WORKING PAPERS AND STUDIES

ISSN 1725-4825

2005 EDITION

Some time series considerations in disaggregation

COPYRIGHT
Luxembourg: Office for Official Publications of the European Communities, 2005

ISSN 1725-4825
ISBN 92-79-01311-4
Cat. No. KS-DT-05-014-EN-N

EUROPEAN COMMISSION

eurostat

EURO INDICATORS

Luxembourg: Office for Official Publications of the European Communities, 2005

ISSN 1725-4825
ISBN 92-79-01311-4

Some time series considerations in disaggregation

By Gudmundur Gudmundsson[1]

Discrete values of flow variables can be regarded as the integrals of unobserved flow functions. The estimated flows are defined as the product of two functions. One is predetermined and selected to eliminate or reduce trend in mean and variance. The other is approximately stationary, selected by smoothness criteria and the requirement that integrals of the flow reproduce the observed discrete values. Disaggregated values are obtained by integration over shorter intervals. The discrete observations contain good information about low-frequency variations of the flow. Joint analysis with related series, observed at shorter intervals, has been employed to introduce high frequency variations of disaggregated values. A fundamental weakness of this approach is that the high-frequency variations of the original series were eliminated by the integration. The relationship with the related series in the relevant frequency range cannot therefore be estimated from the data. Actual examples and comparison with other methods are presented.

Keywords: Disaggregation, Flow variables, Splines, Spectral analysis.

JEL classification: C13, C22, C32.

1 Introduction

Time series analysis has become an important factor in econometric research. Disaggregation is concerned with economic time series, but it has hardly been much influenced by econometric methodology.

Disaggregation has been an aspect of economic research at the Central Bank of Iceland for a long time. The main reason for our interest in this subject was that until 1997 only annual measurements were made of investment and consumption in the national income statistics. When I first came across it in economic research I had some experience of hydrology, where continuous records of river-flow were collected and subsequently converted into daily aggregates. Here the problem was reversed and it seemed normal to approach it by estimating a continuous flow which integrated reproduced the observed aggregates. Disaggregated values can then be obtained by integration over shorter intervals. Estimates of the flow at a given point in time can be relevant in joint analysis with stock data.

Before 1990 inflation in Iceland was very high. Without some arrangements to deal with this, the usual requirement of constant variance in statistical analysis of time series would be badly violated in series measured at current prices. Stationarity has not been a prominent concept in the literature on disaggregation. In econometric analysis of time series the first step in dealing with trend in mean and variance is usually to take logarithms. But this transformation is computationally inconvenient in combination with the requirement that the integral of the flow

[1] Gudmundur Gudmundsson, Central Bank of Iceland, Kalkofnsvegur 1, 105 Reykjavik, Iceland, tel. (+354) 569 9694 e-mail gudmg@centbk.is

reproduces the observed values. This applies also to disaggregation in discrete intervals where the estimated values sum up to the observed aggregates.

Our approach to this was to define the flow as a product of two functions. One is selected to deal with the trend and possibly auxiliary information. The other is approximately stationary and determined by smoothness criteria and the requirement that the flow reproduces the observed aggregates.

Some examples of this are given below and compared with results from widely applied procedures, available in the user-friendly ECOTRIM programs. (Barcelland and Buono, 2002). A more detailed description of the method and other examples were presented by Gudmundsson (2001a and 2001b).

2 Continuous flows

The flow is defined as a product of two functions. Let us write it

$$\eta(t) = w(t)f(t).$$

The function $w(t)$ is predetermined and selected to represent the trend so that $f(t)$ will be approximately stationary. In our applied work $w(t)$ is usually an exponential function where the exponent can be determined by regression of the logarithms of the observed values or simply as the average annual growth from the first to the last year. The function $f(t)$ must be selected so that the integrated values

$$y(t) = \int_{t-1}^{t} \eta(s)ds \qquad (1)$$

coincide with the observed values y_i at the end of year i. For a given $w(t)$ there is an infinite set of functions $f(t)$ which fulfil the requirement to reproduce the observed values.

The aggregation is a filtering operation and in order to assess the limits and possibilities of disaggregation it is illuminating to consider the how the filter affects the power spectrum. (The power spectrum represents the distribution of the variance of a stationary series on frequency). The reduction in power is proportional to the squared modulus of the frequency response of the aggregation, presented in Figure 1 for integration over one year.

