CENTRAL BANK OF ICELAND

WORKING PAPERS No. 13

CALCULATION OF FLOWS AND DISAGGREGATION OF ACCUMULATED VALUES

by

Gudmundur Gudmundsson

August 2001

CENTRAL BANK OF ICELAND

Economics Department

Central Bank of Iceland Working Papers are published by the Economics Department of the Central Bank of Iceland. The views expressed in them are those of their authors and not necessarily the views of the Central Bank of Iceland.

Also available on the Central Bank of Iceland World Wide Web site (http://www.sedlabanki.is)

ISSN: 1028-9445

Calculation of flows and disaggregation of accumulated values

by

Gudmundur Gudmundsson*

August 2001

Abstract

Some important economic flow variables, notably consumption and investment, have only been observed annually in Iceland. We have developed methods to estimate the flow by spline functions or Fourier series. Aggregated values over shorter intervals can be obtained by integration of the flows. A method to obtain quarterly values from annual observations by estimating a quarterly time series model from the observed values is also presented. The source code of Fortran programs and examples of input files are included. Access to the NAG subroutines is necessary to use the programs directly.

Keywords: flow variables, disaggregation, spline functions, Fourier series, Kalman filter

---

* Central Bank of Iceland. The views expressed in this paper are those of the author and do not necesserily reflect the views and policies of the Central Bank of Icland.

1. *Introduction*

Many important concepts in economics have the dimension value per time. They are not observed directly but only as accumulated values over an interval of time. The observations are often called flow variables, distinct from stock variables which can be observed at any moment of time. Exports and consumption are examples of flow variables; prices and money supply are stock variables.

In Iceland investment and consumption have only been observed annually. This is a nuisance in quantitative macroeconomic research involving relationships between these concepts and other economic variables, observed more frequently. For this reason a good deal of research has been carried out here in methods to estimate the flow or disaggregate annual values into shorter time intervals. The first attempts were presented by Guðmundsson and Veneklaas (1976), soon to be replaced by more sophisticated methods (Gudmundsson, 1981). The best method examined in the latter paper, where the flow was represented by the product of a Fourier series and an exponential function, was employed here until 1994.

The survey by Wai-Sum Chan (1993) led to a re-examination of the methodology. Our preferred method for calculation of economic flows is now a spline function, obtained by minimizing squared first derivatives. This is also usually a good choice for disaggregation, but disaggregation by time series methods is sometimes feasible with long series. The theory and examination of accuracy are published in two papers (Gudmundsson, 1999 and 2001).

Our main emphasis has been on estimation of the flow and quarterly values from annual observations. There are some options for introducing auxiliary information; e.g. to take account of monthly or quarterly price indices in the disaggregation of annual aggregates at current prices. Other methods are much more concerned with disaggregation by means of related series. We refer to the papers by Chow and Ling (1971) and Abeysinghe and Lee (1998).

This paper is mainly a presentation of programs for application of these methods. The programs are attached as Pfe files to the internet version of this paper at http://www.sedlabanki.is/interpro/sedlabanki/sedlabanki.nsf/pages/wp. They are written in FORTRAN (as I learned it in the sixties) with frequent calls to NAG routines for various numerical tasks. We take no responsibility for these programs, but provide the source code so that the reader can change and modify them at will. The tasks performed by the NAG routines are explained so that readers who do not have access to them can replace them by other routines.

Readers who merely want to use the programs can skip sections 4-6. They are mainly intended for those who might want to change the programs or add further options.

2. A quick overview of methods.

Our investigations indicate that for economic observations the best way to calculate flows is by splines, derived by minimizing squared first derivatives. (Program Spline1). For extremely smooth flows and accurate observations of the annual totals, better results could be obtained by Fourier series where the period is increased by four intervals. (Second program in Four). I don't think such series are often found in economic observations.

The spline solution is usually also the best choice for disaggregation of economic series. However, if the number of observations is fairly large, (say >30), and the series appears to be stationary after 1st or 2nd differencing of logarithmic values, time series models are a feasible alternative. (Program Tser).

We have mainly used these methods with annual observations, but the flow methods can take into account (fixed) seasonal variation in quarterly observations. In economic time series gradual variation of seasonal patterns is common. In that case the present programs might need some modification, but extension of the series by one or two years at the ends could be adequate. (Programs Spline1s and Four, second program).

Disaggregation of quarterly values into monthly values has not been on our agenda, but time series models might be suitable for that purpose. The number of observations could be adequate for estimation of a satisfactory model in the relevant frequency range. The time series models can be adapted to include auxiliary information in various forms and variable seasonal effects. But this would require modifications of the present program.

3. Caution

Quarterly values, obtained by disaggregation of annual values, are not equivalent with quarterly observations. The accumulation into annual values filters out high frequency variations in quarterly values and this information is not recovered in the disaggregation process. Another point to keep in mind is that the quarterly values produced by these methods contain information from both past and subsequent annual values.

