Section B5: Benchmarking Methods for Time Series

B5.1. Overview of Temporal Disaggregation of Time Series

When we have a low frequency time series (e.g., annual), and we want a higher frequency times series (e.g., monthly or quarterly), we may apply temporal disaggregation methods to obtain an approximation of the desired series. Methods are available in statistical software packages. The basic steps in temporal disaggregation are as follows:

1) Determine a low-frequency series y for which a higher-frequency series is desired.
2) Determine an appropriate preliminary high-frequency series p, sometimes called a *pattern series*. The high-frequency movements of p should be similar to the expected high-frequency movements of the target series.
3) Benchmark p to force consistency between p and y at the lower-frequency level (e.g., the annual averages or sums).

Various methods of identifying or computing pattern series p are discussed in the statistics literature. In the regression-based approaches discussed below, multiple high-frequency series of explanatory variables may be used. When we already have a high-frequency series that we wish to benchmark to a lower-frequency series, we skip step 1. Step 3 above proceeds as follows:

3a) Aggregate p to the lower frequency (e.g., average or sum the quarterly values to obtain annual values).
3b) Compute the differences between the low-frequency values of p and the observed low-frequency values of y.
3c) Distribute the differences among the values of p, to force consistency with y.

B5.2. Types of Temporal Disaggregation Methods

Various methods involve different ways of distributing the differences (step 3c). Temporal disaggregation methods are categorized into (1) mathematical or operations research methods and (2) regression-based or statistical methods.

Operations research methods use a single indicator or pattern series for the preliminary high-frequency series p. To distribute the benchmarking adjustments, these methods optimize an objective function, subject to the benchmarking constraints. Operations research methods include the widely used Denton (1971) methods as well as the Causey-Trager method used at the U.S. Census Bureau.

Regression-based or statistical methods often use multiple indicator series to compute the preliminary series, and they use an estimated or assumed autocorrelation structure to distribute the benchmarking adjustments. Regression-based methods include the Chow-Lin (1971), Fernandez (1981), and Litterman (1983) methods.

Some benchmarking techniques are designed to work with stock variables, while others are designed to work with flow variables. Stock variables measure, at a specific point in time, a quantity that may have accumulated in the past, e.g., volume of natural gas in underground storage, total solar electricity generating capacity. Flow variables measure a quantity over an interval of time, e.g., natural gas consumption per month, solar electricity generation per year.

B5.3. Denton Benchmarking Methods

The Denton benchmarking methods are based on the concept of *movement preservation,* i.e., the benchmarked series θ follows the same pattern of movement over time as the preliminary high frequency series p. There are four basic variants of movement preservation:

1) Additive First Difference Method: Make the additive corrections $(\theta_t - p_t)$ as constant as possible over time so that the series θ is roughly parallel to p. In this approach, we minimize the sum of squared differences of the additive first differences:

$$\min_\theta[(\theta_t - p_t) - (\theta_{t-1} - p_{t-1})] = \min_\theta[\Delta(\theta_t - p_t)]. \tag{B5.3.1}$$

The additive first difference method is Denton's (1971) original method.

2) Additive Second Difference Method: Make the additive corrections as linear as possible by minimizing the sum of squared differences of the additive second differences:

$$\begin{aligned}\min_\theta\{[(\theta_t - p_t) - (\theta_{t-1} - p_{t-1})] - [(\theta_{t-1} - p_{t-1}) - (\theta_{t-2} - p_{t-2})]\}\\ = \min_\theta[\Delta^2(\theta_t - p_t)].\end{aligned} \tag{B5.3.2}$$

In this method, the additive correction $(\theta_t - p_t)$ is close to $2(\theta_{t-1} - p_{t-1}) - (\theta_{t-2} - p_{t-2})$, a linear function of the previous two corrections. This method is often used for stock variables.

3) Proportional First Difference Method: Make the multiplicative corrections $\frac{\theta_t}{p_t}$ as constant as possible by minimizing the sum of squared differences of the ratios $\frac{\theta_t}{p_t}$:

$$\min_\theta\left(\frac{\theta_t}{p_t} - \frac{\theta_{t-1}}{p_{t-1}}\right) = \min_\theta\left[\Delta\left(\frac{\theta_t}{p_t}\right)\right]. \tag{B5.3.3}$$

In this method, $\frac{\theta_t}{p_t}$ is close to $\frac{\theta_{t-1}}{p_{t-1}}$. This method is the most widely used of the Denton methods, because it is robust to scale.

