DISCUSSION OF THE PAPER BY ZENG AND LIN

Yi Li *

Jan, 2007

*Yi Li is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115. This work was supported by U. S. National Cancer Institute grant R01 CA95747.

Zeng and Lin are to be congratulated for a wonderful work on nonparametric maximum likelihood estimation for semiparametric frailty regression models. In this comment, I concentrate on the interpretation of frailties.

In the framework of random effects models, the frailties have been introduced to model the clustering effect and will be useful for prediction as illustrated in Section 5.2. On the other hand, they are meant to model the within-cluster dependence as the variance components of the frailties typically gauge the magnitude of such dependence (Diggle, Liang and Zeger, 1994). This, however, was not elucidated in this paper. This note bridges the frailty parameters with within-cluster dependence measures and highlight a challenge in interpreting these parameters. To convey the idea, consider a (much) simplified version of model (7) in Zeng and Lin for bivariate failure times (T_1, T_2) with no covariates, namely,

$$\Lambda(t|b) = G\left\{\int_0^t e^b d\Lambda(s)\right\} \tag{1}$$

where $b \sim f(\cdot;\gamma)$. Our goal is to link the variance component γ to a 'model free' and standardized dependence measure commonly used for bivariate survival. One such device is the Kendall's coefficient of concordance (Kendall's τ), which can be evaluated by

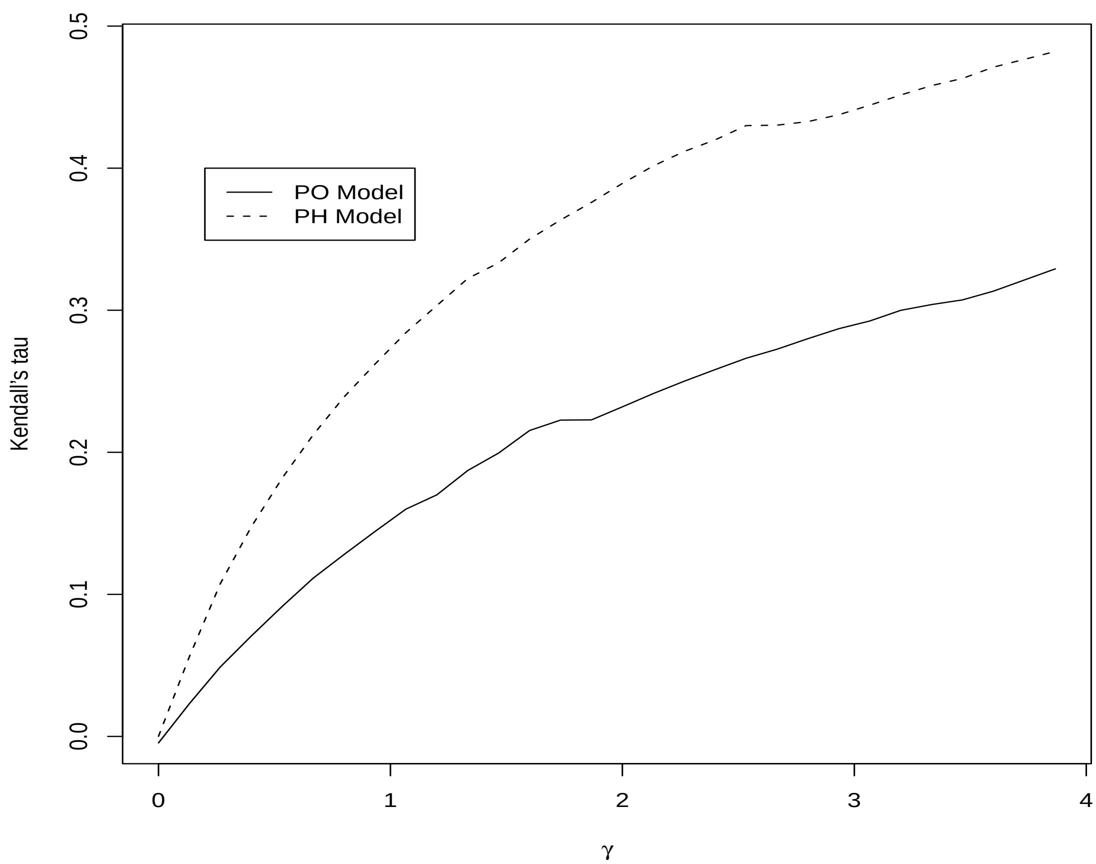
$$\tau = 4\int_0^\infty \int_0^\infty p(t_1, t_2)S(t_1, t_2)dt_1 dt_2 - 1$$

where $p(t_1, t_2)$ and $S(t_1, t_2)$ are the joint bivariate density and survival functions respectively (see, e.g. Hougaard, 2000) . It follows that the joint survival under (1) is $S(t_1, t_2) = \int \exp[-G\{e^b\Lambda(t_1)\} - G\{e^b\Lambda(t_2)\}]f(b;\gamma)db$, and $p(t_1, t_2)$ can be conveniently evaluated by $p(t_1, t_2) = \partial^2 S(t_1, t_2)/\partial t_1 \partial t_2$. Therefore, γ can be viewed to characterize the bivariate dependence through

$$\begin{aligned}\tau \quad = \quad & 4\int_0^\infty \int_0^\infty \left[\int G'(e^b t_1)G'(e^b t_2)e^{2b}\exp\{-G(e^b t_1) - G(e^b t_2)\}f(b;\gamma)db\right] \\ & \times \left[\int \exp\{-G(e^b t_1) - G(e^b t_2)\}f(b;\gamma)db\right] dt_1 dt_2 - 1.\end{aligned}$$

It is worth noting that τ does not depend on the baseline function $\Lambda(t)$ in (1) and its efficient estimate can be obtained by replacing γ with its MLE $\hat{\gamma}$, whose variance estimate will be immediately available via the delta method. In a similar fashion, the relationship of variance component γ with the other global dependence measures, e.g. Spearman correlation, integrated hazard ratio and median concordance, and the local dependence measure, i.e. local cross ratio, can also be established.

However, a serious challenge of interpreting γ as a dependence measure lies in its dependence on the link function G in (1). This can be illustrated by Figure 1, which depicts Kendall's τ against various γ when $b \sim N(0, \gamma)$, under the proportional hazards (PH) model [with $G(x) = x$] and the proportional odds (PO) model [with $G(x) = \log(1 + x)$], respectively. For example, $\gamma = 1.8$ corresponds to Kendall's τ of 0.40 under the PH model, almost twice as much as that of 0.21 under the PO model, begging the cliche question of "how large is large" when viewing the variance component as a measurement for dependence under various transformation models. I would welcome the authors' comments on this issue.



References

Diggle, P., Liang, KY and Zeger, S. (1994) *Analysis of Longitudinal Data.* Clarendon Press, Oxford.

Hougaard, P. (2000) *Analysis of Multivariate Survival Data.* Springer, New York.

Oakes, D. (1989) Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association,* **84,** 487-493.