Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.2, No.5, 2012 www.iiste.org

# Temporal Disaggregation of Time Series Data: A Non-Parametric Approach

Ayodeji S.D.I.* Olagunju M.A.
Department of Mathematics, Obafemi Awolowo University, 220005, Ile-Ife, Nigeria
* E-mail of the corresponding author: idowu.sayo@yahoo.com

**Abstract**

Sometimes it is necessary to disaggregate a given economic time series to obtain values for shorter subperiods. For example, monthly values of GDP may be desired when only a sequence of quarterly data is available. Using non-parametric algorithm, we generalize an existing method so that it can be used to derive values for subperiods of any specified length which are consistent with available data for longer time periods.

**Keywords:** temporal disaggregation, non-parametric, algorithm, Gandolfo

## 1. Introduction

Over the years, interest in modeling, empirical quantification and testing of economic theories and policies have intensified the need for a wide scope of high frequency, high quality and consistent statistics in various sectors of national economies, generally speaking. One frequently finds that for most of the variables actual time series are available on a quarterly basis, whereas for some only annual data can be obtained. The obvious alternatives of (a) omitting those variables for which only yearly observations exist could lead to a mis-specified model; or (b) aggregating all quarterly figures to annual totals would result in a substantial loss of information. Hence, the preferred approach is to utilize some reasonable process for deriving quarterly figures from the available annual observations. This will allow the econometric model to be estimated on a quarterly basis, with all relevant variables included. These processes are called "Temporal disaggregation techniques".

Temporal disaggregation techniques have been widely investigated in the literature. Broadly speaking, methods for disaggregation can be classified into two approaches: (1) one that involves the use of observed related series at the desired higher frequency and (2) one that only relies on pure time series dynamic models and that does not make use of information obtained from other related series. The former has been discussed by a number of authors, one of the first of whom was Friedman (1962). Subsequent contributions have been made by Chow and Lin (1971), Fernandez (1981) and de Alba (1988), to name a few. Procedures based on related variables have been the most popular and the most widely used and successful. Thus, a great number of procedures can be found within this category. When compared with the algorithms that do not use related series, related variable procedures have been assigned two main principal advantages: (i) they present better foundations in the construction hypothesis (which can comparatively affect the results validation); and (ii) they make use of relevant economic and statistical information, being more efficient; although the resulting estimates depend crucially on the indicators chosen. This drawback implies that a special care should be taken in selecting indicators. This limitation, however, far from being resolved, has been severally addressed during decades. For instance, Chang and Liu (1951) tried to establish some criteria the indicators should fulfill (this issue has been repeatedly treated again and again; see Bournay and Laroque (1979), Pavía et. al. (2000) and Nasse (1973) among others). It, however, does not yield that some universally accepted sound criteria have been proposed.

The second approach, explored by Wei and Stram (1990) and Guerrero (1990), depends on the autoregressive integrated moving-average (ARIMA) dynamic structure of the series to be disaggregated. As for the second approach, it only extracts signals from the presumed dynamic pattern of the series but in such a way that no other high frequency related information can be added. The former does not accommodate the possibility of some underlying dynamic structure of time series; it also assumes that there is a full co-integration relation between the nonstationary related series and the unobserved disaggregated time series 'a priori'. In light of the above shortcomings, Huang (2008) evaluated an alternative method, namely the state-space approach, first introduced by Harvey (1989) and later developed by Harvey and Koopman (1997). The state-space representation describes the dynamic structure of disaggregated time series and allows for high frequency related series without imposing the assumption of a co-integration relation. The exact Kalman filtering and smoothing algorithms introduced by Koopman (1997) and Koopman and Durbin (2003) were used to evaluate the likelihood function and to estimate real gross domestic product (GDP) at monthly intervals. Some of the major advantages of using the exact Kalman filter are that it computes the optimal estimates of the latent monthly GDP and provides a way to calculate exact finite-sample forecasts based on the appropriate information set. Apart from this, it avoids the potential 'divergence' problem that may arise from the algorithm proposed by Harvey and Phillips (1979) and Kim and Nelson (1999). Although these approaches both have the potential to be applied to a wide variety of cases, a substantial knowledge of the characteristics of the disaggregated process is assumed. Yet, usually, only aggregate information is available, and it is not sufficient to exactly identify the disaggregated characteristics of interest. In other words, the aforementioned processes mainly rely on undesired and/or arbitrary assumptions which, if violated, will negatively affect the quality of the result so obtained. Hence, in cases where these assumptions cannot be met, non-parametric approaches have also been developed.

