SSC Annual Meeting, June 1998
Proceedings of the Survey Methods Section

A STATISTICAL APPROACH FOR DISAGGREGATING MIXED-FREQUENCY ECONOMIC TIME SERIES DATA

Wai-Sum Chan[1] and Zhao-Guo Chen[2]

ABSTRACT

The problem of mixed-frequency time series data arises from changing the observation frequency. For example, we may have a time series with quarterly observations in the first portion and annual figures in the remainder. We shall call that quarter-year mixed-frequency data. In this paper we suggest a method to disaggregate the annual observations to quarterly values. The proposed method can easily be generalised to the year-quarter, quarter-month, year-month and other mixed-frequency situations; it may avoid difficulties of time series modelling and is easy to implement. A step-by-step algorithm of the method is given so that econometricians not expert in this area can still perform the procedure. The proposed method is illustrated through two real examples. We also conduct a small scale Monte Carlo experiment to compare the proposed procedure with two existing alternative methods. Finally, some concluding remarks are given.

KEY WORDS: Annual, quarterly and monthly data; Autoregression; Best linear forecast; Data source.

RÉSUMÉ

Le problème suscité par la présence de fréquences mixtes dans les séries chronologiques apparaît lorsque la fréquence de la collecte des données change. Par exemple, les observations sont trimestrielles dans la première partie de la série et annuelles par après. Ce sont des données de fréquences mixtes trimestrielles-annuelles. Dans cet article, nous suggérons une méthode de désagrégation des observations annuelles en données trimestrielles. La méthode proposée peut facilement être généralisée pour traiter les cas de données annuelles-trimestrielles, trimestrielles-mensuelles, annuelles-mensuelles et autres cas; de plus elle évite les difficultés reliées à la modélisation des séries chronologiques et sa mise en pratique est simple. L'algorithme de la méthode est conçu de façon à permettre aux économétriciens qui ne sont pas experts dans ce domaine de suivre chaque étape et d'exécuter la procédure. Deux exemples réels sont utilisés pour expliquer la méthode proposée. Nous avons également effectué une petite expérience à l'aide de la méthode de Monte-Carlo pour comparer la procédure proposée à deux autres méthodes. Puis nous concluons.

MOTS CLÉS: Données annuelles, trimestrielles et mensuelles; autorégression; prévisions linéaires optimales; source de données.

Suppose a time series had been observed annually over several decades. Because of the increasing importance of the series, the reporting agency have been collecting and releasing quarterly figures of the variable in recent years. So, we have a time series with annual observations in the first portion and quarterly observation in the remainder. We may call this a year-quarter mixed-frequency time series. Our goal is to disaggregate those annual observations to quarterly figures.

One of the solutions to the year-quarter mixed-frequency problem is temporally aggregating the quarterly portion of the series to annual values, then one may use the method proposed by Wei and Stram (1990), which is statistical model-based, to disaggregate all the available annual totals to quarterly figures. Chan (1993) provided a comparative study of Wei and Stram's method with other existing (numerical) methods in the literature. It seems that there is no significant improvement in the results. In fact, for mixed-frequency time series data, temporal aggregation could lead to a

[1] Department of Economics & Statistics, National University of Singapore, Republic of Singapore 119260

[2] Time Series Research & Analysis Centre, 3-H, R.H. Coats Building, Statistics Canada, Ottawa, ON, Canada, K1A 0T6

considerable loss of information as discussed by some authors, e.g. Tiao (1972), Rossana and Steater (1995). Disaggregation procedures which fully utilize all the information are preferable for the mixed-frequency series. A guideline to solve this problem is first use quarterly data to predict the quarterly figures during the years covered by annual data, then use annual data to adjust these predictions. The procedure we proposed belongs to this category. This procedure can also be used for other types of mixed-frequency data such as quarter-year, month-quarter, year-month etc.

In the following, for convenience of statement, we assume that the data are quarter-year mixed-frequency series, then we may forecast their quarterly values for the years where annual data are observed. By reversing the time, year-quarter data become quarter-year data and backcasting becomes forecasting.

Assume that we have n quarterly observations $y(1), \ldots, y(n)$, followed by N annual observations $Z(T)$. i.e.

$$Z(T)=y(n+4T-3)+y(n+4T-2) +y(n+4T-1)+y(n+4T) \qquad (1)$$

for $T = 1, \ldots, N$. Suppose that the quarterly time series, $y(t)$, follows the model

$$\nabla^{d}\nabla^{d'}_4 y(t)=x(t) \qquad (2)$$

Where, d and d' are non-negative integers; $x(t)$ is a stationary time series; $\nabla = (1 - B)$ and $\nabla_4 = (1 - B^4)$ are the ordinary and the seasonal difference operator respectively, and B is the backshift operator, i.e $B^s\, y(t) = y(t-s)$. We shall only consider the cases of $0 \leq d+d' \leq 2$ and $d' \leq 1$, which are the most commonly encountered situations in economic time series; and we denote $d_0 = d + 4d'$.

