



This module is part of the

# Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

# Method: Chow-Lin Method for Temporal Disaggregation

## Contents

General section.....	3
1. Summary .....	3
2. General description of the method .....	3
2.1 Method.....	3
2.2 Properties.....	4
3. Preparatory phase .....	6
4. Examples – not tool specific.....	6
4.1 Example: The Chow-Lin Regression based approach for the multivariate case .....	6
4.2 Example: The Chow-Lin Regression based approach for the univariate case .....	8
5. Examples – tool specific.....	9
6. Glossary.....	10
7. References .....	10
Specific section.....	11
Interconnections with other modules.....	13
Administrative section.....	14

## General section

### 1. Summary

The Chow-Lin method is a technique used for temporal disaggregation or also known as temporal distribution. Temporal disaggregation is the process of deriving high frequency data (e.g., monthly data) from low frequency data (e.g., annual data).

In addition to the low-frequency data, the Chow-Lin method also uses indicators on the high-frequency data, which contain the short-term dynamics of the time series under consideration. These indicators are time series that are related to the target time series and thus measure a different topic than the time series to be estimated. Since the results of the Chow-Lin method depend on information on a different variable, the method can be considered as an indirect approach. The main goal of temporal disaggregation techniques such as the Chow-Lin method is to create a new time series that is consistent with the low frequency data while keeping the short-term behaviour of the higher frequency indicator series. The Chow-Lin method may be applied to time series, generally aided by one (univariate case) or more indicator series (multivariate case). Presumably, these indicator series should be socio-economic variables deemed to behave like the target variable. In absence of such variables, functions of time can be used, as proposed by Chow and Lin (1976).

Temporal disaggregation is closely related to benchmarking, another data integration technique, see for instance the module “Macro-Integration – Denton’s Method”. The main distinction is that in benchmarking, the sub-annual series to be benchmarked consists of the same variable as the annual benchmarks, while in temporal disaggregation the sub-annual series differ from the annual series.

### 2. General description of the method

In this section we give a non-technical description of the Chow-Lin method. For a more elaborate and technical explanation we refer to Dagum and Cholette (2006) and Chow and Lin (1971).

#### 2.1 Method

Temporal disaggregation is the process of deriving consistent high frequency data from low frequency data. In the past few decades, different methods for disaggregating low frequency data to high frequency data have been developed. These methods can be classified into two types of approaches.

- Models developed without indicator series but which rely upon purely mathematical criteria or time series models to derive a smooth path for the unobserved series;
- Models based on indicator series observed at the desired higher frequency.

In the second case it is assumed that one or more correlated high frequency series are available. The approach includes: the procedure proposed by Denton (1971), which is an adjustment method that does not rely on any statistical model, and ‘optimal’ methods proposed by Chow and Lin (1971) and further developed by Fernandez (1981) and Litterman (1983), which use the best linear unbiased estimator, given some assumed model (Islam, 2009).

The Chow-Lin method is a temporal disaggregation technique that uses a (statistical) relationship between low frequency data and higher frequency indicator variables. This is done by a univariate

regression in the case with one indicator and a multivariate regression if two or more indicators are used. The regression coefficients are estimated at the low frequency level, at this level the target time series and the indicator time series are both available. The indicator series are assumed to be at the high-frequency level, but these data can easily be transformed to the low frequency level, just by aggregating over the high frequency periods.

The results of the regression are used to estimate the high-frequency target series from the high-frequency indicator series. The univariate as well as the multivariate regression consists of two parts. The first part is a normal regression which estimates the monthly values. These values may not be consistent over time with the original low frequency data because a simple regression does not account for consistency. To make up for possible inconsistency, the second part of the regression corrects the estimated monthly values.

## 2.2 Properties

We can classify time series into three different types: stocks series, flows series and index series.

Flows series measure how much of something has happened over a period of time, for instance exports, production and GDP. Stock series measure a quantity existing at some specific time point, for instance inventories of some good. Temporal disaggregation applied on stock variables is also known as *interpolation*, while the application of the method on flow variables is called *distribution*, see also Chow and Lin (1976).

