

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/343308895>

Interpolating a low-frequency time to a high-frequency one: programing and estimation procedure for matlab

Article in *Sri Lanka Journal of Economic Research* · December 2014

DOI: 10.4038/sljer.v2i2.94

CITATIONS

2

READS

451

1 author:



Vinayagathan Thanabalasingam

University of Peradeniya

12 PUBLICATIONS 215 CITATIONS

SEE PROFILE

INTERPOLATING A LOW-FREQUENCY TIME TO A HIGH-FREQUENCY ONE: PROGRAMING AND ESTIMATION PROCEDURE FOR MATLAB

T. Vinayagathan

*Department of Economics and Statistics
University of Peradeniya*

PERSPECTIVES

Sri Lanka Journal of
Economic Research
Volume 2 (2)
December 2014: 93-108
Sri Lanka Forum of
University Economists



Abstract

This study provides an estimation procedures and statistical package programing for temporal disaggregation of time series data. That is, this method is used to disaggregate low frequency data to higher frequency data. Temporal disaggregation can be performed with one or more high frequency indicator series.

Keywords: *High frequency, Low frequency, Programming, Temporal disaggregation.*

INTRODUCTION

Data is a crucial part of responsible research. Whenever the investigators or the research teams start a new research they should concern about issues related to data. If you have a clear plan about your data at the beginning of the research, you save time and effort later on. Also, you are assured that the data you produce will be preserved in a clear, useable format.

Research data are an essential and costly output of the scholarly research process, across all disciplines. They are an important part of the evidence necessary to evaluate research results, and to reconstruct the events and processes leading to them. It is a common problem for researchers and analysts about not having a series at the preferred frequency. For instance, instead of monthly output (gross domestic product: GDP), they only have either quarterly or annual GDP. Even in some time they don't have quarterly GDP. Instead of a daily stock market index, they only have a weekly index. While there is no way to completely make up for the missing time series, there are some useful techniques. That is, using one or more high frequency data series, the low frequency series can be disaggregated into a high frequency series. For example, quarterly imports could help disaggregating annual GDP, and/or monthly investment and monthly exports could help disaggregating the annual output.

In order to maintain the reliability of research, accurate data collection is necessary regardless of the field of study or preference for defining data (quantitative, qualitative). Both the selection of appropriate data collection/disaggregation method and clearly delineated instructions for their correct use reduce the likelihood of errors occurring are essential. The primary motivation for preserving data integrity is to support the detection of errors in the data collection process, whether they are made intentionally or not. Most, Craddick, Crawford, Redican, Rhodes, Rukenbrod, and Laws (2003) explain ‘quality assurance’ and ‘quality control’¹ as two approaches that can preserve data integrity and ensure the scientific validity of study results. Each approach is implemented at different points in the research timeline (Whitney, Lind, and Wahl, 1998). Several researchers are considering above mentioned quality approaches when they interpolate the low frequency data in to high frequency data. For example, Chow and Lin (1971) and Goldberger (1962) used best linear unbiased interpolation method.

Although Litterman (1983), Fernandez (1981) and Chow-Lin (1971) use one or several indicators and perform a regression on the low frequency series, Litterman(1983) and Fernandez (1981) are dealing with non-cointegrated series while Chow and Lin (1971) suited for cointegrated series. Alternatively, Dagum and Cholette (2006) disaggregate a series without an indicator. They primarily concerned with movement preservation, generating a series that is similar to the indicator series whether or not the indicator is correlated with the low frequency series. Pavía-Miralles (2010) classifies and reviews the procedures, provides interesting discussion on the history of the methodological development in this literature and permits to identify the assets and drawbacks of each method, to comprehend the current state of art on the subject and to identify the topics in need of further development.

All of the above mentioned techniques confirm that either the first or the last value (the sum/the average) of the resulting high frequency series is consistent with the low frequency series. Sax and Steiner (2013) state that “they can deal with situations where the high frequency is an integer multiple of the low frequency (e.g. years to quarters, weeks to days), but not with irregular frequencies (e.g. weeks to months)”. The interpolation methods are widely used in official statistics packages. That is, to perform the temporal disaggregation researchers are employing different software packages. For example, R extension by Sax and Steiner (2013), Quilis (2012) used Matlab extension, RATS extension by Doan (2008), Barcellan et al. (2003) employed Ecotrim extension.

Although a very few studies (e.g. Quilis, 2012) provide a programming to interpolate low frequency data in to high frequency one using Matlab software, which is also early version

¹Quality assurance is the activities that take place *before* data collection begins and Quality control is the activities that take place *during* and *after* data collection even it is either a primary or secondary data.