Much econometric modelling which would be feasible with quarterly observations should not be attempted with values obtained by disaggregation. Readers with fair knowledge of time series analysis will not have much difficulty in deciding where disaggregated values might be useful. Others should seek advice before including disaggregated values in their research.

4. Spline functions

The observed accumulated values y_i are regarded as integrals of the flow $h(t)$ over unit intervals:

$$y_j = \int_{j-1}^{j} \eta(t)\,dt, \qquad j=1,2,\ldots,T. \qquad (1)$$

The flow is written as

$$h(t) = f(t)w(t) \qquad (2)$$

where the weight function $w(t)$ represents trend and possibly other variations. This function is determined separately. The function $f(t)$ is determined so that

$$\int_0^T (f^{(n)}(t))^2\,dt$$

is minimum, subject to the requirement that the observations are recovered by the integral in equation (1). The solution for interval j is

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

with local co-ordinates so that each interval is in [0,1]. The necessary equations to determine the $(2n+1)T$ coefficients a_{jk} and b_j are obtained from the following conditions:

y_j reproduced from equation (1) for $j = 1,2,\ldots,T$,
$f^{(n)}$ continuous for $n = 0,1,\ldots,2n-1$.

This provides $(2n+1)T-2n$ equations and the $2n$ remaining conditions are

$f^{(n)}(0) = f^{(n)}(T) = 0$ for $n = n,\ldots,2n-1$.

We found that the best results for economic flows were obtained with $n=1$, i.e. by minimizing the squared first derivative of $f(t)$. The basic version of the program uses the weight function

$w(t) = e^{\lambda t}$

where l is selected by the user. The equations needed to solve the coefficients a_{jk} and b_j when $n=1$ are then:

$$y_j = e^{\lambda(j-1)} \int_0^1 \left[a_{j0} + a_{j1} t + b_j e^{\lambda(j-1)} \int_0^t (1-s) e^{\lambda s} ds \right] e^{\lambda t} dt \qquad j=1,2,\ldots,T$$

$$a_{j,0} = a_{j-1,0} + a_{j-1,1} + b_{j-1} e^{\lambda(j-2)} \int_0^1 (1-s) e^{\lambda s} ds , \qquad j=2,3,\ldots,T$$

$$a_{j,1} = a_{j-1,1} + b_{j-1} e^{\lambda(j-2)} \int_0^1 e^{\lambda s} ds \qquad j=2,3,\ldots,T$$

and the end conditions

$$a_{11} = 0,$$

$$a_{T1} + b_T e^{\lambda(T-1)} \int_0^1 e^{\lambda s} ds = 0.$$

The integrals are all calculated numerically in the program. The NAG routine D01AJF(F,a,b,e1,e2,Y,….) provides the integral of the function F from a to b and delivers the result in Y. The routine D01DAF(a,b,PH1,PH2,F,abs,Y,…) calculates the integral

$$Y = \int_a^b \int_{phi1(t)}^{phi2(t)} F(t,s) ds\, dt .$$

In these equations the variables are a_{jk} and b_j. The program collects their coefficients as a 6×(3T-1) matrix. The routine F07BDF factorises the band-limited matrix and F07BEF solves the system of linear equations.

The program Spline1 has an option to include auxiliary information for disaggregation. Suppose we want to disaggregate T annual values into s values per year. The weight function is then defined as

$$w(t) = e^{\lambda t} \pi(t)^{\gamma}$$

where

$$p(t) = p(j,k); \quad j=1,2,\ldots,T; \quad k=1,2,\ldots,s$$

and $p(j,k)$ takes predetermined value in interval k of year j. The exponent γ is also predetermined and can be used to reduce the impact of $p(j,k)$. Our experience of the application of auxiliary information is rather limited and described by Guðmundsson and Elíasson (1995) and Gudmundsson (1999). There is an extensive literature of disaggregation with more emphasis on connections with observed disaggregated series (Abeysinghe and Lee, 1998; Chow and Lin, 1971).

The program Spline1s also calculates the flow by minimizing squared first derivatives, but the weight function is designed to describe seasonal variations for quarterly values by the weight function

$$w(t) = e^{lt}\{a_0 + a_1 cos\,(pt/2) + a_2 sin(pt/2) + a_3 sin(pt)\}.$$

The parameters are estimated by minimizing $\sum_{1}^{T} \left(ln(y_j / z_j)\right)^2$ where

$$z_j = \int_{j-1}^{j} w(t)dt\,.$$

This is achieved by the NAG routine E04FCF(M,N,LSQFUN,…X,FC,D,FJAC….)

which minimizes the sum of squares of M functions with respect to N parameters. The parameters are in X and the sum of squares in FC. The standard deviation of the parameters is estimated from the Jakobian matrix. F01ABF inverts a positive definite matrix.