In this article we only discuss flows series. For these series to be disaggregated it is necessary to do this consistently in the sense that temporal additivity is observed. This basically means that the sum of the three months of the (estimated) disaggregated series must add up to the value of the associated quarter of the original series. An illustration of the additivity constraint is displayed in Figure 1 and Table 1 using index of manufacturing as high frequency indicator.

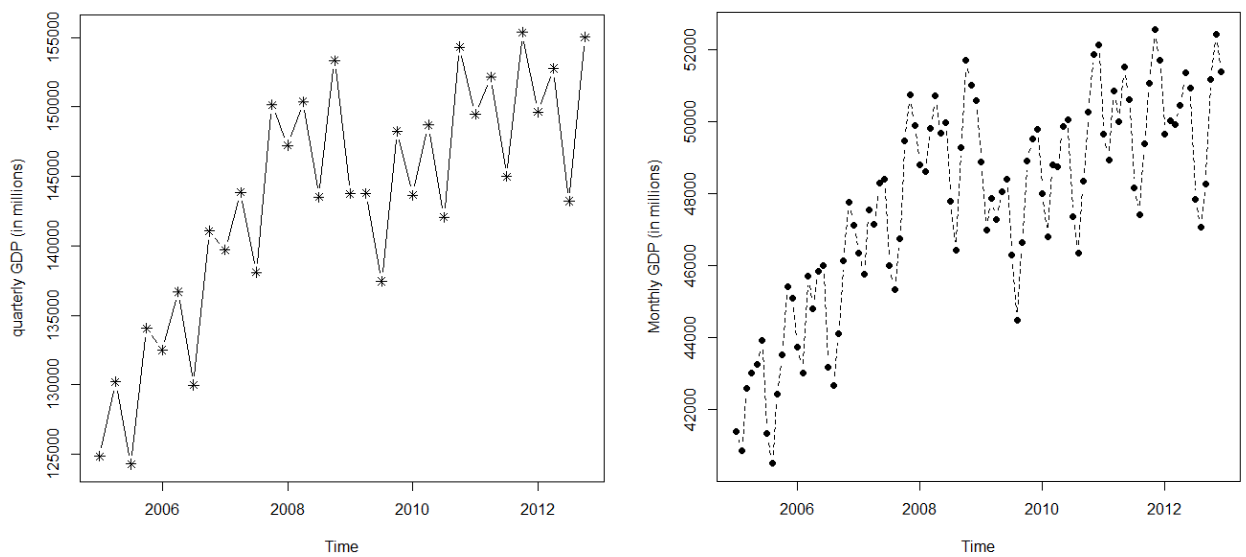


Figure 1. Plots of the quarterly GDP values (left) and estimated monthly values (right)

Table 1. Values of the estimated monthly GDP and quarterly GDP. **Bold** = added quarterly estimates

	Estimated monthly GDP	Quarterly GDP
Jan-05	41398,03	
Feb-05	40865,45	
Mar-05	42587,53	
<b>2005 Q1</b>	<b>124851</b>	124851
Apr-05	43023,87	
May-05	43256,36	
Jun-05	43932,77	
<b>2005 Q2</b>	<b>130213</b>	130213

Table 1 shows the values of the first two quarters in 2005 for the GDP. Also, the estimated monthly GDP values are shown. It is clear that the monthly values add up to the quarterly values.

There is a number of reasons why the Chow-Lin method (or any other disaggregation method) is needed. Quantitative analyses performed by for example National Statistic Institutes (NSIs) rely on statistical data. These data are generally obtained from sample surveys. Due to the need of large resources and high costs to conduct these surveys, NSIs could choose to do this only a few times per year, for example quarterly (low frequency). On the other hand, for efficient statistical and economic analysis and timely decision-making, a higher frequency may be required (Chamberlin, 2010).

Furthermore, low frequency data sources are more precise and describe the long-term movements better than high frequency sources, but the latter provide the only information on the short-term movements. By applying a temporal disaggregation technique one combines the strengths of both frequency types.

There are several variants of the Chow-Lin method. The difference between these variants is the estimation of the covariance matrix of the residuals, which in reality is often unknown. The residuals are defined as the difference between the actual monthly and estimated monthly values. Chow and Lin (1971) propose two assumptions to estimate the covariance matrix.

