Temporal Disaggregation using JDemetra+ (GUI and rjd3bench)

Webinar 2023

María Novás
National Statistical Institute of Spain

September 20th 2023

INE
Instituto Nacional de Estadística

Introduction

Temporal disaggregation Methods: These methods are used to disaggregate a series from low frequency to high-frequency.

When there are high frequency related indicators, these methods provide high frequency estimations for a series whose sums,averages,last or first values are consistent with the observed low frequency series, applying a regression model where it is assumed that monthly observations (if available) of the series to be estimated satisfy a multiple regression with p related series (indicators).

Temporal disaggregation methods

Two types of temporal disaggregation methods:

1. **There is no high frequency related indicator**:
   - **Softened methods**: such as Boot, Fiebes and Lisman (1967). Identical to Denton considering as proxy a series of ones.
   - **Methods of time series**: such as: Wei and Stram (1990).
2. **There is a high frequency related indicator**: these methods are known as optimal methods and among them are Chow-Lin (1971), Fernández (1981), Litterman (1983), Santos Silva and Cardoso, and Proietti.

INE

High frequency model

High frequency model. It is assumed that monthly observations (if available) of the series to be estimated satisfy a multiple regression with p related series:

$$y = X\beta + u \tag{1}$$

where y is monthly observations (if available) of the series to be estimated, X is the matrix with the regression variables in columns, β is a vector of coefficients and u follows:

- **Chow-Lin**: $u_t = \phi u_{t-1} + \epsilon_t$, with $-1 < \phi < 1$ and ϵ_t white noise
- **Fernandez**: $\nabla u_t = \epsilon_t$, with ϵ_t white noise
- **Litterman**: $\nabla u_t = \phi \nabla u_{t-1} + \epsilon_t$, with $-1 < \phi < 1$ and ϵ_t white noise

Low frequency model

Low frequency model. This is the model directly estimated, obtained premultiplying the high frequency model by a matrix:

$$My = MX\beta + Mu \tag{2}$$

where

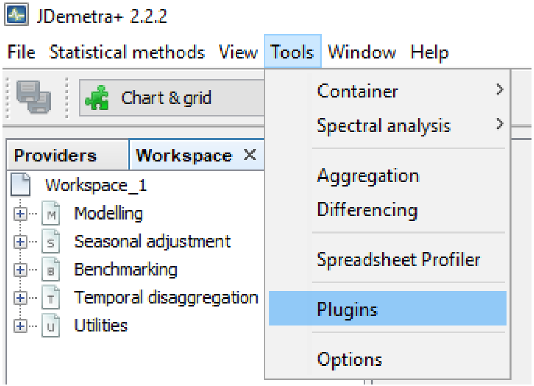
$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \cdots \\ \cdots & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \end{pmatrix}$$

$E(Mu) = ME(u) = 0$ and $Cov(Mu) = MCov(u)M^{\intercal}$

Install the plug-in I

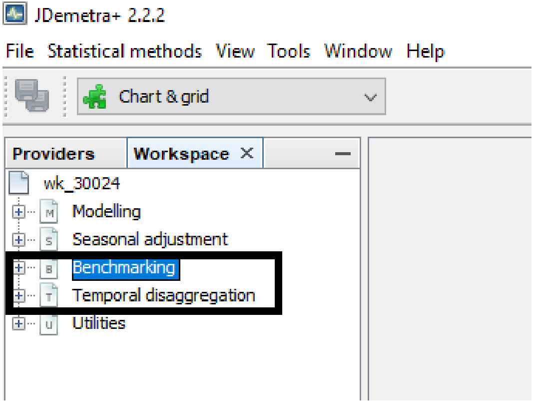
1. **Download the file nbdemetra-benchmarking-2.2.2.nbm**
   https://github.com/nbbrd/jdemetra-benchmarking/releases

