ARC CENTRE OF EXCELLENCE IN
POPULATION AGEING RESEARCH

ARC Centre of Excellence in Population Ageing Research

Working Paper 2015/02

A Multivariate Tweedie Lifetime Model: Censoring and Truncation

Daniel H. Alai[1], Zinoviy Landsman[2] and Michael Sherris[3]

[1]School of Mathematics, Statistics and Actuarial Science University of Kent, CEPAR, email: D.H.Alai@kent.ac.uk
[2]Department of Statistics, University of Haifa, email: landsman@stat.haifa.ac.il
[3]CEPAR, School of Risk and Actuarial Studies, UNSW Business School, UNSW Australia, email: m.sherris@unsw.edu.au

A Multivariate Tweedie Lifetime Model: Censoring and Truncation

Daniel H. Alai[1] Zinoviy Landsman[2] Michael Sherris[3]

School of Mathematics, Statistics and Actuarial Science
University of Kent, Canterbury, Kent CT2 7NF, UK

Departmant of Statistics, University of Haifa
Mount Carmel, Haifa 31905, Israel

CEPAR, Risk and Actuarial Studies, UNSW Business School
UNSW, Sydney NSW 2052, Australia

Abstract

We generalize model calibration for a multivariate Tweedie distribution to allow for censored observations; estimation is based on the method of moments. The multivariate Tweedie distribution we consider incorporates dependence in a pool of lives via a common stochastic component. Pools may be interpreted in various ways, from nation-wide cohorts to employer-based pension annuity portfolios. In general, the common stochastic component is representative of systematic longevity risk, which is not accounted for in standard life tables and actuarial models used for annuity pricing and reserving.

Keywords: systematic longevity risk, dependence, multivariate Tweedie, lifetime distribution, censoring, truncation

[1]d.h.alai@kent.ac.uk
[2]landsman@stat.haifa.ac.il
[3]m.sherris@unsw.edu.au

1 Introduction

A multivariate Tweedie distribution was used to model pools of lifetimes in Alai *et al.* (2015). Dependence was induced via a common stochastic component and model calibration results were obtained for truncated observations. We generalize these results to allow for both truncation and censoring. Without censoring, only pools with no surviving members could be analysed. Insight gained from the analysis could be applied to *active* pools, but not without introducing some form of basis-risk. The allowance for censored observations eliminates this limitation and makes the model more applicable to the study of pools of lives, for example, to price and reserve annuity portfolios. Lifetimes have previously been modelled using the gamma distribution, a subclass of the Tweedie distribution; see e.g. Klein and Moeschberger (1997) and Alai *et al.* (2013). Dependence between lifetimes has previously been investigated in the context of joint-life insurance products in Dhaene *et al.* (2000) and Denuit *et al.* (2001).

A multivariate distribution is formulated using the so-called multivariate reduction method; see Chereiyan (1941) and Ramabhadran (1951) and applications by Mathai and Moschopoulus (1991) and Chatelain *et al.* (2006). This method results in a natural dependency structure for modelling lifetimes of individuals within a pool. It makes use of the fact that the sum of Tweedie random variables with the same canonical parameter follows a Tweedie distribution. The multivariate reduction method applied to the exponential dispersion family (EDF) only yields a multivariate distribution for the Tweedie subclass; see Furman and Landsman (2010).

The Tweedie class, introduced in Tweedie (1984), is widely used in actuarial science; see e.g. Aalen (1992), Jørgensen and De Souza (1994), Smyth and Jørgensen (2002), Wüthrich (2003), Kaas (2005), and Furman and Landsman (2010). Random variable X is said to belong to the EDF of distributions in the additive form if its probability measure $P_{\theta,\lambda}$ is absolutely continuous with respect to some measure Q_λ and can be represented as follows for some function $\kappa(\theta)$ called the cumulant:

$$dP_{\theta,\lambda}(x) = e^{[\theta x - \lambda\kappa(\theta)]} dQ_\lambda(x);$$

see Jørgensen (1997), Section 3.1; for a recent reference see Landsman and Valdez (2005). The parameters θ and λ are called the canonical and dispersion parameters, respectively. The canonical parameter belongs to the set $\Theta = \{\theta \in \mathbb{R} | \kappa(\theta) < \infty\}$ and the dispersion parameter to the set of positive real numbers $\Lambda = (0, \infty) = \mathbb{R}^+$. Let $X \sim ED(\theta, \lambda)$ denote a random variable belonging to the additive EDF.

The cumulant $\kappa(\theta)$ is a twice differentiable function, one-to-one map, and there exists an inverse function

$$\theta = \theta(\mu) = (\kappa')^{-1}(\mu),$$

where

$$\mu = \lambda\kappa'(\theta);$$

see e.g. Landsman and Valdez (2005). $V(\mu) = \kappa''(\theta(\mu))$ is called the unit variance function and determines subclasses of the EDF. In particular, the Tweedie subclass is defined by a power variance structure

$$V(\mu) = \mu^p,$$

with power parameter p. Even further classification is obtained by specific values of p. For example, $p = 0, 1, 2, 3$ yield the normal, overdispersed Poisson, gamma, and inverse Gaussian distributions, respectively. The cumulant $\kappa_p(\theta) = \kappa(\theta)$ for the Tweedie subclass has the general form

$$\kappa(\theta) = \begin{cases} e^{\theta}, & p = 1, \\ -\log(-\theta), & p = 2, \\ \frac{\alpha-1}{\alpha}\left(\frac{\theta}{\alpha-1}\right)^{\alpha}, & p \neq 1, 2, \end{cases}$$

where $\alpha = (p-2)/(p-1)$. Furthermore, the canonical parameter belongs to the set Θ_p, given by

$$\Theta_p = \begin{cases} [0, \infty), & \text{for } p < 0, \\ \mathbb{R}, & \text{for } p = 0, 1, \\ (-\infty, 0), & \text{for } 1 < p \leq 2, \\ (-\infty, 0], & \text{for } 2 < p < \infty; \end{cases}$$

see e.g. McCullagh and Nelder (1989). Let $X \sim Tw_p(\theta, \lambda)$ denote a random variable belonging to the additive Tweedie family.

Remark 1. *Throughout the paper, we derive results for the additive form of the EDF. However, a simple transformation yields the reproductive form. That is, for X, a member of the additive EDF, $Y = X/\lambda$ is a member of the reproductive EDF with probability measure $P^*_{\theta,\lambda}$, absolutely continuous with respect to some measure Q^*_{λ},*

$$dP^*_{\theta,\lambda}(y) = e^{\lambda[\theta y - \kappa(\theta)]} dQ^*_{\lambda}(y).$$

Organization of the paper: Section 2 reviews the multivariate Tweedie lifetime model. In Section 3 we derive the moments of truncated and censored multivariate Tweedie random variables. We make use of the method of moments to formulate parameter estimation algorithms in 4. In Section 5 we apply the estimation procedure to the two most widely used distributions in the Tweedie family, the normal and gamma distributions. Section 6 concludes the paper.

2 Multivariate Tweedie Survival Model

Following Alai *et al.* (2015), we assume M pools of lives. The pools can, in general, be comprised of any collection of lives where the independence assumption is likely to be violated. Applications range from nation-wide cohorts to employer-based pension annuity portfolios. Let $T_{i,j}$ be the survival time of individual $i \in \{1, \ldots, N_j\}$ in pool $j \in \{1, \ldots, M\}$. For simplicity, we assume the number of lives in each pool to be identical; $N_j = N$ for all j. Individual lifetimes are given as follows:

$$T_{i,j} = Y_{0,j} + Y_{i,j},$$

where

- $Y_{0,j}$ follows an additive Tweedie distribution with power parameter p, canonical and dispersion parameters θ_j and λ_0, $Tw_p(\theta_j, \lambda_0)$, $j \in \{1, \ldots, M\}$,
- $Y_{i,j}$ follows an additive Tweedie distribution with power parameter p, canonical and dispersion parameters θ_j and λ_j, $Tw_p(\theta_j, \lambda_j)$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$,
- The $Y_{i,j}$ are independent, $i \in \{0, \ldots, N\}$ and $j \in \{1, \ldots, M\}$.

The pool-specific common component $Y_{0,j}$ impacts the survival of all the individuals of that pool (i.e. $Y_{0,j}$ captures the impact of systematic mortality dependence between the lives in pool j). The pool-specific parameters λ_j and θ_j jointly represent the risk profile of the pool. A consequence of the model is that survival times $T_{i,j}$ are Tweedie distributed with power parameter p, canonical parameter θ_j, and dispersion parameter $\tilde{\lambda}_j = \lambda_0 + \lambda_j$.

3 Moments for the Exponential Dispersion Family

In this section we consider the moments of variables for distributions belonging to the EDF that are truncated and censored.

