Barometr Regionalny
Tom 14 nr 3

Temporal Disaggregation of Time Series with Regularization Term

Sebastian Wójcik
Statistical Office in Rzeszów, Poland

Abstract

Methods of temporal disaggregation are used to obtain high frequency time series from the same low frequency time series — so-called disaggregation — with respect to some additional consistency conditions between low and high frequency series. Conditions depend on the nature of the data — e.g., stack, flow, average and may pertain to the sum, the last value and the average of the obtained high frequency series, respectively. Temporal disaggregation methods are widely used all-over the world to disaggregate for example quarterly GDP. These methods are usually two-stage methods which consist of regression and benchmarking. In this article we propose a method which performs regression and benchmarking at the same time and allows to set a trade-off between them.

Keywords: temporal disaggregation, benchmarking, time series analysis, regularization term

Introduction

Nowadays policy making requires continuous evaluation of decisions they made. Evaluation is often carried out with a use of economic indicators produced by official statistics. Decision makers on national and regional level cannot immediately adjust their strategies if the monitoring indicators are produced with irrelevant frequency. Therefore, official statistics is looking for methods which enable to meet the needs for data of the higher frequency. One of the solutions is to collect data with increased frequency. But it raises burden of interviewees and it is not a desired direction. Thus, official statistics and scientists developed many methods to deal with so-called temporal disaggregation. Basically, we can divide these methods into two groups. The first one consists of smoothing methods (Chamberlin 2010, 108) that do not use any additional indicators (e.g., cubic spines or Boot, Feibes and Lisman (BFL) smoothing method). The second one is based on the use of additional high frequency data (Sax and Steiner 2013, 80) — e.g., Denton (1971), Denton-Cholette (Dagum and Cholette 2006), Chow-Lin (Chow and Lin 1971), Fernandez (1981), Litterman (1983). In the next section we present problem set up and quick review of methods.

1 Methods of temporal disaggregation

Assume that $Y_{l\times 1}$ is a vector of low frequency data and $X_{h\times k}$ is a matrix of k variables of high frequency data. Let $A_{l\times h}$ be an aggregation matrix such that AX is low frequency data. We want to estimate high frequency vector $y_{h\times 1}$ such that linear constrains of the form $Y = Ay$ are met. For example, assume that we want to disaggregate yearly flow data to quarterly flow data of (e.g., GDP). Then $h = 4l$ and we can set $A_{l\times h} = \mathrm{I}_l \otimes 1_4{}'$ where $\otimes$ and $'$ denotes Kronecker product and transposition, respectively. For stack data (e.g., unemployment rate) we would set $A_{l\times h} = \mathrm{I}_l \otimes [1\ 1\ 1\ 1]$.

The first stage of temporal disaggregation is regression. With a use of additional variables we produce preliminary estimate $p_{h\times 1}$ of y. Discrepancy between Y and Ap has form $u = Y - Ap$. Then we distribute among in a following way:

$$y = p + Du, \tag{1}$$

where D is a distribution matrix such that $Y = Ay$ holds.
In Denton and Denton-Cholette methods there is used only one variable $p = X_{h\times 1}$. Particularly, we may use a constant and then it can be considered as simple mathematical method. In fact, there is no regression in Denton-like methods. If we use an indicator, the best choice is to pick a variable which is a component of Y. For example, to disaggregate quarterly exports with a use of monthly exports of goods. But is not always the case.

Let $u = Ae$. Chow-Lin, Fernandez and Litterman use Generalized Least Squares Regression on low frequency data of $k \geq 1$ variables on a model expressed

$$Y = AX\beta + Ae \tag{2}$$

with respect variance-covariance matrix U of u. We have $U = AEA'$. In a result, $p = X\hat{\beta}$, where

$$\hat{\beta}(E) = [X'A'U^{-1}AX]^{-1}X'A'U^{-1}Y. \tag{3}$$

The second stage of temporal disaggregation is benchmarking. It is a solution y of minimizing

$$(y - X\hat{\beta})'E^{-1}(y - X\hat{\beta}) \tag{4}$$

with respect to linear constraint $Y = Ay$. We obtain

$$y = X\hat{\beta} + EA'U^{-1}(Y - AX\hat{\beta}) = X\hat{\beta} + Du. \tag{5}$$

Thus, the distribution matrix D takes form

$$D = EA'U^{-1} \tag{6}$$

and it is common for all methods mentioned above. The methods differ with respect to the way of estimating the variance-covariance matrix E.

