Estimating the Amplitude and Phase in Studies of Seasonality in Serum Cholesterol

Ed Stanek (3/18/99)

Introduction

Seasonal variations have been reported in cross-sectional studies of the general population in serum cholesterol levels ( ). Estimates of the amplitude of the seasonal effects observed in cross-sectional studies (peak to trough) have varied from 8-12mg/dl. The estimates indicate that the serum cholesterol levels peak in the winter for most subjects, with different phases observed for different blood fractions. Such results have yet to be incorporated in recommendations for cholesterol screening, and are at present unexplainable. Factors that may contribute to the seasonality include changes in diet, physical activity, and light exposure.

The Seasons study was conducted in an attempt to identify the contribution each of these sources makes to seasonality in serum cholesterol and other blood fractions, and to more precisely quantify the amplitude and phase of the seasonal effects. The study is longitudinal, with four measures of serum cholesterol made on each of 641 subjects over a period of one year. Subjects are enrolled in the study in a staggered fashion, with the entire data collection phase of the study extending over 3 years.

Practical aspects of study staffing require continuous data collection over the calendar year, and a relatively constant flow of data collection due to the staggered recruitment of subjects. As a result, measures of serum cholesterol will not be made on all subjects at the same time of the year. For example, some subjects may have their first measure made in October, with subsequent measures separated by three month intervals, while other subjects will have their first measure made in November, with a similar staggering of subsequent measures. Assuming an identical sinusoidal seasonal effect for serum cholesterol among several subjects (with a peak serum cholesterol at Jan 1), the different measurement time periods for subjects will result in observed differences in the serum blood pressures measured between subjects, even though all subjects are following the same seasonal pattern.

This document provides details for estimating amplitude and phase for seasonal effects from a mixed model analysis. The details are provided to amplify and clarify the steps needed to construct the estimates in computing algorithms.

Basic Methods for Assessing Seasonal Changes

Several methods can be proposed to estimate of the amplitude (peak to trough distance) of seasonal effects. Perhaps the simplest strategy is to assume that serum cholesterol levels made within a certain window (say plus or minus 30 days) of the zenith in serum cholesterol (assumed, for example, to be at Jan 1), are all made at the zenith, and to similarly assume that similar levels within the azimuth (July 1 plus or minus 30 days) are made at the trough. Paired comparison of the two measures will result in estimates of the amplitude of the seasonal effects for the subjects.
This method has the appeal of simplicity. Limitations of the method include practical difficulty of obtaining serum cholesterol values within the windows for all patients, the approximation of "peak" and "trough" estimates of amplitude, even though the timing of the measures may not coincide with the subject specific peaks and troughs, the arbitrariness of the "window" width, and the lack of utilization of the other serum cholesterol measures at the fall and spring times. Nevertheless, the method is simple and easy to grasp.

A second method that is conceptually easy to use is to classify data into defined time periods (which we call seasons). This strategy was adopted as a preliminary strategy in the Seasons study. The time periods that define the seasons were defined in two ways. First, Light Seasons were defined, where

Light seasons are defined as the following intervals:

- Season 1: Nov. 6 - Feb. 4
- Season 2: Feb. 5 - May 6
- Season 3: May 7 - August 5
- Season 4: August 6 - Nov 5

In addition, seasons were defined in a second manner corresponding to Common Seasons as follows:

Common seasons are defined as the following intervals:

- Season 1 (winter): Dec. 21 - March 20
- Season 2 (spring): March 21- June 20
- Season 3 (summer): June 21 - Sept. 20
- Season 4 (fall): Sept. 21 - Dec 20

The advantage of this second method of defining seasons is that once again, simple comparisons can be made between different quarters. The disadvantages are similar to those for comparison of peak to trough differences.

A third strategy for estimating amplitude is to use the trigonometric functions to decompose the four measures of serum cholesterol into three parameters, with combinations of these parameters representing the amplitude and phase for cholesterol. Such a decomposition is a simple example of a time series approach based on spectral analysis (Koopman, 1974 and Bloomfield, 1976). Defining time in radians such that the Julian day (1 to 365) is converted to "t" where

\[ t = 2 \pi \text{(Julian Day)}/365 \]

then serum cholesterol for the \( i \)th subject at time "t" may be represented in terms of a common amplitude (A) and phase (F) as
\[ y_i = \beta_{i1} + (A/2) \sin[t+(F+\pi/2)] \]

where \( F \) is the phase in radians. Thus, with \( F=0 \), the model will predict a maximum serum cholesterol value on Dec 30, having a value of \( \beta_1 + A/2 \). Using a standard trigonometric identity, we can re-express serum cholesterol for subject \( i \) at time \( t_i \) as

\[ \mu_{it} = \beta_{i1} + [(A/2)\cos(F+\pi/2)] \sin(t_i) + [(A/2)\sin(F+\pi/2)] \cos(t_i) \]

which, setting \( \beta_2 = [(A/2)\cos(F+\pi/2)] \) and \( \beta_3 = [(A/2)\sin(F+\pi/2)] \) can be expressed as the standard regression equation

\[ \mu_{it} = \beta_{i1} + \beta_2 \sin(t_i) + \beta_3 \cos(t_i). \]