The filtering effect is weak in the lowest frequencies, but at half cycle per year only 40% of the power is left. The effects that have been filtered out by the integration cannot be recovered by numerical operations on the aggregated values alone. Seasonal effects are completely wiped out by integration over one year.

For continuous recording of $y(t)$ in equation (1) the squared frequency response represents the information left after the integration. But $y(t)$ is only observed for consecutive non-overlapping intervals which entails further loss of information. Annual values can be reproduced exactly by a

Fourier series in the frequency range from zero to half cycle per year. Actual variations at higher frequencies, which were not completely filtered out by the integration, will therefore appear as variations at lower frequencies. This phenomenon is called aliasing and described in textbooks in time series (e.g. Priestley, 1981). The upper limit, called the Nyquist frequency, for quarterly values is two cycles per year and six for monthly intervals.

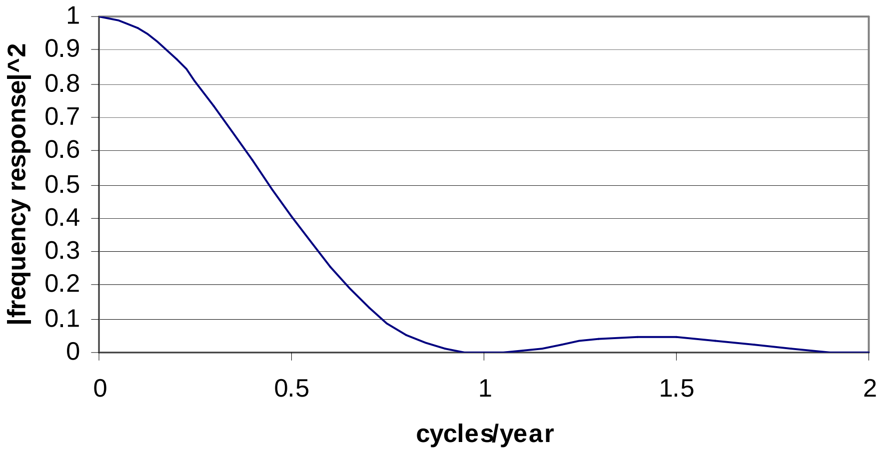

Figure 1: Squared modulus of the frequency response of integration over one year.

As the high frequency effects that were filtered out by the integration cannot be recovered, it is sensible to select, from the infinite set of functions *f(t)* which reproduce the observed aggregated values, a function with little high frequency variations, i.e. a smooth function. An obvious way to achieve this is to represent *f(t)* by a Fourier series. By this means no variations at frequencies above the Nyquist frequency are included. But through aliasing those that were present in the actual flow and not filtered out appear as distortions of the estimated low frequency effects. Calculation of flow functions by Fourier series and numerical examples were presented by Gudmundsson (2001a), but the examples in this paper are based on a different method.

Squared derivatives are widely applied smoothness criteria. We have used the functions obtained by minimizing

$$\int f^{(n)}(t)^2\,dt\,,$$

subject to the requirement that the integration in equation (1) reproduces the observed aggregated values. The order of derivation is denoted by n. Calculus of variation provides spline functions as solutions to this problem. With local co-ordinates so that each interval is in [0, 1] it has the form

$$f_j(t)=\sum_{k=0}^{2n-1} a_{jk}t^k + b_j\int_0^t \frac{(t-s)^{2n-1}}{(2n-1)!}w_j(s)ds$$

in interval j. The functions $f^{(v)}(t)$ are continuous for $v= 0,1,\ldots,2n-1$ with derivatives of the order from n to $2n-1$ equal to zero at the beginning of the first- and the end of the last interval. The necessary equations for calculating the coefficients a_{jk} and b_j are obtained by inserting the observed values y_i in equation (1) and from the continuity requirements and end-conditions.

The relative merits of different smoothness criteria depend upon the properties of the actual process. Fourier series are best for very smooth flows, but in my experience with economic series, spline solutions, obtained by minimizing squared first derivatives, are preferable. These splines produce continuous functions and first derivatives. Splines obtained by minimizing squared second derivatives are continuous up to third derivatives and perform markedly worse.