The program Spline2 calculates flows and disaggregated values with the same weight functions as Spline1 and spline functions derived by minimizing squared values of second derivatives.

5. Fourier transforms

Another method to obtain a smooth function *f(t)* in equation (2) is to represent it by a Fourier series (Gudmundsson, 2001).

$$f(t) = A_0 + \sum_{k=1}^{(T+2n)/2-1} \left(A_k \cos 2pkt/(T+2n) + B_k \sin 2pkt/(T+2n) \right) + B_{(T+2n)/2} \sin pkt \, .$$

In order to prevent spurious high frequency variations the series is extended by 2 or 4 unobserved values (*n = 1 or 2).* This series has *T+2n* unknown coefficients A_j and B_j. Let us define

$$W_k = 2pk/(T+2n).$$

With an exponential weight function the integrals in equation (1) provide *T* linear equations in the coefficients:

$$y_j = \frac{1}{\lambda}(e^{-\lambda j} - e^{-\lambda(j-1)})A_0 +$$

$$\sum_{1}^{(T+2n)/2-1} A_k \frac{e^{\lambda j}}{\lambda^2 + \Omega_k^2} \left[\left(\lambda \cos \Omega_k j + \Omega_k \sin \Omega_k j \right) - e^{-\lambda} \left(\lambda \cos \Omega_k (j-1) + \Omega_k \sin \Omega_k (j-1) \right) \right]$$

$$+ \sum_{1}^{(T+2n)/2} B_k \frac{e^{\lambda j}}{\lambda^2 + \Omega_k^2} \left[\left(\lambda \sin \Omega_k j - \Omega_k \cos \Omega_k j \right) - e^{-\lambda} \left(\lambda \sin \Omega_k (j-1) - \Omega_k \cos \Omega_k (j-1) \right) \right].$$

The *2n* additional equations needed to define the coefficients are provided by the end conditions

$$f'(0) = f'(T) = 0 \qquad \text{for } n=1$$

and

$$f''(0) = f'''(0) = f''(T) = f'''(T) = 0 \qquad \text{for } n=2.$$

Each end condition provides one linear equation in the parameters, e.g. $f''(T)=0$ implies that

$$0 = -\sum_{k=1}^{(T+2n)/2-1} \left(A_k \Omega_k^2 \cos \Omega_k T + B_k \Omega_k^2 \sin \Omega_k T \right) .$$

The program Four provides flows and disaggregated values by this method with $n=1$ and $n=2$. There is also an option to deal with seasonal variation in quarterly observations in the solution with $n=2$. This assumes that $T=4N$ where N is the number of years. The solution has the same form and the equations derived from the integrals are unchanged. But in the equations derived from the end conditions the terms corresponding to the annual variations are left out, i.e. for $k=N+1$ and $k=2N+2$.

The subroutine F04ATF solves the system of linear equations and subroutine D01AHF(a,b,…,FUN..) calculates the integral of the function FUN from a to b.

6. Time series models

Time series models can be applied for disaggregation. Wei and Stram (1990) investigated ARIMA models for this purpose. The present approach is based on state space models, described by Harvey (1989), but modified by including a multiplicative trend. (Gudmundsson, 1999).

In the absence of auxiliary information the quarterly values y_t follow the model

$$y_t = m_{t-1} + d_{1t}, \qquad (3)$$

$$m_t = m_{t-1} + e^{l(t-1)} \; b_{t-1} + d_{2t}$$

$$b_t = b_{t-1} + d_{3t}.$$

The residuals d_{it} have variances $e^{2lt}s_1^2$, $e^{2lt}s_2^2$ and s_3^2 respectively. The annual totals are

$$Y_t = y_t + y_{t-1} + y_{t-2} + y_{t-3}.$$

Every fourth annual total is observed and for given parameters and initial values quarterly values can be estimated from these by the Kalman filter. The exponent λ is estimated directly from the observed annual values by regression.

Gudmundsson (1999) describes calculation of the initial values. The parameters s_i are obtained by maximizing the likelihood function of prediction errors of the observed values as described by Harvey (1989). Subroutine E04UCF is used for the optimization and the programming is largely dictated by it. The routine is described in the documentation of the NAG Fortran Library (Numerical Algorithms Group, 2001). The final estimation of the disaggregated values is obtained by "smoothing", a procedure described in Harvey's book.

The program has an option to include auxiliary information. The quarterly values are then defined as the product of the power of an observed series, z_t^g, and the estimated values so that

$$Y_t = \sum_{i=0}^{3} z_{t-i}^{\gamma}\, y_{t-i}\,.$$

The parameter g can be predetermined or estimated jointly with the other parameters. The value of l is adjusted so that it represents the difference between the growth of the observed annual values and z_t^g.