Interpolating a Low-Frequency Time to a High-Frequency One of it (Matlab 7.6 [R2008a]). Therefore, in this paper we derive best linear unbiased predictor of an individual drawing of Y (low frequency series) given X (X may be one or more high frequency series) in the linear regression model using Matlab 7.14 (R2012a) version. To that aim, we are describing the estimation procedure and manual Matlab programming.

The section 2 discusses the framework of interpolation method. Section 3 presents the estimation procedure and programing using an example. Finally the key results are obtained and discussed in section 4.

THE INTERPOLATION METHOD

The purpose of interpolation is to find out an unknown high frequency series (say Y : monthly GDP), whose averages, sums, first or last values are consistent with a known low frequency series (say annual or quarterly GDP). In order to estimate monthly GDP, one or more other high frequency indicator variables can be used. We collect these high frequency series in a matrix X . Hence, monthly observations of a series can be estimated using either bivariate or multiple regression relationship. Following Chow and Lin (1971) approach, the generalized linear regression model is given by²:

$$Y = X\beta + u \quad (1)$$

where, Y is $12n \times 1$ (or $T \times 1$) vector of regress and observations, X is $12n \times K$ (or $T \times K$) matrix of regressors observations, β is $K \times 1$ vector of coefficients and $u \sim N(0, W)$. Also n is the year, K is the number of high frequency indicator series, T is month ($12 \times n$) and W is the $12n \times 12n$ (or $T \times T$) positive-definite of variance-covariance of disturbances ($W = I_{12n}\sigma^2$). For the purpose of statistical analysis, the indicator series X is going to be treated as fixed in the equation (1).

The equation (1) describes the sample period of $12n$ months relations between regressand and regressors, but we don't have the monthly series of Y instead we have annual data only. Therefore, to converts the $12n$ monthly observations into n annual observations, we need to transform this equation by multiplying compatibility matrix form. The transformed equation is given by:

$$CY = CX\beta + Cu$$

$$\dot{Y} = \dot{X}\beta + \dot{u} \quad (2)$$

with $E(\dot{u}) = 0$, $E(\dot{u}\dot{u}') = \dot{W} = CwC'$

²Throughout this study we will speak about estimating a monthly series given its annual data of that series and monthly data of indicator series. Also, we will provide the Matlab program on estimating monthly series given it annual and quarterly data of that series and monthly series of indicator series.

where, C is $n \times 12n$ matrix. In this case averages, first values or sums are consistent with a known low frequency series (i.e. Y). Therefore, for distribution and interpolation, C matrix can be takes the form as:

$$C_A = \frac{1}{12} \begin{bmatrix} 111111111111 & 000000000000 & 000000000000 & \dots & 000000000000 \\ 000000000000 & 111111111111 & 000000000000 & \dots & 000000000000 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 000000000000 & 000000000000 & 000000000000 & \dots & 111111111111 \end{bmatrix}$$

$$C_S = \begin{bmatrix} 111111111111 & 000000000000 & 000000000000 & \dots & 000000000000 \\ 000000000000 & 111111111111 & 000000000000 & \dots & 000000000000 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 000000000000 & 000000000000 & 000000000000 & \dots & 111111111111 \end{bmatrix}$$

$$C_F = \begin{bmatrix} 100000000000 & 000000000000 & 000000000000 & \dots & 000000000000 \\ 000000000000 & 100000000000 & 000000000000 & \dots & 000000000000 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 000000000000 & 000000000000 & 000000000000 & \dots & 100000000000 \end{bmatrix}$$

Where C_A represents that the averages values are consistent with regressand, C_S denotes that the sums of monthly values are consistent with regressand and C_F indicate that first month value is consistent with the regressand. For the temporal disaggregation we can use any one of these three alternatives.

Since all the data series are in annual basis in equation (2), now we will able to estimate this model by ordinary least square (OLS) method. The estimator of $\hat{\beta}$ is given by:

$$\hat{\beta} = (\dot{X}'\dot{w}^{-1}\dot{X})^{-1}\dot{X}'\dot{w}^{-1}\dot{Y} \quad (3)$$

Where, \dot{X} , \dot{Y} are based on the annual data, and $\dot{w} = CC'\sigma^2$