2. **Add the plugin in the GUI version 2.2.2**



Install the plug-in II

3. **Benchmarking and Temporal Disaggregation options**



Example

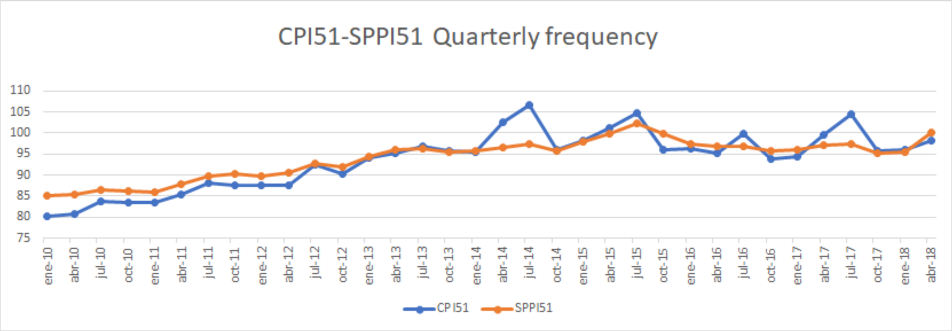
- **Input series**: The Services Producer Price Index without direct taxes with quarterly frequency: Air Transport branch.From the **first quarter of 2010** to the **first quarter of 2018**.
- **Indicator**: The CPI at constant taxes series with monthly frequency: 0733 Air Transport. From **January 2010** to **June 2018.**
- **Output series**: The Services Producer Price Index, air transport branch, with monthly frequency.

Steps to choose the method

1. Graphical Analysis
2. Cointegration Test (optional)
3. Chow-Lin
4. Fernández and Litterman
5. ADL Models: Proietti and Santos Silva and Cardoso
6. Extrapolation

1. Graphical Analysis

Series 51 Air transport: CPI vs SPPI



INE

2. Cointegration test

High frequency model. It is assumed that monthly observations (if available) of the series to be estimated satisfy a multiple regression with p related series:

$$y = X\beta + u \tag{3}$$

where y is monthly observations (if available) of the series to be estimated, X is the matrix with the regression variables in columns, β is a vector of coefficients and u follows:

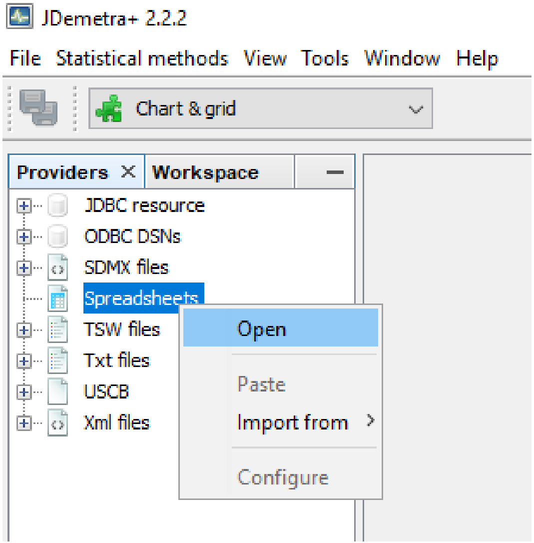
- **Chow-Lin**: $u_t = \phi u_{t-1} + \epsilon_t$, with $-1 < \phi < 1$ and ϵ_t white noise
- **Fernandez**: $\nabla u_t = \epsilon_t$, with ϵ_t white noise
- **Litterman**: $\nabla u_t = \phi \nabla u_{t-1} + \epsilon_t$, with $-1 < \phi < 1$ and ϵ_t white noise

2. Cointegration test

- If the series are **cointegrated in high frequency**, **Chow-Lin** should be applied.
- If the series are **not cointegrated** in high frequency we apply **Fernández and Litterman.**

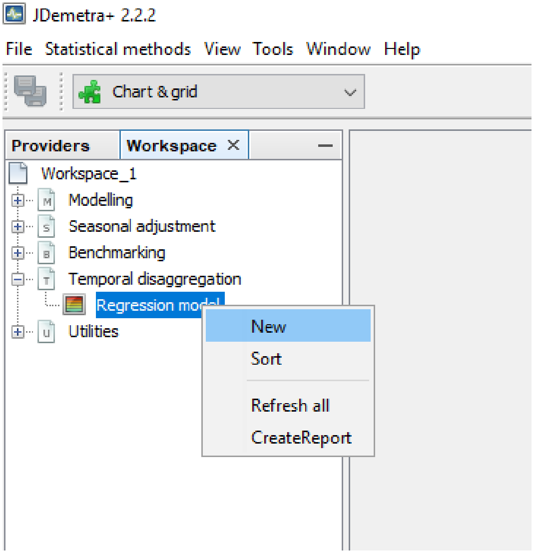
This test can only be done in low frequency and the test conclusions **cannot be scaled from low to high frequency**.