4) Proportional Second Difference Method: Make the multiplicative corrections $\frac{\theta_t}{p_t}$ as linear as possible by minimizing the sum of squared second differences of the ratios $\frac{\theta_t}{p_t}$:

$$\min_\theta\left[\left(\frac{\theta_t}{p_t} - \frac{\theta_{t-1}}{p_{t-1}}\right) - \left(\frac{\theta_{t-1}}{p_{t-1}} - \frac{\theta_{t-2}}{p_{t-2}}\right)\right] = \min_\theta\left[\Delta^2\left(\frac{\theta_t}{p_t}\right)\right]. \tag{B5.3.4}$$

In this method, $\frac{\theta_t}{p_t}$ is close to $\frac{2\theta_{t-1}}{p_{t-1}} - \frac{\theta_{t-2}}{p_{t-2}}$. This method is often used for stock variables.

The derivations of the Denton methods are similar, involving the following basic steps:

i. Write the objective function $f(\boldsymbol{\theta})$ and the benchmarking constraints in vector notation.
ii. Multiply the constraint term by a vector $\boldsymbol{\gamma}$ of Lagrange multipliers, and add it to the objective function; the function becomes $f(\boldsymbol{\theta}, \boldsymbol{\gamma})$.
iii. Differentiate $f(\boldsymbol{\theta}, \boldsymbol{\gamma})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, and set the derivatives equal to $\mathbf{0}$.
iv. Solve for $\boldsymbol{\theta}$. Because $f(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is quadratic, it can be shown that the solution $\widehat{\boldsymbol{\theta}}$ is a global minimum.

The appendix provides the derivation of the additive first difference method. The original Denton methods implicitly incorporate an initial condition,

$$p_1 = \theta_1. \qquad \text{(B5.3.5)}$$

This constraint has been found disadvantageous for many applications, because it can result in a spurious movement at the beginning of the benchmarked series. The Denton-Cholette (1984) methods remove this initial condition and are generally preferred over the original Denton methods.

B5.4. Regression-based Methods

The regression-based temporal disaggregation methods use the following basic steps:

1) Instead of a single indicator series p, we assume we have a $T \times k$ matrix $\mathbf{X}$ consisting of T high frequency (HF) observations of $k \geq 1$ explanatory variables. We aggregate the HF series to the low-frequency (LF) level (e.g, compute annual averages from monthly data) to form the $M \times k$ matrix $\mathbf{JX}$, where $\mathbf{J}$ is an appropriate $M \times T$ aggregation matrix. (See the appendix for a definition of $\mathbf{J}$.)

2) We fit an ordinary (or generalized) least square regression model to the LF series y, using $\mathbf{JX}$ as the matrix of explanatory variables:

$$\mathbf{y} = \mathbf{JX}\beta + \boldsymbol{\varepsilon}. \qquad \text{(B5.4.1)}$$

3) The main assumption behind the regression-based methods is that the linear relationship between the LF series $\mathbf{y}$ and $\mathbf{JX}$ also holds between the HF series $\boldsymbol{\theta}$ and $\mathbf{X}$. We form the preliminary series by setting

$$\mathbf{p} = \mathbf{X}\hat{\beta}. \qquad \text{(B5.4.2)}$$

4) To benchmark the preliminary series, the Chow-Lin method assumes that the true additive corrections follow an autoregressive process of order 1 (AR1) with a constant correlation coefficient ρ, where $|\rho| < 1$:

$$u_t = \theta_t - p_t = \rho u_{t-1} + \epsilon_t, \quad \text{(B5.4.3)}$$

where ϵ_t is a normally distributed random error term with variance σ_ϵ^2.

5) Chow and Lin consider the variance-covariance matrix of the vector $\widehat{\boldsymbol{\theta}} - E(\boldsymbol{\theta})$, in terms of ρ and σ_ϵ^2. They use Lagrange multipliers to minimize the variance of $\widehat{\boldsymbol{\theta}} - E(\boldsymbol{\theta})$, subject to the benchmarking constraints. (The Chow-Lin estimator of $\boldsymbol{\theta}$ is a Best Linear Unbiased Estimator (BLUE)).