Non-parametric procedures which are not confined to any variable type, whether stock or flow, have been applied by Diz (1970) and Gandolfo (1981).They are based on Location or Order Statistics theory and are also distribution-free. Diz (1970) derived the linear interpolation algorithm while Gandolfo (1981) derived the quadratic algorithm. The fundamental assumption is that a flow variable has observed annual values, which are realizations of shorter-period numerical integration of the variable. These non-parametric papers however, restricted their attention to the problem of obtaining quarterly values from given yearly totals; Gandolfo (1981) went further to derive monthly values from quarterly. Now what if observed values are available on annual basis but monthly values are desired or on semi-annual basis but quarterly are needed. The need, therefore, for such a general model arises such that time series could be disaggregated to subperiods of any desired length without being restricted by any arbitrary assumptions or variable-type. The general objective of this paper is to not to discredit the already existing parametric procedures (though each of them has its own flaw(s); see Pavía, 2010 ) but to provide a scientifically - acceptable way out in cases where they cannot be applied. In other words, it provides an algorithm, using a non-parametric approach, for the disaggregation of time series data that are justifiably inputable for any type of quantitative analysis and application, especially in situations in which one needs a robust short-term data that are not available on, for example, weekly, monthly or quarterly basis. The result from this research is important especially for the following reasons: (1) the paper presented, using sound mathematical techniques, a non-parametric approach that does not rely on any underlying distributions or assumptions and also; (2) it proposed a general model that can be used in disaggregating a given time series to obtain consistent values for subperiods of any desired length.

In order to solve the general problem of temporally disaggregating a set of known values for n periods into a pn subperiod values (where p is the number of subperiods per period), the paper generalizes the approach taken by Gandolfo. This will be the focus in the next section. In Section 3, some illustrations of the algorithm developed in Section 2 are discussed.

## 2. The Algorithm

Let $y_{t-1}$, $y_{t-2}$ and $y_{t-3}$ be three consecutive past observations of a continuous flow variable whose current value is $y_t$ at the period t. Using the assumption that the variable of interest can be described by a polynomial function

F(t) of degree 2,

$$F(t) = at^2 + bt + c \qquad \text{-------------} \qquad (2.1)$$

where a,b and c are coefficients to be determined.

The approximation of an observation for three periods $y_{t-1}$, $y_t$ and $y_{t+1}$ are given by the definite integrals from initial period 0 to period 3:

$$\left.\begin{aligned} y_{t-1} &= \int_0^1 \left(at^2 + bt + c\right)dt \\ y_t &= \int_1^2 \left(at^2 + bt + c\right)dt \\ y_{t+1} &= \int_2^3 \left(at^2 + bt + c\right)dt \end{aligned}\right\} \qquad \text{--------------------} \qquad (2.2)$$

Integrating out and solving for the coefficients gives

$$\left.\begin{aligned} a &= \frac{1}{2}y_{t-1} - y_t + \frac{1}{2}y_{t+1} \\ b &= -2y_{t-1} + 3y_t - y_{t+1} \\ c &= \frac{11}{6}y_{t-1} - \frac{7}{6}y_t + \frac{1}{3}y_{t+1} \end{aligned}\right\} \qquad \text{---------------------} \qquad (2.3)$$

The subperiod figures within any period, t, can be generated using the following condition in order statistics theory:

$$y_t^{(i)} = \int_{\frac{p+i-1}{p}}^{\frac{p+i}{p}} \left(at^2 + bt + c\right)dt; \qquad \text{----------------------} \qquad (2.4)$$

$$i = 1, 2, \cdots, p, \qquad and \qquad p \geq 2$$

By the required additive property of flow statistics:

$$y_t = \sum_{i=1}^{p} y_t^{(i)} \qquad \text{----------------------}$$

(2.5)

Equation (2.5) assures consistency, that is, the sum of the subperiod values for a period is equal to the corresponding given value of that period. It should be noted that the procedures described do not have any underlying assumptions about the distribution or parameters of the variables, hence, they are regarded as non-parametric and distribution-free, which are desirable properties required of input data for further econometric or regression analysis.

