For Official Use **STD/STESEG(2003)35**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **18-Jun-2003**

English - Or. English

STATISTICS DIRECTORATE

STD/STESEG(2003)35
For Official Use

OECD Short-Term Economic Statistics Expert Group

STESEG TASK FORCE ON DATA PRESENTATION AND SEASONAL ADJUSTMENT

SAMPLING ERROR AND SEASONAL ADJUSTMENT

Paris, 26 - 27 June 2003

Prepared by Monique Graf
Swiss Federal Statistical Office

Submitted to the Expert Group under item 4 of the draft agenda

Disclaimer: This paper reports the views of the author and does not necessarily reflect those of the Swiss Federal Statistical Office.

English - Or. English

JT00146331

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

SAMPLING ERROR AND SEASONAL ADJUSTMENT

Monique Graf, SFSO[1]

June 2003

Introduction

The search for national experiences on the issue of presentation of sampling error estimates in relation with seasonal adjustment has shown that very active research is conducted on the subject. Nevertheless except for aggregate data, no estimates of variance or CV are published in general for short term indexes. This short review will thus have 2 parts: 1. presentation of sampling errors in repeated surveys, and 2. a tentative summary of the research results on the problem of bringing together the frameworks of time series and survey design. This report only contains ideas on the subject, as they can be found from easily accessible published research and papers, and is certainly not exhaustive. The presentation of the precision of yearly aggregates is out the scope of this paper.

Scott and Smith (1974), and Scott, Smith and Jones (1977) distinguish between the primary and secondary analysis of sampling errors. While primary analysis is concerned with the direct estimation of sampling error variance and correlations based on the sampling design, in a secondary analysis a time series model on the error sequence is stipulated.

Presentation of sampling errors in repeated surveys (primary analysis)

Let t be the time index, y_t the observed series, Y_t the corresponding population series and e_t the sampling error. In a design-based perspective an additive model is assumed

$$y_t = Y_t + e_t \qquad (1)$$

and in principle the estimated design-based variances and covariances of the survey errors can be computed. These estimated variances and covariances also determine the variance of the estimate of change and by linearization the variance of a ratio.

The sampling design can be quite complex, incorporating stratification, clusters and rotation groups or renewal procedure. The way the sample is renewed across time has an impact on the autocorrelations of the observed series and thus on the estimated variance of change. Different methods of estimation have been proposed: design-based variances and covariances or correlations, jackknife approximations or Taylor linearization. Many technical problems arise about the way to handle the common and specific parts of the samples taken at different time points in the computation of the design-based covariance. Four approaches are found.

[1] Disclaimer: This paper reports the views of the author and does not necessarily reflect those of the Swiss Federal Statistical Office.

1. Chambers et al. (2000) introduce different formulae regarding the computation of the standard deviation of the Average Earning Index in the UK and estimates of change. The difference relates to whether the finite population correction factor is computed from the size of the common sample on two occasions, or of the sample of those units at one occasion that were alive on the other occasion. If the series are broken down by strata or domains, there are occasionally not enough data to compute variances or covariances directly. This problem can be approached by imputation. Problems can arise with non-positiveness of the resulting covariance matrix. If sampling coefficients of variation had to be provided for each individual series, the amount of information would be overwhelming on the one hand and of a doubtful quality for the small strata or domains on the other hand. Therefore this approach would lead to the publication of sampling coefficient of variation for aggregates. Kokic (1998, 2000) gives bootstrap estimates for the precision of the estimator of change for the UK Index of Production (see also his references). According to the 2003 paper « Short Term Output Indicators Review (STOIR), implementation plan »[2] the ONS has deferred the regular production of sampling error measure for the IoP, estimates of change and seasonally adjusted indicators.

2. The Canadian Labour Force Survey is published on a CD along with tables of the 3rd quartile of CV's by area (Canada or Province) and size measure of the estimate for a total. The estimate of the variability of change between times 1 and 2 is given by a simple approximate formula :

$$CV(y_1-y_2) = (1-\rho)^{1/2}[y_1^2 CV(y_1)^2 + y_2^2 CV(y_2)^2]^{1/2} / |y_1-y_2| \qquad (2)$$

where ρ is the sample overlap between the 2 periods. Words of caution are given about the approximate nature of the provided CV's.

3. The exact computation of sampling variances using design based formulae can be intractable, so that replication methods are of advantage. This approach is also taken by the U.S. Bureau of Labor Statistics for the Current Population Survey (Evans et al., 1993)[3]. Because the resulting standard errors can be very variable, the method used at BLS is to provide the user with an estimating (smoothing) equation that represents the observed coefficients of variation as a function of stratum or domain information and level of the series. To help the user summary tables of the fitted values are provided. These tables are published with warnings about the approximate nature of the deduced CV's. The time series approach of Bell and Hillmer (1990) is also applied to the CPS (see below).