6) The solution takes the form

$$\widehat{\boldsymbol{\theta}} = \mathbf{X}\hat{\beta} + \boldsymbol{\Sigma}\mathbf{J}'(\mathbf{J}\boldsymbol{\Sigma}\mathbf{J}')^{-1}(\mathbf{J}\mathbf{X}\hat{\beta} - \mathbf{y}), \quad \text{(B5.4.4)}$$

where $$\boldsymbol{\Sigma} = \frac{\sigma_\epsilon^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix}.$$

Calculating $\widehat{\boldsymbol{\theta}}$ involves estimating the autoregressive parameter ρ. Chow and Lin suggest an iterative approach. Other methods have also been presented in the literature, e.g., Dagum and Cholette (2006) developed a method that allows users to specify ρ. The Fernandez and Litterman methods, useful for disaggregating stock variables, are similar to the Chow-Lin method, but they assume that the additive corrections follow a non-stationary autoregressive process:

$$u_t = \theta_t - p_t = \theta_{t-1} - p_{t-1} + v_t, \quad \text{(B5.4.5)}$$

where $v_t = \rho v_{t-1} + \epsilon_t$.

B5.5. Comparing Benchmarking Methods

The Denton proportional difference methods are generally robust to scale and provide good results when one good pattern series is available. They are, however, sensitive to outliers. The Chow-Lin method (and the other regression-based methods) assume that the strength of the correlation between the LF series holds for the HF series. If the LF series are weakly correlated, the regression-based methods don't preserve movement. The main advantage of the regression-based methods is that they can incorporate multiple pattern series. They are also fairly robust to outliers.

B5.6. Appendix: Deriving the Denton First Difference Method

<u>Notation</u>

Let M represent the number of low-frequency (LF) time periods (e.g., years) in the data series. Let L represent the number of high-frequency (HF) periods in each LF period (e.g., for monthly data, $L = 12$). Let $m = 1, \dots, M$ represent the LF time periods and denote the HF periods in each LF period m by $t_{m,1}, \dots, t_{m,L}$. Let $T = ML$, the total number of HF time periods in the data series.

The summing (or averaging) weights $j_{m,t}$ form an MxT matrix **J:**

$$\mathbf{J} = \begin{bmatrix} j_{11} & j_{12} \cdots & j_{1T} \\ j_{21} & j_{22} \cdots & j_{2T} \\ \vdots & \vdots \ \vdots & \vdots \\ j_{M1} & j_{M2} \cdots & j_{MT} \end{bmatrix}. \quad \text{(B5.6.1)}$$

For example, when benchmarking two years of quarterly data to annual sums, we set

$$\mathbf{J} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}. \quad \text{(B5.6.2)}$$

For annual averages of quarterly benchmarked data,

$$\mathbf{J}\boldsymbol{\theta} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \theta_{13} \\ \theta_{14} \\ \theta_{21} \\ \theta_{22} \\ \theta_{23} \\ \theta_{24} \end{bmatrix} = \begin{bmatrix} \frac{1}{4}\sum_{l=1}^{4} \theta_{1l} \\ \frac{1}{4}\sum_{l=1}^{4} \theta_{2l} \end{bmatrix}. \quad \text{(B5.6.3)}$$

The benchmarking constraints are therefore written as $\mathbf{y} - \mathbf{J\theta} = \mathbf{0}$.

All of the Denton methods use a single indicator series for the preliminary high-frequency series p. The indicator series may be a HF series that we'd like to benchmark to a low-frequency series or, in the case of temporal disaggregation, a series whose short-term movements we believe to correlate with the target HF series.

For each HF time period t, let p_t and θ_t represent the values of the preliminary and benchmarked series, respectively. Then

$$p_t = \theta_t + e_t, \quad \text{(B5.6.4)}$$

where e_t is an error term to be estimated. The benchmarking constraints may be expressed as $\mathbf{y} - \mathbf{J\theta} = \mathbf{0}$ or

$$y_m = \sum_{t=t_{m,1}}^{t_{m,L}} j_{t,m}\theta_t, \quad \text{(B5.6.5)}$$

for $m = 1, \dots, M$, where y_m is the LF series to which we wish to benchmark the p_t (e.g., y_m may represent annual averages), and the $j_{t,m}$ are weights (e.g., $\frac{1}{12}, \frac{1}{4}, 1$).