Now the problem is, how to estimate the monthly observations on the dependent variables. To that aim, now assume that we estimate a vector of Z , which is identical with Y in the case of temporal disaggregation. Therefore, the regression model is given by:

$$Z = Y = X_Z\beta + u_Z \quad (4)$$

where, X_Z and u_Z are identical with X and u of the equation (1) for the interpolation and distribution. Using some $m \times p$ matrix A , a linear unbiased estimator \hat{z} of z satisfies as:

$$\hat{Z} = A\dot{Y} = A(\dot{X}\beta + \dot{u}) \quad (5)$$

where,

$$A = M\dot{X}'\dot{w}^{-1} + \dot{w}_Z\dot{w}^{-1}$$

$$M = X_Z(\dot{X}'\dot{w}^{-1}\dot{X})^{-1} - (\dot{w}_Z\dot{w}^{-1})\dot{X}(\dot{X}'\dot{w}^{-1}\dot{X})^{-1}$$

Hence,

$$A = \left[X_Z(\dot{X}'\dot{w}^{-1}\dot{X})^{-1} - (\dot{w}_Z\dot{w}^{-1})\dot{X}(\dot{X}'\dot{w}^{-1}\dot{X})^{-1} \right] \dot{X}'\dot{w}^{-1} + \dot{w}_Z\dot{w}^{-1}$$

After solving this, the estimated value of Z is:

$$\hat{Z} = A\dot{Y} = x_Z\hat{\beta} + (\dot{w}_Z\dot{w}^{-1})\hat{u} \quad (6)$$

$$\text{where} \quad \hat{u} = \dot{Y} - \dot{X}\hat{\beta} \quad (7)$$

For the temporal disaggregation, we assume that

$$\dot{w}_Z = E(\dot{u}\dot{u}') = C'\sigma^2$$

where the definition of C is given by either C_F or C_A or C_S .

As in Chow and Lin, to compute the estimated equation of (6) without any difficulties, we took three assumptions. The first case is to assume that monthly regression residuals are serially uncorrelated. In this case, to estimate the monthly observation of the dependent variable, one can easily assume that the term $(\dot{w}_Z\dot{w}^{-1})\hat{u} = C'_F\hat{u}$ in equation (6) assigns the regression residual for any year to the first month of that year. On the other hand if we assume that the term $(\dot{w}_Z\dot{w}^{-1})\hat{u} = C'_A\hat{u}$ in equation (6) assigns the regression residual for any year to the average values of 12 month of that year. Moreover, if we assume that the term $(\dot{w}_Z\dot{w}^{-1})\hat{u} = C'_S\hat{u}$ in equation (6) assigns the regression residual for any year to the summation of 12 month of that year. Second case is to assume that monthly residuals follow first order auto-regression

$$u_t = \alpha u_{t-1} + \varepsilon_t \text{ with } E(\varepsilon_t \varepsilon_s) = \theta_{ts} \sigma^2 \quad (7)$$

In this case, for the interpolation we needs

$$\dot{w}_Z\dot{w}^{-1} = wC'(CwC')^{-1} = AC'(CAC')^{-1} \quad (8)$$

In the third case, to estimate the w, we assume that although monthly residual series are serially uncorrelated, but variances are proportional to a certain linear combination of independent variables or to a known function of a regressors. Then, w will be the proportional and diagonal to a given matrix.

THE ESTIMATION PROCEDURE AND PROGRAMMING

Let assume that we have annual data of gross domestic product (GDP), which is taken as dependent variable (named as Y). Let X be monthly data matrix, which is considered as explanatory variables. The matrix X consists of inflation rate, interest rate, exchange rate and narrow money supply. These all data are representing the Sri Lanka's economic situation. GDP covers the period of 1978–2011 and all the explanatory variables cover the period of 1978M₁ –2011M₁₂. The data come primarily from International Financial Statistics (IFS) for the following variables: inflation rate computed as Colombo consumer price index in 2005 based year; interest rate (IR), i.e. Interbank call money market rate in percentage, exchange rate (ER) measured by per U.S. dollar; narrow money (M0) in million U.S. dollar. The data of GDP in U.S. dollar in constant 2000 prices come from World Development Indicator (WDI) database.

Following the methodology described in the section 2, now we start to write a Matlab programing to interpolate GDP data from annual series to monthly series.

Before starting the programing, we should have data in proper format. Hence, since number of observations is differing between dependent and explanatory variables keep two separated excel/txt data file; one for dependent variable (low frequency data series) and another for explanatory variable/s (high frequency indicator series).

