercices et les années civiles, c'est-à-dire de la notion d'*« année de référence »*. À une extrémité, et suivant un concept d'année de référence semblable à celui de l'EUE, où un exercice correspond à l'année civile durant laquelle il prend fin, le Calindex et la fiscalité apparente sont maximisés. À l'autre extrémité, une année de référence faisant correspondre les exercices aux années civiles pour lesquelles il y a le plus de chevauchement aurait tendance à minimiser le Calindex, et donc la fiscalité vraisemblable. Ainsi, certains concepts d'année de référence comportent une correspondance plus étroite entre l'exercice et l'année civile. De même, il est important de calculer le Calindex en fonction du même concept d'année de référence lorsque l'on compare la fiscalité d'une branche d'activité ou d'une région géographique à l'autre. Les principales limites du Calindex sont qu'il ne tient pas compte de la durée de l'exercice et qu'il laisse de côté les mouvements de la tendance cyclique.

Le tableau 1 résume, en fonction du concept d'année de référence de l'EUE, les valeurs Calindex pondérées selon le revenu pour certains groupes du SCIAN, pour tout le Canada, de même que pour les provinces comportant des valeurs Calindex minimales et maximales. On y trouve également les dates moyennes pondérées de début et de fin d'exercice selon le revenu et l'effet éventuel de la calendrialisation, à supposer que la série du PIB aux prix du marché soit utilisée comme indicateur de la calendrialisation. Bien sûr, le degré de fiscalité peut varier selon le secteur et la région géographique. Par conséquent, si l'on envisage des estimations à un niveau autre que celui de l'ensemble des branches d'activité à l'échelle nationale, on devrait prévoir une évaluation de la « fiscalité » à ce niveau plus détaillé.

TABLEAU 1. Les Calindex, les dates moyennes pondérées de début et de fin d'exercice et l'effet prévu de la calendrialisation pour les échantillons de l'EUE selon certains groupes du SCIAN

Groupe du SCIAN		Val. extr. prov. du Calindex		Calindex pour tout le Canada	Dates moyennes pondérées de début et de fin d'exercice		Différence année civile – exercice %
Code	Descripteur	Min.	Max.		Début	Fin	
1125	Aquaculture animale	0,07	0,73	0,19	19/10/96	18/10/97	1,12 %
2312	Construction de bâtiments	0,13	0,55	0,32	19/08/96	19/08/97	1,50 %
2313	Travaux de génie	0,22	0,62	0,32	15/08/96	15/08/97	1,54 %
2321	Préparation du terrain	0,03	0,57	0,36	25/08/96	25/08/97	1,43 %
2325	Installation d'équipements techniques	0,30	0,68	0,36	06/08/96	04/08/97	2,24 %
4853	Services de taxi et de limousine	0,00	0,69	0,33	13/09/96	11/09/97	1,82 %
4921	Messageries	0,12	0,55	0,23	16/10/96	15/10/97	1,16 %
5311	Bailleurs de biens immobiliers	0,11	0,76	0,26	06/10/96	04/10/97	1,56 %
5312	Bureaux d'agents et de courtiers immobiliers	0,00	0,50	0,34	07/09/96	06/09/97	1,59 %
7221	Restaurants à service complet	0,23	0,51	0,43	08/08/96	03/08/97	3,12 %

Peu importe ce que le Calindex peut indiquer, l'effet prévu de la calendrialisation (EI) se calcule comme suit :

$$EI = (EC - EF) 100 / EF;$$

où EC, c'est-à-dire la valeur prévue de l'année civile, représente la somme des observations de l'indicateur infra-annuel pour l'année civile cible, et EF, c'est-à-dire la valeur prévue de l'exercice, représente la somme des observations du même indicateur infra-annuel pour l'exercice défini par les dates moyennes pondérées de début et de fin de l'exercice. Un tel calcul presuppose que l'on applique une calendrialisation proportionnelle plutôt qu'additive et que la série indicatrice soit appropriée. À noter que l'EI comporterait une certaine erreur, car on perd beaucoup d'information sur les périodes comptables lorsque l'on calcule les dates moyennes. Enfin, l'EF, et donc l'EI, se laissent améliorer quelque peu si l'on a recours à la modification de Tarassoff (voir ci-dessous) de façon que la somme des indicateurs comprend l'activité des mois partiels.

