

Constrained retropolation of high-frequency data using related series: A simple dynamic model approach ¹

Un semplice modello dinamico per la retropolazione vincolata di dati a cadenza temporale elevata

Tommaso Di Fonzo

Dipartimento di Scienze Statistiche

Università di Padova

difonzo@stat.unipd.it

Riassunto: Il classico approccio di Chow e Lin (1971, 1976) per la disaggregazione temporale di serie economiche mediante indicatori di riferimento viene esteso al caso di un semplice modello dinamico ed applicato per retropolare una serie storica nel rispetto del vincolo di aggregazione temporale dato da osservazioni disponibili a cadenza più bassa.

Keywords: temporal disaggregation, related series, dynamic model.

1. Position of the problem

Let $\mathbf{y}_h$ be the $(n \times 1)$ vector of high-frequency data for a variable Y and suppose that over the first n_1 high-frequency periods ($n_1 < n$) only m_1 low-frequency observations are available. Let s be the number of sub-periods in each low-frequency time interval ($s = 4$ in the quarterly-annual case and $s = 3$ in the monthly-quarterly case) and $n_2 = n - n_1$ (for algebraic convenience we assume that $m_1 = \frac{n_1}{s}$ is an integer).

We assume that Y is linked to k (strictly exogenous) related variables by the regression equation formulated at high-frequency level

$$y_{h,t} = \mathbf{x}'_{h,t} \boldsymbol{\beta} + u_{h,t}, \quad t = 1, \dots, n \quad (1)$$

where $\mathbf{x}_{h,t}$ is a $(k \times 1)$ vector of high-frequency observations on k related series, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of unknown parameters and $u_{h,t}$ is a zero-mean stochastic disturbance not correlated with $\mathbf{x}_{h,t}$. Furthermore, let us write $\mathbf{y}_h$ as $\mathbf{y}_h = [\mathbf{y}_h^1, \mathbf{y}_h^2]'$, where $\mathbf{y}_h^1$ and $\mathbf{y}_h^2$ have dimensions $(n_1 \times 1)$ and $(n_2 \times 1)$ respectively, $\mathbf{y}_h^1$ being unknown whilst $\mathbf{y}_h^2$ is observed. Let $\mathbf{y}_l^1$ be the $(m_1 \times 1)$ vector of low-frequency observed values linked to $\mathbf{y}_h^1$ by $\mathbf{y}_l^1 = \mathbf{C}_1 \mathbf{y}_h^1$, $\mathbf{C}_1$ being the $(m_1 \times n_1)$ aggregation matrix $\mathbf{C}_1 = \mathbf{I}_{m_1} \otimes \mathbf{c}'$, where $\mathbf{c}$ is a $(s \times 1)$ vector of known constants whose values depend on the nature of the data to be estimated². In matrix form we have

$$\begin{bmatrix} \mathbf{y}_h^1 \\ \mathbf{y}_h^2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_h^1 \\ \mathbf{X}_h^2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{u}_h^1 \\ \mathbf{u}_h^2 \end{bmatrix}, \quad (2)$$

¹Lavoro stato svolto con fondi MURST 60%.

² $\mathbf{c} = (1, 1, \dots, 1)'$ for a distribution problem of flows, $\mathbf{c} = \frac{1}{s}(1, 1, \dots, 1)'$ for index data, $\mathbf{c} = (0, 0, \dots, 1)'$ for an interpolation problem of end-of-period stock data and, finally, $\mathbf{c} = (1, 0, \dots, 0)'$ for an interpolation problem of begin-of-period stock data.

where $\mathbf{X}_h^1$ and $\mathbf{X}_h^2$ are matrices $(n_1 \times k)$ and $(n_2 \times k)$, respectively. The stochastic disturbance vector has been split in two parts, $\mathbf{u}_h^1$ and $\mathbf{u}_h^2$, respectively, each with zero mean and covariance matrix given by

$$E(\mathbf{u}_h \mathbf{u}_h') = \mathbf{V}_h = \begin{bmatrix} \mathbf{V}_{h,11} & \mathbf{V}_{h,12} \\ \mathbf{V}_{h,12}' & \mathbf{V}_{h,22} \end{bmatrix}.$$

Premultiplying (2) by the $((m_1 + n_2) \times n)$ matrix $\mathbf{C}$ given by

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_2} \end{bmatrix},$$

we get the (observed) transformed model, containing both low and high-frequency observed data:

$$\mathbf{y}_a = \mathbf{X}_a \boldsymbol{\beta} + \mathbf{u}_a, \quad (3)$$

where $\mathbf{y}_a = \mathbf{C} \mathbf{y}_h = [\mathbf{y}_l^1, \mathbf{y}_h^2]'$ and $\mathbf{u}_a$ is a zero-mean stochastic disturbance with covariance matrix $\mathbf{V}_a = \mathbf{C} \mathbf{V}_h \mathbf{C}'$.