First, we have to load the data file in to Matlab software. To do that, follow as below:

```
clear all; % will clear the memory of the work file
           and start freshly
format bank; % or you can write "format short"/ "format long"
              as well
[num, txt, raw] = xlsread('explanatory.xlsx'); % num is initialized with all the numbers
                                                % txt is initialized with all the text
                                                % raw is a cell matrix with all the numbers & text
numbers=cell2mat(raw(2:end,2:end)); % This returns the matrix with all the numbers
headings=cell2mat(raw(1:1,1:end)); % This returns the headings of your matrix
text=raw(2:end,1); % This returns the first column.
data1=numbers(:,:); % Defining the monthly data as a matrix
data2=xlsread('gdp.xlsx'); % loading the annual data to Matlab file
```

Second, we are assigning the row and column for data series. To do that, follow as below:

```
[p1 q1]=size(data1); % defining the row and column for data1
[p2 q2]=size(data2); % defining the row and column for data2
```

Third, define the dependent and independent variables as in matrix/vector form:

```

x=data1(:,:);
y=data2(:,2);

```

% defining explanatory variables as a matrix
 % defining the dependent variable as a vector

Fourth, since variables y and x has different frequencies, now we will construct matrix “C” to convert the regression model (1) from monthly to annual to maintain consistent number of observation. In this case, we use C_F matrix, which first month value of a particular year is consistent the annual series.

```

n=34;
m=34*12;
C_F=zeros(n,m);

j=1;
for i=1:n
    C_F(i,j)=1;
    j=j+12;
end

```

% defining the total number of year
 % defining the total number of month
 % preallocating zero value to C matrix

Fifth, converting the eq. (1) as (2): monthly to annual series (particularly for x variables)

```

X=C_F*x;
Y=y;
sigma_sq=eye(n);

C=C_F*C_F';
W=C*sigma_sq;

```

% assume that $X=x.C_F*x$
 % assume that $Y=y$.
 % to define the variance-covariance matrix of \hat{u}
 (i.e. \hat{w}) we first estimate σ^2 by preallocating
 n-by-n identity matrix
 % defining the variance-covariance matrix of \hat{u}
 ($W = \hat{w}$)

Finally, we estimate equation (3) and (6) to obtain the interpolated series.

```

b_hat=inv(X'*inv(W)*X)*(X'*inv(W)*Y);
U_hat=Y-(X*b_hat);
z_hat=x*b_hat;
W_z=C_F'*sigma_sq;
(W_z*inv(W))*U_hat==C_F'*U_hat;
z_hat=x*b_hat+(W_z*inv(W))*U_hat;

```

% this returns the $\hat{\beta}$ from eq (3).
 % to estimate the eq (6), first we need to
 estimate residual. Assume that $\hat{U} = \hat{u}$
 % here we multiply the b_hat by x matrix not
 X matrix to get the monthly series
 % estimating the first part of eq. (6) this
 equation is based on monthly series]
 % estimating and defining second part of eq
 (6). Where $(\hat{w}_z \hat{w}^{-1}) \hat{u} = c_F' \hat{u}$ [$\hat{w}_z = W_z$, $\hat{w} =$
 W , $\hat{u} = \hat{U}$]
 % return the interpolated GDP data from eq.(6)

Now we check whether the interpolated series (first month value of that year) and original data (annual data) are consistent using following Matlab commands.

```
figure;
plot(z_hat(1:12:end),'*')           % this plot the interpolated data
hold on;
plot(data2(:,2),'r')                % this Plot the initial data
hold off;
xlabel('Year','FontSize',13)
ylabel('GDP','FontSize',13)
legend('Interpolated Data', 'Original Data')
```

% To obtain the interpolated data only (whole monthly value not the first monthly value of the each year) we use below commands.

```
figure;
plot(z_hat(:,1),'*')                % This plot the interpolated data
xlabel('Year','FontSize',13)
ylabel('GDP','FontSize',13)
legend('Interpolated Data')
title('Interpolated data of Sri Lankan GDP:1978M1-2011M12')
```

Alternative Method I

In this section, we use monthly average value which should be consistent with annual series using C_A matrix.