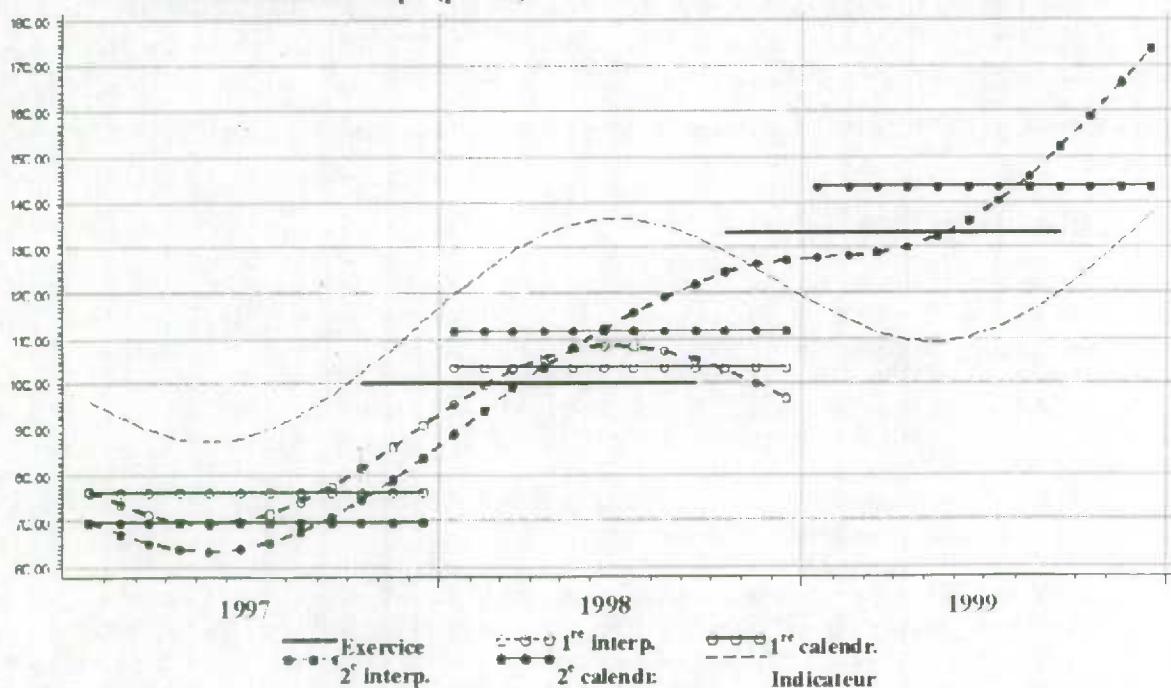
L'EI dépend beaucoup de la série indicatrice. En effet, la calendrialisation réelle des données sur l'aquaculture pour l'année de référence 1997 a donné un effet allant de -21,53 % à +3,80 %, selon la province. Bien que l'effet de la calendrialisation soit souvent de l'ordre de quelques points de pourcentage, la transformation de la calendrialisation est orthogonale relativement à d'autres sources d'erreur. Par conséquent, compte tenu des ressources disponibles, il est probablement mieux de calendrialiser que de ne pas calendrialiser.

2.2 La calendrialisation de Cholette-Denton

En termes non mathématiques, la calendrialisation de Cholette-Denton consiste à ajuster une série d'interpolations infra-annuelles à des jalons de période comptable, puis à agréger les séries interpolées relativement aux périodes civiles cibles. La série infra-annuelle peut être préparée soit directement à partir des jalons eux-mêmes, soit en forçant une série indicatrice infra-annuelle à correspondre aux jalons. La méthode est telle que les interpolations infra-annuelles suivent le mouvement de l'indicateur, surtout aux extrémités. La méthode permet de traiter les variables de stock aussi bien que de flux. Grâce à la modification de Tarassoff qui consiste à utiliser des fractions tirées des poids de jours ouvrables afin de répartir les valeurs de période civile, il est désormais possible de traiter des exercices dont le début ou la fin ne correspond pas aux bouts d'une période civile, donc des données réparties de façon irrégulière (Cholette, 2000).