Notation

We briefly provide some necessary notation. Denote with $\alpha_k(X)$ and $\mu_k(X)$ the k^{th}, $k \in \mathbb{Z}^+$, raw and central (theoretical) moments of random variable X, respectively.

$$\begin{aligned} \alpha_k(X) &= E[X^k], \\ \mu_k(X) &= E[(X - \alpha_1(X))^k]. \end{aligned}$$

The raw sample moments for random sample $\mathbf{X} = (X_1, \ldots, X_n)'$ are given by

$$a_k(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k \in \mathbb{Z}^+.$$

The raw sample moments of an identically distributed sample are unbiased estimators of the corresponding raw moments of X_1.

$$E[a_k(\mathbf{X})] = \alpha_k(X_1).$$

Finally, *adjusted* second and third central sample moments are given by

$$\begin{aligned} \widetilde{m}_2(\mathbf{X}) &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - a_1(\mathbf{X}))^2, \\ \widetilde{m}_3(\mathbf{X}) &= \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(X_i - a_1(\mathbf{X}))^3. \end{aligned}$$

The adjusted central sample moments of an independent and identically distributed sample are unbiased and consistent estimators of the corresponding central moments of X_1.

$$E[\widetilde{m}_2(\mathbf{X})] = \mu_2(X_1) \qquad \text{and} \qquad E[\widetilde{m}_3(\mathbf{X})] = \mu_3(X_1).$$

3.1 Moments for Truncated Variables

Consider truncated variables ${}_{\tau_j}T_{i,j} = T_{i,j}|T_{i,j} > \tau_j$ with known truncation point τ_j. We make the simplifying assumption that all pools are subject to the same truncation point; $\tau_j = \tau$ for all j.

We recall a useful result regarding the theoretical moments of truncated variables; please refer to Theorem 1 in Alai *et al.* (2015).

Theorem 1. *Consider $Y \sim ED(\theta, \lambda)$ with probability density and survival function denoted $f(y, \theta, \lambda)$ and $\overline{F}(y, \theta, \lambda)$, respectively. Define the associated truncated random variable ${}_{\tau}Y = Y|Y > \tau$. The first raw moment and the second and third central moments of ${}_{\tau}Y$ are given by*

$$\begin{aligned} \alpha_1({}_{\tau}Y) &= \alpha_1(Y) + g_1(\tau), \\ \mu_2({}_{\tau}Y) &= \mu_2(Y) + g_2(\tau) - g_1(\tau)^2, \\ \mu_3({}_{\tau}Y) &= \mu_3(Y) + g_3(\tau) - 3g_2(\tau)g_1(\tau) + 2g_1(\tau)^3, \end{aligned}$$

where

$$g_k(\tau) = g_k(\tau; \theta, \lambda) = \frac{1}{\overline{F}(\tau, \theta, \lambda)} \frac{\partial^k \overline{F}(\tau, \theta, \lambda)}{\partial \theta^k}, \quad k = 1, 2, 3.$$

In Theorem 1, the g_k take the interpretation of *additive truncation adjustments* requisite for transforming non truncated into truncated moments.

3.2 Moments for Truncated and Censored Variables

We presently consider truncated and censored variables ${}^{\upsilon_j}_{\tau_j}T_{i,j} = \min(T_{i,j}, \upsilon_j)|T_{i,j} > \tau_j$ with known truncation and censoring points τ_j and υ_j. For ease of presentation and simplicity, we assume all pools are subject to the same truncation and censoring points; $\tau_j = \tau$ and $\upsilon_j = \upsilon$ for all j.

The following theorem gives the theoretical moments of truncated and censored variables as a function of the theoretical moments of truncated (only) variables.

Theorem 2. *Consider Y with distribution and survival function denoted $F(y)$ and $\overline{F}(y)$, respectively. Define the associated truncated random variable*

$${}_{\tau}Y = Y|Y > \tau,$$

and the truncated and censored random variable

$${}^{\upsilon}_{\tau}Y = \min(Y, \upsilon)|Y > \tau,$$

where $\upsilon \geq \tau$. The first raw moment and the second and third central moments of ${}^{\upsilon}_{\tau}Y$ are given by

$$\begin{aligned} \alpha_1({}^{\upsilon}_{\tau}Y) &= \alpha_1({}_{\tau}Y) + h_1(\tau, \upsilon), \\ \mu_2({}^{\upsilon}_{\tau}Y) &= \mu_2({}_{\tau}Y) + h_2(\tau, \upsilon) - h_1(\tau, \upsilon)^2, \\ \mu_3({}^{\upsilon}_{\tau}Y) &= \mu_3({}_{\tau}Y) + h_3(\tau, \upsilon) - 3h_2(\tau, \upsilon)h_1(\tau, \upsilon) + 2h_1(\tau, \upsilon)^3, \end{aligned}$$

where, for $k = 1, 2, 3$,

$$h_k(\tau, \upsilon) = \Big\{ H_k(\Upsilon, {}_{\tau}Y) - H_k({}_{\upsilon}Y, {}_{\tau}Y) \Big\} \frac{\overline{F}(\upsilon)}{\overline{F}(\tau)},$$

where Υ is degenerate with value υ, and for any two random variables X, Z, and $j \in \mathbb{Z}^{+}$, $H_j(X, Z)$ is given by

$$H_j(X, Z) = \Big(\alpha_1(X) - \alpha_1(Z)\Big)^j + \sum_{i=1}^{j} \binom{j}{i} \Big(\alpha_1(X) - \alpha_1(Z)\Big)^{j-i} \Big(\mu_i(X) - \mu_i(Z)\Big).$$

The latter may be considered as a measure of distance between random variables X and Z, based on the central and raw moments. We note that $\mu_1(\cdot) = 0$.

Proof. Consider the first raw moment.

$$\begin{aligned} \alpha_1({}_{\tau}^{\upsilon}Y) &= \frac{\int_{\tau}^{\upsilon} y dF(y) + \upsilon \overline{F}(\upsilon)}{\overline{F}(\tau)} = \frac{\int_{\tau}^{\infty} y dF(y) + \upsilon \overline{F}(\upsilon) - \int_{\upsilon}^{\infty} y dF(y)}{\overline{F}(\tau)} \\ &= \frac{\int_{\tau}^{\infty} y dF(y)}{\overline{F}(\tau)} + \frac{\upsilon \overline{F}(\upsilon)}{\overline{F}(\tau)} - \frac{\int_{\upsilon}^{\infty} y dF(y)}{\overline{F}(\upsilon)} \frac{\overline{F}(\upsilon)}{\overline{F}(\tau)} \\ &= \alpha_1({}_{\tau}Y) + \Big\{ \upsilon - \alpha_1({}_{\upsilon}Y) \Big\} \frac{\overline{F}(\upsilon)}{\overline{F}(\tau)}. \end{aligned}$$

Notice that

$$h_1(\tau, \upsilon) = \Big\{ \upsilon - \alpha_1({}_{\upsilon}Y) \Big\} \frac{\overline{F}(\upsilon)}{\overline{F}(\tau)},$$

which agrees with $h_k(\tau, \upsilon)$ given above with $k = 1$. The second central moment is given by:

$$\begin{aligned} \mu_2({}_{\tau}^{\upsilon}Y) &= \frac{\int_{\tau}^{\upsilon} (y - \alpha_1({}_{\tau}^{\upsilon}Y))^2 dF(y) + (\upsilon - \alpha_1({}_{\tau}^{\upsilon}Y))^2 \overline{F}(\upsilon)}{\overline{F}(\tau)} \\ &= \frac{\int_{\tau}^{\infty} (y - \alpha_1({}_{\tau}^{\upsilon}Y))^2 dF(y) + (\upsilon - \alpha_1({}_{\tau}^{\upsilon}Y))^2 \overline{F}(\upsilon) - \int_{\upsilon}^{\infty} (y - \alpha_1({}_{\tau}^{\upsilon}Y))^2 dF(y)}{\overline{F}(\tau)}. \end{aligned}$$

We have

$$\begin{aligned} \frac{\int_{\tau}^{\infty} (y - \alpha_1({}_{\tau}^{\upsilon}Y))^2 dF(y)}{\overline{F}(\tau)} &= (\alpha_1({}_{\tau}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))^2 + \mu_2({}_{\tau}Y), \\ \frac{\int_{\upsilon}^{\infty} (y - \alpha_1({}_{\tau}^{\upsilon}Y))^2 dF(y)}{\overline{F}(\upsilon)} &= (\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))^2 + \mu_2({}_{\upsilon}Y). \end{aligned}$$

Notice that $(\alpha_1({}_{\tau}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))^2 = h_1(\tau,\upsilon)^2$, this implies

$$\begin{aligned} \mu_2({}_{\tau}^{\upsilon}Y) &= \mu_2({}_{\tau}Y) + h_1(\tau,\upsilon)^2 \\ &+ \Big\{(\upsilon - \alpha_1({}_{\tau}^{\upsilon}Y))^2 - \Big((\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))^2 + \mu_2({}_{\upsilon}Y)\Big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)}. \end{aligned}$$

We have

$$\begin{aligned} &(\upsilon - \alpha_1({}_{\tau}^{\upsilon}Y))^2 - (\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))^2 \\ &\qquad = (\upsilon - \alpha_1({}_{\tau}Y))^2 - (\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))^2 - 2h_1(\tau,\upsilon)(\upsilon - \alpha_1({}_{\upsilon}Y)), \end{aligned}$$

which finally implies that

$$\begin{aligned} \mu_2({}_{\tau}^{\upsilon}Y) &= \mu_2({}_{\tau}Y) + h_1(\tau,\upsilon)^2 - 2h_1(\tau,\upsilon)\Big\{\upsilon - \alpha_1({}_{\upsilon}Y)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)} \\ &+ \Big\{(\upsilon - \alpha_1({}_{\tau}Y))^2 - \big((\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))^2 + \mu_2({}_{\upsilon}Y)\big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)} \\ &= \mu_2({}_{\tau}Y) - h_1(\tau,\upsilon)^2 \\ &+ \Big\{(\upsilon - \alpha_1({}_{\tau}Y))^2 - \big((\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))^2 + \mu_2({}_{\upsilon}Y)\big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)}. \end{aligned}$$