4. The computation of sampling variances and covariances is also largely treated in the literature on panel data. In this approach the succession of observations at the unit level is considered, and the dependence structure is modelled by a hierarchical model with fixed and random effects. Pfefferman (1998), Feder, Nathan and Pfeffermann (2000) relate the panel approach to the time series approach (see below).

Time series approach of sampling errors (secondary analysis)

When series are viewed at a detailed level resulting in relatively small sample size, large sampling errors can occur. Several authors have investigated the use of time series modeling and signal extraction to borrow information over time for improving the estimates. As far as we could find out, these very

[2] Short Term Output Indicators Review (STOIR) (2000), White paper *Building Trust in Statistics*, ONS and the 2003 implementation plan. http://www.statistics.gov.uk/methods_quality/qualityreview/downloads/STOIR2000_econ.pdf

[3] See also BLS Handbook of Methods (1997), chap 1 and 4.

promising but complex methods are not yet applied routinely by statistical agencies. Nevertheless numerous research papers contain applications to real series like LFS.

If the sampling design can be supposed to be uninformative for the time series model, e.g. the error process { e_t } is of constant variance over time and uncorrelated with the true signal { Y_t }, then the observed series is design unbiased for { Y_t } (Bell and Hillmer, 1987). This hypothesis is in many instances not very realistic. Bell and Hillmer (1990) introduce a variance stabilizing transformation (log.) and approximate the variance of log(y_t) by CV$(y_t)^2$. Assuming e_t has mean 0 and does not require differencing, they observe that { Y_t } and { y_t } will require the same differencing order and have the same mean function. Thus the ARIMA model obtained for { y_t } ignoring the sampling errors, dictates the form of the model for { Y_t }. This is the basic assumption underlying the U.S. Census Bureau's REGCMPNT program. Nguyen et al. (2002) describe the methodology and apply it to the monthly series "Value of construction Put in Place". Binder and Dick (1990) assume a heteroscedastic error process. The time-dependent variances γ_t are estimated by the design-based methods and e_t is represented by

$$e_t = \gamma_t^{1/2} e^*_t \qquad (3)$$

where { e^*_t } is second-order stationary and modeled by an ARMA process. They then use a Kalman filter to obtain a smoothed version of { y_t } on which the seasonal adjustment procedure is applied. They apply their method to data from the Canadian Labour Force Survey.

Bell and Kramer (1999) start with a similar approach and extend the variance results to the X-11 seasonally adjusted series.

Sutcliffe (1993) discusses the effects of (additive) sampling error on the multiplicative X-11 seasonal adjustment with application to the Australian Monthly Retail Turnover. McLaren and Steel (1997) study the effects of different rotation patterns on the sampling variance of X-11 seasonally adjusted and trend. They concentrate on monthly labour force surveys and suppose that the variance of the sampling errors is constant over time. They assume that the autocorrelations of the sampling errors depend on the number of dwellings per selected first stage unit, the overlap factor, the autocorrelation for the same dwelling and the autocorrelation between dwellings within the same first stage unit. The last 2 autocorrelation functions were obtained from the Australian Labour Force Survey for the proportion of persons employed and unemployed.

Pfefferman (1989) develops state space models for repeated survey data, using a basic structural model and takes design features into account when estimating the seasonal effects. Feder, Nathan and Pfeffermann (2000) pursue the approach. They propose a multilevel model (households and individuals in their application to the quarterly Israel Labor Force Survey) with fixed and random effects. Weights (inverse inclusion probabilities at the 2 levels) are introduced to account for non-ignorable sampling design. Next the authors model the time series relationships of the coefficients and random effects. The seasonal adjustment problem is not considered in this paper.

Discussion

We can regard the SA series as a series of weighted sums of elements from the original series. (The weights are determined by the seasonal adjustment procedure). In order to compute a sampling variance for the SA series, we see that we not only need the variances of the original series, but also all the covariances between values at different times. The variance of change between two successive periods is not sufficient. The aim of the primary analysis is to estimate these sampling variances and covariances. If the covariance matrix of values of the original series is available then in principle the variance of the SA series can be computed.

In the secondary analysis, the sampling errors are treated as another unobserved component and filtered out. The resulting smoothed original series is seasonally adjusted in a second step. In this approach, there is ideally no sampling error left in the smoothed original series, thus there is no sampling variance of the SA series. The time series approach (secondary analysis) is to the author's knowledge, for the time being only experimental.

The two approaches raise an interesting alternative: in a primary analysis the aim to incorporate information about the precision of the original series, whereas in a secondary analysis time series techniques are used to try to filter out the sampling errors.

Conclusion

The topic relating the sampling variance of a repeated survey and the seasonal adjustment is extremely interesting, but technically demanding. So far the actual presentation for users of sampling errors in this context seems to confine to the original series (not seasonally adjusted).