Denton method enables several approaches to define E:

- $E = I$ in a case of minimizing squared absolute deviation from indicator series (additive method)
- $E = \mathrm{Diag}(X)\mathrm{Diag}(X)$ in a case of minimizing squared relative deviation from indicator series (proportional method)
- $E = (H'H)^{-1}$ in a case of minimizing squared absolute deviation from differenced indicator series, where has H 1 on main diagonal, -1 on the first subdiagonal and 0 elsewhere
- $E = \mathrm{Diag}(X)(H'H)^{-1}\mathrm{Diag}(X)$ in a case of minimizing squared relative deviation from differenced indicator series

Last two approaches focus on preserving period-to-period changes, additive and relative, respectively.

Chow-Lin, Fernandez and Litterman methods model high frequency time series. They assume that

$$e_t = \delta e_{t-1} + v_t, \tag{7}$$

$$v_t = \theta v_{t-1} + \varepsilon_t, \tag{8}$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon)$ is a white noise process.
For Chow-Lin $\theta = 0,\ |\delta| < 1$, thus e_t is autoregressive process of the first order. Covariance matrix has a form $E = \left(\sigma_\varepsilon^2/(1-\delta)\right)R$, where $R = [r_{ij}]$ and $r_{ij} = \delta^{i+j-1}$ for $i \neq j$ and $r_{ii} = 1$. Litterman assumes that $\theta = 0,\ \delta = 1$ and $E = \sigma_\varepsilon^2(H'H)^{-1}$. Fernandez assumes that $|\theta| < 1,\ \delta = 1$ and $E = \sigma_\varepsilon^2(H'Q'QH)^{-1}$, where Q has 1 on main diagonal, $-\theta$ on first subdiagonal and 0 elsewhere.

Autoregressive parameter is estimated usually from low frequency time series of residuals u_t. There are two approaches to estimate autoregressive parameter. The first is based on maximizing likelihood

$$L(\delta, \sigma_\varepsilon^2, \beta) = \frac{e^{-\frac{1}{2}u'U^{-1}u}}{(2\pi)^{\frac{l}{2}}(\det U)^{\frac{1}{2}}}. \tag{9}$$

In the second one we minimize sum of squares of residual with respect to variance-covariance matrix $U = AEA'$

$$RSS(\delta, \sigma_\varepsilon^2, \beta) = u'U^{-1}u. \tag{10}$$

Chow-Lin, Fernandez and Litterman methods distribute residuals in very similar way as additive Denton method. As it will be shown further, additive distribution may result with unacceptable outcomes (e.g., negative values in a case of poor regression stage of temporal disaggregation). Thus, it may be a disadvantage of those methods.

2 Proposition

Clearly, estimated high frequency error term $e = Du$ depends on previous estimation of β. Poor results of regression may lead to unacceptable benchmarking outcome. All of the methods mentioned above are based on that two-stage procedure. There is no possibility to set a trade-off between regression an benchmarking. We propose a method that allows to estimate β and e at the same time and to control where we put an emphasis—on regression or benchmarking.

2.1 Additive approach

At first, we start from the equation (2). Let B be a variance-covariance matrix of β. We shall minimize

$$e'E^{-1}e + \beta'B^{-1}\beta \tag{11}$$

with respect to linear constraint $Y = AX\beta + Ae$. The term $\beta'B^{-1}\beta$ is a regularization term. In a case $B = I$, this term is equal to $\|\beta\|_2^2$—squared Euclidean norm of vector β. Regularization term is often used to prevent overfitting. In this case it enables to change focus from regression to benchmarking. In this approach β does not stem from minimizing sum of squared residuals $u'u$, but just from regularization term and benchmarking. Using Lagrange multipliers method we obtain

$$e = EA'(AEA' + AXBX'A')^{-1}Y, \tag{12}$$

$$\beta = BX'A'(AEA' + AXBX'A')^{-1}Y. \tag{13}$$

In consequence

$$y = X\beta + e = (XBX'A' + EA')(AEA' + AXBX'A')^{-1}Y. \tag{14}$$

What is worth to notice, (12) and (13) shows that estimates of β and e depend on a common component $(AEA' + AXBX'A')^{-1}$, which defines the trade-off between regression and benchmarking. Checking the constraint (2) we have