3. Chow-Lin

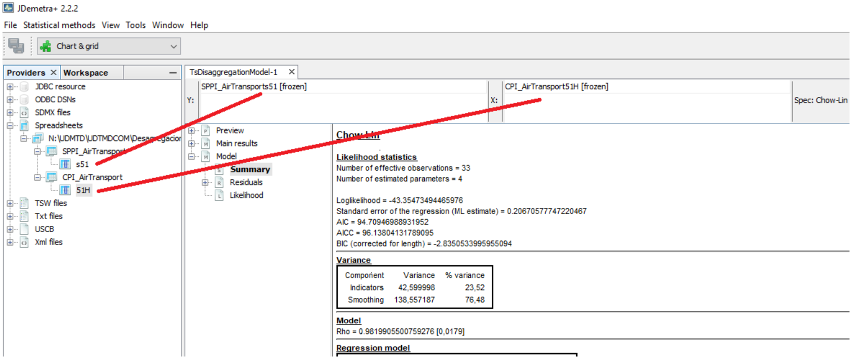


JDemetra+ 2.2.2
File Statistical methods View Tools Window Help
Chart & grid
Providers
Workspace
JDBC resource
ODBC DSNs
SDMX files
Spreadsheets
TSW files
Txt files
USCB
Xml files
Open
Paste
Import from
Configure

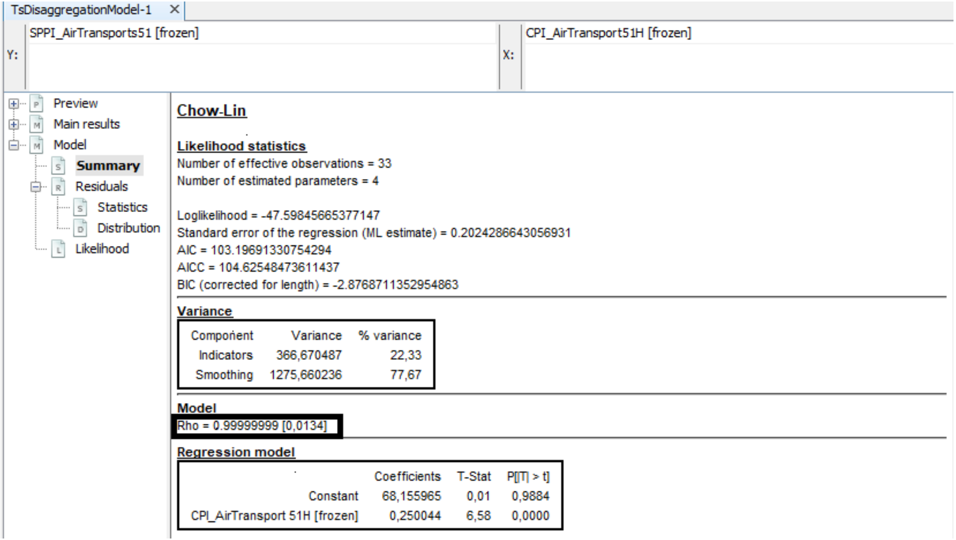
3. Chow-Lin


JDemetra+ 2.2.2
File Statistical methods View Tools Window Help
Chart & grid
Providers
Workspace
Workspace_1
Modelling
Seasonal adjustment
Benchmarking
Temporal disaggregation
Regression model
Utilities
New
Sort
Refresh all
CreateReport