- There is no serial correlation amongst the residuals. This means that when the monthly values are estimated, the levels of the high frequency indicator are used.
- The residuals are serially correlated, which means that the changes and fluctuations of the high frequency indicators are taken in the monthly estimates instead of the level.

The unknown covariance matrix needs to be estimated and one can estimate this matrix based on the first or second assumption dependent on empirical evidence or presumption of the user. The assumption of no serial correlation in the residuals of monthly estimates is generally not supported by empirical evidence. Chow and Lin propose a method to estimate the covariance matrix under the second assumption that the residuals follow a first order autoregressive AR(1) process. We will not elaborate here, as it is out of the scope of this text, but this covariance matrix is eventually used to estimate high frequency values. A slightly other covariance matrix will cause different high frequency values, although the additivity constraint will still hold independent of the variants used. In short, if the user presumes that the residuals are serially correlated which implies that the fluctuations are more important than the levels, one should use an AR(1) model for the residuals. If not, the residuals need

not be modelled and so the Chow-Lin method boils down to Ordinary Least Squares (a normal regression).

### 3. Preparatory phase

### 4. Examples – not tool specific

#### 4.1 Example: The Chow-Lin Regression based approach for the multivariate case

To illustrate the Chow-Lin method for the multivariate case, we use a dataset obtained from Statistics Netherlands. Our dependent variable will be the Dutch GDP measured quarterly from 2005 till 2012. We will use three monthly indicators as explanatory variables where all three have higher frequency (monthly); index of manufacturing, inflation rate and the unemployment rate, as one may assume a correlation between these variables and GDP.

Hence we have three times more indicator observations than GDP observations. Before we continue this example, we want to stress that we do not use this technique at Statistics Netherlands. This example is purely for illustration of the Chow-Lin method.

In Figure 2 the plots of the GDP and the other monthly economic indicators are shown. The results of the regression used to derive the monthly GDP estimates are stated in Table 2. The unemployment rate and index of manufacturing are able to explain a lot more than the inflation as the latter is statistically not significant.