$$Ay = (AXBX'A' + AEA')(AEA' + AXBX'A')^{-1}Y = Y. \tag{15}$$

One may pose a question why not to obtain β with classical GLS method. Suppose we are interested in deriving β from minimizing sum of squared residuals $u'u$ instead using regularization term. Let $U = AEA'$. In this approach we drop regularization term. We shall minimize

$$e'E^{-1}e + u'U^{-1}u. \tag{16}$$

It is equivalent to

$$e'E^{-1}e + (Y - AX\beta)'U^{-1}(Y - AX\beta). \tag{17}$$

Let $V = (X'A'U^{-1}AX)^{-1}$. Using Lagrange multipliers method we obtain

$$e = EA'(U + AXVX'A')^{-1}(AXVX'A'U^{-1} + I)Y, \tag{18}$$

$$\beta = \left(\left(V - VX'A'(U + AXVX'A')^{-1}AXV\right)X'A'U^{-1} - VX'A'\left(U + AXVX'A'\right)^{-1}\right)Y. \tag{19}$$

Using the fact that if K and L are symmetric, then $KL = LK$, we obtain

$$\begin{aligned} Ae &= AEA'(U + AXVX'A')^{-1}(AXVX'A'U^{-1} + I)Y \\ &= U(U + AXVX'A')^{-1}(AXVX'A'U^{-1} + I)Y \\ &= (U + AXVX'A')^{-1}(UAXVX'A'U^{-1} + U)Y \\ &= (U + AXVX'A')^{-1}(AXVX'A' + U)Y \\ &= Y \end{aligned} \quad . \tag{20}$$

Thus $AX\beta = 0$ what can be derived explicitly from (15). In a result we obtained Denton additive method with constant as an explanatory variable which can be concerned as the least valuable method.

2.2 Proportional approach

Picking $E = I$ gives similar distribution results as additive Denton method. In fact, almost all of the methods mentioned previously distribute errors quite uniformly what will be presented in the result section. One may be interested in proportional distribution of error term. In Denton approach it is done by picking $E = \text{Diag}(X)\text{Diag}(X)$. For other methods we usually have more than one explanatory variable. This problem can be solved two-fold:

- picking one of the explanatory variables — available for all methods but not always suitable,
- picking $X\hat{\beta}$ — after obtaining estimate $\hat{\beta}$ of β we may define $E = \text{Diag}(X\hat{\beta})\text{Diag}(X\hat{\beta})$ or just which refer to squared Euclidean and squared chi-square distance function. Then we estimate β and e as in Additive approach.

3 Comparison of Methods — Case of Gross Value Added in Construction

We disaggregated quarterly data on gross value added in construction (current prices) with a use of monthly data on number of dwellings completed covering 2002–2015 period. Quarterly data on number of dwellings completed is definitely connected with gross value added in construction but moderately correlated — correlations coefficient is 0,544.[1] It is just exercise to show performance of the methods — not to disaggregate data as well as possible — and to reveal meaningfulness of our proposition. Methods proposed in this paper were applied using “Matrix” package in RStudio program and own code. All of the computation for others methods were made using “tempdisagg” package which is designed for temporal disaggregation process. These packages are available on R-CRAN project website.

For Denton method we used single indicator (number of dwellings completed) while for other methods we added constant term for regression stage of temporal disaggregation. For approach proposed and investigated in this paper we picked $B = I$ and $E = I$ in Additive approach, $E = \text{Diag}(X\hat{\beta})\text{Diag}(X\hat{\beta})$ in Proportional approach and simulate results for $E = \lambda I$ for different $\lambda > 0$ choosing one of the reasonable outcomes in Calibrated additive approach. Results of the methods are presented in table 1. Last three rows pertain to methods described in this paper. Disaggregated time series can be divided into two groups: Denton-Cholette, Proportional approach, Denton proportional and the rest. The second big group is very homogenous. Each two time series are linearly correlated with at least 0,94 of correlation coefficient. The first group is rather heterogeneous. It may be the result of picking different E matrix — for Denton it is based on X while for Proportional approach it is based on $X\hat{\beta}$. The table shows the crucial moment of temporal disaggregation — the third quarter of 2003. Let us have a look at what happens in July. Number of dwellings completed reached 43 000 which is four times greater than average in the 2002–2015 period.