For $\mathbf{V}_h$, and thus $\mathbf{V}_a$, known, the classical (BLU) solution of Chow and Lin (1971, 1976) to this missing value problem is given by

$$\hat{\mathbf{y}}_h = \mathbf{X}_h \hat{\boldsymbol{\beta}} + \mathbf{L} \hat{\mathbf{u}}_a, \quad (4)$$

where $\hat{\boldsymbol{\beta}}$ is the GLS estimate of $\boldsymbol{\beta}$ from model (3), $\mathbf{L} = \mathbf{V}_h \mathbf{C}' \mathbf{V}_a^{-1}$ and $\hat{\mathbf{u}}_a = \mathbf{y}_a - \mathbf{X}_a \hat{\boldsymbol{\beta}}$.

Only the retropolated values, that is $\hat{\mathbf{y}}_h^1$, are of interest for our purposes. It can be shown that they are given by³:

$$\hat{\mathbf{y}}_h^1 = \mathbf{X}_h^1 \hat{\boldsymbol{\beta}} + \mathbf{L}_{1,1} \hat{\mathbf{u}}_l^1 + \mathbf{L}_{1,2} \hat{\mathbf{u}}_h^2, \quad (5)$$

where $\mathbf{L}_{1,1}$ and $\mathbf{L}_{1,2}$ are matrices $(n_1 \times m_1)$ and $(n_1 \times n_2)$, respectively, given by

$$\mathbf{L}_{1,1} = \mathbf{V}_{h,11,2} \mathbf{C}_1' \mathbf{V}_{a,11,2}^{-1} \quad \mathbf{L}_{1,2} = (\mathbf{V}_{h,12} - \mathbf{V}_{h,11,2} \mathbf{C}_1' \mathbf{V}_{a,11,2}^{-1} \mathbf{V}_{a,12}) \mathbf{V}_{h,22}^{-1},$$

with $\mathbf{V}_{h,11,2} = \mathbf{V}_{h,11} - \mathbf{V}_{h,12} \mathbf{V}_{h,22}^{-1} \mathbf{V}_{h,21}$,

$$\mathbf{V}_{a,11,2} = \mathbf{V}_{a,12} - \mathbf{V}_{a,12} \mathbf{V}_{h,22}^{-1} \mathbf{V}_{a,21} = \mathbf{C}_1 \mathbf{V}_{h,11} \mathbf{C}_1' - \mathbf{C}_1 \mathbf{V}_{h,12} \mathbf{V}_{h,22}^{-1} \mathbf{V}_{h,21} \mathbf{C}_1',$$

$\hat{\mathbf{u}}_l^1 = \mathbf{y}_l^1 - \mathbf{X}_l^1 \hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}_h^2 = \mathbf{y}_h^2 - \mathbf{X}_h^2 \hat{\boldsymbol{\beta}}$.

Furthermore, the estimates are coherent with the available (both aggregated and disaggregated) information, as it is easily proven by pre-multiplying (4) by the transformation matrix $\mathbf{C}$. Finally, the error covariance matrix of the estimated values is

$$E(\hat{\mathbf{y}}_h^1 - \mathbf{y}_h^1)(\hat{\mathbf{y}}_h^1 - \mathbf{y}_h^1)' = (\mathbf{I}_{n_1} - \mathbf{L}_{1,1} \mathbf{C}_1) \mathbf{V}_{h,11,2} + \mathbf{M} \mathbf{X}_h (\mathbf{X}_a' \mathbf{V}_a^{-1} \mathbf{X}_a)^{-1} \mathbf{X}_h' \mathbf{M}', \quad (6)$$

where $\mathbf{M}$ is the $(n_1 \times n)$ matrix given by $\mathbf{M} = [\mathbf{I}_{n_1} - \mathbf{L}_{1,1} \mathbf{C}_1 | -\mathbf{L}_{1,2}]$.

³The proofs of these results are contained in the extended version of the paper.

