APPENDIX 4A: Disaggregation Procedures

Disaggregation

Suppose we have a sequence of numbers

$$y_1, y_2, y_3, \ldots, y_{2n}$$

Let

$$\left.\begin{aligned} s_1 &= (y_1 + y_2)/2 \\ s_2 &= (y_3 + y_4)/2 \\ &\vdots \\ s_n &= (y_{2n-1} + y_{2n})/2 \end{aligned}\right\} \longrightarrow A$$

Problem: Given the numbers $s_1, s_2, \ldots, s_n$, recover the numbers $y_1, y_2, \ldots, y_{2n}$.

Remark. Clearly there is no unique solution to the problem since the equations in A constitute n equations in $2n$ unknowns. Therefore additional assumptions or information is required to carry out the disaggregation. We shall assume that any consecutive set of $2k + 2$ values $y_{2r+1}, y_{2r+2}, \ldots, y_{2(r+k)+1}, y_{2(r+k)+2}$ of the y sequence lie "approximately" on a polynomial curve of degree k.

We will investigate here only the cases $k = 0, 1, 2, 3$.

CASE 1: $k = 0$.

Here we are assuming that every pair of adjacent y values is "approximately" equal. This leads to the estimates

$$y_{2i+1} = y_{2i} = s_i$$

because

$$y_1 = y_2 \text{ and } s_1 = \frac{y_1 + y_2}{2} \Longrightarrow y_1 \approx s_1,\ y_2 \approx s_1.$$

CASE 2: $k = 1$.

Consider y_1, y_2, y_3, y_4. They lie on a straight line "approximately".

So

$$\left.\begin{array}{rcl} y_1 & \approx & a+b \\ y_2 & \approx & a+2b \\ y_3 & \approx & a+3b \\ y_4 & \approx & a+4b \end{array}\right. \quad \text{for some } a, b.$$

So

$$s_1 = a + \frac{3}{2}b, \; s_2 = a + \frac{7}{2}b \Longrightarrow 2b = s_2 - s_1 \text{ or } b = \frac{s_2 - s_1}{2}$$

and

$$a = s_1 - \frac{3}{4}(s_2 - s_1) = \frac{7s_1 - 3s_2}{4}.$$

So

$$y_1 = a + b = \frac{7s_1 - 3s_2}{4} + \frac{2s_2 - 2s_1}{4} = \frac{5s_1 - s_2}{4}$$

$$y_2 = a + 2b = \frac{7s_1 - 3s_2 + 4s_2 - 4s_1}{4} = \frac{3s_1 + s_2}{4}$$

$$y_3 = a + 3b = \frac{7s_1 - 3s_2 + 6s_2 - 6s_1}{4} = \frac{s_1 + 3s_2}{4}$$

$$y_4 = a + 4b = \frac{7s_1 - 3s_2 + 8s_2 - 8s_1}{4} = \frac{-s_1 + 5s_2}{4}.$$

Note that

$$y_1 + y_2 = 2s_1$$

$$y_3 + y_4 = 2s_2.$$

Hence the disaggregation procedure may be carried out as follows:

$$\left.\begin{array}{rcl} y_1 & = & \frac{5s_1 - s_2}{4} \\ y_2 & = & \frac{3s_1 + s_2}{4} \end{array}\right\} \text{based on the sequence } \underbrace{y_1, y_2}_{s_1}, \underbrace{y_3, y_4}_{s_2}.$$

$$\left.\begin{array}{rcl} y_3 & = & \frac{5s_2 - s_3}{4} \\ y_4 & = & \frac{3s_2 + s_3}{4} \end{array}\right\} \text{based on the sequence } \underbrace{y_3, y_4}_{s_2}, \underbrace{y_5, y_6}_{s_3}.$$

For $2 \leq r \leq n-1$ carry out,

$$\left.\begin{array}{rcl} y_{2r-1} & = & \frac{5s_r - s_{r+1}}{4} \\ y_{2r} & = & \frac{3s_r + s_{r+1}}{4} \end{array}\right\} \text{based on the sequence } \underbrace{y_{2r-1}, y_{2r}}_{s_r}, \underbrace{y_{2r+1}, y_{2r+2}}_{s_{r+1}}.$$

And finally

$$\left.\begin{array}{lcl} y_{2n-1} & = & \frac{s_{n-1}+3s_n}{4} \\ y_{2n} & = & \frac{-s_{n-1}+5s_n}{4} \end{array}\right\} \begin{array}{l}\text{based on the sequence} \\ \underbrace{y_{2n-3}, y_{2n-2}}_{s_{n-1}}, \underbrace{y_{2n-1}, y_{2n}}_{s_n}\end{array}$$

Alternatively for $2 \leq r \leq n-1$, we could have taken

$$\left.\begin{array}{lcl} y_{2r-1} & = & \frac{s_{r-1}+3s_r}{4} \\ y_{2r} & = & \frac{-s_{r-1}+5s_r}{4} \end{array}\right\} \begin{array}{l}\text{based on the sequence} \\ \underbrace{y_{2r-3}, y_{2r-2}}_{s_{r-1}}, \underbrace{y_{2r-1}, y_{2r}}_{s_r}.\end{array}$$

Note that for $1 \leq i \leq n$

$$2s_i = y_{2i} + y_{2i-1}.$$

The same reasoning is used to derive the disaggregation formulae for $k = 2$ and $k = 3$. They are simply stated below.