3. Chow-Lin


JDemetra+ 2.2.2
File Statistical methods View Tools Window Help
Chart & grid
Providers Workspace
JDBC resource
ODBC DSNs
SDMX files
Spreadsheets
SPPI_AirTransport
s51
CPI_AirTransport
51H
TSW files
Txt files
USCB
Xml files
TsDisaggregationModel-1
Y: SPPI_AirTransports51 [frozen]
X: CPI_AirTransport51H [frozen]
Spec: Chow-Lin
Preview
Main results
Model
Summary
Residuals
Likelihood
Chow-Lin
Likelihood statistics
Number of effective observations = 33
Number of estimated parameters = 4
Loglikelihood = -43.35473494465976
Standard error of the regression (ML estimate) = 0.20670577747220467
AIC = 94.70946988931952
AICC = 96.13804131789095
BIC (corrected for length) = -2.8350533995955094
Variance
Component Variance % variance
Indicators 42,599998 23,52
Smoothing 138,557187 76,48
Model
Rho = 0.9819905500759276 [0,0179]
Regression model

3. Chow-Lin



TsDisaggregationModel-1

Y: SPPI_AirTransports51 [frozen]

X: CPI_AirTransport51H [frozen]

- Preview
- Main results
- Model
  - Summary
  - Residuals
    - Statistics
    - Distribution
  - Likelihood

Chow-Lin

Likelihood statistics

Number of effective observations = 33
Number of estimated parameters = 4

Loglikelihood = -47.59845665377147
Standard error of the regression (ML estimate) = 0.2024286643056931
AIC = 103.19691330754294
AICC = 104.62548473611437
BIC (corrected for length) = -2.8768711352954863

Variance

Component	Variance	% variance
Indicators	366,670487	22,33
Smoothing	1275,660236	77,67

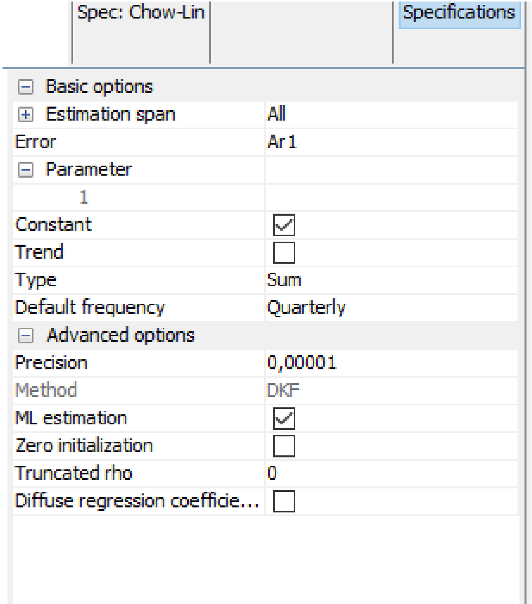
Model

Rho = 0.99999999 [0,0134]