Dans le cadre du traitement d'une enquête, il existe deux étapes notionnelles de calendrialisation pour les données de chaque année de référence, comme l'indique la figure 1. À la première étape, l'année de référence en cours est également l'année de référence la plus récente, de sorte que le jalon de l'exercice ne recouvre pas nécessairement toute l'année civile cible. La série d'interpolations infra-annuelles doit donc être extrapolée en fonction des mouvements de la série indicatrice infra-annuelle, ou encore en fonction d'un ensemble de paramètres de convergence. La deuxième étape a lieu lorsque le prochain jalon d'exercice recouvrant l'année civile devient disponible. À cette étape, la calendrialisation dépend beaucoup moins de la série indicatrice. En résumé, tandis que la première étape permet de produire des données calendrialisées provisoires de façon relativement périfiable, la deuxième étape suppose une révision moins périfiable. Toutefois, l'actualité des données de la calendrialisation de la première étape est obtenue au prix d'une dépendance à l'égard d'une série indicatrice n'ayant possiblement qu'une valeur d'approximation.

FIGURE 1. Interpolations et valeurs calendrialisées obtenues à la première étape (cercles) et à la deuxième étape (points).



Nota : À des fins de représentation graphique, les données sont divisées par le nombre de mois couverts.

Les indicateurs de la calendrialisation indiquent le mouvement qui se reflète dans les interpolations infra-annuelles, et ils ont une importance particulière lors de la calendrialisation de la première étape. Les six caractéristiques souhaitables des indicateurs sont les suivantes : le fait de franchir l'union des périodes de l'exercice et de l'année civile en question; le fait d'inclure la tendance cyclique et les composantes saisonnières; le fait de refléter le même contenu que les variables à calendrialiser; la finesse de la résolution temporelle; l'étroitesse du rapport avec les entreprises à calendrialiser ou du moins avec les combinaisons secteur-région pour lesquelles il s'agit de préparer des estimations.

Malheureusement, la disponibilité des séries indicatrices infra-annuelles est limitée, et il faut souvent avoir recours à des substituts. Il est donc possible que certaines séries indicatrices aient une pertinence inconnue ou douteuse. Une façon d'évaluer les séries indicatrices infra-annuelles serait de régresser leurs séries d'agrégations annuelles contre des séries annuelles correspondantes pour les unités qu'il s'agit de calendrialiser, bien que la meilleure solution soit d'assurer la calendrialisation de la deuxième étape et sa révision correspondante.

Certaines variables peuvent être jugées relativement constantes au fil des ans, contrairement aux séries indicatrices pertinentes, et ne seront donc peut-être pas transformées par calendrialisation. De même, différents indicateurs peuvent être utilisés pour diverses variables. Compte tenu de ces deux faits, un calage postérieur à la calendrialisation est essentiel si l'on veut s'assurer de l'addition convenable des variables.

Avant toute application de contraintes de période cible, la mise en œuvre de la méthode de Cholette-Denton par Cholette dans le langage SAS produit un enregistrement pour chaque période civile chevauchée par l'exercice, même si l'unité en question ne fait pas partie de l'échantillon pour les deux périodes de référence. Même si ces enregistrements « primés » ne sont pas utilisés actuellement pour l'échantillonnage et l'estimation, ils offrent tout de même une occasion d'accroissement des échantillons de périodes de référence adjacentes.

3. Préalables de la calendrialisation, problèmes de données d'enquête et solutions

3.1 Microdonnées et macrodonnées

La méthode de calendrialisation de Cholette-Denton s'applique de façon optimale aux *microdonnées*. Malgré la pratique bien répandue de l'étalonnage des *macrodonnées*, et malgré la vraisemblable efficacité du traitement de séries moins nombreuses à un niveau d'agrégation plus élevé, cette façon de procéder peut entraîner des biais. Ceux-ci résulteraient du fait que, puisque la plupart des *macrodonnées* représentent des agrégations de diverses périodes comptables, les *macrodonnées* ne peuvent être délimitées que très rarement par une seule paire de dates. Même si l'on réussissait à délimiter les *macrodonnées* à l'aide des tendances centrales des répartitions des dates de début et de fin d'exercice, on aurait les mêmes erreurs que ci-dessus relativement à l'effet prévu de la calendrialisation. Encore une fois, cela s'expliquerait par la perte d'information au sujet d'exercices particuliers résultant du calcul des dates moyennes.