Definitively, regression lead to very high estimate of gross value added in construction in July. In consequence, aggregated value added for the third quarter before benchmarking reached — e.g., 38 000 (29 in July, 4 in August, 3 in September) in Chow-Lin-maxlog method, while true value was 14 000. When the discrepancy of 24 000 was distributed quite uniformly then each monthly

1. [In the journal European practice of number notation is followed — for example, 36 333,33 (European style) = 36 333.33 (Canadian style) = 36,333.33 (US and British style). — Ed.]

Tab. 1. Results of the methods for 2003

	Month											
Method	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Dwellings completed	9 607	9 554	8 492	9 573	10 524	18 585	43 492	6 458	8 174	9 548	9 214	19 465
Denton-Cholette	2 848	2 020	1 703	2 444	2 959	5 114	9 261	1 878	3 184	4 913	5 051	9 606
Denton additive	2 960	2 508	1 103	1 896	1 491	7 130	28 547	−9 101	−5 123	1 389	3 794	14 386
Denton proportional	2 853	2 019	1 699	2 443	2 959	5 115	9 261	1 878	3 184	4 913	5 051	9 606
Fernzandez	3 002	2 278	1 291	2 194	2 166	6 155	20 816	−4 755	−1 738	3 149	4 801	11 620
Litterman-maxlog	3 002	2 278	1 291	2 194	2 166	6 155	20 816	−4 755	−1 738	3 149	4 801	11 620
Litterman-Linrss	2 644	2 363	1 564	2 409	1 849	6 258	25 301	−7 479	−3 499	1 867	4 307	13 396
Chow-Lin-minrss-ecotrim	2 998	2 274	1 298	2 107	2 185	6 225	20 418	−4 469	−1 626	3 341	4 838	11 391
Chow-Lin-minrss-quilis	3 001	2 283	1 287	2 184	2 154	6 178	20 955	−4 831	−1 801	3 122	4 782	11 665
Additive approach	2 433	2 400	1 738	1 436	2 028	7 053	19 806	−3 276	−2 207	4 532	4 324	10 713
Proportional approach	2 144	2 153	2 274	3 702	3 819	2 996	6 176	3 667	4 481	5 322	5 164	9 083
Calibrated additive approach	2 322	2 304	1 945	2 384	2 705	5 428	12 921	411	991	5 444	5 331	8 794