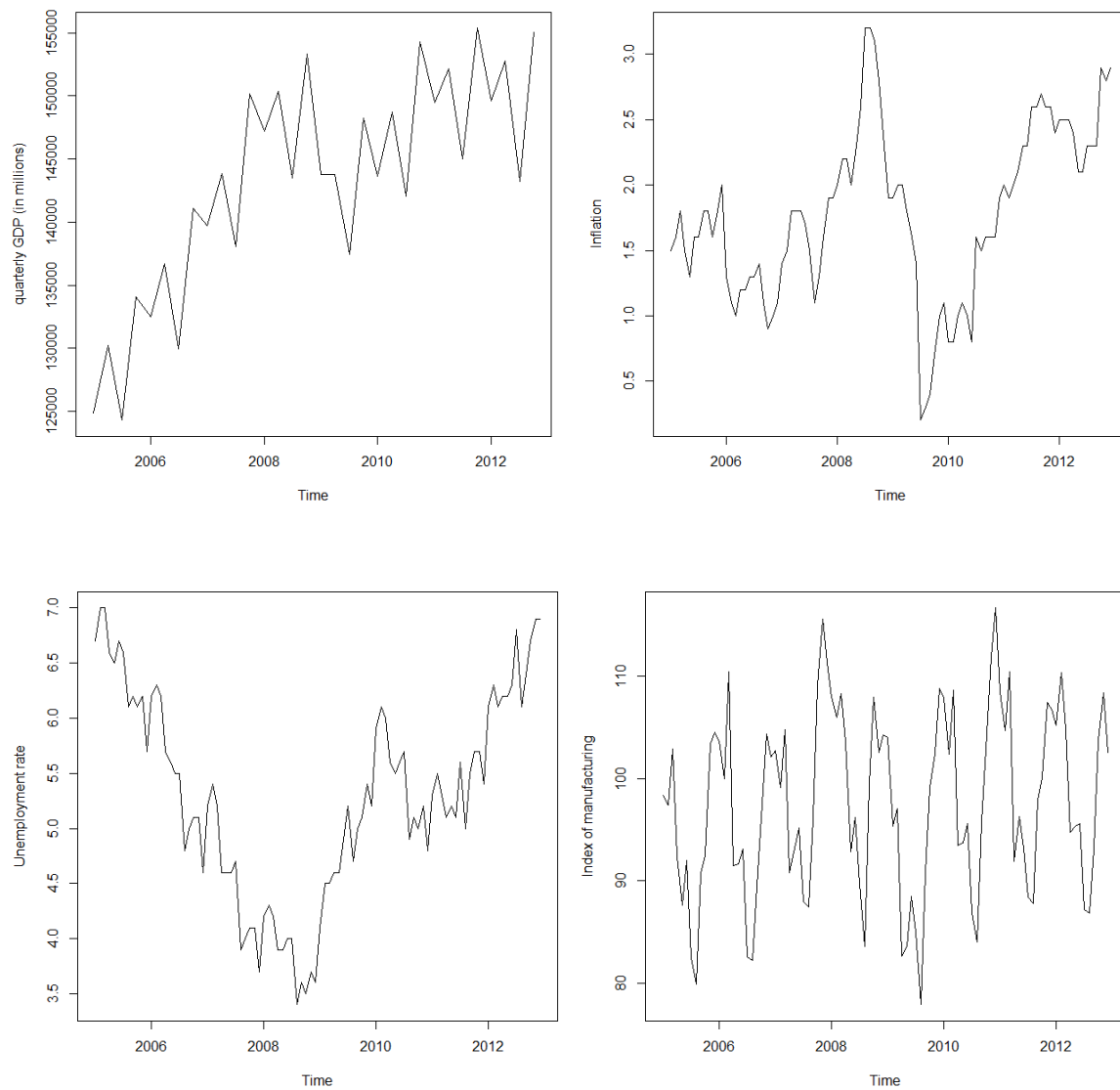
Table 2. Results regression

	Estimate	Standard Error	t-statistic
<i>Intercept**</i>	36751.07	5176.65	7.099
Inflation	1177.68	728.31	1.617
Unemployment *	-1246.98	560.83	-2.223
Index of Manufacturing**	158.83	43.25	3.672

\*Significance level  $\alpha = 0.05$ , \*\*  $\alpha = 0.01$

Figure 3 presents the monthly GDP estimates from the Chow-Lin regression approach, where we choose to use an AR(1) model to allow for serial correlation in the residuals as proposed by Chow and Lin (1971). In this example all three explanatory variables displayed in Table 2 are used, although inflation could have been left out as it is insignificant.

For comparison purposes, the original quarterly series (from Figure 1) is also plotted in Figure 3. It is clear that both frequencies approximately have the same behaviour but the monthly estimates show more detail in terms of fluctuations in the short term behaviour.