Notice that

$$h_2(\tau,\upsilon) = \Big\{(\upsilon - \alpha_1({}_{\tau}Y))^2 - \big((\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))^2 + \mu_2({}_{\upsilon}Y)\big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)},$$

which agrees with $h_k(\tau,\upsilon)$ given above with $k = 2$. The third central moment is given by:

$$\begin{aligned} \mu_3({}_{\tau}^{\upsilon}Y) &= \frac{\int_{\tau}^{\upsilon}(y - \alpha_1({}_{\tau}^{\upsilon}Y))^3 dF(y) + (\upsilon - \alpha_1({}_{\tau}^{\upsilon}Y))^3\overline{F}(\upsilon)}{\overline{F}(\tau)} \\ &= \frac{\int_{\tau}^{\infty}(y - \alpha_1({}_{\tau}^{\upsilon}Y))^3 dF(y) + (\upsilon - \alpha_1({}_{\tau}^{\upsilon}Y))^3\overline{F}(\upsilon) - \int_{\upsilon}^{\infty}(y - \alpha_1({}_{\tau}^{\upsilon}Y))^3 dF(y)}{\overline{F}(\tau)}. \end{aligned}$$

We have

$$\begin{aligned} \frac{\int_{\tau}^{\infty}(y - \alpha_1({}_{\tau}^{\upsilon}Y))^3 dF(y)}{\overline{F}(\tau)} &= (\alpha_1({}_{\tau}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))^3 \\ &+ 3(\alpha_1({}_{\tau}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))\mu_2({}_{\tau}Y) + \mu_3({}_{\tau}Y), \\ \frac{\int_{\upsilon}^{\infty}(y - \alpha_1({}_{\tau}^{\upsilon}Y))^3 dF(y)}{\overline{F}(\upsilon)} &= (\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}^{\upsilon}Y))^3 \\ &+ 3\mu_2({}_{\upsilon}Y)(\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}^{\upsilon}Y)) + \mu_3({}_{\upsilon}Y). \end{aligned}$$

Substituting $h_1(\tau, \upsilon)$ where possible, we obtain

$$\begin{aligned}
\mu_3({}_\tau^\upsilon Y) &= \mu_3({}_\tau Y) - 3\mu_2({}_\tau Y) h_1(\tau,\upsilon) - h_1(\tau,\upsilon)^3 \\
&+ \Big\{(\upsilon - \alpha_1({}_\tau^\upsilon Y))^3 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau^\upsilon Y))^3 \\
&- 3(\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau^\upsilon Y))\mu_2({}_\upsilon Y) - \mu_3({}_\upsilon Y)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)}.
\end{aligned}$$

We have

$$\begin{aligned}
&(\upsilon - \alpha_1({}_\tau^\upsilon Y))^3 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau^\upsilon Y))^3 \\
&\quad = (\upsilon - \alpha_1({}_\tau Y))^3 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y))^3 + 3h_1(\tau,\upsilon)^2(\upsilon - \alpha_1({}_\upsilon Y)) \\
&\quad - 3h_1(\tau,\upsilon)\big((\upsilon - \alpha_1({}_\tau Y))^2 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y))^2\big),
\end{aligned}$$

which implies

$$\begin{aligned}
\mu_3({}_\tau^\upsilon Y) &= \mu_3({}_\tau Y) - 3\mu_2({}_\tau Y) h_1(\tau,\upsilon) - h_1(\tau,\upsilon)^3 + 3h_1(\tau,\upsilon)^3 \\
&+ \Big\{(\upsilon - \alpha_1({}_\tau Y))^3 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y))^3 - \mu_3({}_\upsilon Y) \\
&- 3h_1(\tau,\upsilon)\big((\upsilon - \alpha_1({}_\tau Y))^2 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y))^2\big) \\
&- 3(\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau^\upsilon Y))\mu_2({}_\upsilon Y)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)}.
\end{aligned}$$

Rewriting the last two terms in the equation directly above gives:

$$\begin{aligned}
&\Big\{-3(\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau^\upsilon Y))\mu_2({}_\upsilon Y) \\
&\quad -3h_1(\tau,\upsilon)\big((\upsilon - \alpha_1({}_\tau Y))^2 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y))^2\big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)} \\
&\quad = \Big\{-3\Big(\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y) + \alpha_1({}_\tau Y) - \alpha_1({}_\tau^\upsilon Y)\Big)\mu_2({}_\upsilon Y) \\
&\qquad\qquad - 3h_1(\tau,\upsilon)\big((\upsilon - \alpha_1({}_\tau Y))^2 - (\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y))^2\big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)} \\
&\quad = \Big\{-3(\alpha_1({}_\upsilon Y) - \alpha_1({}_\tau Y))\mu_2({}_\upsilon Y)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)} - 3h_2(\tau,\upsilon)h_1(\tau,\upsilon).
\end{aligned}$$

This implies that

$$\begin{aligned}
\mu_3({}^{\upsilon}_{\tau}Y) &= \mu_3({}_{\tau}Y) - 3h_2(\tau,\upsilon)h_1(\tau,\upsilon) + 2h_1(\tau,\upsilon)^3 \\
&+ \Big\{(\upsilon - \alpha_1({}_{\tau}Y))^3 - 3(\upsilon - \alpha_1({}_{\upsilon}Y))\mu_2({}_{\tau}Y) - 3(\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))\mu_2({}_{\upsilon}Y) \\
&- \Big((\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))^3 + \mu_3({}_{\upsilon}Y)\Big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)} \\
&= \mu_3({}_{\tau}Y) - 3h_2(\tau,\upsilon)h_1(\tau,\upsilon) + 2h_1(\tau,\upsilon)^3 \\
&+ \Big\{(\upsilon - \alpha_1({}_{\tau}Y))^3 - 3(\upsilon - \alpha_1({}_{\tau}Y))\mu_2({}_{\tau}Y) - \Big((\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))^3 \\
&+ 3(\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))(\mu_2({}_{\upsilon}Y) - \mu_2({}_{\tau}Y)) + \mu_3({}_{\upsilon}Y)\Big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)}.
\end{aligned}$$

Notice that

$$\begin{aligned}
h_3(\tau,\upsilon) &= \Big\{(\upsilon - \alpha_1({}_{\tau}Y))^3 - 3(\upsilon - \alpha_1({}_{\tau}Y))\mu_2({}_{\tau}Y) - \Big((\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))^3 \\
&+ 3(\alpha_1({}_{\upsilon}Y) - \alpha_1({}_{\tau}Y))(\mu_2({}_{\upsilon}Y) - \mu_2({}_{\tau}Y)) + \mu_3({}_{\upsilon}Y)\Big)\Big\}\frac{\overline{F}(\upsilon)}{\overline{F}(\tau)},
\end{aligned}$$

which agrees with $h_k(\tau,\upsilon)$ given above with $k = 3$. ■

In Theorem 2, the h_k take the interpretation of *additive censoring adjustments* requisite for transforming uncensored to censored moments. Theorem 2 holds for any general distribution. We formulate this theorem for the exponential dispersion family below. However, we first provide a useful lemma.

Lemma 1. *For $j = 1, 2, 3$, and random variables A, B, C, D,*

$$H_j(A,C) - H_j(B,C) = \sum_{i=0}^{j} \binom{j}{i} \Big(H_i(A,D) - H_i(B,D)\Big) H_{j-i}(D,C),$$

where $H_0 \equiv 1$.

Please see the appendix for a proof.