3.2 Délimitation des exercices

Les exercices se rapportant aux données à calendrialiser devraient être délimités de façon que chaque jalon d'exercice comporte une date de début et de fin, et que ces dates soient mesurées avec précision (au jour près) et sans erreur. Les périodes d'activité saisonnière qui diffèrent de l'exercice devraient également être bien délimitées, tout comme les périodes d'interruption temporaire appréciable.

Les dates d'exercice représentent le paramètre le plus important de la calendrialisation, car elles indiquent au cours de quelle période l'activité en question a eu lieu dans le cas d'une variable de flux, ou à quel moment un niveau quelconque a été atteint dans le cas d'une variable de stock. La surestimation ou la sous-estimation de la durée d'une activité résultant de dates manquantes ou incorrectes peut avoir un effet appréciable sur la calendrialisation. Lorsque la calendrialisation est un objectif, les dates ont donc une plus grande importance et ne jouent plus simplement le rôle d'indiquer si une réponse d'enquête donnée respecte la période de référence.

Par conséquent, il importe de vérifier et d'imputer de façon plus rigoureuse les champs liés aux dates. En plus des valeurs manquantes, il faut surveiller des erreurs comme les dates qui à l'origine ne relevaient pas de la période de référence et qui ont été « corrigées » de façon à correspondre à cette période, les chiffres de date ambigus des questionnaires, les corrections de date ambiguës apportées par les répondants, les dates de période saisonnière déclarées comme des dates d'exercice annuel, de même que les jours et les mois qui sont inversés lors de la saisie des données ou par les répondants eux-mêmes. Bien sûr, les données démographiques sur les entreprises sont également une source utile de renseignements sur les événements pouvant modifier les régimes de date.

Parfois, les unités d'échantillonnage peuvent différer des unités statistiques, par exemple lorsqu'un questionnaire unique englobe un groupe d'établissements. Le cas échéant, les dates de l'exercice figurant dans le questionnaire s'appliquent vraisemblablement à la plupart des établissements du groupe, mais il est possible que certains établissements comportent des exercices différents qui ne sont pas mesurés. Une possibilité serait d'inclure des champs de dates d'exercice pour toute feuille de travail supplémentaire annexée en vue de la répartition de données de l'entité de collecte au niveau de l'établissement.

Habituellement, des dates d'événement comme la date de vente devraient servir à limiter la période civile cible, afin que les interpolations et les extrapolations ne dépassent pas la ou les périodes au cours desquelles l'unité a dû exister ou être active. Toutefois, cela s'applique également aux interactions de type calendrialisation-estimation. Certains événements ne peuvent pas être traités de façon convenable par stratification *a posteriori* (p. ex. la vente des établissements d'une entreprise à tirage complet au cours de l'exercice). Compte tenu du délai type qui accompagne la mise à jour d'une base de sondage, ou même malgré ce délai, les établissements vendus dont l'identité ne change pas continuent d'exister, mais sans se prêter à un échantillonnage et sans influencer les chiffres d'une stratification *a posteriori*. En l'absence de calendrialisation, leurs exercices tronqués entraîneraient une sous-estimation de l'activité commerciale. En présence de calendrialisation, par contre, leur activité pourrait être répartie au-delà de la date de vente.

3.3 Durée des séries chronologiques et calendrialisation parallèle

La méthode de calendrialisation de Cholette-Denton s'applique de façon optimale à des séries chronologiques comportant au moins deux observations consécutives. Idéalement, chaque période civile cible devrait être couverte entièrement par des observations d'exercices adjacents. Toutefois, même si elles doivent à terme fournir des séries chronologiques de macrodonnées, plusieurs enquêtes-entreprises sont en réalité conçues et traitées d'un point de vue plus transversal. Dans le cas d'échantillons non longitudinaux, la plupart des unités sont liées à des séries chronologiques à observation unique. Le niveau des variables pour les parties de la période civile cible qui ne sont pas couvertes par les exercices disponibles doit alors être extrapolé à l'aide de séries indicatrices infra-annuelles, dans le cadre d'une espèce de calendrialisation de première étape perpétuelle.