*Figure 2. Plots of the low frequency (GDP) and high frequency series.*

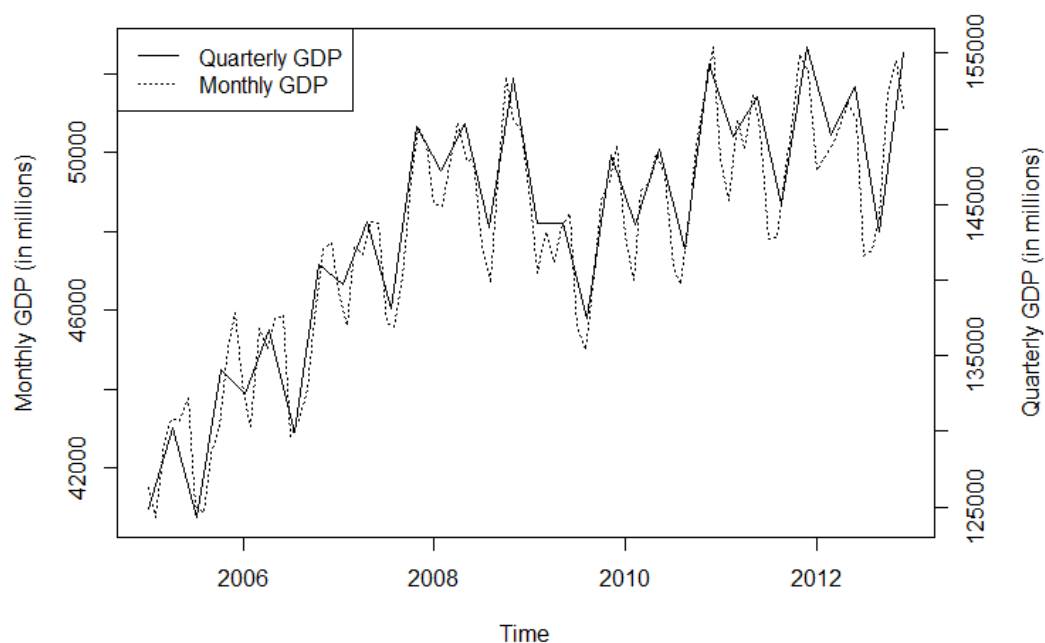


Figure 3. Monthly GDP estimates

#### 4.2 Example: The Chow-Lin Regression based approach for the univariate case

Here we look at the univariate case of the Chow-Lin method. For this example we only use one high frequency indicator series, consumer credit, to estimate the values of the GDP month-to-month changes (low frequency series). The original GDP series is measured in terms of quarter-to-quarter changes (in percentages). In Figure 4, the GDP (left) and the consumer credit (right) are plotted.

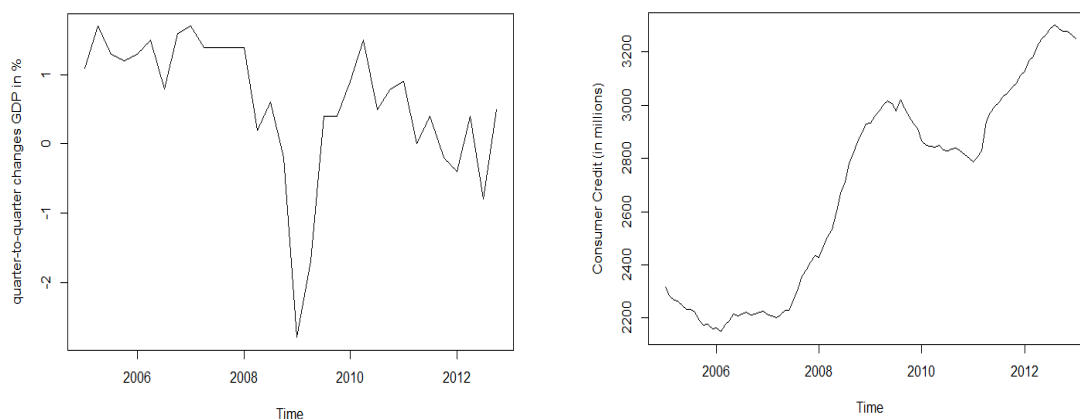


Figure 4. Plots of the GDP (left) and the consumer credit (right).



In Table 3, the result of the univariate regression is shown. Once again, we model the serial correlation in the residuals with the means of an AR(1) model. From the results stated in Table 3 we can see that the consumer credit has an significant effect on the GDP. With a  $R^2$  of 0.27, consumer credit explains a reasonable part of the GDP's movement.

Table 3. Regression results of the univariate regression

	Estimate	Standard Error	t-statistic
<i>Intercept**</i>	1.7171	0.4421	3.884
Consumer Credit**	-0.0006	0.0002	-3.461

\*\*Significance level  $\alpha = 0.01$

When we make a plot of the estimated high frequency GDP and compare this to the original quarterly GDP series, it is clear from Figure 5 that the monthly GDP has the same fluctuation pattern as the quarterly GDP series. The only difference is that the estimated monthly GDP is more smooth.