Theorem 3. *Consider $Y \sim ED(\theta, \lambda)$ with probability density and survival function denoted $f(y,\theta,\lambda)$ and $\overline{F}(y,\theta,\lambda)$, respectively. Define the associated truncated random variable*

$${}_{\tau}Y = Y | Y > \tau,$$

and the truncated and censored random variable

$${}^{\upsilon}_{\tau}Y = \min(Y, \upsilon) | Y > \tau,$$

where $\upsilon \geq \tau$. The first raw moment and the second and third central moments of ${}^{\upsilon}_{\tau}Y$ are given by

$$\begin{aligned}
\alpha_1({}^{\upsilon}_{\tau}Y) &= \alpha_1({}_{\tau}Y) + h_1(\tau, \upsilon), \\
\mu_2({}^{\upsilon}_{\tau}Y) &= \mu_2({}_{\tau}Y) + h_2(\tau, \upsilon) - h_1(\tau, \upsilon)^2, \\
\mu_3({}^{\upsilon}_{\tau}Y) &= \mu_3({}_{\tau}Y) + h_3(\tau, \upsilon) - 3h_2(\tau, \upsilon)h_1(\tau, \upsilon) + 2h_1(\tau, \upsilon)^3,
\end{aligned}$$

where, for $k = 1, 2, 3$,

$$h_k(\tau, \upsilon) = \left\{ \sum_{i=0}^{k} \binom{k}{i} \left(\frac{f^{(i)}(\upsilon)}{f(\upsilon)} - \frac{\overline{F}^{(i)}(\upsilon)}{\overline{F}(\upsilon)} \right) \left(\overline{L}_{k-i}(\tau) \right) \right\} \frac{\overline{F}(\upsilon)}{\overline{F}(\tau)},$$

where all differentiation is with respect to θ, and for $j \in \mathbb{Z}^*$,

$$\overline{L}_j(x) = \sum_{i=0,\, i \neq 1}^{j} \binom{j}{i} \left(L^{(i)}(x) \right) \left(L^{(1)}(x) \right)^{j-i},$$

with

$$\begin{aligned}
L^{(0)}(x) &= 1, \\
L^{(k)}(x) &= \frac{-\partial^k \ln \overline{F}(x)}{\partial \theta^k}, \qquad k \in \mathbb{Z}^+.
\end{aligned}$$

Proof. In order to prove this theorem, we must show that for $j = 1, 2, 3$,

$$\begin{aligned}
\frac{f^{(j)}(x)}{f(x)} &= H_j(\Xi, Y), \\
\frac{\overline{F}^{(j)}(x)}{\overline{F}(x)} &= H_j({}_xY, Y), \\
\overline{L}_j(x) &= H_j(Y, {}_xY),
\end{aligned}$$

where Ξ is degenerate with value x. The rest follows directly from Theorem 2 and Lemma 1. To show the above is a simple matter of taking derivatives with respect to the density function, the survival function, and the function $\overline{L}$, which is a composition of derivatives of the logarithm of the survival function. The density is given by

$$f(x; \theta, \lambda) = e^{[\theta x - \lambda \kappa(\theta)]} q_\lambda(x).$$

We take the first derivative with respect to θ and normalize:

$$\frac{f^{(1)}(x)}{f(x)} = \frac{1}{f(x)} \frac{\partial e^{[\theta x - \lambda \kappa(\theta)]} q_\lambda(x)}{\partial \theta} = \frac{(x - \lambda \kappa'(\theta)) e^{[\theta x - \lambda \kappa(\theta)]} q_\lambda(x)}{f(x)} = x - \alpha_1(Y).$$

Recall that $\lambda\kappa'(\theta) = \alpha_1(Y)$. Similarly, we obtain

$$\begin{aligned}
\frac{f^{(2)}(x)}{f(x)} &= \Big(x - \alpha_1(Y)\Big)^2 - \mu_2(Y), \\
\frac{f^{(3)}(x)}{f(x)} &= \Big(x - \alpha_1(Y)\Big)^3 - 3\Big(x - \alpha_1(Y)\Big)\mu_2(Y) - \mu_3(Y).
\end{aligned}$$

Recall that $\lambda\kappa''(\theta) = \mu_2(Y)$ and $\lambda\kappa'''(\theta) = \mu_3(Y)$. Furthermore, notice that $\alpha_1(\Xi) = x$, $\mu_2(\Xi) = 0$, and $\mu_3(\Xi) = 0$. Therefore, the above agrees with $H_j(\Xi, Y)$ for $j = 1, 2, 3$. Differentiation with respect to the survival function follows similarly; we may refer the reader to the proof of Theorem 1 in Alai *et al.* (2015). We present the results,

$$\begin{aligned}
\frac{\overline{F}^{(1)}(x)}{\overline{F}(x)} &= \alpha_1({}_xY) - \alpha_1(Y), \\
\frac{\overline{F}^{(2)}(x)}{\overline{F}(x)} &= \Big(\alpha_1({}_xY) - \alpha_1(Y)\Big)^2 + \Big(\mu_2({}_xY) - \mu_2(Y)\Big), \\
\frac{\overline{F}^{(3)}(x)}{\overline{F}(x)} &= \Big(\alpha_1({}_xY) - \alpha_1(Y)\Big)^3 + 3\Big(\alpha_1({}_xY) - \alpha_1(Y)\Big)\Big(\mu_2({}_xY) - \mu_2(Y)\Big) \\
&+ \Big(\mu_3({}_xY) - \mu_3(Y)\Big).
\end{aligned}$$

Therefore, the above agrees with $H_j({}_xY, Y)$ for $j = 1, 2, 3$. Finally, we present the differentiation of $\ln \overline{F}(x)$ with respect to θ. We have

$$\begin{aligned}
L^{(1)}(x) &= \frac{-\partial \ln \overline{F}(x)}{\partial\theta} = -\frac{\overline{F}^{(1)}(x)}{\overline{F}(x)} = \alpha_1(Y) - \alpha_1({}_xY), \\
L^{(2)}(x) &= -\frac{\overline{F}^{(2)}(x)}{\overline{F}(x)} + \left(\frac{\overline{F}^{(1)}(x)}{\overline{F}(x)}\right)^2 = \mu_2(Y) - \mu_2({}_xY), \\
L^{(3)}(x) &= -\frac{\overline{F}^{(3)}(x)}{\overline{F}(x)} + 3\frac{\overline{F}^{(2)}(x)}{\overline{F}(x)}\frac{\overline{F}^{(1)}(x)}{\overline{F}(x)} - 2\left(\frac{\overline{F}^{(1)}(x)}{\overline{F}(x)}\right)^3 = \mu_3(Y) - \mu_3({}_xY).
\end{aligned}$$

Notice that the functions of $L^{(j)}(x)$ given by $\overline{L}_k(x)$ yield $H_k({}_xY, Y)$.

$$\begin{aligned}
\overline{L}_1(x) &= L^{(1)}(x) = \alpha_1(Y) - \alpha_1({}_xY), \\
\overline{L}_2(x) &= L^{(1)}(x)^2 + L^{(2)}(x) = \Big(\alpha_1(Y) - \alpha_1({}_xY)\Big)^2 + \Big(\mu_2(Y) - \mu_2({}_xY)\Big), \\
\overline{L}_3(x) &= L^{(1)}(x)^3 + 3L^{(2)}(x)L^{(1)}(x) + L^{(3)}(x) \\
&= \Big(\alpha_1(Y) - \alpha_1({}_xY)\Big)^3 + 3\Big(\alpha_1(Y) - \alpha_1({}_xY)\Big)\Big(\mu_2(Y) - \mu_2({}_xY)\Big) \\
&+ \Big(\mu_3(Y) - \mu_3({}_xY)\Big).
\end{aligned}$$

■

We explore the truncated and censored lifetime ${}^{\upsilon}_{\tau}T_{i,j}$ by separating it into its component parts: the systematic $Y_{0,j}$ and the idiosyncratic, truncated and censored $Y_{i,j}$. We provide the details below using a conditional argument and the property that for $a, b, c \in \mathbb{R}$, $\min(a+b, a+c) = a + \min(b, c)$.

$$\begin{aligned}
{}^{\upsilon}_{\tau}T_{i,j}|Y_{0,j} &= \Big\{\min(Y_{0,j} + Y_{i,j}, \upsilon)|(Y_{0,j} + Y_{i,j} > \tau)\Big\}\Big|Y_{0,j} \\
&= \Big\{Y_{0,j} + \min(Y_{i,j}, \upsilon - Y_{0,j})|(Y_{i,j} > \tau - Y_{0,j})\Big\}\Big|Y_{0,j} \\
&= \Big\{Y_{0,j} + {}^{\upsilon'}_{\tau'}Y_{i,j}\Big\}\Big|Y_{0,j} \\
&= Y_{0,j} + {}^{\upsilon'}_{\tau'}Y_{i,j}|Y_{0,j}, \qquad (1)
\end{aligned}$$

where $\tau' = \tau - Y_{0,j}$ and $\upsilon' = \upsilon - Y_{0,j}$. We consider the general case and a simplified case and obtain systems of equations for both. For the simplified case, we provide algorithms that facilitate parameter estimation.

4 Parameter Estimation

Consider given truncated and censored samples ${}^{\upsilon}_{\tau}\mathbf{T}_1, \ldots, {}^{\upsilon}_{\tau}\mathbf{T}_M$, where ${}^{\upsilon}_{\tau}\mathbf{T}_j = ({}^{\upsilon}_{\tau}T_{1,j}, \ldots, {}^{\upsilon}_{\tau}T_{N,j})'$. From each pool j, we estimate θ_j and λ_j, and predict the value of $Y_{0,j}$. Define ${}^{\upsilon'}_{\tau'}\mathbf{Y}_j = ({}^{\upsilon'}_{\tau'}Y_{1,j}, \ldots, {}^{\upsilon'}_{\tau'}Y_{N,j})'$. We have

$$a_1({}^{\upsilon}_{\tau}\mathbf{T}_j)|Y_{0,j} = Y_{0,j} + \frac{1}{N}\sum_{i=1}^{N} {}^{\upsilon'}_{\tau'}Y_{i,j}|Y_{0,j} = Y_{0,j} + a_1({}^{\upsilon'}_{\tau'}\mathbf{Y}_j)|Y_{0,j},$$

and

$$\begin{aligned}
\widetilde{m}_2({}^{\upsilon}_{\tau}\mathbf{T}_j)|Y_{0,j} &= \widetilde{m}_2({}^{\upsilon'}_{\tau'}\mathbf{Y}_j)|Y_{0,j}, \\
\widetilde{m}_3({}^{\upsilon}_{\tau}\mathbf{T}_j)|Y_{0,j} &= \widetilde{m}_3({}^{\upsilon'}_{\tau'}\mathbf{Y}_j)|Y_{0,j}.
\end{aligned}$$

Note that $Y_{0,j}$ is present in the truncation and censoring points τ', υ'.

