Bayesian Analysis (2012) **7**, Number 4, pp. 801–804

Comment on Article by Kim et al.

Bruno Scarpa *

I really enjoyed reading this very interesting paper. I find human fecundity a very intriguing topic in biometrical analysis, and the authors propose a clever approach to model fecundability on the mainstream of extensions of the seminal Barrett and Marshall model. As well described in this paper, it is quite challenging to develop a good model which takes into account all the characteristics of the available data, which are very complex, so that models need to include latent and unobserved variables. I would like to congratulate the authors on their success in achieving this.

However, my role requires me to raise some points for discussion. I group them in two parts, one related to the model and the other to the real data analysis.

1 The model

I find both simple and clever the idea of using a cumulative distribution function with positive support as the link function, to relate all available explanatory variables, including acts of intercourse, to the probability of conception, as described in models (3) and (5). However, if I want to use this model, I would have difficulty in choosing the 'right' F for my data. Following Czado and Santner (1992), the authors emphasize the importance of choosing a good link function, but I did not find in the paper any suggestion or guideline aiding me, when I am defining the model (i.e., before actual data analysis), to identify the link I should use. Are there any guidelines for it?

For example, recalling the results of Czado and Santner (1992), it seems that the highest effect of the choice of the link function is observed when a skewed link function is used to fit data with actual symmetric links. So I wonder if the authors can indicate when a skewed link function is more appropriate than a symmetric one. Along these lines, it would also be interesting, in the simulation and/or application, to see how the model fits the data when a skew link function is used (considering, for example, the Box-Cox link function used by Czado and Santner 1992, or the distribution function of a skew-normal variable; see, e.g., Azzalini 1985).

A data-driven approach would further generalize the model and 'let data suggest' the right link function, – for example, by choosing a non-parametric model for it. Clearly, this might be a different model, more computationally demanding, and with different theoretical characteristics.

A second aspect in the model definition regards prior distributions. The authors choose widely dispersed priors, since they want to include very little information in addition to data. Is it possible to think about informative priors – for example, for day-specific probability parameters? In particular, considering the dataset analysed in

*Department of Statistical Sciences, University of Padua, Padua, Italy scarpa@stat.unipd.it

 DOI:10.1214/12-BA726A