Malheureusement, la révision que l'on peut obtenir au cours de la deuxième étape d'une calendrialisation pour une série à plusieurs observations n'est pas directement disponible pour une série à une seule observation. Toutefois, une calendrialisation parallèle axée sur des séries administratives à plusieurs

observations pourrait représenter une solution. Il s'agirait alors d'assurer le couplage de l'unité d'enquête et de ses homologues administratifs, d'obtenir et de calendrialiser les séries administratives de ces dernières unités, puis d'appliquer la matrice résultante des poids de calendrialisation à l'observation unique. Une stratégie permettant d'effectuer une telle calendrialisation parallèle à l'aide de séries administratives est en voie d'élaboration; une autre stratégie axée sur la deuxième étape d'une calendrialisation des variables disparues à l'aide des mouvements de variables connexes est également envisagée.

3.4 Structure constante des entreprises

La méthode de calendrialisation de Cholette-Denton s'applique de façon optimale aux structures d'entreprise qui sont constantes au fil des ans, le même nombre d'établissements, par exemple, étant présent pour chaque observation successive d'une entreprise donnée. Néanmoins, puisqu'une acquisition peut être considérée comme une autre dimension de la croissance, à part la croissance inhérente, ce préalable peut offrir une certaine souplesse. Par contre, cela pourrait entraîner certaines répercussions méthodologiques, y compris une exigence voulant que les acquisitions soient reflétées par un changement de niveau soudain plutôt que graduel pour les séries interpolées.

Si une entreprise relève d'enquêtes à la fois infra-annuelles et annuelles, les données de celles-là peuvent servir à calendrialiser ces dernières. Toutefois, il faudrait alors tenir compte explicitement de la structure de l'entreprise. Il en est ainsi parce que l'observation annuelle pour l'exercice correspondrait à la structure en fin d'exercice, tandis que la série infra-annuelle refléterait l'activité d'une série de structures d'entreprise. Le problème ressemble à celui de l'établissement de panels d'entreprise longitudinaux. Lavallée et Struijs (1998) peuvent servir de base à une solution, bien qu'il ne s'agisse que d'une seule transition structurelle.

Une question importante qui n'est pas résolue est de savoir quelle structure adopter comme norme. Devrait-il s'agir de la structure à la fin de l'année civile cible, de celle à la fin de l'exercice, de la structure qui a duré le plus longtemps au cours de l'exercice ou de l'année civile, ou encore de la structure qui a marqué le plus long chevauchement entre l'exercice et l'année civile? Le besoin d'une structure d'entreprise constante du point de vue longitudinal et le fait que la répartition de données d'une structure de niveau plus élevé à un niveau élémentaire comporte un aspect temporel laissent supposer qu'il existe des points communs entre la répartition et la calendrialisation, et qu'une stratégie intégrée pourrait être fructueuse.

3.5 Contenu stable des variables

La méthode de calendrialisation de Cholette-Denton s'applique de façon optimale à des variables qui sont mesurées tout au long de l'intervalle de temps en question, c'est-à-dire à des données d'entreprise dont le cliché d'article ne change pas au fil des ans. Toutefois, les questionnaires et les ensembles de métadonnées qu'ils recouvrent peuvent évoluer au fil des ans. Ainsi, trois variables détaillées du revenu pour une année donnée peuvent correspondre à deux variables l'année suivante, et à quatre pour l'année subséquente. Aux fins de la calendrialisation, une série devrait se rapporter au même contenu tout au long du processus. Ainsi, et dans l'exemple ci-dessus, il faudrait envisager l'agrégation provisoire des variables de chaque année en fonction d'une base commune, prévoir la calendrialisation de cette série continue mais ad hoc, puis assurer le calage du détail non calendrialisé de chaque année en fonction de la valeur calendrialisée correspondante de la série ad hoc. On élabore actuellement un algorithme permettant de déterminer le moindre dénominateur commun pour ce qui est du contenu continu, à partir des correspondances entre variables.

4. Conclusions et perspectives d'avenir

La calendrialisation permet d'améliorer la qualité des données d'enquête-entreprise grâce à une certaine uniformité aussi bien au cours du temps que d'une branche d'activité à l'autre, tout en transformant les données en une meilleure représentation du concept de période civile normalement invoquée. Statistique Canada adopte actuellement la calendrialisation, celle-ci ayant déjà été appliquée aux données canadiennes de 1997 sur l'aquaculture et les dépenses en investissement.