To optimize the objective function and solve for the vector $\boldsymbol{\theta}$, we write the function as a quadratic form in matrix notation and add the linear benchmarking constraints with a vector $\boldsymbol{\gamma}$ of Lagrange multipliers:

$$f(\boldsymbol{\theta},\boldsymbol{\gamma}) = (\boldsymbol{\theta}-\mathbf{p})'\mathbf{D}'\mathbf{D}(\boldsymbol{\theta}-\mathbf{p}) - 2\boldsymbol{\gamma}'(\mathbf{y}-\mathbf{J}\boldsymbol{\theta}), \tag{B5.6.6}$$

where the matrix $\mathbf{D}$ is the first-difference operator matrix of order T:

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ -1 & 1 & 0 & \cdots \\ 0 & -1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{B5.6.7}$$

The matrix $\mathbf{D}'\mathbf{D}$ is symmetric, and the factor 2 in the second term of $f(\boldsymbol{\theta},\boldsymbol{\gamma})$ is applied for computational convenience. We compute partial derivatives of $f(\boldsymbol{\theta},\boldsymbol{\gamma})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ and set them equal to 0:

$$\begin{gathered}\frac{\partial}{\partial\boldsymbol{\theta}} f(\boldsymbol{\theta},\boldsymbol{\gamma}) = 2\mathbf{D}'\mathbf{D}\boldsymbol{\theta} - 2\mathbf{D}'\mathbf{D}\mathbf{p} + 2\mathbf{J}'\boldsymbol{\gamma} = \mathbf{0} \\ \Rightarrow \mathbf{D}'\mathbf{D}\boldsymbol{\theta} + \mathbf{J}'\boldsymbol{\gamma} = \mathbf{D}'\mathbf{D}\mathbf{p}.\end{gathered} \tag{B5.6.8}$$

Similarly,

$$\begin{gathered}\frac{\partial}{\partial\boldsymbol{\gamma}} f(\boldsymbol{\theta},\boldsymbol{\gamma}) = \mathbf{2J}\boldsymbol{\theta} - 2\mathbf{y} = \mathbf{0} \\ \Rightarrow \mathbf{J}\boldsymbol{\theta} = \mathbf{J}\mathbf{p} + (\mathbf{y}-\mathbf{J}\mathbf{p}).\end{gathered} \tag{B5.6.9}$$

The equations

$$\mathbf{D}'\mathbf{D}\boldsymbol{\theta} + \mathbf{J}'\boldsymbol{\gamma} = \mathbf{D}'\mathbf{D}\mathbf{p} \quad \text{and} \quad \mathbf{J}\boldsymbol{\theta} = \mathbf{J}\mathbf{p} + (\mathbf{y}-\mathbf{J}\mathbf{p}) \tag{B5.6.10}$$

can be written in partitioned matrix form as

$$\begin{bmatrix} \mathbf{D}'\mathbf{D} & \mathbf{J}' \\ \mathbf{J} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{D}'\mathbf{D} & \mathbf{0} \\ \mathbf{J} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ (\mathbf{y}-\mathbf{J}\mathbf{p}) \end{bmatrix}, \tag{B5.6.11}$$

which gives the solution

$$\begin{bmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{D}'\mathbf{D} & \mathbf{J}' \\ \mathbf{J} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}'\mathbf{D} & \mathbf{0} \\ \mathbf{J} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ (\mathbf{y}-\mathbf{J}\mathbf{p}) \end{bmatrix}. \tag{B5.6.12}$$

Statistical software packages perform the matrix inversion numerically to compute $\hat{\boldsymbol{\theta}}$.

<u>References</u>

Cholette, P. A., 1984. "Adjusting Sub-annual Series to Yearly Benchmarks," *Survey Methodology, Vol. 10*, pp. 35–49.

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, Nov. 1971. pp. 80-83. Available at http://www.oecd.org/std/21779751.pdf.

Dagum, E. B. and Cholette, P. A. (2006). *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series. Lecture Notes in Statistics.* Springer-Verlag, 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*, pp. 99–102, March 1971.

Fernández, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics, 63(3*) pp. 471–476.

Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics, 1(2)*, pp. 169–173.

Sax, Christoph; and Steiner, Peter (2013). Temporal Disaggregation of Time Series. Contributed research article, *The R Journal Vol. 5/2*, December 2013

Sayal, Homesh; Aston, John A. D.; Duncan, Elliott; and Ombao, Hernando (2017). An introduction to applications of wavelet benchmarking with seasonal adjustment. *Journal of the Royal Statistical Association, Statistics in Society (Series A)*. Available at https://www.repository.cam.ac.uk/handle/1810/260183.