First, we load the .xlsx. data file to Matlab using following commands.

```
clearall;
formatbank;
[num, txt, raw] = xlsread('explanatory.xlsx'); % num is initialized with all the numbers
                                                % txt is initialized with all the text
                                                % raw is a cell matrix with all the numbers & text
numbers=cell2mat(raw(2:end,2:end)); % This returns the matrix with all the numbers
headings=cell2mat(raw(1:1,1:end)); % This returns the headings of your matrix
text=raw(2:end,1); % This returns the first column.
data1=numbers(:,:); % Defining the monthly data as a matrix
data2=xlsread('gdp.xlsx'); % loading the annual data to Matlab file
```

Second, we assign the row and column for data series using below commands.

```
[p1 q1]=size(data1); % defining the row and column for data1
[p2 q2]=size(data2); % defining the row and column for data2
```

Third, define the dependent and independent variables as in matrix/vector form:

```
x=data1(:,:); % defining explanatory variables as a matrix
y=data2(:,2); % defining the dependent variable as a vector
```

Fourth, we will construct C_A matrix, in which average monthly value is equal to low frequency data, to convert the eq. (1) as annual series equation.

```

n=34;                                % defining the total number of years
m=34*12;                             % defining the total number of month
C=zeros(n,m);                        % preallocating zero value to C_A matrix

j=1;
for i=1:n
    C(i,j)=1;
    C(i,j+1)=1;
    C(i,j+2)=1;
    C(i,j+3)=1;
    C(i,j+4)=1;
    C(i,j+5)=1;
    C(i,j+6)=1;
    C(i,j+7)=1;
    C(i,j+8)=1;
    C(i,j+9)=1;
    C(i,j+10)=1;
    C(i,j+11)=1;
    j=j+12;
end

C_A=(1/12).*C;                      %this returns the average value of 12 months
                                     %which is consistent with annual data

```

Fifth, converting the eq. (1) as (2): monthly to annual series (particularly for x variables using C_A matrix).

```

X=C_A*x;                            % assume that  $X=x.C\_A$ 
Y=y;                                % assume that  $Y=y$ .
sigma_sq=eye(n);                    % to define the variance-covariance matrix of
                                      $\dot{u}$ (i.e.  $\dot{w}$ ) we first estimate  $\sigma^2$  by preallocating n-
                                     by-n identity matrix

C=C_A'*C_A;
W=C*sigma_sq;                       % defining variance-covariance matrix of  $\dot{u}$ 
                                     ( $W = \dot{w}$ )

```

Finally, estimating the equation (3) and (6) to obtain the interpolated series

```

b_hat=inv(X'*inv(W)*X)*(X'*inv(W)*Y); % this returns the  $\hat{\beta}$  from eq (3).
U_hat=Y-(X*b_hat);                  % to estimate the eq. (6), first we need to
                                     estimate residual. Assume that  $\hat{U} = \hat{\hat{u}}$ 

```

```

z_hat=x*b_hat; % here we multiply the b_hat by x matrix not X
                % matrix to get the monthly series
W_z=C_A'*sigma_sq; % estimating the first part of eq. (6) [this
                    % equation is based on monthly series]
(W_z*inv(W))*U_hat==12*C_A'*U_hat; % estimating and defining second part of
                                   % eq. (6). Where  $(\hat{w}_z \hat{w}^{-1}) \hat{u} = c'_A \hat{u}$  [ $\hat{w}_z =$ 
                                   %  $W_z$ ,  $\hat{w} = W$ ,  $\hat{u} = \hat{U}$ ]
z_hat=x*b_hat+(W_z*inv(W))*U_hat; % this returns the interpolated GDP data from
                                   % eq.(6)

```

Now we check whether the interpolated series (average month value of that year) and original data (annual data) are consistent using following Matlab commands.

```

figure;
plot(z_hat(1:12:end),'*') % This plot the interpolated data
hold on;
plot(data2(:,2),'r') % This Plot the initial data
hold off;
xlabel('Year','FontSize',13)
ylabel('GDP','FontSize',13)
legend('Interpolated Data', 'Original Data')

```

% To obtain the interpolated data only (considering normal monthly value not average monthly value of the each year) we use below commands.

```

figure;
plot(z_hat(:,1),'*') % This plot the interpolated data
xlabel('Year','FontSize',13)
ylabel('GDP','FontSize',13)
legend('Interpolated Data')
title('Interpolated data of Sri Lankan GDP:1978M1-2011M12')

```

Alternative Method II

In this section, we attempt to make sum of the monthly value is equal to the annual series rather than average value or first value of the month using C_S matrix.

```

clear all;
format bank;

```

First, we load the `xlsx` data file to Matlab using following commands.

```

[num, txt, raw] = xlsread('explanatory.xlsx'); % num is initialized with all the numbers
                                                % txt is initialized with all the text
                                                % raw is a cell matrix with all the numbers & text

```

Interpolating a Low-Frequency Time to a High-Frequency One

```

numbers=cell2mat(raw(2:end,2:end));% This returns the matrix with all the numbers
headings=cell2mat(raw(1:1,1:end)); % This returns the headings of your matrix
text=raw(2:end,1);                % This returns the first column.
data1=numbers(:,:);               % Defining the monthly data as a matrix
data2=xlsread('gdp.xlsx');        % Loading the annual data to Matlab file

