$7^{th}$-$8^{th}$ November 2024

# A General Multinomial Model for Cause of Death

Adam Shao, Keen Liang Chee

Biometric Risk Modelling | SCOR

SCOR The Art & Science of Risk | Life

# Contents

# Introduction

- Aggregate mortality data are widely used to predict mortality trends.
- Estimating mortality trend based on aggregate mortality data can omit many important details, which are usually available in cause of death data.
- Competing risks in a cause of death modelling framework are not necessarily independent: a reduction in one cause of death can possibly have transfer of deaths to other causes (Li & Lu, 2019; Alai et al., 2015, 2018; Kaakai et al., 2018; Ulcinaite, 2023).
- The Multinomial logistic model and Clayton copula are used in the literature to model multiple decrements allowing for the dependence between competing risks (e.g. see Alai et al., 2015; Li & Lu, 2019; Park et al., 2006; Murray et al., 2006; Shahraz et al., 2013).
- Relationship between the Multinomial logistic model and the Clayton copula model is not well exploited.

# Research Summary

- We establish theoretical relationship between the Clayton copula model and the Multinomial logistic model: the latter is a special case of the former.
- Inspired by the crude-net intensity relationship in Clayton copula, we propose a **general Multinomial model** that allows for dependence between cause of deaths, which is simple to implement.
- The proposed general Multinomial model is calibrated to US CoD death. The calibrated value of $\theta$ has important insights into death transfers between different causes.

# Contents

# Multinomial Logistic Model

Following notation in Alai et al. (2015)

$$\begin{aligned} \ln \frac{q_i(x,t)}{p(x,t)} &= X(x,t)\beta_i \\ \sum_k q_k(x,t) + p(x,t) &= 1 \end{aligned}$$

- $q_i(x,t)$: the probability of death due to cause $i$ for age $x$ at time $t$
- $p(x,t)$: the all cause survival probability for age $x$ at time $t$
- $X(x,t)$: explanatory variables of age and time
- $\beta_i$: coefficient vector for death due to cause $i$

$$\begin{aligned} q_i(x,t) &= \frac{\exp\left(X(x,t)\beta_i\right)}{1+\sum_k \exp\left(X(x,t)\beta_k\right)} \\ p(x,t) &= \frac{1}{1+\sum_k \exp\left(X(x,t)\beta_k\right)} \end{aligned}$$

# Multinomial Logistic Model: Mortality Shock

A multiplicative reduction shock on death probability by cause $i$ only, i.e.
$\rho_i < 1, \;\; \rho_j = 1$ for $\forall\, j \neq i$

$$q_i(x,t)' = \frac{\rho_i \exp\Big(X(x,t)\beta_i\Big)}{1 + \rho_i \exp\Big(X(x,t)\beta_i\Big) + \sum_{j\neq i} \rho_j \exp\Big(X(x,t)\beta_j\Big)}$$

$$\frac{q_i(x,t)'}{q_i(x,t)} = \frac{\rho_i}{1 - \Big(1-\rho_i\Big) q_i(x,t)}$$

$$\frac{p(x,t)'}{p(x,t)} = \frac{q_j(x,t)'}{q_j(x,t)} = \frac{1}{1 - \Big(1-\rho_i\Big) q_i(x,t)} > 1$$

- Saved deaths are re-distributed to other causes, including survival as a "dummy cause"
- Re-distribution is proportional to the death probability due to each cause

# Contents

# Clayton Copula

The joint survival probability of death times due to different causes (Li & Lu, 2019)

$$S(t_1, t_2, \cdots, t_m) = \left[S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} + \cdots + S_m(t_m)^{-\theta} - m + 1\right]^{-1/\theta},$$

- $S(t_1, t_2, \cdots, t_m)$ denotes the joint survival probability due to $m$ causes
- $S_j(t)$ denotes the net survival probability up to time $t$
- $i = 1, 2, \cdots, m$ denotes cause of death
- $\theta \in [-1, 0) \cup (0, +\infty)$ is a parameter that captures the dependence
  - $\theta \uparrow$: stronger positive dependence (more death transfers) between death causes
  - $\theta \to 0$, causes are becoming independent (almost no death transfers)