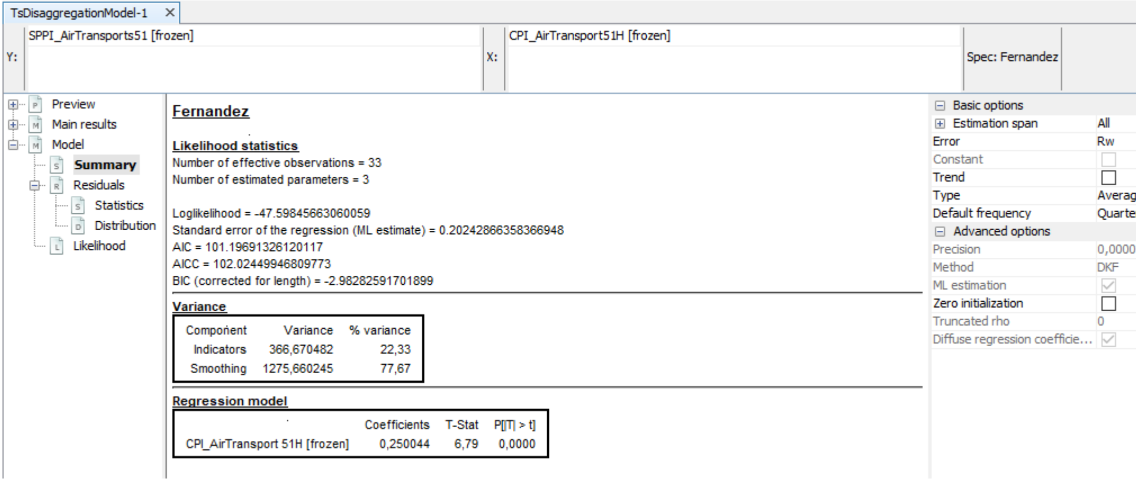
Regression model

	Coefficients	T-Stat	P[|T| > t]
Constant	68,155965	0,01	0,9884
CPI_AirTransport 51H [frozen]	0,250044	6,58	0,0000

3. Chow-Lin


Spec: Chow-Lin
Specifications
Basic options
Estimation span
All
Error
Ar1
Parameter
1
Constant
Trend
Type
Sum
Default frequency
Quarterly
Advanced options
Precision
0,00001
Method
DKF
ML estimation
Zero initialization
Truncated rho
0
Diffuse regression coefficie...

4. Fernandez


TsDisaggregationModel-1
Y:
SPPI_AirTransports51 [frozen]
X:
CPI_AirTransport51H [frozen]
Spec: Fernandez
Preview
Main results
Model
Summary
Residuals
Statistics
Distribution
Likelihood
Fernandez
Likelihood statistics
Number of effective observations = 33
Number of estimated parameters = 3
Loglikelihood = -47.59845663060059
Standard error of the regression (ML estimate) = 0.20242866358366948
AIC = 101.19691326120117
AICC = 102.02449946809773
BIC (corrected for length) = -2.98282591701899
Variance
Component Variance % variance
Indicators 366,670482 22,33
Smoothing 1275,660245 77,67
Regression model
Coefficients T-Stat P[|T| > t]
CPI_AirTransport 51H [frozen] 0,250044 6,79 0,0000
Basic options
Estimation span All
Error Rw
Constant
Trend
Type Averag
Default frequency Quarte
Advanced options
Precision 0,0000
Method DKF
ML estimation
Zero initialization
Truncated rho 0
Diffuse regression coefficie...

4. Litterman

Preview
Main results
Model
Summary
Residuals
Likelihood

Litterman

Likelihood statistics

Number of effective observations = 33
Number of estimated parameters = 4

Loglikelihood = -42.97055716264021
Standard error of the regression (ML estimate) = 0.15681419618642797
AIC = 93.94111432528042
AICC = 95.36968575385185
BIC (corrected for length) = -3.387522952751936

Variance

Component	Variance	% variance
Indicators	350,623615	21,19
Smoothing	1304,120243	78,81

Model

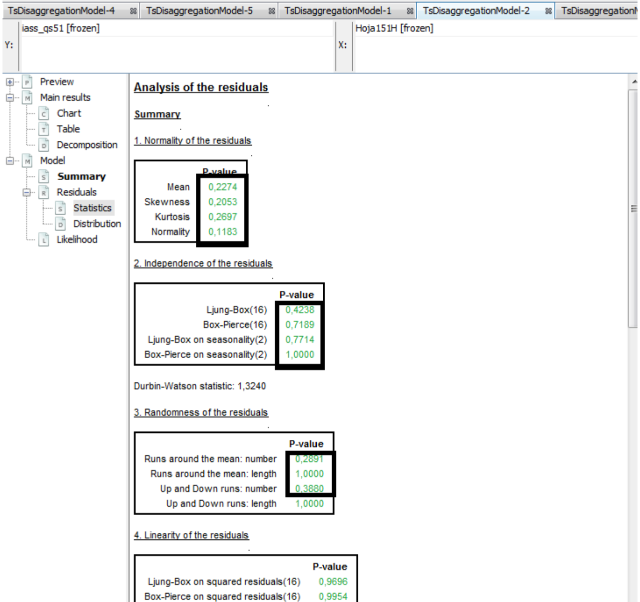
Rho = 0.29166666374999994 [0,3218]