<u>CASE 3:</u> $k = 2$.

$$\left.\begin{array}{lcl} y_1 & = & \frac{s_3-4s_2+11s_1}{8} \\ y_2 & = & \frac{-s_3+4s_2+5s_1}{8} \end{array}\right\} \begin{array}{l}\text{based on the sequence} \\ \underbrace{y_1, y_2}_{s_1}, \underbrace{y_3, y_4}_{s_2}, \underbrace{y_5, y_6}_{s_3}.\end{array}$$

For $2 \leq r \leq n-1$

$$\left.\begin{array}{lcl} y_{2r-1} & = & \frac{s_{r-1}+8s_r-s_{r+1}}{8} \\ y_{2r} & = & \frac{-s_{r-1}+8s_r+s_{r+1}}{8} \end{array}\right\} \begin{array}{l}\text{based on the sequence} \\ \underbrace{y_{2r-3}, y_{2r-2}}_{s_{r-1}}, \underbrace{y_{2r-1}, y_{2r}}_{s_r}, \underbrace{y_{2r+1}, y_{2r+1}}_{s_{r+1}}.\end{array}$$

Finally

$$\left.\begin{array}{lcl} y_{2n-1} & = & \frac{5s_n+4s_{n-1}-s_{n-2}}{8} \\ y_{2n} & = & \frac{11s_n-4s_{n-1}+s_{n-2}}{8} \end{array}\right\} \begin{array}{l}\text{based on the sequence} \\ y_{2n-5}, y_{2n-4}, y_{2n-3}, y_{2n-2}, y_{2n-1}, y_{2n} \\ \text{(i.e. } s_{n-2}, s_{n-1}, s_n).\end{array}$$

<u>CASE 4a:</u> $k = 3$ and n is even.

$$\left.\begin{array}{lcl} y_1 & = & \frac{-5s_4+23s_3-47s_2+93s_1}{64} \\ y_2 & = & \frac{5s_4-23s_3+47s_2+35s_1}{64} \end{array}\right\} \begin{array}{l}\text{based on the sequence} \\ y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8 \\ \text{(i.e. } s_1, s_2, s_3, s_4).\end{array}$$

For $r = 2, 4, 6, \ldots$,etc. up to $r = n - 2$ carry out,

$$\left.\begin{array}{lcl} y_{2r-1} & = & \frac{3s_{r+2}-17s_{r+1}+73s_r+5s_{r-1}}{64} \\ y_{2r} & = & \frac{-3s_{r+2}+17s_{r+1}+55s_r-5s_{r-1}}{64} \\ y_{2r+1} & = & \frac{5s_{r+2}+55s_{r+1}+17s_r-5s_{r-1}}{64} \\ y_{2r+2} & = & \frac{5s_{r+2}+73s_{r+1}-17s_r+3s_{r-1}}{64} \end{array}\right\}$$

based on the sequence $y_{2r-3}, y_{2r-2}, y_{2r-1}, y_{2r}, y_2r + 1, y_{sr+2}, y_{2r+3}, y_{2r+4}$ (i.e. $s_{r-1}, s_r, s_{r+1}, s_{r+2}$).

Ending with

$$\left.\begin{array}{lcl} y_{2n-1} & = & \frac{35s_n+47s_{n-1}-23s_{n-2}+5s_{n-3}}{64} \\ y_{2n} & = & \frac{93s_n-47s_{n-1}+23s_{n-2}-5s_{n-3}}{64} \end{array}\right\}$$

based on the sequence $y_{2n-7}, y_{2n-6}, y_{2n-5}, y_{2n-4},, y_{2n-3}, y_{2n-2}, y_{2n-1}, y_{2n}$ (i.e. $s_{n-3}, s_{n-2}, s_{n-1}, s_n$).

CASE 4b: $k = 3$ and n is odd.

y_1, y_2 same as above.

$y_{2r-1}, y_{2r}, y_{2r+1}, y_{2r+2}$ same as above for $r = 2, 4, 6, \ldots, n - 3$.

$$\left.\begin{array}{lcl} y_{2n-3} & = & \frac{-5s_n+55s_{n-1}+17s_{n-2}-3s_{n-3}}{64} \\ y_{2n-2} & = & \frac{5s_n+73s_{n-1}-17s_{n-2}+3s_{n-3}}{64} \end{array}\right\}$$

based on the sequence $y_{2n-7}, y_{2n-6}, y_{2n-5}, y_{2n-4}, y_{2n-3}, y_{2n-2}, y_{2n-1}, y_{2n}$ (i.e. $s_{n-3}, s_{n-2}, s_{n-1}, s_n$).

y_{2n-1}, y_{2n} as above.

Conclusions

These disaggregation procedures need to be tested under a variety of conditions to evaluate their performance characteristics. Disaggregation of 12 hour data to 6 hour or 48 to 24 may work better than disaggregation of 24 hour data to 12 hour data. The reason is the diurnal variation which has not been explicitly taken into account. However, procedures are available to account for this.