The ${}^{\upsilon'}_{\tau'}Y_{1,j}, \ldots, {}^{\upsilon'}_{\tau'}Y_{N,j}$ are independent and identically distributed given $Y_{0,j}$. Consequently, the first raw sample moment is an unbiased estimator of $\alpha_1({}^{\upsilon'}_{\tau'}Y_{1,j}|Y_{0,j})$. Moreover,

$$a_1({}^{\upsilon}_{\tau}\mathbf{T}_j|Y_{0,j}) \xrightarrow{P} Y_{0,j} + \alpha_1({}^{\upsilon'}_{\tau'}Y_{1,j}|Y_{0,j}),$$

and the (adjusted) second and third central sample moments are unbiased and consistent estimators of $\mu_2({}^{\upsilon'}_{\tau'}Y_{1,j}|Y_{0,j})$ and $\mu_3({}^{\upsilon'}_{\tau'}Y_{1,j}|Y_{0,j})$, respectively. Theorems 1 and 2 yield the following system:

$$\begin{aligned}
E[a_1({}^{\upsilon}_{\tau}\mathbf{T}_j)|Y_{0,j}] &= Y_{0,j} + E[a_1({}^{\upsilon'}_{\tau'}\mathbf{Y}_j)|Y_{0,j}] \\
&= Y_{0,j} + \alpha_1({}^{\upsilon'}_{\tau'}Y_{1,j}|Y_{0,j}) \\
&= Y_{0,j} + \lambda_j\kappa'(\theta_j) + g_1(\tau') + h_1(\tau', \upsilon'), \qquad (2) \\
E[\widetilde{m}_2({}^{\upsilon}_{\tau}\mathbf{T}_j)|Y_{0,j}] &= E[\widetilde{m}_2({}^{\upsilon'}_{\tau'}\mathbf{Y}_j)|Y_{0,j}] = \mu_2({}^{\upsilon'}_{\tau'}Y_{1,j}|Y_{0,j}) \\
&= \lambda_j\kappa''(\theta_j) + g_2(\tau') - g_1(\tau')^2 \\
&\quad + h_2(\tau', \upsilon') - h_1(\tau', \upsilon')^2, \qquad (3) \\
E[\widetilde{m}_3({}^{\upsilon}_{\tau}\mathbf{T}_j)|Y_{0,j}] &= E[\widetilde{m}_3({}^{\upsilon'}_{\tau'}\mathbf{Y}_j)|Y_{0,j}] = \mu_3({}^{\upsilon'}_{\tau'}Y_{1,j}|Y_{0,j}) \\
&= \lambda_j\kappa'''(\theta_j) + g_3(\tau') - 3g_2(\tau')g_1(\tau') + 2g_1(\tau')^3 \\
&\quad + h_3(\tau', \upsilon') - 3h_2(\tau', \upsilon')h_1(\tau', \upsilon') + 2h_1(\tau', \upsilon')^3, \qquad (4)
\end{aligned}$$

where $g_k(\tau') = g_k(\tau'; \theta_j, \lambda_j), k = 1, 2, 3$, as defined in Theorem 1; and $h_k(\tau', \upsilon') = h_k(\tau', \upsilon'; \theta_j, \lambda_j), k = 1, 2, 3$, as defined in Theorem 2.

4.1 The Simplified Case

We assume that $\theta_j = \theta$ and $\lambda_j = \lambda$ for all j. This assumption implies the pools of lives have the same risk profile. In the case of heterogeneous pools, the groups should be reduced to homogeneous sub-groups.

Define ${}^{\upsilon}_{\tau}\mathbf{T} = ({}^{\upsilon}_{\tau}T_{1,1}, \ldots, {}^{\upsilon}_{\tau}T_{N,M})'$, the vector of all lifetimes. The components of ${}^{\upsilon}_{\tau}\mathbf{T}$ are identically distributed, although not independent. This implies that the raw sample moments of ${}^{\upsilon}_{\tau}\mathbf{T}$ are unbiased estimators of the raw moments of ${}^{\upsilon}_{\tau}T_{1,1}$. Recall that $T_{i,j} \sim Tw_p(\theta, \tilde{\lambda} = \lambda_0 + \lambda)$. Then ${}^{\upsilon}_{\tau}T_{i,j}$ has a truncated and censored $Tw_p(\theta, \tilde{\lambda})$ distribution. Utilizing the first raw and second central moments, we obtain the following from Theorems 1 and 2:

$$\begin{aligned}
E[a_1({}^{\upsilon}_{\tau}\mathbf{T})] &= \alpha_1({}^{\upsilon}_{\tau}T_{1,1}) = \tilde{\lambda}\kappa'(\theta) + g_1(\tau; \theta, \tilde{\lambda}) + h_1(\tau, \upsilon; \theta, \tilde{\lambda}), \qquad (5) \\
E[\widetilde{m}_2({}^{\upsilon}_{\tau}\mathbf{T})] &\approx \mu_2({}^{\upsilon}_{\tau}T_{1,1}) = \tilde{\lambda}\kappa''(\theta) + g_2(\tau; \theta, \tilde{\lambda}) - g_1(\tau; \theta, \tilde{\lambda})^2 \\
&\qquad + h_2(\tau, \upsilon; \theta, \tilde{\lambda}) - h_1(\tau, \upsilon; \theta, \tilde{\lambda})^2. \qquad (6)
\end{aligned}$$

Equation (5) arises from the fact that, for the simplified case, the raw sample moments of ${}_{\tau}^{\upsilon}\mathbf{T}$ are unbiased estimators of the raw moments of ${}_{\tau}^{\upsilon}T_{1,1}$. This does not hold for central moments, yielding the approximation given by (6). Furthermore, notice that we no longer condition on a single $Y_{0,j}$. This is due to the fact that ${}_{\tau}^{\upsilon}\mathbf{T}$ contains M different realizations from the $Tw_p(\theta, \lambda_0)$ distribution, rather than one. It is, therefore, a viable option to take expectations with respect to the $Y_{0,j}$. To facilitate presentation, we introduce the *cumulative* adjustment functions c_k, defined as follows:

$$\begin{aligned} c_k(\tau, \upsilon; \theta, \lambda) &= g_k(\tau; \theta, \lambda) + h_k(\tau, \upsilon; \theta, \lambda), \\ c_k^{(2)}(\tau, \upsilon; \theta, \lambda) &= g_k(\tau; \theta, \lambda)^2 + h_k(\tau, \upsilon; \theta, \lambda)^2, \end{aligned}$$

for $k = 1, 2$.

In order to solve the non-linear system given by equations (5) and (6), we apply an iterative algorithm. This algorithm is a modified version of Algorithm 1 in Alai *et al.* (2015). To apply it we first notice that

$$\frac{\kappa'(\theta)}{\kappa''(\theta)} = \begin{cases} 1, & p = 1, \\ \frac{\theta}{\alpha - 1}, & p \neq 1. \end{cases}$$

Then system of equations (5) and (6) for $p \neq 1$ can be reduced to the following:

$$\theta = \frac{(\alpha - 1)(\alpha_1({}_{\tau}^{\upsilon}T_{1,1}) - c_1(\tau, \upsilon; \theta, \tilde{\lambda}))}{\mu_2({}_{\tau}^{\upsilon}T_{1,1}) - c_2(\tau, \upsilon; \theta, \tilde{\lambda}) + c_1^{(2)}(\tau, \upsilon; \theta, \tilde{\lambda})}, \tag{7}$$

$$\tilde{\lambda} = \frac{\alpha_1({}_{\tau}^{\upsilon}T_{1,1}) - c_1(\tau, \upsilon; \theta, \tilde{\lambda})}{\kappa'(\theta)}. \tag{8}$$

Algorithm 1.

1. *Assume starting values for θ and $\tilde{\lambda}$, denote them $\theta(1)$ and $\tilde{\lambda}(1)$.*

2. *Substitute $\theta(r)$ and $\tilde{\lambda}(r)$ into equations (7) and (8) to obtain parameter estimators $\theta(r + 1)$ and $\tilde{\lambda}(r + 1)$ as follows:*

$$\begin{aligned} \theta(r + 1) &= \frac{(\alpha - 1)(a_1({}_{\tau}^{\upsilon}\mathbf{T}) - c_1(\tau, \upsilon; \theta(r), \tilde{\lambda}(r)))}{\widetilde{m}_2({}_{\tau}^{\upsilon}\mathbf{T}) - c_2(\tau, \upsilon; \theta(r), \tilde{\lambda}(r)) + c_1^{(2)}(\tau, \upsilon; \theta(r), \tilde{\lambda}(r))}, \\ \tilde{\lambda}(r + 1) &= \frac{a_1({}_{\tau}^{\upsilon}\mathbf{T}) - c_1(\tau, \upsilon; \theta(r), \tilde{\lambda}(r))}{\kappa'(\theta(r + 1))}, \end{aligned}$$

where the sample moments of ${}_{\tau}^{\upsilon}\mathbf{T}$ are used to estimate the theoretical moments.