The end conditions are important in determining the disaggregated values in the first and last period. If the time interval included in the analysis starts at $t = 0$ and ends at $t = T$, the end conditions of the spline solution with $\kappa = 1$ imply that

$$\eta^{(1)}(0) = f(0)w^{(1)}(0) \quad \text{and} \quad \eta^{(1)}(T) = f(T)w^{(1)}(T),$$

i.e. the first derivatives at the ends are determined by the slope of the trend. This is sensible if we know nothing about the series except the observed aggregates, but in practice we often know more and this should be taken into account, formally or by some ad hoc arrangements. Simple means for this purpose are including one observed disaggregated value before the first period where disaggregated values are needed, and extending the disaggregated series by prediction. These arrangements are not implemented in the examples in this paper.

Various methods have been suggested to disaggregate series by means of related series, observed at shorter intervals, maintaining the requirement that the estimated disaggregated values reproduce the observed aggregates. Linear relationships are assumed in these methods. This implies that variations at each frequency are only determined by variations at the same frequency in the other series. The low frequency components of the estimated series are already largely determined by the observed aggregates. In some cases the relationship between the series in the higher frequency range may be known. Another possibility is that both series have been observed at shorter intervals for a period of time so that the relationship can be estimated and assumed to hold for the period when only aggregated values are observed of one of the series (Harvey and Pierse, 1984). But in the absence of other knowledge than the observations of the two series, there are no means of estimating the relationship in the relevant frequency range from the data; the necessary information in the aggregated series has been filtered and aliased out.

In view of the filtering effect of the aggregation process I have not tried to implement any estimation of the weight to give to auxiliary series. But it is possible to include an observed series, $x(t)$, in the trend function,

$$w(t) = x(t)^{\gamma}\, v(t). \qquad (2)$$

In this formulation of $w(t)$ the parameter γ is predetermined to give desired weight to the information in $x(t)$, and $v(t)$ is a new trend function, selected so that $w(t)$ accounts for the main trend in the observed series. The only form of this implemented in our present programs is to let $x(t)$ be a step function, equal to the observed value in respective interval, and $v(t)$ an exponential function, representing the difference in the trend of $y(t)$ and $x(t)^{\gamma}$. This implies that the flow is not continuous. However, with the present capacity of desk computers there would be little practical problems in converting *x(t)* into a continuous function, representing the observed values, if a continuous flow of *y(t)* is of interest.

Stram and Wei (1986) performed disaggregation by time series modelling. State space time series models, emphasizing the introduction of auxiliary series for disaggregation, were presented by Moauro and Savio (2005). Gudmundsson (1999) included multiplicative trends in state space models for disaggregation, but we have not used this much, mainly because the available data tend to be inadequate for good estimation of the time series models.

3 Examples

In the following examples we use data from the International Financial Statistics to demonstrate some of the effects described above. Quarterly values are estimated from annual totals, with or without auxiliary series, and compared with observed quarterly values by the root mean square error of logarithmic values. The estimates with continuous flows were based on an exponential trend function and spline functions where squared first derivatives were minimized. This criterion was also employed when other series were taken into account according to equation (2). The results are compared with methods from the ECOTRIM programs where squared first differences were minimized when only annual totals were included and methods by Litterman (1983) and Fernandez (1980) employed for including auxiliary series.

3.1 Consumption in Turkey

Consumption in Turkey from 1987-2003 at nominal prices is an example of a series with a strong trend. The average annual increase is about 70%. The observed values were seasonally adjusted for comparison with values estimated from the annual aggregates. The spline solution with exponential trend function in the last five years is presented in Figure 2, together with the observed annual values and the quarterly values (multiplied by 4). Results from comparison of observed and estimated quarterly values are presented in Table 1. The root mean square error (rmse.) of the logarithms of the quarterly values, obtained by integrating the spline functions, was 0.0345. The rmse. of quarterly values, obtained by minimizing squared first differences, was 0.0316. (In this example, where the last observed quarterly value was in fact lower than the previous one, the end condition spoilt the spline estimates).

When inflation is much higher than growth in production and population, variations in consumption at current prices will presumably be positively related with prices. (In low inflation the negative relationship between prices and quantity might dominate). When quarterly averages of the consumer price index are included as *x(t)* in equation (2) a small reduction in rmse. is observed, but estimates by Litterman’s and Fernandez methods are less accurate than results obtained without auxiliary series. The reason for this is that the linear relationships, assumed between the aggregated- and the related series in these procedures, are inappropriate for series, dominated by trends that are exponential rather than linear. Considering the prevalence of the logarithmic transformation in econometric work it seems likely that such procedures are often inappropriate or sub-optimal.

