Reviewer's report

Title: Closed-form estimators of the intensity of seasonal occurrence.

Version: 1 Date: 22 January 2008

Reviewer: David Dickey

Reviewer's report:

(1) General comments

Working with ratios statistically is always tricky. Often it can be helpful to take the logarithm of a ratio thus rendering it a difference of logarithms which can often lead to better behaved estimators. Nam (1995) appears to use this attack on the problem and might be a nice article to reference in addition to the Nam reference already given.

Statistically, a logical modern approach would be to fit this with a generalized linear model, using the Poisson or other distribution for the counts, and letting, for example, the Poisson intensity parameter be a sinusoidal function of the season as in Edwards’s model. The link function is logarithmic which goes nicely with the comments above. Also we know that the maximum likelihood estimator of a function of parameters is that function of the maximum likelihood estimates of the parameters themselves so the amplitude and resulting seasonal maximum to minimum ratio could be thus estimated using a well known approach with well documented theory backing it. Using both a sine and cosine of the proper period allows the computation of the amplitude without worrying about the 0.5 shift in expression (1). Further, the intercept term and amplitude can be estimated separately in such an approach, providing a more general model. In the simulations, a similar approach labeled MLE is used, but as mentioned in the paper, it is not based on the likelihood from which the data were generated. I appreciated the code given and noticed that a general distribution was employed in this calculation. It would also be interesting to know how a Poisson regression would fare on data generated by the assumed Poisson model with sinusoidal intensity function, the above suggested generalized linear model. Perhaps better convergence rates would be found. My main reaction is that I would prefer this approach. Nevertheless a quick check of the literature shows that there are users of Edwards’s method and for them, the modification of this paper may hold an interest.

Edwards’s calculation is based on placing masses of weight square root of Ni around the rim of a wheel of radius Alpha. The relationship of the center of gravity of the weighted wheel to the wheel’s hub (at the origin (0,0)) is a function of Alpha and leads to an estimator. The current paper replaces this by the placement of unit weights along equiangular spaced vectors at distances Ni from the origin on a plane and computing the distance of the center of gravity of the thus weighted plane to the (0,0) origin. The expected distance is, as in
Edwards, a function of \( \bar{i} \) and leads to an estimate equivalent to one proposed by Nam. A two step weighted least squares approach is suggested for the estimation but is not iterated as it might be in practice. A second new estimate uses the expected square of the distance just described, this again being a function of \( \bar{i} \). This estimator can fall outside the parameter space and a remedy is suggested. This results in a somewhat complex but not prohibitive methodology.

It would help a lot if we could be told why, based on distributional considerations or otherwise, one would expect scattering such unit weights at different distances to be a better approach than scattering different weights around a circle as did Edwards. I suppose one can appeal to the simulations and say â##it seems to workâ## but a motivating sentence would help those of us who seek logic behind our methodology.

Minor essential revisions

(1) A couple of typos seem to have made their way into the paper. The first expression in section 2.1 is a constant function of \( t \), cycling through a full circle with every unit increment of \( t \). That canâ##t be right.

(2) Referring to the title of section 2.4, one does not compute confidence intervals for estimates (estimates are known), but rather for the parameters being estimated. No caret should appear atop the \( R \).

(3) My next two paragraphs contain some relatively minor points. In other papers (see for example the St.Leger reference) Edwardsâ## angles are described as \( 2\pi(2i-1)/(2k) \) whereas in the current paper they are \( \theta_i = 2\pi(i/k) \) (line 3 section 2.2), a shift of 15 degrees for monthly data. Formula (1) seems closer to St Leger. Related to this, the mean function in (1) is a continuous function of \( i \) with extremum when the cosine function becomes \( \cos(0) \), that is, at \( i=tp + 0.5 \), not \( i=tp \) The time of peak incidence would thus be \( tp + 0.5 \) if we use (1). Although \( R \) depends only on \( \alpha \), the estimate of \( \alpha_E \) is computed using the angle \( \theta_i \) so there would be an effect, likely of small magnitude, on Edwardsâ## estimator and others. If one is going to go to the effort of this centering by 0.5 then one might as well also express the point of maximization exactly.

The statement is made on page 6 that the ratio \( R \) does not depend on \( t_p \). That seems OK, but in light of the comment about the 0.5 in expression 1, \( R \) does (slightly) depend on whether that 0.5 is there or not, that is, it depends on a correct definition of \( t_p \). If I fit a cosine with 0 phase to a slightly shifted cosine, I will misestimate the amplitude.

(4) I suppose it is obvious in 2.2 that \( N \) is the sum of the \( N_i \), but perhaps this should be stated in this section rather than later (first occurrence of \( N \) is in the variance expression in 2.2). Another detail is the description of \( (n/k) \) in (1). For monthly data \( k = 12 \) and \( n \) is described as the â##total expected number of events across all yearsâ## which if taken literally would mean, over a 10 year period, the sum of 120 expected monthly occurrences. In that case the mean in
expression 1 is the expected total number of occurrences added up across all years for, say, January rather than the expected occurrences in a particular January, and t runs from 1 to 12 rather than 1 to 120 in this example. It would be nice if this were clarified. If we are talking about overall total occurrences then isn’t n the same as N?

Discretionary Revisions

(1) Near the bottom of page 10, a comment is made about the effect of leaving out those cases where the MLE did not converge. It left me wondering how one could figure out the effect of leaving out things that are not there in the first place. To what are the results without those cases being compared?

(2) In the appendix, page 14, I am confused by the expression (line 3 e.g.) “â##distance from the origin of the center of gravity.â## At first I thought it was a typo and should be “â##to the centerâ## but it appears several additional times so I am not sure what the origin of the center of gravity would be.

Additional reference:


What next?: Accept after minor essential revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare no competing interests.