3. Return to Step 2 with $r = r+1$ until parameter estimates are stable.

From Algorithm 1, we obtain parameter estimate $\widehat{\theta}$. We return to individual pool j. We reconsider equations (2) and (3), this time, utilizing both $\widehat{\theta}$ and the cumulative adjustment functions c_k.

$$\begin{aligned} E[a_1({}_{\tau}^{\upsilon}\mathbf{T}_j)|Y_{0,j}] &\approx Y_{0,j} + \lambda\kappa'(\widehat{\theta}) + c_1(\tau', \upsilon'; \widehat{\theta}, \lambda), & (9)\\ E[\widetilde{m}_2({}_{\tau}^{\upsilon}\mathbf{T}_j)|Y_{0,j}] &\approx \lambda\kappa''(\widehat{\theta}) + c_2(\tau', \upsilon'; \widehat{\theta}, \lambda) - c_1^{(2)}(\tau', \upsilon'; \widehat{\theta}, \lambda), & (10)\end{aligned}$$

where $\tau' = \tau - Y_{0,j}$ and $\upsilon' = \upsilon - Y_{0,j}$; see equation (1). Again, we are presented with a non-linear system of equations. We apply the following iterative algorithm, which is a modification of Algorithm 2 developed in Alai *et al.* (2015).

Algorithm 2.

1. *Assume starting values for $Y_{0,j}$ and λ, denote them $Y_{0,j}(1)$ and $\lambda(1)$.*
2. *Substitute $Y_{0,j}(r)$ and $\lambda(r)$ into equation (10) to obtain $\lambda(r+1)$,*

$$\lambda(r+1) = \frac{\widetilde{m}_2({}_{\tau}^{\upsilon}\mathbf{T}_j) - c_2(\tau'(r), \upsilon'(r); \widehat{\theta}, \lambda(r)) + c_1^{(2)}(\tau'(r), \upsilon'(r); \widehat{\theta}, \lambda(r))}{\kappa''(\widehat{\theta})},$$

where $\tau'(r) = \tau - Y_{0,j}(r)$ and $\upsilon'(r) = \upsilon - Y_{0,j}(r)$.

3. *Substitute $\lambda(r+1)$ into equation (9) to obtain $Y_{0,j}(r+1)$,*

$$Y_{0,j}(r+1) = a_1({}_{\tau}^{\upsilon}\mathbf{T}_j) - \lambda(r+1)\kappa'(\widehat{\theta}) - c_1(\tau'(r), \upsilon'(r); \widehat{\theta}, \lambda(r+1)).$$

4. *Return to Step 2 with $r = r+1$ until parameter estimates are stable.*

Finally, we set

$$\widehat{\lambda} = \frac{1}{M}\sum_{j=1}^{M}\widehat{\lambda}^{(j)}, \quad \text{and} \quad \widehat{\lambda}_0 = \frac{1}{M}\sum_{j=1}^{M}\frac{\widehat{Y}_{0,j}}{\kappa'(\widehat{\theta})},$$

where $\widehat{\lambda}^{(j)}$ and $\widehat{Y}_{0,j}$ are the estimate of λ and predicted value of $Y_{0,j}$, respectively, obtained using Algorithm 2 on pool j.

5 The Normal and Gamma Distributions

We apply the theory of the preceding sections to the normal and gamma distributions. We find simplified versions of the adjustment functions g and h for these distributions. We subsequently test the performance of Algorithms 1 and 2.

5.1 Normally Distributed Lifetimes

Suppose $Y \sim Tw_p(\theta, \lambda)$ with $p = 0$. Equivalently, Y may be represented using the standard parametrization of the normal distribution, $Y \sim N(\mu = \theta\lambda, \sigma^2 = \lambda)$.

This equivalence yields

$$\begin{aligned} f_Y(x, \theta, \lambda) &= \varphi((x - \theta\lambda)/\sqrt{\lambda})/\sqrt{\lambda}, \\ \overline{F}_Y(x, \theta, \lambda) &= \bar{\Phi}((x - \theta\lambda)/\sqrt{\lambda}), \end{aligned}$$

where $\varphi(x)$ and $\bar{\Phi}(x)$ are the standard normal density and survival functions, respectively.

Functions g_1 and g_2 are given in Alai *et al.* (2015).

$$\begin{aligned} g_1(\tau; \theta, \lambda) &= \sqrt{\lambda}\, \frac{\varphi((\tau - \lambda\theta)/\sqrt{\lambda})}{\bar{\Phi}((\tau - \lambda\theta)/\sqrt{\lambda})}, \\ g_2(\tau; \theta, \lambda) &= -\lambda\, \frac{\varphi'((\tau - \lambda\theta)/\sqrt{\lambda})}{\bar{\Phi}((\tau - \lambda\theta)/\sqrt{\lambda})}. \end{aligned}$$

Functions h_1 and h_2 are given by

$$\begin{aligned} h_1(\tau, \upsilon; \theta, \lambda) &= \left\{ \frac{-\sqrt{\lambda}\varphi'((\upsilon - \lambda\theta)/\sqrt{\lambda})}{\varphi((\upsilon - \lambda\theta)/\sqrt{\lambda})} - g_1(\upsilon; \theta, \lambda) \right\} \frac{\bar{\Phi}((\upsilon - \theta\lambda)/\sqrt{\lambda})}{\bar{\Phi}((\tau - \theta\lambda)/\sqrt{\lambda})} \\ &= \left\{ (\upsilon - \lambda\theta) - g_1(\upsilon; \theta, \lambda) \right\} \frac{\bar{\Phi}((\upsilon - \theta\lambda)/\sqrt{\lambda})}{\bar{\Phi}((\tau - \theta\lambda)/\sqrt{\lambda})}, \\ h_2(\tau, \upsilon; \theta, \lambda) &= \left\{ 2\left((\upsilon - \lambda\theta) - g_1(\upsilon; \theta, \lambda) \right) \left(-g_1(\tau; \theta, \lambda) \right) \right. \\ &+ \left. \left(\frac{\lambda\varphi''((\upsilon - \lambda\theta)/\sqrt{\lambda})}{\varphi((\upsilon - \lambda\theta)/\sqrt{\lambda})} - g_2(\upsilon; \theta, \lambda) \right) \right\} \frac{\bar{\Phi}((\upsilon - \theta\lambda)/\sqrt{\lambda})}{\bar{\Phi}((\tau - \theta\lambda)/\sqrt{\lambda})} \\ &= \left\{ 2\left((\upsilon - \lambda\theta) - g_1(\upsilon; \theta, \lambda) \right) \left(-g_1(\tau; \theta, \lambda) \right) \right. \\ &+ \left. \left(\left((\upsilon - \lambda\theta)^2 - \lambda \right) - g_2(\upsilon; \theta, \lambda) \right) \right\} \frac{\bar{\Phi}((\upsilon - \theta\lambda)/\sqrt{\lambda})}{\bar{\Phi}((\tau - \theta\lambda)/\sqrt{\lambda})}. \end{aligned}$$

Suppose $Y_{i,j} \sim Tw_p(\theta, \lambda)$ and $Y_{0,j} \sim Tw_p(\theta, \lambda_0)$, with $p = 0$. Equivalently, $Y_{i,j} \sim N(\theta\lambda, \lambda)$ and $Y_{0,j} \sim N(\theta\lambda_0, \lambda_0)$. This implies $T_{i,j} \sim Tw_{p=0}(\theta, \tilde{\lambda} = \lambda + \lambda_0) \equiv N(\theta\tilde{\lambda}, \tilde{\lambda})$. Note that $\kappa_{p=0}(\theta) = \theta^2/2$, $\kappa'_{p=0}(\theta) = \theta$, $\kappa''_{p=0}(\theta) = 1$, and $\alpha = 2$. Together with functions g and h, Algorithms 1 and 2 are easily implemented for truncated and censored multivariate normal lifetimes.

Numerical Results

To roughly mirror human mortality, we simulate truncated and censored multivariate normal lifetimes where

$$\begin{aligned} Y_{i,j} &\sim Tw_{p=0}(\theta = 0.2, \lambda = 375) \equiv N(\theta\lambda = 75, \lambda = 375), \\ Y_{0,j} &\sim Tw_{p=0}(\theta = 0.2, \lambda_0 = 25) \equiv N(\theta\lambda_0 = 5, \lambda_0 = 25). \end{aligned}$$

Consequently, each individual lifetime is normally distributed with mean 80 and standard deviation 20,

$$T_{i,j} \sim Tw_{p=0}(\theta = 0.2, \tilde{\lambda} = \lambda + \lambda_0 = 400) \equiv N(\theta\tilde{\lambda} = 80, \tilde{\lambda} = 400).$$

The truncation point was chosen to reflect a retirement-age sample and the censoring point with deferred-annuity type products in mind. The performance of Algorithm 1 applied to the normal distribution is shown in Table 1. The performance is judged by the accuracy of the the estimate of θ. Each column of Table 1 represents a scenario with various numbers of pools and individuals, we find that with the exception of the first scenario, θ is well estimated.