```

Second, we assign the row and column for data series using below commands.

```

[p1 q1]=size(data1);              % defining the row and column for data1
[p2 q2]=size(data2);              % defining the row and column for data2

```

Third, define the dependent and independent variables as in matrix/vector form:

```

x=data1(:,:);                     % defining explanatory variables as a matrix
y=data2(:,2);                     % defining the dependent variable as a vector

```

Fourth, we will construct C_S matrix, in which sum of monthly value is consistent with annual series, to convert the eq. (1) as annual series equation.

```

n=34;                             % defining the total number of year
m=34*12;                           % defining the total number of month
C_S=zeros(n,m);                    % preallocating zero value to C matrix

```

```

j=1;
for i=1:n
    C_S(i,j)=1;
    C_S(i,j+1)=1;
    C_S(i,j+2)=1;
    C_S(i,j+3)=1;
    C_S(i,j+4)=1;
    C_S(i,j+5)=1;
    C_S(i,j+6)=1;
    C_S(i,j+7)=1;
    C_S(i,j+8)=1;
    C_S(i,j+9)=1;
    C_S(i,j+10)=1;
    C_S(i,j+11)=1;
    j=j+12;
end

```

Fifth, converting the eq (1) as (2): monthly to annual series (particularly for x variables using C_S matrix).

```

X=C_S*x;                          % assume that X=x.=C_S*x
Y=y;                              % assume that Y=y.

```

```
sigma_sq=eye(n); % to define the variance-covariance matrix of  $\hat{u}$ 
                    (i.e.  $\hat{w}$ ) we first estimate  $\sigma^2$  by preallocating
                    n-by-n identity matrix
```

```
C=C_S*C_S';
W=C*sigma_sq; % defining the variance-covariance matrix of  $u$ 
               dot (i.e.  $w$  dot)
```

Finally, estimating the eq. (3) and (6) to obtain the interpolated series

```
b_hat=inv(X'*inv(W)*X)*(X'*inv(W)*Y); % this returns the  $\hat{\beta}$  from eq (3)
U_hat=Y-(X*b_hat); % to estimate the eq. (6), first we need to
                    estimate residual. Assume that  $\hat{U} = \hat{u}$ 
z_hat=x*b_hat; % here we multiply the  $\hat{\beta}$  by "x" matrix not "X"
               matrix to get the monthly series
```

```
W_z=C_S'*sigma_sq; % estimating the first part of eq. (6)[this equation
                    is based on monthly series]
```

```
(W_z*inv(W))*U_hat==12*C_S'*U_hat; % estimating and defining second part of eq.
                                     (6). Where  $(\hat{w}_z \hat{w}^{-1}) \hat{u} = c'_s \hat{u}$  [ $\hat{w}_z = W_z$ ,
                                      $\hat{w} = W$ ,  $\hat{u} = \hat{U}$ ]
z_hat=x*b_hat+(W_z*inv(W))*U_hat; % this returns the interpolated GDP data from
                                     eq. (6)
```

% now we Sum the interpolated monthly series to annual

```
k=1;
count=1;
for i=12:12:size(z_hat,1) % i iterates from 1,12,24 etc.
    result(k)=sum(z_hat(count:i));
    k=k+1; % incrementing k from 0->34
    count=i+1; % this is used to keep the previous value.
end
b=[result]' % result vector is composed of the additions of
             the 34 arrays
```

Now we check whether the interpolated series (sum of the month value) and original data (annual data) are consistent using following Matlab commands:

```
figure;
plot(b(:,1),'*') % this plot the interpolated data
hold on;
plot(data2(:,2),'r') % this plot the initial data
hold off;
```

```
xlabel('Year','FontSize',13)
ylabel('GDP','FontSize',13)
legend('Interpolated Data', 'Original Data')
```