A further option is to include cyclical variation. The variable $\Psi_{1\tau}$ is then added to the right hand side of equation (3) and defined as

$$\begin{bmatrix} \psi_{1t} \\ \psi_{2t} \end{bmatrix} = \rho \begin{bmatrix} cos\,2\pi f & sin\,2\pi f \\ -\,sin\,2\pi f & cos\,2\pi f \end{bmatrix} \begin{bmatrix} \psi_{1t-1} \\ \psi_{2t-1} \end{bmatrix} + \begin{bmatrix} \delta_{4t} \\ \delta_{5t} \end{bmatrix}$$

where $|\rho| < 1$ and d_{4t} and d_{5t} are independent and $N(0;\sigma_4^2)$.

The estimation is based on the likelihood function of the prediction errors of the observed annual values. Unfortunately the value of g which produces the best predictions may not be the value which produces the best fit between $z_t^g\, y_t$ and the actual quarterly values.

7. Application of programs

The programs Spline1 and Spline2 need similar input files. An example is presented in Sinn.dat. The first line contains the number of aggregated values, NA, (T in the text) and the number of disaggregated values, NP. The second line contains the values of the exponent V (λ in the text) and the parameter VP, corresponding to γ in the text. Then comes a logical variable VIS. The aggregated values (y_i in the text) start in the fourth line, read in free format, separated by space, comma or new line. If VIS=t the auxiliary series P(I,J) (=p_{ij} in text) is read as NA×NP values in NA lines. The data in Sinn.dat are from Ginsburg (1973) and consist of 10 annual observations and a corresponding set of quarterly values of auxiliary information. The observed quarterly values are also included in the data file, but not used in the estimation.

The program Spline1s reads NA,NS and NP in the first line. NS is the number of seasons per year so that T=NA*NS. The next line provides values for 4 control variables for subroutine E04FCF which estimates the trend and seasonal variation by non-linear least squares. This is followed by the observed accumulated values. Example of an input file is in Seas.dat.

The program Four reads the number of years, seasons and disaggregated values per year (NA,NS and NP) in the first line. The exponent V (=λ in text) is in the second line. The third line contains the logical variable SEAS which controls whether seasonal adjustment is applied to the end conditions in the second program (with $n=2$). The value of NS=1 except when SEAS=t when NS=4. The observed values start in the 4th line. Example of an input file is in Finn.dat.

The time series program Tser reads the number of years, NA, and number of disaggregated values per year NS (=4) in the first line. (The present version of the program is only designed to disaggregate annual values into quarterly values). The next three lines contain input values for estimation of parameters. The first line has 10 values of the integer array ITEST which take values 1 if respective parameter is estimated and 0 if it is predetermined. The next line reads the vector YC with the initial values for estimated parameters or fixed values of parameters which are not estimated, depending on respective value of ITEST.

The optimization works best if the magnitude of the parameters does not differ too much and they should not be very large. The program sets an upper limit of 100 on the estimated value of σ_i but with large aggregated values this may be too low. This can be adjusted by the vector SCALE in the next line after YC. The actual value is the estimated value multiplied by respective value of SCALE. In our time series model the value of σ_3 is normally much lower than the parameters σ_1 and σ_2 so it is often suitable to use a smaller value of SCALE(3) than SCALE(1) and SCALE(2). The 4^{th} to 6^{th} parameters are for the cyclical variations, i.e. σ_4, ρ and f. The 7^{th} parameter is *g.* Example of an input file is in Tser.dat.

References

Abeysinghe, T. and Lee, C. (1998). Best linear unbiased disaggregation of annual GDP to quarterly figures: The case of Malaysia. *Journal of Forecasting, 17,* 527-537.

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

Ginsburg, V.A. (1973). A further note on the derivation of quarterly figures consistent with annual data. *Applied Statistics 22*, 268-374.

Gudmundsson.G. (1981*). Flows and derivatives of economic time series*. Presented at the European meeting of the Econometric Society, Amsterdam.

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

Gudmundsson.G. (2001). Estimation of continuous flows from observed aggregates. *The Statistician, 50*.

Guðmundur Guðmundsson and Lúðvík Elíasson (1995). Disaggregation of flow variables with multiplicative trends. *Unpublished manuscript*. Central Bank of Iceland.

Guðmundur Guðmundsson og Frank Veneklaas. Flæði og heildarframleiðsla. *Fjármálatíðindi (1976) 23*, 59-67.

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

Numerical Algorithms Group (2000). *Nag Fortran Library, Mark 19.* NAG Ltd., Oxford.

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

Wei, W.W.S. and Stram, D.O. (1990). Disaggregation of time series models. *J. R. Statist. Soc. B. 52*, 453-467.