The ECOTRIM programs provide various common tests from regression analysis with the Litterman and Fernandez estimates. These provide no serious indications of bad fit or misspecification. But the programs also present confidence bands. I don't know the details of how they are calculated, but they are of similar size for the whole interval and orders of magnitude to wide in the first years and much too narrow in the last years.

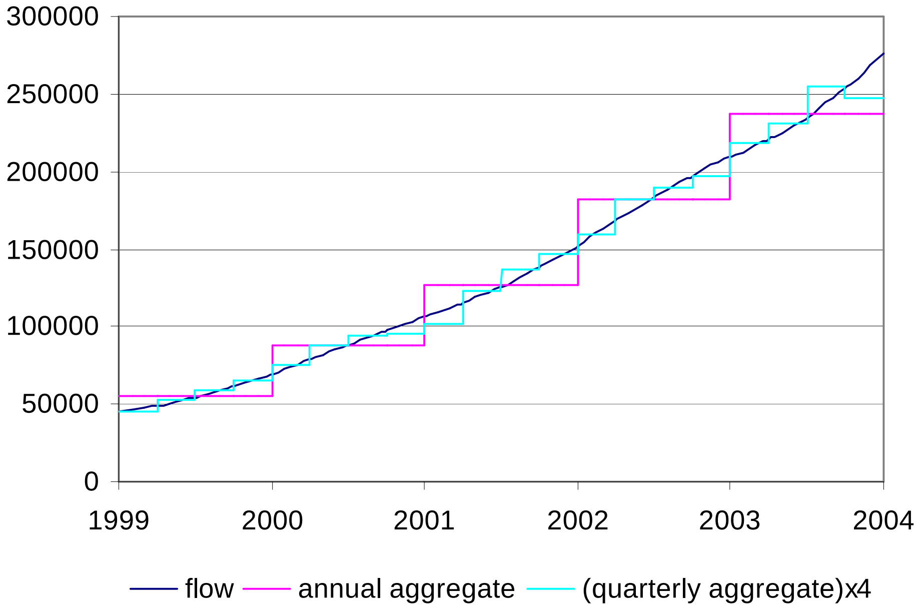

Figure 2: Observed values and estimated flow function of private consumption at current prices in Turkey. (Units: trillions of Liras and trillions of Liras/year, multiplied by 10^{-9}).

Table 1: Consumption in Turkey at current prices 1987-2003. Root mean square errors of log-values of quarterly values, estimated from annual totals, compared with observed values. "Discrete" estimates minimize squared first differences, "Splines" minimize squared first derivatives. *CPI* is the consumer price index.

Estimates without CPI	Rmse.
Discrete*	0.0316
Splines (γ=0)	0.0345
Estimates including CPI	
Litterman*	0.0396
Fernandez*	0.0352
Splines, γ=0.25	0.0309
Splines, γ=0.50	0.0290
Splines, γ=0.75	0.0293
Splines, γ=1.0	0.0318

* Calculated with ECOTRIM.

3.2 US GDP

The fit of estimates of quarterly values of US GDP from 1958-2003 at constant and current prices is presented in Tables 2 and 3. The average trend is about 3.3% per year for the constant price series and 7.1% with current prices. The index of industrial production is employed as auxiliary information with both GDP series and the GDP deflator also for GDP at current prices.

The relationship between the GDP at constant prices and the index of industrial production, estimated by the Litterman procedure, greatly improves the estimates of quarterly values, compared with estimates where only the annual aggregates are employed. A similar improvement is attained in the multiplicative model by inserting the index as *x(t)* in equation (2), provided a suitable value of γ is selected.

Table 2: US GDP at constant prices 1957-2003. Root mean square errors of quarterly log-values, estimated from annual totals, compared with observed values. "Discrete" estimates minimize squared first differences, "Splines" minimize squared first derivatives. I is the index of industrial production.

Estimates without I	Rmse
Discrete*	0.00519
Splines (γ=0)	0.00508
Estimates including I	
Litterman*	0.00332
Fernandez*	0.00425
Splines, γ=0.25	0.00358
Splines, γ=0.50	0.00361
Splines, γ=0.75	0.00518
Splines, γ=1.0	0.00737

* Calculated with ECOTRIM.