N	1,000	100,000	10,000	1,000
M	1	1	50	1,000
τ	60	60	60	60
υ	85	85	85	85
$\tilde{\lambda}$	400	400	400	400
$\widehat{\tilde{\lambda}}$	474	382	401	400
θ	0.200	0.200	0.200	0.200
$\widehat{\theta}$	0.163	0.209	0.198	0.199

Table 1: Simulation results to test Algorithm 1 using the normal distribution.

The performance of Algorithm 2 applied to the normal distribution is shown in Table 2. Algorithm 2 requires θ known (estimated, practically speaking), and produces $\widehat{\lambda}$ and $\widehat{Y_0}$ for one pool. In our simulation, we focus on one pool of various sizes, stipulate $Y_0 = 5$, which is its expected value, and use the true θ. It is evident from the results that accurate prediction of the systematic component requires a large sample.

5.2 Gamma Distributed Lifetimes

Suppose $Y \sim Tw_p(\theta, \lambda)$ with $p = 2$. Equivalently, Y may be represented using the standard parametrization of the gamma distribution, $Y \sim \Gamma(\lambda, \beta = -\theta)$, where λ, β, are the shape and rate parameters, respectively.

N	100	1,000	10,000	100,000	1,000,000
τ	60	60	60	60	60
υ	85	85	85	85	85
θ	0.2	0.2	0.2	0.2	0.2
Y_0	5.000	5.000	5.000	5.000	5.000
$\widehat{Y}_0$	11.193	14.012	8.960	2.964	5.453
λ	375.000	375.000	375.000	375.000	375.000
$\widehat{\lambda}$	346.352	333.347	356.411	384.955	372.916

Table 2: Simulation results to test Algorithm 2 using the normal distribution.

This equivalence yields

$$\begin{aligned} f_Y(x,\theta,\lambda) &= \frac{(-\theta)^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{\theta x} \equiv \gamma(x;\lambda,-\theta), \\ \overline{F}_Y(x,\theta,\lambda) &= \int_x^\infty f_Y(t,\theta,\lambda)dt \equiv \overline{\Gamma}(x;\lambda,-\theta). \end{aligned}$$

where $\gamma(x)$ and $\overline{\Gamma}(x)$ are the gamma density and survival functions, respectively.

Functions g_1 and g_2 are given in Alai *et al.* (2015).

$$\begin{aligned} g_1(\tau;\theta,\lambda) &= \frac{\lambda}{\theta}\Big(1 - K_1(\tau;\theta,\lambda)\Big), \\ g_2(\tau;\theta,\lambda) &= \frac{\lambda}{\theta^2}\Big((\lambda-1) - 2\lambda K_1(\tau;\theta,\lambda) + (\lambda+1)K_2(\tau;\theta,\lambda)\Big), \end{aligned}$$

where

$$K_k(x;\theta,\lambda) = \frac{\overline{\Gamma}(x;\lambda+k,-\theta)}{\overline{\Gamma}(x;\lambda,-\theta)}, \quad k = 1,2.$$

Additionally, we have that

$$\begin{aligned} \frac{f^{(1)}(x)}{f(x)} &= x + \frac{\lambda}{\theta} = \frac{\lambda}{\theta}\Big(1 - k_1(x;\theta,\lambda)\Big), \\ \frac{f^{(2)}(x)}{f(x)} &= \Big(x + \frac{\lambda}{\theta}\Big)^2 - \frac{\lambda}{\theta^2} = \frac{\lambda}{\theta^2}\Big((\lambda-1) - 2\lambda k_1(x;\theta,\lambda) + (\lambda+1)k_2(x;\theta,\lambda)\Big), \\ \bar{L}_1(x) &= L^{(1)}(x) = -\frac{\overline{F}^{(1)}(x)}{\overline{F}(x)} = -g_1(x;\theta,\lambda), \end{aligned}$$

where

$$k_k(x;\theta,\lambda) = \frac{\gamma(x;\lambda+k,-\theta)}{\gamma(x;\lambda,-\theta)}, \quad k = 1,2.$$

We obtain

$$\begin{aligned}
h_1(\tau,\upsilon;\theta,\lambda) &= \left\{\frac{\lambda}{\theta}\Big(1-k_1(\upsilon;\theta,\lambda)\Big)-\frac{\lambda}{\theta}\Big(1-K_1(\upsilon;\theta,\lambda)\Big)\right\}\frac{\overline{\Gamma}(\upsilon;\lambda,-\theta)}{\overline{\Gamma}(\tau;\lambda,-\theta)}\\
&= \left\{\frac{\lambda}{\theta}\Big(K_1(\upsilon;\theta,\lambda)-k_1(\upsilon;\theta,\lambda)\Big)\right\}\frac{\overline{\Gamma}(\upsilon;\lambda,-\theta)}{\overline{\Gamma}(\tau;\lambda,-\theta)},\\
h_2(\tau,\upsilon;\theta,\lambda) &= \Bigg\{\frac{2\lambda}{\theta}\Big(K_1(\upsilon;\theta,\lambda)-k_1(\upsilon;\theta,\lambda)\Big)\Big(-g_1(\tau;\theta,\lambda)\Big)\\
&+ \Bigg(\frac{\lambda}{\theta^2}\Big((\lambda-1)-2\lambda k_1(\upsilon;\theta,\lambda)+(\lambda+1)k_2(\upsilon;\theta,\lambda)\Big)\\
&- \frac{\lambda}{\theta^2}\Big((\lambda-1)-2\lambda K_1(\upsilon;\theta,\lambda)+(\lambda+1)K_2(\upsilon;\theta,\lambda)\Big)\Bigg)\Bigg\}\frac{\overline{\Gamma}(\upsilon;\lambda,-\theta)}{\overline{\Gamma}(\tau;\lambda,-\theta)}\\
&= \Bigg\{\frac{2\lambda}{\theta}\Big(K_1(\upsilon;\theta,\lambda)-k_1(\upsilon;\theta,\lambda)\Big)\Big(-g_1(\tau;\theta,\lambda)\Big)\\
&+ \frac{\lambda}{\theta^2}\Bigg(2\lambda\Big(K_1(\upsilon;\theta,\lambda)-k_1(\upsilon;\theta,\lambda)\Big)\\
&- (\lambda+1)\Big(K_2(\upsilon;\theta,\lambda)-k_2(\upsilon;\theta,\lambda)\Big)\Bigg)\Bigg\}\frac{\overline{\Gamma}(\upsilon;\lambda,-\theta)}{\overline{\Gamma}(\tau;\lambda,-\theta)}.
\end{aligned}$$

Suppose $Y_{i,j} \sim Tw_p(\theta,\lambda)$ and $Y_{0,j} \sim Tw_p(\theta,\lambda_0)$, with $p = 2$. Equivalently, $Y_{i,j} \sim \Gamma(\lambda,-\theta)$ and $Y_{0,j} \sim \Gamma(\lambda_0,-\theta)$. This implies that $T_{i,j} \sim Tw_{p=2}(\theta,\tilde{\lambda} = \lambda+\lambda_0) \equiv \Gamma(\tilde{\lambda},-\theta)$. Note that $\kappa_{p=2}(\theta) = -\ln(-\theta)$, $\kappa'_{p=2}(\theta) = -1/\theta$, $\kappa''_{p=2}(\theta) = 1/\theta^2$, and $\alpha = 0$. Together with functions g and h, Algorithms 1 and 2 easily implemented for truncated and censored multivariate gamma lifetimes.

Numerical Results

We simulate truncated and censored multivariate gamma lifetimes where

$$\begin{aligned}
Y_{i,j} &\sim Tw_{p=2}(\theta=-0.2,\lambda=15) \equiv \Gamma(\lambda=15,\beta=0.2),\\
Y_{0,j} &\sim Tw_{p=2}(\theta=-0.2,\lambda_0=1) \equiv \Gamma(\lambda_0=1,\beta=0.2).
\end{aligned}$$

Each individual lifetime is gamma distributed with mean 80 and standard deviation 20,

$$T_{i,j} \sim Tw_{p=2}(\theta=-0.2,\tilde{\lambda}=\lambda+\lambda_0=16) \equiv \Gamma(\tilde{\lambda}=16,\beta=0.2).$$

The performance of Algorithm 1 applied to the gamma distribution is shown in Table 3. Again, save the first scenario, we find that θ is well estimated.

N	1,000	100,000	10,000	1,000
M	1	1	50	1,000
τ	60	60	60	60
υ	85	85	85	85
$\tilde{\lambda}$	16.00	16.00	16.00	16.00
$\widehat{\tilde{\lambda}}$	24.15	17.04	15.84	15.97
θ	-0.200	-0.200	-0.200	-0.200
$\widehat{\theta}$	-0.306	-0.213	-0.199	-0.201

Table 3: Simulation results to test Algorithm 1 using the gamma distribution.

The performance of Algorithm 2 applied to the gamma distribution is shown in Table 4. As before, we focus on one pool of various sizes, stipulate $Y_0 = 5$, and use the true θ. Again, accurate prediction of the systematic component requires a relatively large sample.

N	100	1,000	10,000	100,000	1,000,000
τ	60	60	60	60	60
υ	85	85	85	85	85
θ	-0.2	-0.2	-0.2	-0.2	-0.2
Y_0	5.000	5.000	5.000	5.000	5.000
$\widehat{Y}_0$	9.449	15.879	1.134	7.417	4.946
λ	15.000	15.000	15.000	15.000	15.000
$\widehat{\lambda}$	14.605	13.213	15.796	14.542	15.016

Table 4: Simulation results to test Algorithm 2 using the gamma distribution.