Regression model

	Coefficients	T-Stat	P[|T| > t]
Hoja1 51H [frozen]	0,244512	7,01	0,0000

INE

4. Analysis of the residuals



4. Disaggregated series (Fernández)

TsDisaggregationModel-1

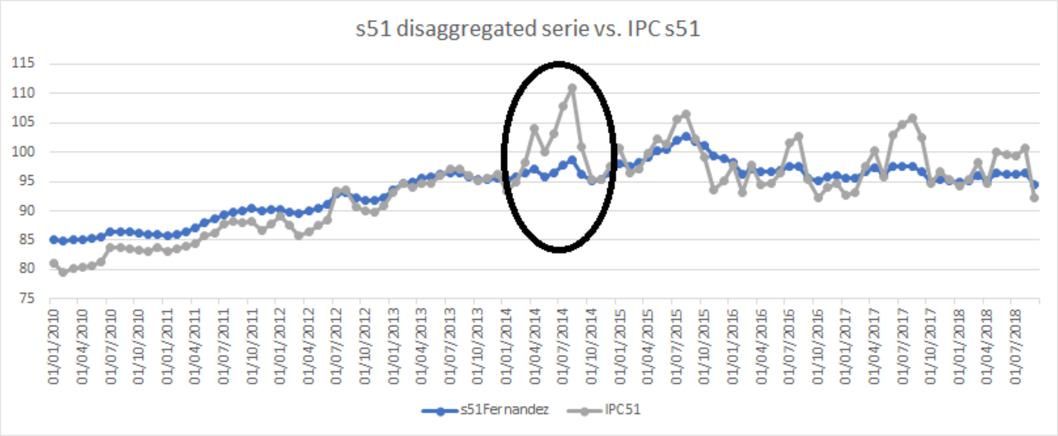
Y: SPPI_AirTransports51 [frozen]

X: CPI_AirTransport51H [frozen]

- Preview
- Main results
  - Chart
  - Table
  - Decomposition
- Model
  - **Summary**
  - Residuals
    - Statistics
    - Distribution
  - Likelihood

	disaggregation	disggregationError
1-2010	85,188	0,425
2-2010	84,84	0,292
3-2010	85,055	0,397
4-2010	85,131	0,373
5-2010	85,305	0,285
6-2010	85,636	0,371
7-2010	86,433	0,371
8-2010	86,488	0,285
9-2010	86,459	0,37
10-2010	86,29	0,37
11-2010	86,101	0,285
12-2010	86,094	0,37
1-2011	85,734	0,37
2-2011	85,936	0,285
3-2011	86,423	0,37
4-2011	87,166	0,37
5-2011	88,071	0,285
6-2011	88,612	0,37
7-2011	89,338	0,37

4. Results: Air transport (Fernández)



rjd3bench

- https://github.com/rjdemetra/rjd3bench/releases
- temporaldisagggregation function:

```
function (series, constant = T, trend = F, indicators = NULL,
    model = c("Ar1", "Rw", "RwAr1"), freq = 4, conversion = c("Sum",
        "Average", "Last", "First", "UserDefined"), conversion.obsposition = 1,
    rho = 0, rho.fixed = F, rho.truncated = 0, zeroinitialization = F,
    diffuse.algorithm = c("SqrtDiffuse", "Diffuse", "Augmented"),
    diffuse.regressors = F)
```

5. ADL Models

In **Santos Silva and Cardoso** the model is:

$$y_t = \phi y_{t-1} + x_t\beta + \epsilon_t \tag{4}$$

In **Proietti** the model is:

$$y_t = \phi y_{t-1} + m + gt + x_t\beta + x_{t-1}\gamma + \epsilon_t \tag{5}$$

where ϵ_t is a white noise process with variance σ^2

5. ADL Models

- Likelihood ratio test:
  - Proietti and Santos Silva Cardoso
  - Proietti and Chow-Lin
- Analysis of the residuals

6. Extrapolation

Using 8 quarters of extrapolated data.

- Chow-Lin:
$$RMSE = 0{,}60 \tag{6}$$

- Fernandez:
$$RMSE = 0{,}59 \tag{7}$$

- Fernandez:
$$RMSE = 0{,}70 \tag{8}$$