Table 3: US GDP at current prices 1957-2003. Rmse. of quarterly log-values, estimated from annual totals, compared with observed values. "Discrete" estimates minimize squared first differences, "Splines" minimize squared first derivatives. I is the index of industrial production, P is the DGP deflator.

Estimates without I or P	Rmse.
Discrete*	0.00527
Splines (γ=0)	0.00513
Estimates including I and P	
Litterman, I*	0.00635
Litterman, P*	0.00508
Litterman , I and P*	0.00790
Litterman, I*P*	0.00320
Litterman, I*$e^{\lambda t}$ *	0.00332
Splines, I, γ=0.25	0.00352
Splines, I, γ=0.5	0.00345
Splines, P, γ=0.25	0.00507
Splines, P, γ=0.5	0.00505
Splines, I*P, γ=0.5	0.00345

* Calculated with ECOTRIM.

When no auxiliary information is introduced, the multiplicative model and the discrete model produce estimates of similar accuracy for the GDP at current prices. Introduction of the production index as *x(t)* leads to a similar improvement as for the constant price values in the multiplicative model. But the Litterman procedure for discrete values produces worse estimates than estimation without auxiliary information. Introduction of the GDP deflator has negligible effect upon the fit in both the discrete- and the multiplicative method. When it is introduced together with the production index with the Litterman procedure the fit is even worse than with the index alone. However, by introducing the product of the index and the deflator as an auxiliary series, great improvement in fit is obtained. A similar effect is produced by adjusting the trend of the production index by an exponential trend as in equation (2) to match the trend in GDP at current prices.

The various t- and χ^2 tests provided by the ECOTRIM programs are not of much use in judging whether an auxiliary series actually produces an improved fit. This is not surprising in view of the filtering effect of the aggregation.

4. Conclusions

Apart from the end values there is little difference between disaggregation, based on minimizing squared first differences, and estimates based on a continuous flow with multiplicative trend and minimizing the squared first derivative of the stationary factor of the flow.

One reason why minimizing squared first derivatives produces better estimates than Fourier series or minimizing squared higher derivatives is probably that the effects of variations or measurement errors in one year have less effect upon the values in other years because the flow is only continuous up to the first derivative. The rapidly diminishing influence of the observed aggregated value in one year upon the disaggregated values in other years may also be an important factor in producing good results from minimizing squared first differences, even in series with strong exponential trends.

It is not possible to decide by numerical analysis of an aggregated series, *y*, and a disaggregated series, *x*, whether series *x* contains useful information about the disaggregated values of series *y*. The linear models that have been used for this purpose are obviously inappropriate when the series contain an exponential trend. Adjusting the trend of *x* to match approximately the trend of *y* would often be an advantage, but does not guarantee that the estimation of the disaggregated values will be improved by including *x.* In my opinion disaggregation should not be based on auxiliary series unless the relationship, estimated or otherwise determined, is supported by strong economic arguments. The aggregated values provide us with good information about the low frequency variations and it may be best to stay content with this.

References

Barcellan, R., and D. Buono, (2002): *ECOTRIM Interface. User Manual. Eurostat.* The Statistical Office of the European Communities.

Fernandez, R.B. (1980): "A methodological note on the estimation of time series", *The Review of Economics and Statistics*, 63, 471-476.

Gudmundsson, G. (1999): "Disaggregation of annual flow data with multiplicative trends", *Journal of Forecasting*, 18, 33-37.

Gudmundsson G. (2001,a): "Estimation of continuous flows from observed aggregates", *The Statistician*, 50, 285-393.

Gudmundsson G. (2001,b): "Calculation of flows and disaggregation of accumulated values", Working Papers No. 13. Central Bank of Iceland (www.sedlabanki.is).

Harvey, A.C., and R.G. Pierse (1984): "Estimating missing observations in economic time series", *Journal of the American Statistical Association*, 79, 125-131.

Litterman, R.B. (1983): "A random walk, Markov model for the distribution of time series", *Journal of Business and Economic Statistics*, 1, 169-173.

Moauro, F. and G. Savio (2005): "Temporal disaggregation using multivariate structural time series models", *The Econometrics Journal, 8,* 214-234.

Priestley, M.B. (1981): *Spectral Analysis and Time Series. Volume 1.* Academic Press.

Stram, D.O. and W.W.S. Wei (1986): "A methodological note on the disaggregation of time series totals", *Journal of Time Series Analysis*, 7, 293-302.