Values of the phase and amplitude can be recovered from the regression coefficients using the expression:

\[ F+\pi/2 = \text{arcTan}(\beta_3/\beta_2) \quad \text{and} \quad A = \text{Abs}[2\beta_2/cos(F+\pi/2)]. \]

If the sign of the coefficients \( \beta_2 \) and \( \beta_3 \) are the same, then \( F \) must be negative so that the peak amplitude is prior to Dec. 30. If one coefficient is positive, and the other negative, then the peak amplitude is after Dec. 30, in the Spring of the year.

The phase, \( F \), will be expressed in radians, and can be positive or negative. Note that \( F=0 \) corresponds to Dec. 30. Similarly, \( F=\pi \) corresponds to a Julian day of 182.5. A positive value of \( F \) indicates that the peak response actually occurs in Jan or Feb (after Dec 30.). A negative \( F \) indicates that the peak response occurs prior to Dec 30. To convert the phase to a date where the amplitude peaks, we first express the phase as a julian date via the expression:

\[ F^* = \text{Phase(in Julian days)} = \text{INT}(365.25*F/2\pi) \]

where INT indicates the integer part of the expression. When \( F^* \) is positive, it corresponds to the Julian date of peak amplitude. When \( F^* \) is negative, the peak amplitude is on Julian date \( (365+F^*) \).

**Constructing Estimates in SAS**

The estimates of the amplitude and phase in SAS are constructed using trig functions. First, an angle is calculated corresponding to the \( \tan^{-1}(\beta_3/\beta_2) \). This angle will vary from \( \pi/2 \) to \( -\pi/2 \), with the angle being negative if the quotient is negative. Since the phase can vary over \( 2\pi \), the arc tangent will not uniquely identify the location of the peak response. Instead, it will identify either the peak, or trough in the response. A simple simulation varying the phase illustrates this relationship. We vary the phase for the true peak response over a calendar year, and record the estimates of the amplitude, the regression coefficients, and the value of \( F+\pi/2 \), as produced by the function \( \text{atan}(\beta_3/\beta_2) \). In addition, we calculate the amplitude as \( A=2\beta_2/cos(F+\pi/2) \). These results
were obtained using the program SNE99P39.SAS.

<table>
<thead>
<tr>
<th>True Peak Response (Julian Day)</th>
<th>Amplitude $2\beta_2/\cos(\pi/2)$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$(\pi/2)$ in Radians from $\text{atan}(\beta_3/\beta_2)$</th>
<th>Estimated Peak Response (Julian Day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-80</td>
<td>-0.009</td>
<td>40.1</td>
<td>-1.570</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>80.1</td>
<td>28</td>
<td>28</td>
<td>0.7969</td>
<td>45</td>
</tr>
<tr>
<td>90</td>
<td>80.2</td>
<td>40</td>
<td>0.9</td>
<td>0.023</td>
<td>90</td>
</tr>
<tr>
<td>135</td>
<td>80.2</td>
<td>29</td>
<td>-27</td>
<td>-0.747</td>
<td>135</td>
</tr>
<tr>
<td>180</td>
<td>80</td>
<td>2</td>
<td>-39</td>
<td>-1.52</td>
<td>180</td>
</tr>
<tr>
<td>210</td>
<td>-79</td>
<td>-17</td>
<td>-35</td>
<td>1.102</td>
<td>209</td>
</tr>
<tr>
<td>240</td>
<td>-79</td>
<td>-33</td>
<td>-21</td>
<td>0.582</td>
<td>240</td>
</tr>
<tr>
<td>270</td>
<td>-79</td>
<td>-39</td>
<td>-2.5</td>
<td>0.0626</td>
<td>270</td>
</tr>
<tr>
<td>300</td>
<td>-79</td>
<td>-36</td>
<td>17</td>
<td>-0.454</td>
<td>300</td>
</tr>
<tr>
<td>330</td>
<td>-80</td>
<td>-23</td>
<td>33</td>
<td>-0.969</td>
<td>330</td>
</tr>
</tbody>
</table>

Using these results, we can determine an expression to estimate the amplitude and phase for peak response. These expressions are:

When $\beta_2 > 0$,

\[
\text{Julian Peak} = -\tan^{-1}\left(\frac{\beta_3}{\beta_2}\right) - \frac{\pi}{2} + \frac{365.25}{2\pi} \left(\text{atan}\left(\frac{\beta_3}{\beta_2}\right) + \frac{\pi}{2}\right) - \frac{365.25}{2\pi}.
\]

and when $\beta_2 < 0$,

\[
\text{Julian Peak} = 365.25 + \left[\tan^{-1}\left(\frac{\beta_3}{\beta_2}\right) - \frac{\pi}{2}\right] - \frac{365.25}{2\pi} \left(\text{atan}\left(\frac{\beta_3}{\beta_2}\right) + \frac{\pi}{2}\right).
\]
Estimating the Variance of Amplitude and Phase

We estimate the variance of the amplitude and phase using the Delta method based on a first order Taylor Series expansion. Basically, the estimates are constructed as a linear transformation of the estimated regression coefficients, where the linear transformation is defined by the matrix of first derivatives, evaluated at the estimated values.

In our example, let

\[
F = \begin{pmatrix}
    f(\boldsymbol{\beta}) \\
    g(\boldsymbol{\beta})
\end{pmatrix} = \begin{pmatrix}
    \frac{2\beta_2}{\cos\tan^{-1}(\beta_3/\beta_2)} \\
    -\tan^{-1}(\beta_3/\beta_2) - \frac{2\pi}{2}\frac{365.25}{2\pi}
\end{pmatrix} = \begin{pmatrix}
    \text{amplitude} \\
    \text{phase}
\end{pmatrix}
\text{where } \boldsymbol{\beta} = \begin{pmatrix}
    \beta_2 \\
    \beta_3
\end{pmatrix}.
\]

Then define

\[
C = \begin{pmatrix}
    \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_2} & \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_3} \\
    \frac{\partial g(\boldsymbol{\beta})}{\partial \beta_2} & \frac{\partial g(\boldsymbol{\beta})}{\partial \beta_3}
\end{pmatrix}.
\]

Then a first order Taylor Series approximation to the variance of \(F\) is given by \(\text{var}(F) = CVC\). To evaluate this expression, we need to evaluate the derivatives of each term. We perform this evaluation step by step below.

A. Evaluating \(\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_2}\).

Let

\[
f(\boldsymbol{\beta}) = \frac{u}{v} = \frac{2\beta_2}{\cos\tan^{-1}(\beta_3/\beta_2)} \text{ where } u = 2\beta_2 \text{ and } v = \cos\tan^{-1}(\beta_3/\beta_2).
\]
Then

\[
\frac{\partial f(\beta)}{\partial \beta_2} = \frac{v \frac{\partial u}{\partial \beta_2} - u \frac{\partial v}{\partial \beta_2}}{v^2}
\]

where

\[\frac{\partial u}{\partial \beta_2} = 2, \quad \text{and} \quad \frac{\partial v}{\partial \beta_2} = \frac{\partial v \, \partial w \, \partial x}{\partial \beta_2} \]

and

\[v = \cos(w); \quad \frac{\partial v}{\partial w} = -\sin(w) = -\sin\left[\tan^{-1}\left(\frac{\beta_3}{\beta_2}\right)\right]\]

\[w = \tan^{-1}(x); \quad \frac{\partial w}{\partial x} = (1 + x^2)^{-1} = \frac{1}{1 + \left(\frac{\beta_3}{\beta_2}\right)^2}, \quad \text{and} \quad x = \frac{\beta_3}{\beta_2}, \quad \frac{\partial x}{\partial \beta_2} = -\frac{\beta_3}{\beta_2^2}.
\]

Combining terms,

\[
\frac{\partial f(\beta)}{\partial \beta_2} = \frac{2\cos[\tan^{-1}(\beta_3/\beta_2)] - 2\beta_2\left\{-\sin[\tan^{-1}(\beta_3/\beta_2)]\right\} - \frac{\beta_3}{\beta_2}}{\cos[\tan^{-1}(\beta_3/\beta_2)]^2}
\]
B. Evaluating $\frac{\partial f(\beta)}{\partial \beta_3}$.

Let

$$f(\beta) = \frac{u}{v} = \frac{2\beta_2}{\cos\tan^{-1}(\beta_3/\beta_2)} \quad \text{where} \quad u = 2\beta_2 \quad \text{and} \quad v = \cos\tan^{-1}(\beta_3/\beta_2).$$