The conclusion is that no recommendations can be made for the time being on the dissemination of sampling error estimates in relation with seasonal adjustment and that more work is needed on the subject.

REFERENCES

Bell, William R., and Hillmer, Steven C. (1990), "The time series approach to estimation for repeated surveys", *Survey Methodology*, **16**, 195-215

Bell, William R., and Kramer, Matthew (1999), "Toward variances for X-11 seasonal adjustments", *Survey Methodology*, **25**, 13-29

Binder, David A., and Dick, J. Peter (1989), "Implications of survey designs for estimating seasonal ARIMA models", *ASA Proceedings of the Section on Survey Research Methods*, 11-15

Binder, David A., and Dick, J. Peter (1990), "A method for the analysis of seasonal ARIMA models", *Survey Methodology*, **16**, 239-253

Chamber, R., Weale, M., and Youll, R. (2000), "The average earnings index", *The Economic Journal*, **110**, 461, 100-121

Cleveland, William P. (2002), "Estimated variance of seasonally adjusted series", *ASA Proceedings of the Joint Statistical Meetings*, 578-584

Cholette, P.A et Dagum, E. B. (1994). "Benchmarking time series with autocorrelated sampling errors", *International Statistical Review*, 62, 365-377

Evans, Thomas D., Tiller, Richard B. and Zimmermann, Tamara S. (1993), "Time series models for state labor force estimates", *ASA Proceedings of the Section on Survey Research Methods*, 358-363

Feder, William, Gad, Nathan, and Pfeffermann, Danny (2000), "Multilevel modelling of complex survey longitudinal data with time varying random effects", *Survey Methodology*, 26, 53-65

Hausman, Jerry A., and Watson, Mark W. (1985), "Errors in variables and seasonal adjustment procedures", *Journal of the American Statistical Association*, **80**, 531-540

Hausman, Jerry A., and Watson, Mark W. (1990), "Seasonal adjustment of preliminary data", *ASA Proceedings of the Section on Survey Research Methods*, 640-648

Kokic, P. N. (1998), "Estimating the sampling variance of the UK Index of Production", *Journal of Official Statistics*, **14**, 163-179

Kokic, P.N. (2000), "Incorporation of the stock adjustment in the sampling variance of the IoP". In "Short-term output indicators review ", *White Paper Series Building Trust in Statistics*, ONS

McLaren, Craig H., and Steel, David G. (1997), "The effect of different rotation patterns on the sampling variance of seasonal trend filters", *ASA Proceedings of the Section on Survey Research Methods*, 790-795

McLaren, C. H., and Steel, D. G. (2000), "The impact of different rotation patterns on the sampling variance of seasonally adjusted and trend estimates", *Survey Methodology*, **26** (2), 163-172

McLaren, C. H., and Steel, D. G. (2001), "Rotation patterns and trend estimation for repeated surveys using rotation group estimates", *Statistica Neerlandica*, **55** (2), 221-238

Nguyen, Thuy Trang, Bell, William R., and Gomish, James M. (2002), "Investigating model-based time series methods to improve estimates from monthly value of construction put-in-place surveys", *ASA Proceedings of the Joint Statistical Meetings*, 2470-2475

Pfeffermann, D. (1989), "Estimation and seasonal adjustment of population means using data from repeated surveys", *ASA Proceedings of the Section on Survey Research Methods*

Pfeffermann, D. (1994), "A general method for estimating the variances of X-11 seasonally adjusted estimators", *Journal of Time Series Analysis*, **15**, 85-116

Pfeffermann, Danny, Morry, Marietta, and Wong, Paul (1995), "Estimation of the variances of X-11 ARIMA seasonally adjusted estimators for a multiplicative decomposition and heteroscedastic variances", *International Journal of Forecasting*, **11**, 271-283

Pfeffermann, Danny, and Scott, Stuart (1997), "Variance measures for X-11 seasonally adjusted estimators: Some new developments with application to labor force series", *ASA Proceedings of the Business and Economic Statistics Section*, 211-216

Steel, David G., and McLaren, Craig H. (2000), "The effect of different rotation patterns on the revisions of trend estimates", *Journal of Official Statistics*, **16** (1), 61-76

Scott A. J., and Smith, T.M.F. (1974), "Analysis of repeated surveys using time series methods", *Journal of the American Statistical Association*, **69**, 674-678

Scott A. J., Smith, T.M.F., and Jones, R.G. (1977), "The application of time series methods to the analysis of repeated surveys", *International Statistical Review*, **45**, 13-28

Sutcliffe, Andrew (1993), "X 11 times series decomposition and sampling errors", *Australian Bureau of Statistics*, Working paper no 93/2

Tiller, Richard B. (1992), "Time Series Modeling of Sample Survey Data from the U.S. Current Population Survey," *Journal of Official Statistics*, **8**, 149-166