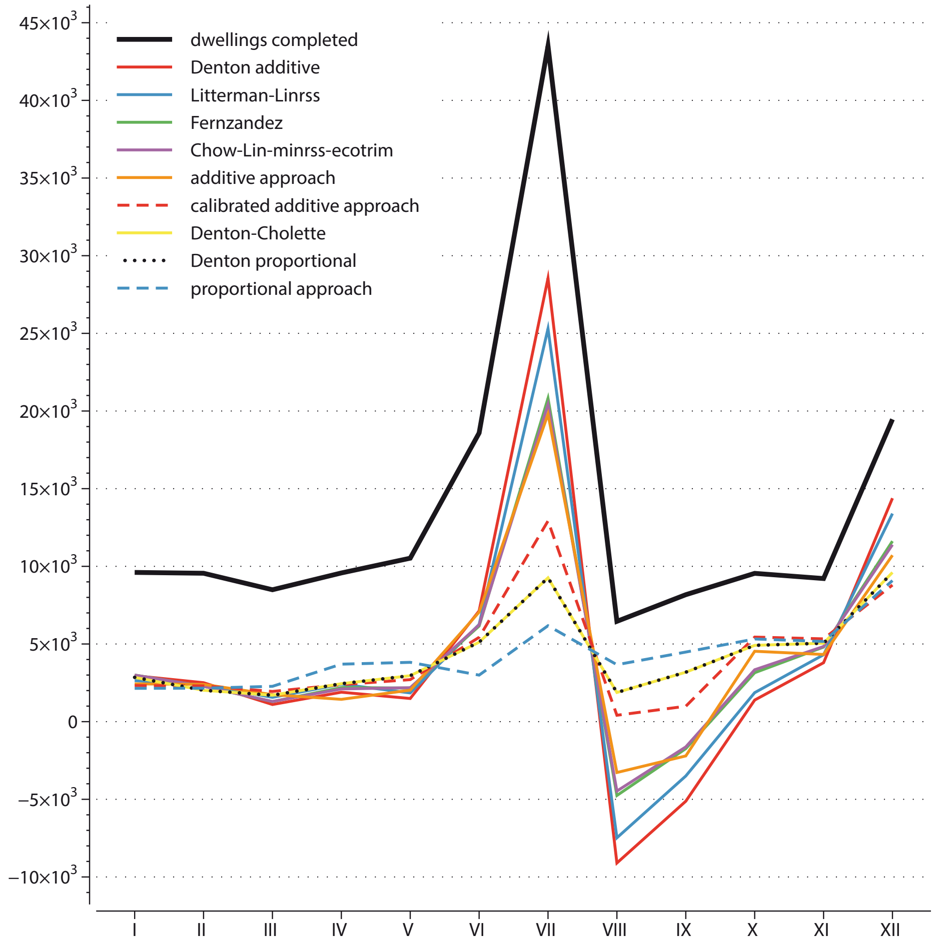

Fig. 1. Graphical representation of selected methods results

estimate was decreased by 8 000 on average. Most of the methods produced negative results in the third quarter what is unacceptable. Figure 1 shows how the selected methods reflected pick of dwellings completed in July. What is worth to notice, in a case of moderate and low correlation between low frequency data variance-covariance matrix based on $X\hat{\beta}$ produces "flatter" time series than matrix based on what is visible comparing Proportional approach and Denton proportional. X affects error term e as it would be high correlation between low frequency data while it is not a case. $X\hat{\beta}$ reflects it in a better way. In a case of high correlation $X\hat{\beta}$ and X would produce similar distribution of error term.

Additive approach produced very similar outcomes to — for example — Chow-Lin and Fernandez. It is especially interesting due to the fact that $\hat{\beta}$ was not obtained classically from GLS regression for low frequency data but just from minimizing RRS for error term and regularization term.

Calibrated additive approach seems to be most suitable. It allows to avoid negative values. Moreover, changing lambda parameter smoothly changes outcomes from Additive approach towards Proportional approach. For extremely high values of lambda parameter we obtain Denton additive approach with constant as an explanatory variable. The advantage over Proportional approach is that it is still one-stage procedure while Proportional approach involves obtaining $\hat{\beta}$ first and then performing whole procedure. That is, in fact, two-stage procedure. Lambda parameter allows to simply set the trade-off between regression and benchmarking.

Summary

Methods of temporal disaggregation will be paid more and more attention due to the possibility of obtaining high frequency data without increasing burden of interviewers and interviewees. At this moment most of the methods are two-stages consisting of regression and benchmarking. In a case of poor performance of regression stage, benchmarking may lead to unacceptable results. In this article we propose a method which performs regression and benchmarking at the same time and allows to set a trade-off between them.

References

Chamberlin, G. 2010. "Methods Explained: Temporal Disaggregation." *Economic & Labour Market Review* no. 4 (11):106–121. doi: 10.1057/elmr.2010.157.

Chow, G.C., and A.L. Lin. 1971. "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series." *Review of Economics and Statistics* no. 53 (4): 372–375. doi: 10.2307/1928739.

Dagum, E.B., and P.A. Cholette. 2006. *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series, Lecture Notes in Statistics*. New York: Springer.

Denton, F.T. 1971. "Adjustment of Monthly or Quarterly Series to Annual Totals — Approach Based on Quadratic Minimization." *Journal of the American Statistical Association* no. 66 (333):99–102.

Di Fonzo, T. 2003. "Temporal Disaggregation of a System of Time Series when the Aggregate is Known: Optimal vs. Adjustment Methods." In *Paris–Bercy 05/12/1994 – 06/12/1994. Workshop on Quarterly National Accounts*, edited by R. Barcellan and G.L. Mazzi, 63–77. Luxembourg: Office for Official Publications of the European Communities.

Fernandez, R.B. 1981. "A Methodological Note on the Estimation of Time-Series." *Review of Economics and Statistics* no. 63 (3):471–476. doi: 10.2307/1924371.

Litterman, R.B. 1983. "A Random Walk, Markov Model for the Distribution of Time Series" *Journal of Business & Economic Statistics* no. 1 (2):169–173.

Sax, C., and P. Steiner. 2013. "Temporal Disaggregation of Time Series." *R Journal* no. 5 (2):81–88.