SCOR The Art & Science of Risk | Life

# Crude and Net Intensities

$$\begin{aligned}\mu_j(t) &= -\frac{\partial}{\partial t_j} \log S(t_1, t_2, \cdots, t_m) \mid_{t_1=t_2=\cdots=t} \\ &= \frac{S_j(t)^{-\theta}}{S_1(t)^{-\theta} + S_2(t)^{-\theta} + \cdots + S_m(t)^{-\theta} - m + 1} \lambda_j(t)\end{aligned}$$

- Crude intensity $\mu_j(t)$: mortality intensity due to cause $j$ at age $t$ with presence of other causes of death
- Net intensity $\lambda_j(t)$: mortality intensity due to cause $j$ at age $t$, if $j$ is the only cause of death
- Net intensities are independent; Crude intensities incorporate dependence between different causes
- $\theta \to 0: \ \mu_j(t) = \lambda_j(t)$
- $\theta \uparrow$: difference between crude and net intensities is larger

# Mortality Shock

Apply a reduction shock $\rho_i$ to *annual* (from $t-1$ to $t$) mortality rate by cause $i$, i.e. $\rho_i < 1$, $\rho_j = 1$ for $\forall\ j \neq i$

$$\begin{aligned}
\mu_i'(t)/\mu_i(t) &= \rho_i \left(\pi(\theta, \rho_i, t) S_j(t)^{(1-\rho_i)}\right)^{\theta} \\
\mu_j'(t)/\mu_j(t) &= \left(\pi(\theta, \rho_i, t)\right)^{\theta} \\
p'(t)/p(t) &= \pi(\theta, \rho_i, t)
\end{aligned}$$

- $\pi(\theta, \rho_i, t)$ measures proportional "saved" deaths, i.e. $\frac{S'(t)}{S(t)}$
- $\theta \to 0$: $\left(\pi(\theta, \rho_i, t)\right)^{\theta} \to 1$, i.e no death transfers to other causes
- $\theta = 1$: death transfers **proportional** to mortality rate, including survival probability as a "dummy cause"
- $\theta \uparrow$: more saved deaths transfer to other death causes, and fewer transfer to survival

# Contents

# Insights from Clayton Copula

- How can we extend the Multinomial logistic model to allow for more flexible dependence between causes?
- The crude-net intensity relationship has important insights.
- Assuming net intensity $\lambda_j(x,t)$ is a function of age and time, $f_j(x,t)$, and using Taylor expansion of net survival probability, we have:

$$\begin{aligned}\mu_j(t) &= \frac{S_j(t)^{-\theta}}{S_1(t)^{-\theta}+S_2(t)^{-\theta}+\cdots+S_m(t)^{-\theta}-m+1}\lambda_j(x,t)\\ &\approx \frac{f_j(x,t)}{1+\theta\sum_k f_k(x,t)}\end{aligned}$$

SCOR The Art & Science of Risk | Life

# General Multinomial Model

- Apply the crude-net relationship to death probabilities:

$$
\begin{aligned}
q_j(x,t) &= \frac{f_j(x,t)}{1+\theta\sum_k f_k(x,t)} \\
p(x,t) &= \frac{1+(\theta-1)\sum_k f_k(x,t)}{1+\theta\sum_k f_k(x,t)}
\end{aligned}
$$

- The regression formula is:

$$
\ln\left(\frac{q_j(x,t)}{\theta p(x,t)-\theta+1}\right) = \ln\Big(f(x,t)\Big) = X(x,t)\beta_j
$$

- $X(x,t)$ is vector of explanatory variables of age and time
- When $\theta = 1$, the above model reduces to the Multinomial logistic regression model, i.e.

$$
\ln\left(\frac{q_j(x,t)}{p(x,t)}\right) = X(x,t)\beta_j
$$

SCOR The Art & Science of Risk | Life

# General Multinomial Model: Mortality Shocks

- We apply a reduction shock $\rho_i$, where $0 < \rho_i < 1$, to $f_j(x,t)$. Relative transfer to other causes:

$$\Delta_\theta \quad = \quad \frac{q_j'(x,t)}{q_j(x,t)} \mid \theta = \frac{1+\theta\sum_j f_j(x,t)}{1+\theta\sum_{j\neq i} f_j(x,t) + \theta\rho_i f_i(x,t)}$$

- Changes in mortality rates and survival probabilities:

$$\begin{aligned} \frac{q_i'(x,t)}{q_i(x,t)} &= \rho_i\Delta_\theta \\ \frac{q_j'(x,t)}{q_j(x,t)} &= \Delta_\theta \\ \frac{p'(x,t)}{p(x,t)} &= \frac{\Delta_\theta}{\Delta_{\theta-1}} \end{aligned}$$

- As $\theta$ increases, transfer to other causes increases, and transfer to survival decreases.