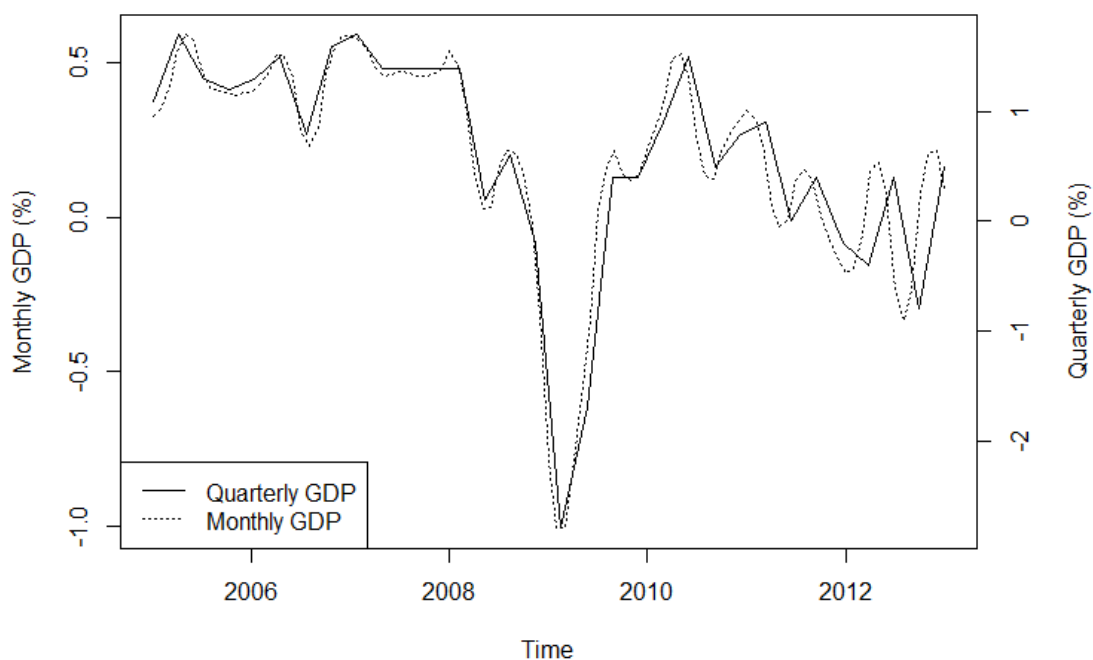


Figure 5. Comparison of the estimated monthly GDP with the quarterly GDP

## 5. Examples – tool specific

## 6. Glossary

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

## 7. References

- Barcellan, R. and Buono, D. (2002), *ECOTRIM interface user manual*. Eurostat.
- Chamberlin, G. (2010), Methods Explained: Temporal disaggregation. *Economic and Labour Market Review* **4**, 106–121.
- Chow, G. and Lin, A. (1971), Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics* **53**, 372–375.
- Chow, G. and Lin, A. (1976), Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association* **71**, 719–721.
- Dagum, E. B. and Cholette, P. A. (2006), *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series*. Springer, New York.
- Denton, F. T. (1971), Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the American Statistical Association* **66**, 99–102.
- Fernández, R. (1981), A methodological note on the estimation of time series. *The Review of Economics and Statistics* **63**, 471–478.
- Islam, M. R. (2009), Evaluation of different temporal disaggregation techniques and an application to Italian GDP. *BRAC University Journal* **VI**, No. 2, 21–32.
- Litterman, R. (1983), A random walk, Markov model for the distribution of time series. *Journal of Business and Economic Statistics* **1**, 169–173.
- Mitchell, J., Smith, R. J., Weale, M. R., Wright, S., and Salazar, E. L. (2005), An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth. *Economic Journal, Royal Economic Society* **115**, F108–F129.
- Rizk, M. (2010), Temporal Disaggregation of the Quarterly Real GDP Series: Case of Egypt. Research Department Working Paper, Central Bank of Egypt.

## **Specific section**

### **8. Purpose of the method**

The method is used for temporal disaggregation which disaggregates a low frequency time series to a higher frequency series.

### **9. Recommended use of the method**

1. This method can be applied to any dataset that contains variables with different frequencies and where the user needs a dataset with the same frequency for all variables.
2. The method is useful if one wants to combine the strengths of the short-term (high frequency) and the long-term data (low frequency).
3. The method is more applicable for flows series as these are used for economic statistics.
4. Application of this method is only possible if besides the low frequency series, other high frequency indicators are available.