Toutefois, la mise en œuvre de la calendrialisation dans le cadre d'une enquête-entreprise présente certains défis pour lesquels on envisage les solutions que voici : un processus amélioré de vérification et d'imputation des dates, une calendrialisation parallèle à l'aide de données administratives, un rapprochement structurel et la détermination du moindre dénominateur commun pour ce qui est du contenu continu. Il y a lieu de poursuivre les recherches pour ce qui est de l'évaluation et de la sélection des séries indicatrices de la calendrialisation, et de la mesure de l'effet des séries indicatrices de la calendrialisation sur les coefficients de variation transversaux au niveau des estimations d'enquête. Un autre but à très long terme est la calendrialisation intégrée de données fiscales annuelles et l'étalonnage de données infra-annuelles. Traditionnellement, on a séparé ces deux processus, même s'ils partagent essentiellement les mêmes données et les mêmes méthodes. D'autres travaux à long terme ont trait à l'insertion rétrospective d'observations (année civile) « primes » de l'échantillon d'une année donnée dans l'échantillon de l'année antérieure, de même qu'à l'intégration de la calendrialisation et de la répartition.

5. Remerciements

L'auteur tient à reconnaître les idées et l'apport de George Beelen, Zhao-Guo Chen, Pierre Cholette, Michel Hidioglu, Stuart Pursey et Philip Smith. Faute d'espace, il est impossible de fournir de plus amples détails.

6. Bibliographie

- Beaucage, Y., and P.S. Tarassoff (1998), "La calendrialisation dans le cadre de l'enquête unifiée sur les entreprises," *1998 Proceedings of the Survey Methods Section, Statistical Society of Canada*, pp.95-101.
- Boot, J.C.G., W. Feibes, and J.H.C. Lisman (1967). "Further Methods of Derivation of Quarterly Figures from Annual Data," *Applied Statistics*, **16**, pp.65-75.
- Cholette, P.A. (2000a), "Interpolating Monthly Values from Data Covering Various Fiscal Periods," Draft paper, Ottawa, Canada: Business Survey Methods Division, Statistics Canada.
- Cholette, P.A. (2000b), "Formulae for the Cal-Index [sic] Statistics," Internal note, Ottawa, Canada: Business Survey Methods Division, Statistics Canada.
- Cholette, P.A., and Baldwin, A. (1989), "Converting Fiscal Year Data into Calendar Year Values," *Working Paper No. TSRA-89-007E*, Ottawa, Canada: Time Series Research and Analysis Division, Methodology Branch, Statistics Canada.
- Cholette, P.A., and Chhab, N.B. (1991), "Converting Aggregates of Weekly Data into Monthly Values," *Applied Statistics*, **40**, pp.411-422.

Cholette, P.A., and J.-L. Coster (1998), "Interpolating Time Series from Administrative Data with Diverse Reference Periods," *Symposium 1997, New Directions in Surveys and Censuses: Proceedings*, Ottawa, Canada: Statistics Canada, pp.233-238.

Cholette, P.A., and J. Higginson (1987), "A Note On Aggregating Data Referring to Different Financial Years," *Working Paper No. TRSA-87-015E*, Ottawa, Canada: Time Series Research and Analysis Division, Methodology Branch, 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*, **66**, pp.99-102.

Lavallée, P., and P. Struijs (1998), "Treatment of Enterprise Restructuring in Business Panels," *1998 Proceedings of the Survey Methods Section, Statistical Society of Canada*, pp.205-212.

Série technique - Index

11 septembre, 2000

Dans le cadre du PASEP, on a réimprimé une série de documents techniques, principalement pour usage interne. Voici la liste des réimpressions disponibles. Pour obtenir des copies communiquez avec Bonnie Bercik au (613) 951-6790 ou Diane Proulx au (613) 951-7192, numéro de télécopieur (613) 951-0411 ou écrire à Statistique Canada, 13^e étage, Immeuble Jean Talon, Parc Tunney, Ottawa, Ontario, K1A 0T6