As before,

$$\frac{\partial f(\beta)}{\partial \beta_3} = \frac{v \frac{\partial u}{\partial \beta_3} - u \frac{\partial v}{\partial \beta_3}}{v^2}$$

where

$$\frac{\partial u}{\partial \beta_3} = 0, \quad \text{and} \quad \frac{\partial v}{\partial \beta_3} = \frac{\partial v}{\partial \beta} \frac{\partial \beta}{\partial \beta_3} \frac{\partial \beta_3}{\partial \beta_3} \quad \text{where}$$

$$v = \cos(w); \quad \frac{\partial v}{\partial \beta} = -\sin(w) = -\sin\left[\tan^{-1}\left(\frac{\beta_3}{\beta_2}\right)\right]$$

$$w = \tan^{-1}(x); \quad \frac{\partial w}{\partial x} = \left(1 + x^2\right)^{-1} = \frac{1}{1 + \left(\frac{\beta_3}{\beta_2}\right)^2}; \quad \text{and} \quad x = \frac{\beta_3}{\beta_2}; \quad \frac{\partial x}{\partial \beta_3} = \frac{1}{\beta_2}.$$ 

Combining terms,

$$\frac{\partial f(\beta)}{\partial \beta_3} = -2\beta_2 \left[\frac{-\sin[\tan^{-1}(\beta_3/\beta_2)]}{1 + (\beta_3/\beta_2)^2} \right] \left(\frac{1}{\beta_2}\right)$$

$$= \frac{-2\beta_2 \left[\frac{-\sin[\tan^{-1}(\beta_3/\beta_2)]}{1 + (\beta_3/\beta_2)^2} \right]}{\left(\frac{1}{\beta_2}\right)}.$$
C. Evaluating $\frac{\partial g(\beta)}{\partial \beta_2}$.

Let

$$g(\beta) = -\left[w - \frac{\pi}{2}, \frac{365.25}{2\pi}\right] \quad \text{where} \quad w = \tan^{-1}(x) \quad \text{and} \quad x = \frac{\beta_3}{\beta_2}.$$  

Then

$$\frac{\partial g(\beta)}{\partial \beta_2} = \frac{\partial w}{\partial \beta_2} \frac{\partial x}{\partial \beta_2} = \left(-\frac{365.25}{2\pi}\right) \frac{1}{\left[1 + \left(\frac{\beta_3}{\beta_2}\right)^2\right]^2} \left[1 + \left(\frac{\beta_3}{\beta_2}\right)^2\right] = \frac{365.25\beta_3}{2\pi\beta_2^2}.$$  

D. Evaluating $\frac{\partial g(\beta)}{\partial \beta_3}$.

Let

$$g(\beta) = -\left[w - \frac{\pi}{2}, \frac{365.25}{2\pi}\right] \quad \text{where} \quad w = \tan^{-1}(x) \quad \text{and} \quad x = \frac{\beta_3}{\beta_2}.$$  

Then

$$\frac{\partial g(\beta)}{\partial \beta_3} = \frac{\partial w}{\partial \beta_3} \frac{\partial x}{\partial \beta_3} = \left(-\frac{365.25}{2\pi}\right) \frac{1}{\left[1 + \left(\frac{\beta_3}{\beta_2}\right)^2\right]^2} \left[1 + \left(\frac{\beta_3}{\beta_2}\right)^2\right] = \frac{-365.25}{2\pi\beta_2^2}.$$
Combining terms,

\[
C = \left( \begin{array}{ccc}
2\cos[\tan^{-1}(\beta_3/\beta_2)] & \frac{2\sin[\tan^{-1}(\beta_3/\beta_2)] \beta_3}{1 + (\beta_3/\beta_2)^2} & \frac{2\sin[\tan^{-1}(\beta_3/\beta_2)]}{1 + (\beta_3/\beta_2)^2} \\
\cos[\tan^{-1}(\beta_3/\beta_2)]^2 & 365.25 \beta_3 & -365.25 \\
2\pi \beta_2^2 \left[ 1 + \left( \frac{\beta_3}{\beta_2} \right)^2 \right] & 2\pi \beta_2 \left[ 1 + \left( \frac{\beta_3}{\beta_2} \right)^2 \right] & 2\pi \beta_2^2 \left[ 1 + \left( \frac{\beta_3}{\beta_2} \right)^2 \right]
\end{array} \right)
\]

The variance can then be constructed by a first order Taylor Series approximation to the variance of \(F\) given by \(\text{var}(F) = C V C\), where \(V = \text{var}(\beta)\). We estimate the variance by replacing the terms in the expressions by their estimates. In particular,

\[
\text{var}(F) = \begin{pmatrix}
  c_{11} & c_{12} \\
  c_{21} & c_{22}
\end{pmatrix}
\begin{pmatrix}
  v_{11} & v_{12} \\
  v_{12} & v_{22}
\end{pmatrix}
\begin{pmatrix}
  c_{11} & c_{21} \\
  c_{12} & c_{22}
\end{pmatrix}
= \begin{pmatrix}
  c_{11}^2 v_{11} + 2c_{11}c_{12}v_{12} + c_{12}^2 v_{22} & \cdots \\
  \cdots & \cdots \\
  c_{21}^2 v_{11} + 2c_{21}c_{22}v_{12} + c_{22}^2 v_{22}
\end{pmatrix}
\]