SCOR The Art & Science of Risk | Life

# Contents

# Data

- US cause of death data from the Human Cause-of-Death Data series
- Calendar years: 1979-2019
  - Training set: 1979-2014
  - Test set: 2015-2019
- Ages: 20-89
- Group death causes into 5: Accidents, Cancer, Cardiovascular, Respiratory, and Other

# General Multinomial Model: Parameter Estimation

- We decompose $f_j(x,t)$ into two effects: age and time. The regression formula:

$$\ln\left(\frac{q_j(x,t)}{\theta p(x,t)-\theta+1}\right)=\ln\left(f_j(x,t)\right)=\beta_{1,x,j}+\beta_{2,t,j}$$

- Parameters are estimated using MLE. The likelihood function is:

$$L=\prod^{x,t}\binom{E_{x,t}}{D_{x,t,1},\ldots,D_{x,t,m},S_{x,t}}q_{x,t,1}^{D_{x,t,1}}\ldots q_{x,t,m}^{D_{x,t,m}}p_{x,t}^{S_{x,t}}$$

  - $E_{x,t}$ is the number of lives at age $x$ in year $t$
  - $D_{x,t,i}$ is the number of deaths due to cause $i$ for lives at age $x$ in year $t$
  - $S_{x,t}$ is the number of survivals for lives at age $x$ in year $t$

SCOR The Art & Science of Risk | Life

# General Multinomial Model: Estimated value of $\theta$

| Gender | Male | Female |
|---|---|---|
| $\theta$ | 5.5137 | to be inserted |

Table 1: Estimated $\theta$ by gender

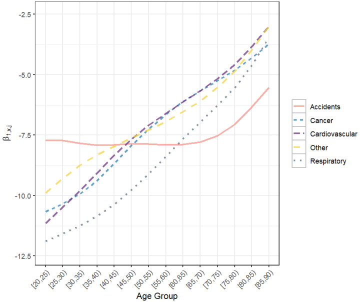
$\theta$ captures the correlation strength between causes of deaths.

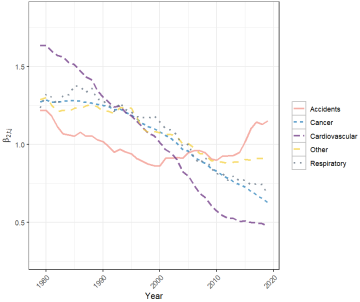
- Higher $\theta$ implies higher correlation between each cause of death.
- Values of $\theta$ in many studies (Li & Lu, 2019) were input based on expert judgment. Most widely used values are 2 and 5.
- Our empirical results show that $\theta = 5$ is a reasonable input.

SCOR The Art & Science of Risk | Life

# General Multinomial Model: Estimated parameters

Estimated parameter values that capture age effect (left) and time effect (right)



# Model comparison

- We compare the following three models:
  - General Multinomial (GM) Model: $\theta$ is estimated from the US historical CoD data
  - Multinomial Logistic (ML) Model: death transfers are proportional to death probabilities due to each cause
  - Independent (I) Model: Independence between causes of deaths, i.e. saved deaths all transfer to survival
- General Multinomial model provides better fit