It should also be noted that setting p = 4 in the system of equations yields the quadratic interpolation algorithm of Gandolfo:

$$\left.\begin{aligned} y_t^{(1)} &= \frac{7}{128}y_{t-1} + \frac{15}{64}y_t - \frac{5}{128}y_{t+1} \\ y_t^{(2)} &= \frac{1}{128}y_{t-1} + \frac{17}{64}y_t - \frac{3}{128}y_{t+1} \\ y_t^{(3)} &= -\frac{3}{128}y_{t-1} + \frac{17}{64}y_t - \frac{1}{128}y_{t+1} \\ y_t^{(4)} &= -\frac{5}{128}y_{t-1} + \frac{15}{64}y_t + \frac{7}{128}y_{t+1} \end{aligned}\right\} \quad \text{------------------------} \quad (2.6)$$

Where $y_{t-1}$, $y_t$ and $y_{t+1}$ are three successive annual observations of a continuous flow variable y(t). And, of course, setting p = 3 in the same system of equations gives his algorithm for interpolating monthly values from quarterly:

$$\left.\begin{aligned} y_t^{(1)} &= \frac{10}{162}y_{t-1} + \frac{52}{162}y_t - \frac{8}{162}y_{t+1} \\ y_t^{(2)} &= -\frac{2}{162}y_{t-1} + \frac{58}{162}y_t - \frac{2}{162}y_{t+1} \\ y_t^{(3)} &= -\frac{8}{162}y_{t-1} + \frac{52}{162}y_t + \frac{10}{162}y_{t+1} \end{aligned}\right\} \quad \text{-----------------------} \quad (2.7)$$

Where $y_{t-1}$, $y_t$ and $y_{t+1}$ are three successive quarterly observations of a continuous flow variable y(t). There are some restrictions on the smallest values permissible for p: $p \geq 2$. Obviously, p = 1 is an uninteresting case since no disaggregation takes place.

It is good to note, as Gandolfo (1981) pointed out, that the treatment described so far is valid for flow variables. In dealing with stock variables, he suggested to take first differences to obtain a series of changes in stocks over the period (as changes in stocks can be treated as flow variables), then, interpolate changes over the intermediate period using the algorithm developed. Also, for non-stock instantaneous variables, he suggested a standard cubic interpolation formula.

## 3. Illustration

The algorithm developed in Section 2 permits one to temporally disaggregate any consecutive number n of period values to obtain values for any number p of subperiod values per period. Although the initial observations $y_{t-1}$, $y_{t-2}$ and $y_{t-3}$ on which the disaggregation is based may relate to periods of any arbitrary length, for the illustrative examples they are interpreted as yearly values. Accordingly, the subdivisions into semi-annual, quarterly, monthly and weekly are the most interesting ones, and the respective values for p are p = 2, 4, 12 and 52.

In order to illustrate this algorithm, first, generate (by integration) five annual values (years 1 to 5) using the equation

$$F(t) = 4t^2 + 5.2\sin t + 50 \qquad \text{------------} \qquad (3.1)$$

One year's length corresponds to an increment of t+1. Using this same procedure we generate actual month and week values, where one month's length and one week's length correspond to an increment of (t+1/12) and (t+1/52) respectively. Then using functions (2.2) to (2.5), the generated annual values are disaggregated to two periods - month and week values. Figure A.1 shows the graphs of actual and disaggregated month figures, and Figure A.2, the graph of actual and disaggregated week figures. The plots show very close fit in both cases.

One issue of concern that may arise while implementing this algorithm is the effect of growing p on the disaggregated values. That is, whether or not growing p affects the quality of the disaggregated values. Again, the function (3.1) has been used to generate five annual values (years 1 to 5) which then were disaggregated into semi-annual, quarter, month and week values. In Figure A.3, all the disaggregated values for the three middle years (years 2 to 4) have been plotted together (in terms of annual rates). The results do not indicate any deterioration for growing p.

From figures A.1, A.2 and A.3, one can conclude that the proposed approach to temporal disaggregation is useful when generalized to arbitrary values of p.

**References**

Bournay J. & Laroque G. (1979). Réflexions sur le Méthode d'Élaboration des Comptes Trimestriels. Annales de L'Insée, 36, 3-29.

Chang C. G. & Liu T. C. (1951). Monthly Estimates of Certain National Product Components, 1946-1949', Review of Economics and Statistics, 33(3), 219-227.

Chow, G. C. & Lin A. L. (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series. Review of Economics and Statistics. 53, 372 - 375.

De Alba, E. (1988). Disaggregation and forecasting: A Bayesian analysis. Journal of Business and Economic Statistic. 6, 197 - 206.

Diz, A. (1970). Money and Prices in Argentina 1935-1965 in Varieties of Monetary Experiences (1st Ed.). University of Chicago Press, Chicago.

Fernandez, R. (1981). A methodological note on the estimation of time series. Review of Economics and Statistics. 63, 471 - 478.

Friedman, M., 1962. The interpolation of time series by related series. Journal of the American Statistical Association. 57, 729 - 757.

Gandolfo, G. (1981). Qualitative Analysis and Econometric Estimation of Continuous Time Dynamic Models. North Holland, Amsterdam. 1, 2

Guerrero, V. M. (1990). Temporal disaggregation of time series: An ARIMA based approach. International Statistical Review. 58, 29 - 46.