1. Programme unifié des statistiques sur les entreprises – Projet d'amélioration des statistiques économiques provinciales – le 5 mai 1997 – Programme de Gestionnaire du PASEP.
2. Cadre d'évaluation du PASEP – le 15 septembre 1997 – Philip Smith.
3. Rapport de l'Enquête unifiée sur les entreprises et les modalités de déclaration – Consultations auprès des entreprises – août 1997 – Guy Gellatly, Larry Murphy et Junior Smith.
4. Bilan de l'évolution du PASEP : Notes d'une séance d'information à l'intention des représentants fédéral et provinciaux des finances, qui a eu lieu à Halifax (Nouvelle-Écosse), le 12 mars 1997 – Philip Smith.
5. Aperçu du Projet d'amélioration des statistiques économiques provinciales – novembre 1997 – George Beelen, Francine Hardy et Don Royce.
6. Des bases de données pour la conception, la génération et le stockage des questionnaires-entreprises à Statistique Canada – le 5 novembre 1997 – Alana M. Boltwood.
7. La statistique des entreprises : sa raison d'être – janvier 1999 – Elise Mennie. (*Diffusion interne seulement*)
8. Bilan du PASEP 15 mois après son lancement – le 24 avril 1998 – Philip Smith.
9. Programme des gestionnaires des répondants clés (GRC) – Rapport annuel pour 1997-1998 – mai 1998 – Vicki Crompton.
10. Un cadre de planification de la collecte des données de l'Enquête unifiée sur les entreprises – le 28 octobre 1998 – Alana Boltwood.
11. Répercussions du financement du PASEP sur le programme et les réalisations de la Division des services en 1997-1998 – avril 1998 – Gordon Baldwin. (*Diffusion interne seulement*)
12. L'organisation et la structure décisionnelle du PASEP – le 17 août 1998 – Philip Smith. (*Diffusion interne seulement*)
13. Les buts principaux du PASEP – le 17 novembre 1997 – Philip Smith.
14. Terminologie et cadre de référence du questionnaire de l'Enquête unifiée sur les entreprises – Révisé en mars 1999 – Philip Smith.
15. Amélioration de la qualité des statistiques économiques provinciales et mesure des changements apportés – août 1998 – Philip Smith.
16. Rapport annuel 1997-1998 – Médiateur – Fardeau de réponse de la petite entreprise – juillet 1998 – Michael Issa. (*Diffusion interne seulement*)
17. Le processus décisionnel du PASEP le 1^{er} octobre 1998 – Philip Smith.
18. Groupe de travail sur la déclaration électronique des données (DED) – avril 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. Trousse d'information sur le PASEP – octobre 1998 – Philip Smith.
20. L'EUE et les secteurs non commerciaux – le 17 septembre 1997 – Art Ridgeway.

21. Le guide de l'utilisateur du SASC – avril 1998 – Janet Howatson. (*Diffusion interne seulement*) Non disponible
22. Compte rendu de la collecte et la saisie de données DOI pour l'EUE de 1997 – le 3 septembre 1998 – Anne Ladouceur. (*Diffusion interne seulement*)
23. Contribution prévue de la DDS au PASEP, 1998-1999 – septembre 1998 – Shirley Dolan.
24. La formule de répartition des recettes de la taxe de vente harmonisée – août 1998 – Karen Hall. (*Diffusion interne seulement*)
25. Groupe de travail sur l'acquisition des données auprès des entreprises – le 22 juillet 1998 – François Maranda et Don Royce.
26. Rôles et responsabilités dans le cadre du Programme unifié des statistiques sur les entreprises – le 15 décembre 1998 – George Andrusiak, Richard Bamabé, Albert Meguerditchian, Ray Ryan et Philip Smith. (*Diffusion interne seulement*)
27. Document sur le Projet d'amélioration des statistiques économiques Provinciales de la conférence mixte de l'AISE/AISO – le 22 juillet 1998 – Don Royce.
28. Groupe de travail sur les relations avec les répondants – le 5 mars 1999 – Wayne Smith.
29. Enquête de suivi et d'analyse des réponses – mars 1999 – Kristen Underwood.
30. Dossier d'information sur le partage des données – mars 1999 – John Crysdale. (*Diffusion interne seulement*)
31. Analyse de cohérence – Étude de cas du programme des Gestionnaires des répondants clés – le 23 avril 1999 – Rachel Bernier et Julie Mandeville.
32. Évaluation des documents de soutien de la collecte utilisés durant l'Enquête unifiée sur les entreprises de 1997 – le 16 novembre 1998 – Yvele Paquette.
33. Trousse d'information sur les renonciations – mai 1999 – John Crysdale. (*Diffusion interne seulement*)
34. Plan du PASEP pour 1999-2000 – le 14 juin 1999 – Philip Smith. (*Diffusion interne seulement*)
35. Forum SCE + Mesures de suivi de la conférence – avril 1999 – Participants de la conférence de Cornwall.
36. Rapport du groupe d'étude des sources d'information sur les entreprises – mars 1999 – Vicki Crompton et Mark Marcogliese.
37. Rapport du Groupe de travail du secteur 5 sur l'amélioration de la dotation générique – août 1999 – Mel Jones.
38. Étude des questionnaires des enquêtes auprès des entreprises – juin 1999 – Jason Gilmore.
39. Échelle de complexité des questionnaires sur les entreprises – juin 1999 – Jason Gilmore.
40. Rapport sur l'avancement du PASEP – septembre 1999 – Philip Smith.
41. Seuils d'exclusion et méthodes particulières d'échantillonnage pour les enquêtes-entreprises – Stratégie de mise en oeuvre – septembre 1999 – L'Équipe de la stratégie de la mise en oeuvre.
42. Utilisation des données fiscales pour la production des statistiques économiques provinciales – octobre 1999 – Peter Bissett.
43. Enquête sur la qualité des données de 1996 – mars 1999 – Ed Bunko. (*Diffusion interne seulement*)