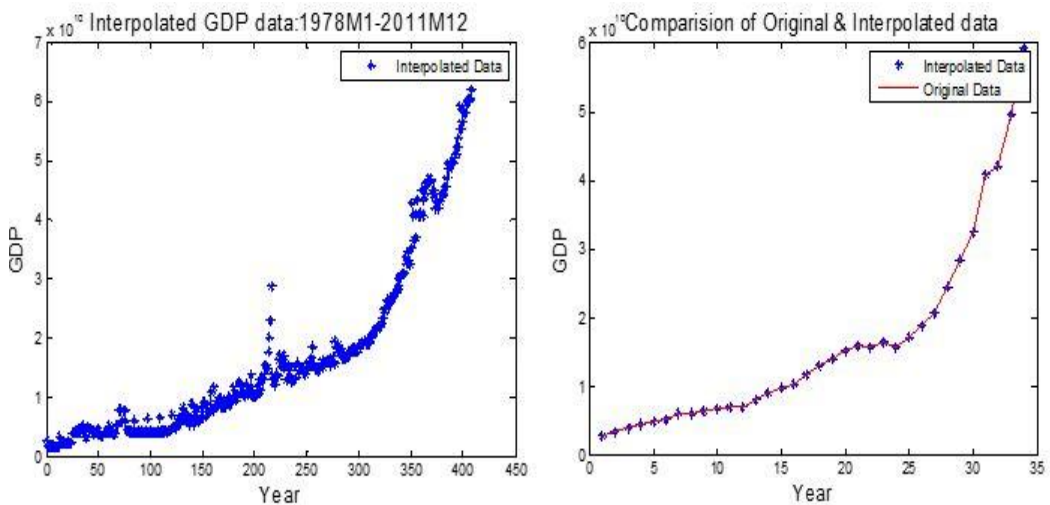
Next to obtain the interpolated data only (considering normal monthly value sum of the monthly value of the each year) we use below commands:

```
figure;
plot(z_hat(:,1),'*') % this plot the interpolated data
xlabel('Year','FontSize',13)
ylabel('GDP','FontSize',13)
legend('Interpolated Data')
title('Interpolated data of Sri Lankan GDP:1978M1-2011M12')
```

ESTIMATION RESULTS

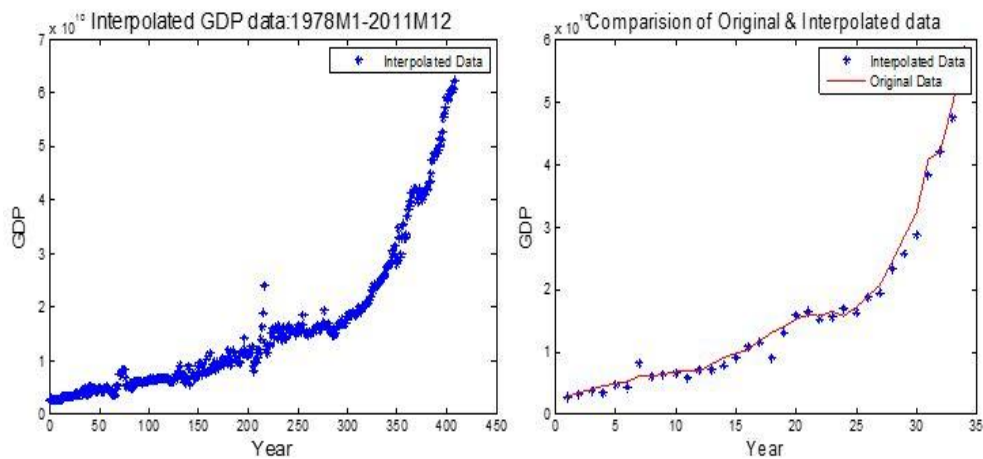
We used three methods such as first month value, average monthly value and sum of the monthly value to disaggregate low frequency data into high frequency data following Chow and Lin (1971) approach. There is indeed one to one relationship between initial data and interpolated data among three methods. The results are given below:

Figure 1: Temporal Disaggregation Using C_F matrix



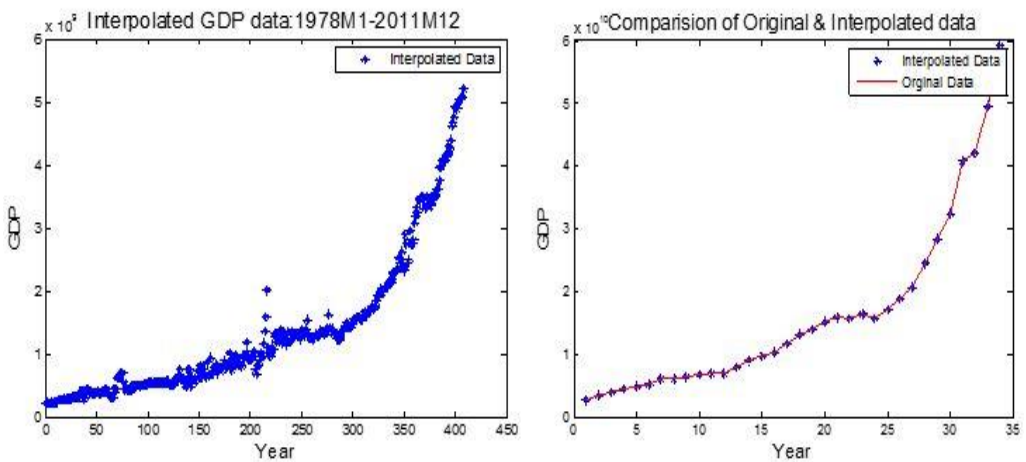
Source : Authors' Estimations

Figure 2: Temporal Disaggregation Using C_A matrix



Source : Authors' Estimations

Figure 3: Temporal Disaggregation Using C_S matrix



Source : Authors' Estimations

1st panel of the Figure 1, 2, and 3 shows the monthly data of GDP from 1978M1 to 2011M12

After the temporal disaggregation and 2nd panel represents the comparison of original low frequency data (i.e. original annual GDP) and interpolated GDP data. According to the 1st panel, we can observe the similar pattern among three methods regardless of the matrix forms of C , that is either first month value (C_F) or average monthly value (C_A) or sum of the monthly value (C_S), that we used to convert regression model (1) to (2). This implies

that our estimates satisfy the Best Linear Unbiased property as shown in Chow and Lin (1971).

In our example, we actually know the true data on annual GDP, so we can compare the interpolated values to the true values. With an indicator series, Chow-Lin procedures produce the series with one to one relationship among all three methods (Panel 2 of Figure 1, 2 and 3). This is, of course, due to fact that in this example, our estimate satisfies the Best Linear Unbiased property.

CONCLUDING AND REMARKS

This study attempt offers estimation procedure and Matlab programing to disaggregate a low frequency time series into a higher frequency series, using either first or the average, or the sum value of the resulting high frequency series is consistent with the low frequency series. Although temporal disaggregation can be performed with or without the help of one or more high frequency indicators, here we used more than one high frequency indicator series to the disaggregation. If good indicators are and estimation procedures are at hand, the resulting series may be close to the true series.^{2nd} panel of Figures 1, 2 and 3 proof this statement and we found one to one relationship between resulting series and true series among all three method suggesting that empirical researchers can use one of these method to disaggregate their data from low frequency to high frequency and proceed their work.

References:

- Barcellan, R., DiFonzo, T., Raffaele, D., Staplehurst, V. and Buono, D. (2003). Ecotrim: A Program for temporal disaggregation of time series.
URL <https://circabc.europa.eu/w/browse/c6049bc0-c633-4cab-9811-b476ffe08370>.
Version 1.01.
- Chow, G. C. and Lin, A. L.(1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4):372–375, pp 80-83.
- Dagum, E. B. and Cholette, P. A. (2006). Benchmarking, temporal distribution & reconciliation methods for time series. *Lecture Notes in Statistics*. Springer-Verlag, New York, pp 80-82.
- Doan, T. (2008). Disaggregate: A General procedure for interpolation.
URL www.estima.com/procs_perl/disaggregate.src. RATS library version.

- Fernández, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, 63(3): 471–476.
- Goldberger, A. S. (1962). Best linear unbiased prediction in generalized linear regression model. *Journal of the American Statistical Association*, 57(298): 369-375.
- Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics*, 1(2):169–173.
- Most, M.M., Craddick, S., Crawford, S., Redican, S., Rhodes, D., Rukenbrod, F., and Laws, R. (2003). Dietary quality assurance processes of the DASH-Sodium controlled diet study. *Journal of the American Dietetic Association*, 103(10): 1339-1346.
- Pavía-Miralles, J. M. (2010). A survey of methods to interpolate, distribute and extrapolate time series. *Journal of Service Science and Management*, 449-463.
- Quilis, E. M. (2012). Temporal Disaggregation Library.
URL [www.mathworks.com/matlabcentral/fileexchange/24438-Matlab library version](http://www.mathworks.com/matlabcentral/fileexchange/24438-Matlab-library-version).
- Sax, C. and Steiner, P. (2013). tempdisagg: Methods for temporal disaggregation and interpolation of time series,
URL <http://CRAN.R-project.org/package=tempdisagg>. R package version 0.22.
- Sax, C., and Steiner, P. (2013). Temporal disaggregation of time series. *The R Journal*, 5(2):80–87.
- Whitney, C.W., Lind, B.K., and Wahl, P.W. (1998). Quality assurance and quality control in longitudinal studies. *Epidemiologic Reviews*, 20(1): 71-80.