2. A simple dynamic model

The retropolation problem can be worked out in the framework of a simple dynamic model, more realistic than the static one depicted in (1). Following the procedure described by Santos Silva and Cardoso (2001) (see Di Fonzo, 2001, for a critical review), we show how to estimate $\hat{\mathbf{y}}_h^1$ according to the model

$$y_{h,t} = \phi y_{h,t-1} + \mathbf{x}_{h,t}' \boldsymbol{\beta} + \varepsilon_{h,t}, \quad t = 1, \dots, n \quad (7)$$

where $|\phi| < 1$. This model is enough general to encompass several interesting relationships between Y and X 's. Namely, if Y and the X 's are (not stationary but) cointegrated, model (7) can be seen as a re-parameterization of an error correction model. For $\phi \neq 0$, model (7) can be written in matrix form as follows (Di Fonzo, 2001):

$$\mathbf{D}_\phi \mathbf{y}_h = \mathbf{X}_h \boldsymbol{\beta} + \mathbf{q} \eta + \boldsymbol{\varepsilon}_h = \mathbf{Z}_h \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_h,$$

where $\mathbf{D}_\phi$ is the $(n \times n)$ matrix

$$\mathbf{D}_\phi = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{bmatrix},$$

$\mathbf{q} = (\phi, 0, 0, \dots, 0)'$ is a $(n \times 1)$ vector, $\mathbf{Z}_h = [\mathbf{X}_h | \mathbf{q}]$ is a $(n \times (k+1))$ matrix and $\boldsymbol{\gamma} = [\boldsymbol{\beta}' | \eta]'$ is a $((k+1) \times 1)$ vector. According to Harvey (1990), we treat $\eta = E(y_0)$ as a fixed (and unknown) parameter to be estimated. Given that the inverse of $\mathbf{D}_\phi$ is the lower triangular matrix

$$\mathbf{D}_\phi^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \phi & 1 & 0 & \cdots & 0 & 0 \\ \phi^2 & \phi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & \phi & 1 \end{bmatrix},$$

the model can be expressed as

$$\mathbf{y}_h = \mathbf{D}_\phi^{-1} \mathbf{Z}_h \boldsymbol{\gamma} + \mathbf{D}_\phi^{-1} \boldsymbol{\varepsilon}_h = \mathbf{Z}_{h,\phi} \boldsymbol{\gamma} + \mathbf{u}_h. \quad (8)$$

ML estimates of the parameters of the transformed model $\mathbf{y}_a = \mathbf{Z}_{a,\phi} \boldsymbol{\gamma} + \mathbf{u}_a$, where $\mathbf{Z}_{a,\phi} = \mathbf{C} \mathbf{Z}_{h,\phi}$, can be obtained through a grid search estimation strategy as described in Di Fonzo (2001), and thus estimates of $\mathbf{y}_h^1$ and of its error covariance matrix can be calculated according to (5) and (6).

3. An application

In this section we present an application of the retropolation method to the same data set used by Santos Silva and Cardoso (2001). We estimate US quarterly personal consumption data for five years (1953:1–1957:4) assuming the availability of (i) the quarterly

consumption series from 1958:1 to 1984:4, (ii) the annual (temporally aggregated) consumption series from 1953 to 1957, and (iii) the quarterly disposable income series from 1953:1 to 1984:4 used as related indicator of personal consumption. The estimated parameters are presented in table 1. The estimated autoregressive parameter, $\hat{\phi} = 0.788$, and the estimated long run effect, $\frac{\hat{\beta}_x}{1 - \hat{\phi}} = 0.930$, where $\hat{\beta}_x$ is the estimated regression coefficient of the related series, are quite good.

Parameter	Estimate	Std. Dev.	t-stat	p-value
Intercept	0.924	1.709	0.541	0.295
β_x	0.197	0.002	88.621	0.000
η	356.3	14.295	24.924	0.000

Table 1: Estimated parameters

The visual inspection of estimated and actual values (the graphs are not reported here for space needs) and the performance indicators on levels, quarterly growth rates and four-quarters growth rates (correlation between observed and estimated data, standard errors, min. and max. of the errors, see table 2), confirm the reasonability of the procedure.

	Correlation	Std. Dev.	Min	Max
Levels	0.997	0.412	-0.940	0.872
Quarterly rate of change	0.806	0.410	-0.810	0.657
Annual rate of change	0.949	0.676	-1.579	1.446

Table 2: Performance indicators of the retropolation results

References

- Chow, G. 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**, 372–375.
- Chow, G. and Lin, A. L. (1976). Best linear unbiased estimation of missing observations in an economic time series, *Journal of the American Statistical Association* **71**, 719–721.
- Di Fonzo, T. (2001). *Su una procedura di disaggregazione temporale di serie economiche mediante un semplice modello dinamico*, mimeo.
- Harvey, A. C. (1990). *The econometric analysis of time series*, Philip Allan, Deddington, Oxford.
- Santos Silva, J. and Cardoso, F. (2001). The Chow-Lin method using dynamic models, *Economic Modelling* **18**, 269–280.