Harvey, A. C. (1989). Forecasting, Structural Time Series and the Kalman Filter. Cambridge: Cambridge University Press.

Harvey, A. C. & Koopman S. J. (1997). Multivariate structural time series models, in System Dynamics. In C. Heij, J.M. Shumacher, B. Hanzon & C. Praagman, Chichester (Eds.) Economic and Financial Models (pp. 269 - 298), John Wiley and Sons.

Harvey, A. C. & Phillips G. D. A. (1979). The estimation of regression models with autoregressive-moving average disturbances. Biometrika, 66, 49 - 58.

Huang, Y. L. 2008. Estimating Taiwans Monthly GDP in an Exact Kalman Filter Framework. Department of Quantitative Finance, National Tsing Hua University.

Kim, C.-J. & Nelson C. R. (1999). State-Space Models with Regime Switching: Classical and Gibbs-Sampling

 

Approaches with Applications. Cambridge: MIT Press.

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. Journal of the American Statistical Association, 92, 1630 - 1638.

Koopman, S. J. & Durbin J. (2003). Filtering and smoothing of state vector for diffuse state-space models. Journal of Time Series Analysis, 24, 85 - 98.

Nasse P. (1973). Le Système des Comptes Nationaux Trime- strels. Annales de L'Inssée, 14, 127-161.

Pavía, J. M., Cabrer B. & Felip J. M. (2000). Estimación del VAB Trimestral No Agrario de la Comunidad Valenciana. Valencia, Generalitat Valenciana.

Pavía, J. M. (2010). A Survey of Methods to Interpolate, Distribute and Extrapolate Time Series. J. Service Science & Management, 449 – 463.

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

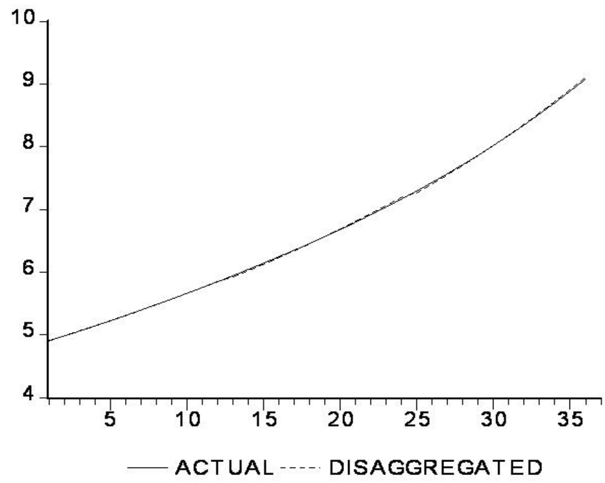


Figure A.1: Actual and Disaggregated Month values

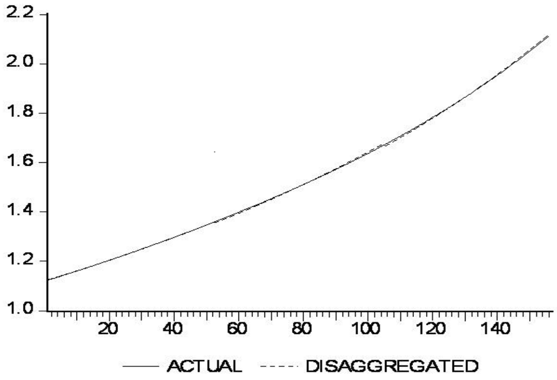


Figure A.2: Actual and Disaggregated Week values

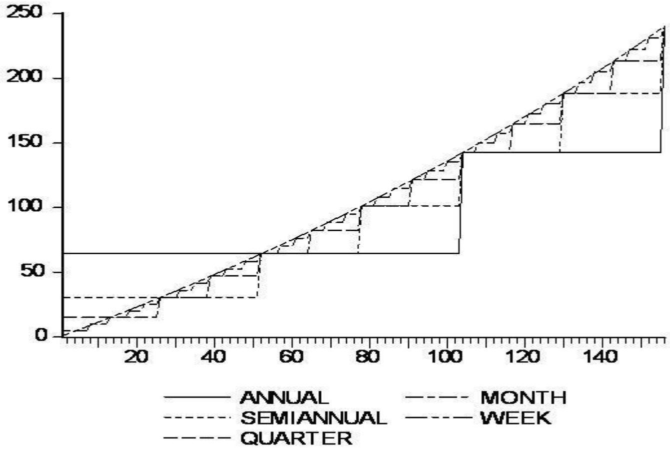


Figure A.3: Actual Annual values and Disaggregated Semi-annual, Quarter, Month and Week values at Annual rates.