As the forecasting of $y(t)$ is concerned, in the literature, most authors (e.g. Guerrero, 1989; Trabelsi and Hillmer, 1989) suggest using ARIMA modelling, where within (2), an ARMA model is fitted for $x(t)$. This forecasting avenue is often not convenient because of model identification and parameter estimation (these become more difficult when data size is small). Also, a special forecasting formula have to be derived for the fitted model in each case, and iteration is always involved in the forecasting procedure when $d_0>0$. Here, we suggest an alternative forecasting avenue for model (2) without assuming ARMA model for $x(t)$ but only stationarily. Then, at time n, the general formula to forecast the unobserved $y(n+t)$, $t > 0$, can be expressed with its past observations $y^* = (y(n-d_0+1), \ldots, y(n))'$ and observations' differences $x^*=(x(n), x(n-1), \ldots, x(d_0+1))'$ as follows:

$$\hat{y}(n+t|n)=f(t, y^*)+\sum_{j=1}^{t} a(t,j)\hat{x}(n+j|n), \qquad (3)$$

where, $\hat{x}(n+j|n)$ is a forecast of $x(n+j)$ from x^*; f is a simple function of t and y^*, $a(t,j)$ is a simple function of t and j (different formulae depend on different d and d'; for details, see Table 1 in Chan and Chen, 1998).

The advantage of (3) is giving some explicit relationship between the forecast of nonstationary series and the forecast of stationary series, $\hat{x}(n+j|n)$, which can be derived via some much easier ways, say, simply using autocovariance to project $x(n+j)$ on x^* which gives the best linear unbiased prediction of $x(n+j)$ (see, e.g. Whittle, 1963). The autocovariance of series $x(t)$ can be very easily estimated from data. Moreover, when a non-linear model is more suitable for $x(t)$, using non-linear forecast we can get better $\hat{x}(n+j|n)$.

If the covariance of the prediction errors of $\hat{x}(n+i|n)$ and of $\hat{x}(n+j|n)$ is $V_x(i,j)$, then the covariance of the prediction errors of $\hat{y}(n+s|n)$ and $\hat{y}(n+t|n)$ is

$$V_y(s,t)=\sum_{i=1}^{s}\sum_{j=1}^{t} a(s,i)\,V_x(i,j)\,a(t,j). \qquad (4)$$

Using the annual data as "benchmarks", the forecasted values given by (3) as the "original observations", and the covariance of the prediction error given by (4) as that of the "survey error", the "benchmarked values" can be obtained by standard benchmarking formula (see e.g. Guerrero, 1989, or Trabelsi and Hillmer, 1989). These "benchmarked values" are the disaggregated quarterly figures. It can be shown that, if the forecast is the best linear unbiased prediction of unknown quarterly values from known quarterly data, then the disaggregated figures are the best linear unbiased prediction of the true quarterly values from the whole data set (mixed-frequency time series).

The details of the above sketched method is presented in Chan and Chen (1998). Simulations and comparisons with other disaggreation methods for artificial and real data are also presented in there.

REFERENCES

Chan, W.S. (1993). Disaggregation of annual time-series data to quarterly figures: A comparative study. Journal of Forecasting, **12**, 677-688.

Chan, W.-S. and Chen, Z.-G. (1998). A Statistical approach for disaggregating mixed-frequency economic time series data. Advances in Econometrics: Messy data - Missing Observations, Outliers, and Mixed-Frequency Data. **13**. To appear.

Guerrero, V.M. (1989). Optimal conditional ARIMA forecasts. Journal of Forecasting, **8**, 215-229.

Rossana, R.J. and Steater, J.J. (1995). Temporal aggregation and economic time series. Journal of Business and Economic Statistics, **13**, 441-451.

Tiao, G.C. (1972). Asymptotic behavior of temporal aggregates of time series. Biometrika, **59**, 525-531.

Trabelsi, A. and Hillmer, S.C. (1989). A benchmarking approach to forecast combination. Journal of Business and Economic Statistics, **7**, 353-362.

Wei, W.W.S. and Stram, D.O. (1990). Disaggregation of time series models. Journal of the Royal Statistical Society, Series B, **52**, 453-467.

Whittle, P. (1963). Prediction and Regulation by Linear Least-Square Methods. Minneapolis: University of Minnesota Press.