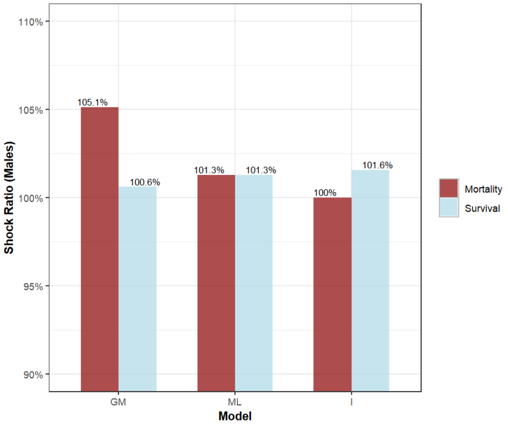
| Models | AIC | BIC |
|---|---|---|
| GM | $8.13253709 \times 10^8$ | $8.13255349 \times 10^8$ |
| ML | $1.009914167 \times 10^9$ | $1.009915807 \times 10^9$ |
| I | $1.130406416 \times 10^9$ | $1.130406416 \times 10^9$ |

Table 2: Model goodness of fit based on AIC and BIC

# Shock scenario: cancer mortality is halved

$$\text{Shock Ratio} = \frac{\text{Non-cancer mortality (Survival) in shock scenario}}{\text{Non-cancer Mortality (Survival) without shock}}$$

- Below chart shows an example based on mortality rates for males aged 75-80 in year 1990



# Gains in life expectancy if cancer mortality is halved

Below table shows gains in (residual) life expectancy for a cohort of 60-year-old males in year 2019:

| Model | Gains in life expectancy (years) |
|---|---|
| GM | 0.711 |
| ML | 0.831 |
| I | 0.857 |

Table 3: Gains in life expectancy if cancer mortality is halved

- Lower gains in life expectancy in the GM model compared to other models, as more "saved deaths" are transferred to other causes of deaths in the GM model.

# Contents

# Conclusions

- The Multinomial logistic model is a special case of the Clayton copula model, where the copula parameter $\theta$ is equal to 1.
  - The Multinomial logistic model is based on annual death probabilities; the Clayton copula model is based on annual death rates.
- Inspired by the crude-net intensity relationship in Clayton copula, we propose a General Multinomial (GM) model, which is calibrated to the US CoD data.
- Calibrated value of $\theta$ in the proposed GM is around 5, which is in line with prior studies' assumption based on expert judgment.

# Extensions

- The GM model can be used to explore how the calibrated value of $\theta$ changes with respect to different age-period set-up, such as Lee-Carter, APC.
- We can use the GM model to investigate the calibrated value of $\theta$ for different cohorts, e.g. old vs. young, general vs. insured populations.
- The GM model can be extended to allow for more flexible dependence structures, e.g. strong dependence between cancer and CVD deaths and weak dependence between accident and all other deaths.

# References


Alai, D. H., Arnold-Gaille, S., Bajekal, M., & Villegas, A. M. (2018). Mind the gap: A study of cause-specific mortality by socioeconomic circumstances. *North American Actuarial Journal*, 22(2), 161–181.

Alai, D. H., Arnold-Gaille, S., & Sherris, M. (2015). Modelling cause-of-death mortality and the impact of cause-elimination. *Annals of Actuarial Science*, 9(1), 167–186.

Kaakai, S., Labit-Hardy, H., Arnold-Gaille, S., & Karoui, N. E. (2018). How can a cause-of-death reduction be compensated for in the presence of heterogeneity? A population dynamics approach based on English data by deprivation. working paper or preprint.

Li, H. & Lu, Y. (2019). Modeling cause-of-death mortality using hierarchical archimedean copula. *Scandinavian Actuarial Journal*, 2019(3), 247–272.

Murray, C. J. L., Kulkarni, S. C., & Ezzati, M. (2006). Understanding the coronary heart disease versus total cardiovascular mortality paradox: a method to enhance the comparability of cardiovascular death statistics in the united states. *Circulation*, 113(17), 2071–2081.

Park, Y., Choi, J. W., & Lee, D.-H. (2006). A parametric approach for measuring the effect of the 10th revision of the international classification of diseases. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55(5), 677–697.

Shahraz, S., Bhalla, K., Lozano, R., Bartels, D., & Murray, C. J. L. (2013). Improving the quality of road injury statistics by using regression models to redistribute ill-defined events. *Injury Prevention*, 19(1), 1–5.

Ulcinaite, A. (2023). Prospective mortality by cause of death modelling. Master's thesis, SCOR, Link: https://www.institutdesactuaires.com/se-documenter/memoires/memoires-d-actuariat-4651?id=277846f82d635f430eba48db8a7544b8.