### **10. Possible disadvantages of the method**

1. The results rely on a covariance matrix, which is often unknown in practice and needs to be estimated on the basis of assumptions.

### **11. Variants of the method**

1. Based on the Chow-Lin method these variants are expansions of the Chow-Lin method. The only difference is the computation of the covariance matrix, which is used for the monthly estimates.
  - 1.1 Fernandez (1981)
  - 1.2 Litterman (1983)

### **12. Input data**

1. A low frequency time series (i.e., quarterly).
2. One or more high frequency indicator time series (i.e., monthly).
3. The value for serial correlation in the covariance matrix; if the user knows this beforehand or assumes a value. If not specified, the value for serial correlation can be estimated.

### **13. Logical preconditions**

1. Missing values
  1. High frequency indicator may not contain missing values.
  2. Low frequency series may contain missing values. When the low frequency value for the last period is missing, the Chow-Lin method can be used to forecast that missing value (Mitchell et al, 2004).
2. Erroneous values

1. Erroneous low frequency input values will be preserved in the output.
2. Erroneous high frequency input values can result in invalid high-frequency output series. But the high-frequency series that is obtained will still be consistent with the low-frequency data.
3. Other quality related preconditions
  1. One quarterly period covers a whole number of monthly periods. As for annual periods; these cover a whole number of sub-annual periods.
  2. For the short-term movements to be incorporated in the estimated target series, proper high frequency indicators should be used in the sense that they have a (statistical) relationship with the low frequency series.
4. Other types of preconditions
  - 1.
  - 2.

#### **14. Tuning parameters**

#### **15. Recommended use of the individual variants of the method**

1. Different time series are called co-integrated, if they do not drift apart during time. For example, consumption and income are co-integrated, if consumption is roughly a constant proportion of income over the long term. If the target series and the high frequency series are not co-integrated, the Fernandez (1981) or Litterman (1983) variant is preferred (Rizk, 2010).

#### **16. Output data**

1. Ds-output1 = a dataset with estimated high frequency values derived from the original low frequency series and high frequency indicators.

#### **17. Properties of the output data**

1. The estimated high frequency data add up to the original low frequency values. In that sense, the output is consistent.
2. The short-term movement of the high frequency indicators is incorporated.

#### **18. Unit of input data suitable for the method**

Processing full datasets.

#### **19. User interaction - not tool specific**

1. Before execution of the method, the input datasets must be specified.
2. During operation no user interaction is needed.
3. After use of the method the quality indicators should be inspected.

**20. Logging indicators**

- 1.

**21. Quality indicators of the output data**

1. The consistency of the high frequency estimates should be checked, i.e., the values should add up to the original low frequency values. Furthermore, the (short-term) behaviour is of particular interest. The short-term behaviour of in- and output time series should somewhat have the same behaviour. This can be checked fast and easy by graphical means.

**22. Actual use of the method**

1. Intensively used by National Statistical Institutes, especially in France, Italy, Belgium, Portugal and Spain (Barcellan and Bueno, 2002).

**Interconnections with other modules**

**23. Themes that refer explicitly to this module**

1. Macro-Integration – Main Module

**24. Related methods described in other modules**

1. Macro-Integration – Denton’s Method

**25. Mathematical techniques used by the method described in this module**

1. Linear regression with linear constraints

**26. GSBPM phases where the method described in this module is used**

1. GSBPM phase 6.2 “Validate Outputs”

**27. Tools that implement the method described in this module**

1. Matlab
2. R
3. ECOTRIM – Eurostat has made an application program named “ECOTRIM” for the use of several temporal disaggregation methods.

**28. Process step performed by the method**

Temporal disaggregation

## Administrative section

### 29. Module code

Macro-Integration-M-Chow-Lin

### 30. Version history

Version	Date	Description of changes	Author	Institute
0.1	28-05-2013	first version	Feysel Negash	CBS
0.2	04-10-2013	review comments Roberto Iannaccone adopted	Feysel Negash Jacco Daalmans	CBS
0.3	30-10-2013	review comments Editorial Board adopted	Jacco Daalmans	CBS
0.3.1	31-10-2013	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

### 31. Template version and print date

Template version used	
Print date	