44. Estimations des coûts d'information pour les entreprises répondantes, 1998 – le 16 septembre 1999 – Linda Grant et Michael Issa. (*Diffusion interne seulement*)
45. Groupe de travail de la sécurité des données – janvier 1999 – Mark Steski. (*Diffusion interne seulement*)
46. Programme des gestionnaires des répondants clés – Rapport de programme pour 1998-1999 – juin 1999 – Monique Gaudreau.
47. Traitement des sièges sociaux et autres unités de support dans l'EUE – le 28 octobre 1999 – Charles Delorme.
48. Tour d'horizon de l'EUE – le 6 décembre 1999 – Philip Smith.
49. Stratégie de l'EUE en vue de combiner les données au niveau de l'entreprise, de l'établissement et de l'entité légale – le 29 novembre 1999 – Philip Smith.
50. La qualité des données de l'Enquête-pilote unifiée de 1997 auprès des entreprises (EUE) – le 10 janvier 2000 - Stuart Pursey.
51. Options pour l'Enquête sur les dépenses des ménages – janvier 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. (*Diffusion interne seulement*) EN ATTENTE DE L'APPROBATION DE MIKE SHERIDAN
52. Budget projeté du PASEP pour les exercices 2001-2002 et ultérieurs – le 28 mars 2000 – Équipe de gestion de projet. (*Diffusion interne seulement*)
53. Statistiques sur les entreprises non constituées en société du Programme des estimations fiscales – le 8 mars 2000 – Philip Smith.
54. Le PASEP et l'utilisation des statistiques à des fins administratives – le 12 avril 2000 – Philip Smith.
55. Manuel d'information des intervieweurs – Enquête de 1999 sur l'industrie de la construction – février 2000 – Claude Grenier. (*Diffusion interne seulement*)
56. Évolution conceptuelle et pratique du questionnaire intégré de l'Enquête unifiée auprès des entreprises – mai 2000 – Mario A. Vella et Annette Laurent.
57. Stratégie générale de Statistique Canada relativement à la statistique des entreprises – le 25 mai 2000 – Philip Smith.
58. La calendrialisation des données des enquêtes-entreprises – juin 2000 – Peter S. Tarassoff.

Autres documents

La structure du projet du PASEP – Révisé en juillet 1999

La confidentialité des données statistiques – avril 1995 – D. Binder et L. Desramaux

The Business Register Hellerman Autocoder Project – novembre 1999 – Division du registre des entreprises
(version anglaise seulement)

Le pourquoi des enquêtes-entreprises – mars 2000



Projet d'amélioration des statistiques
économiques provinciales

La calendrialisation des données des enquêtes- entreprises

Project to Improve Provincial Economic
Statistics

Calendarizing Business Survey Data

Série technique

Numéro 58

Technical Series

Number 58



Internet : www.statcan.ca
Intranet : <http://pasep>



Statistique
Canada Statistics
Canada

Canadä

STATISTICS CANADA LIBRARY



1010311081

BIBLIOTHÈQUE STATISTIQUE DU CANADA

C. C. S