Statistica Sinica Preprint No: SS-2015-0369R3

Title	Additive mean residual life model with latent variables under right censoring
Manuscript ID	SS-2015.0369.R3
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202015.0369
Complete List of Authors	Xinyuan Song Haijin He Deng Pan and Liuquan Sun
Corresponding Author	Xinyuan Song
E-mail	xysong@sta.cuhk.edu.hk
Notice: Accepted version subject to English editing.	

2 Model

Let $\mathbf{V}_i$ $(i = 1, 2, \cdots, n)$ be a $p \times 1$ vector of observed variables, and $\boldsymbol{\xi}_i$ be a $q \times 1$ vector of latent variables. The latent variables in $\boldsymbol{\xi}_i$ are measured by the observed variables in $\mathbf{V}_i$ via a distribution-free factor analysis model

$$\mathbf{V}_i = \mathbf{B}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where $\mathbf{B}$ is $p \times q$ factor loading matrix, $\boldsymbol{\xi}_i$ has mean zero and covariance matrix $\boldsymbol{\Phi}$, $\boldsymbol{\epsilon}_i$ is a $p \times 1$ vector of random errors independent of $\boldsymbol{\xi}_i$, and $\boldsymbol{\epsilon}_i$ is assumed to have mean zero and diagonal covariance matrix $\boldsymbol{\Psi}_\epsilon$. In this study, we consider model (1) as a confirmatory factor analysis model, where the numbers of observed variables and latent factors, p and q, as well as the structure of the factor loading matrix, $\mathbf{B}$, are pre-determined based on substantive theory, expert knowledge, and/or existing literature (Bollen, 1989; Lee, 2007). In substantive research, if such information is unavailable, one can conduct an exploratory factor analysis to determine p, q, and the structure of $\mathbf{B}$ based on the data (Jöreskog and Sörbom, 1996).

Let $\mathbf{Z}_i$ be a $s \times 1$ vector of observed covariates. To investigate the effects of $\mathbf{Z}_i$ and $\boldsymbol{\xi}_i$ on the failure time T_i, we propose an additive MRL model:

$$m(t|\mathbf{Z}_i, \boldsymbol{\xi}_i) = m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + \boldsymbol{\gamma}^T\boldsymbol{\xi}_i, \tag{2}$$

where $m_0(t)$ is the unspecified baseline MRL function, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $s \times 1$ and $q \times 1$ vectors of unknown regression parameters. Assume that $m_0(t)+\boldsymbol{\beta}^T\mathbf{Z}_i+\boldsymbol{\gamma}^T\boldsymbol{\xi}_i$ is nonnegative. The joint model defined by (1) and (2) inherits from the additive MRL model the following features: (i) it preserves the embedded constraint of the MRL function that $m(t|\mathbf{Z}_i, \boldsymbol{\xi}_i) + t = E[T_i|T_i > t]$ is nondecreasing (Chen and Cheng, 2006); and (ii) the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ directly explain the effects of $\mathbf{Z}_i$ and $\boldsymbol{\xi}_i$ on the MRL function. Unlike random effects that are mainly used to address the dependence of responses in conventional mixed-effect models, $\boldsymbol{\xi}_i$ in (2) includes latent traits (e.g., lipid) that truly exist but cannot be characterized by a single observed variable.

Notably, it is possible to use other techniques such as principle component analysis (PCA) to project groups of correlated variables into independent basis functions. In various cases,