6 Conclusion

We model dependence within a portfolio of lives using a common stochastic component. In this structure, the common component represents systematic mortality improvements; pools may represent a variety of situations, from nation-wide cohorts to employer-based pension annuity portfolios. We develop parameter estimation in the presence of truncated and censored observations. Previous work considered truncation only. The allowance for censoring makes the model much more applicable to the study of pools of lives, since it eliminates basis-risk. From an annuity provider perspective, the model provides the means to actively manage systematic longevity risk. The introduction of censoring leads to some interesting theoretical results; censoring adjustments are derived and added to previously formulated truncation adjustments. Parameter estimation is developed using modifications of established algorithms. Finally, explicit solutions are derived for the normal and gamma distributions, the two most widely-used members of the Tweedie family of distributions.

Acknowledgements

The authors would like to acknowledge the financial support of ARC Linkage Grant Project LP0883398 Managing Risk with Insurance and Superannuation as Individuals Age with industry partners PwC, APRA and the World Bank as well as the support of the Australian Research Council Centre of Excellence in Population Ageing Research (project number CE110001029). The authors also wish to thank the Israel Zimmerman Foundation for financial support.

References

Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, **2**(4), 951–972.

Alai, D. H., Landsman, Z., and Sherris, M. (2013). Lifetime dependence modelling using a multivariate gamma distribution. *Insurance: Mathematics and Economics*, **52**(3), 542–549.

Alai, D. H., Landsman, Z., and Sherris, M. (2015). Multivariate Tweedie lifetimes: The impact of dependence. To appear in *Scandinavian Actuarial Journal*.

Chatelain, F., Tourneret, J.-Y., Inglada, J., and Ferrari, A. (2006). Parameter estimation for multivariate gamma distributions. In Proceedings of EUSIPCO, Florence, Italy, September 2006.

Chereiyan, K. C. (1941). A bivariate correlated gamma-type distribution function. *Journal of the Indian Mathematical Society*, **5**, 133–144.

Denuit, M., Dhaene, J., Le Bailly de Tilleghem, C., and Teghem, S. (2001). Measuring the impact of a dependence among insured lifelengths. *Belgian Actuarial Bulletin*, **1**, 18–39.

Dhaene, J., Vanneste, M., and Wolthuis, H. (2000). A note on dependencies in multiple life statuses. *Bulletin of the Swiss Association of Actuaries*, **1**, 19–34.

Furman, E. and Landsman, Z. (2010). Multivariate Tweedie distributions and some related capital-at-risk analysis. *Insurance: Mathematics and Economics*, **46**(2), 351–361.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall, London.

Jørgensen, B. and De Souza, M. C. P. (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial J.*, **1**, 69–93.

Kaas, R. (2005). Compound Poisson distributions and GLM's – Tweedie's distribution. Lecture, Royal Flemish Academy of Belgium for Science and the Arts.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer.

Landsman, Z. and Valdez, E. (2005). Tail conditional expectation for exponential dispersion models. *ASTIN Bulletin*, **35**(1), 189–209.

Mathai, A. M. and Moschopoulus, P. G. (1991). On a multivariate gamma. *Journal of Multivariate Analysis*, **39**(1), 135–153.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, 2nd edition.

Ramabhadran, V. R. (1951). A multivariate gamma-type distribution. *Journal of Multivariate Analysis*, **38**, 213–232.

Smyth, G. K. and Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *ASTIN Bulletin*, **32**(1), 143–157.

Tweedie, M. C. K. (1984). An index which distinguishes some important exponential families. in *Statistics: Applications and New Directions*, J. K. Ghosh and J. Roy (eds), Indian Statistical Institute, Calcutta, pp. 579–604.

Wüthrich, M. V. (2003). Claims reserving using Tweedie's compound Poisson model. *ASTIN Bulletin*, **33**(2), 331–346.

A Proof of Lemma 1

We restate the lemma:

Lemma. *For $j = 1, 2, 3$, and random variables A, B, C, D,*

$$H_j(A,C) - H_j(B,C) = \sum_{i=0}^{j} \binom{j}{i} \Big(H_i(A,D) - H_i(B,D)\Big) H_{j-i}(D,C),$$

where $H_0 \equiv 1$.

Proof. The idea of the proof is simple. First, using the definition of H, we expand the left hand side. Second, for all differences of moments, we add and subtract the appropriate moment for random variable D. Lastly, we expand all powers, collect terms, and rewrite as functions of H.

For $j = 1$, we have:

$$\begin{aligned}
H_1(A,C) &- H_1(B,C) \\
&= \Big(\alpha_1(A) - \alpha_1(C)\Big) - \Big(\alpha_1(B) - \alpha_1(C)\Big) \\
&= \Big(\alpha_1(A) - \alpha_1(D) + \alpha_1(D) - \alpha_1(C)\Big) \\
&- \Big(\alpha_1(B) - \alpha_1(D) + \alpha_1(D) - \alpha_1(C)\Big) \\
&= H_1(A,D) - H_1(C,D) - \Big(H_1(B,D) - H_1(C,D)\Big) \\
&= H_1(A,D) - H_1(B,D).
\end{aligned}$$

For $j = 2$, we have:

$$\begin{aligned}
H_2(A,C) &- H_2(B,C) \\
&= \Big(\alpha_1(A) - \alpha_1(C)\Big)^2 + \Big(\mu_2(A) - \mu_2(C)\Big) \\
&- \Big(\alpha_1(B) - \alpha_1(C)\Big)^2 - \Big(\mu_2(B) - \mu_2(C)\Big) \\
&= \Big(\alpha_1(A) - \alpha_1(D) + \alpha_1(D) - \alpha_1(C)\Big)^2 + \Big(\mu_2(A) - \mu_2(D)\Big) \\
&- \Big(\alpha_1(B) - \alpha_1(D) + \alpha_1(D) - \alpha_1(C)\Big)^2 - \Big(\mu_2(B) - \mu_2(D)\Big) \\
&= 2\Big((\alpha_1(A) - \alpha_1(D)) - (\alpha_1(B) - \alpha_1(D))\Big)\Big(\alpha_1(D) - \alpha_1(C)\Big) \\
&+ \Big(\alpha_1(A) - \alpha_1(D)\Big)^2 + \Big(\mu_2(A) - \mu_2(D)\Big) \\
&- \Big(\alpha_1(B) - \alpha_1(D)\Big)^2 - \Big(\mu_2(B) - \mu_2(D)\Big) \\
&= 2\Big(H_1(A,D) - H_1(B,D)\Big)H_1(D,C) + \Big(H_2(A,D) - H_2(B,D)\Big).
\end{aligned}$$

For $j = 3$, we have:

$$
\begin{aligned}
H_3(A,C) &- H_3(B,C)\\
&= \Big(\alpha_1(A)-\alpha_1(C)\Big)^3 + 3\Big(\alpha_1(A)-\alpha_1(C)\Big)\Big(\mu_2(A)-\mu_2(C)\Big) + \Big(\mu_3(A)-\mu_3(C)\Big)\\
&- \Big(\alpha_1(B)-\alpha_1(C)\Big)^3 - 3\Big(\alpha_1(B)-\alpha_1(C)\Big)\Big(\mu_2(B)-\mu_2(C)\Big) - \Big(\mu_3(B)-\mu_3(C)\Big)\\
&= \Big(\alpha_1(A)-\alpha_1(D)+\alpha_1(D)-\alpha_1(C)\Big)^3 + \Big(\mu_3(A)-\mu_3(D)\Big)\\
&+ 3\Big(\alpha_1(A)-\alpha_1(D)+\alpha_1(D)-\alpha_1(C)\Big)\Big(\mu_2(A)-\mu_2(D)+\mu_2(D)-\mu_2(C)\Big)\\
&- \Big(\alpha_1(B)-\alpha_1(D)+\alpha_1(D)-\alpha_1(C)\Big)^3 - \Big(\mu_3(A)-\mu_3(D)\Big)\\
&- 3\Big(\alpha_1(B)-\alpha_1(D)+\alpha_1(D)-\alpha_1(C)\Big)\Big(\mu_2(B)-\mu_2(D)+\mu_2(D)-\mu_2(C)\Big)\\
&= 3\Big(\big(\alpha_1(A)-\alpha_1(D)\big)^2 - \big(\alpha_1(B)-\alpha_1(D)\big)^2\Big)\Big(\alpha_1(D)-\alpha_1(C)\Big)\\
&+ 3\Big(\big(\alpha_1(A)-\alpha_1(D)\big) - \big(\alpha_1(B)-\alpha_1(D)\big)\Big)\Big(\alpha_1(D)-\alpha_1(C)\Big)^2\\
&+ 3\Big(\alpha_1(A)-\alpha_1(D)+\alpha_1(D)-\alpha_1(C)\Big)\Big(\mu_2(A)-\mu_2(D)+\mu_2(D)-\mu_2(C)\Big)\\
&- 3\Big(\alpha_1(B)-\alpha_1(D)+\alpha_1(D)-\alpha_1(C)\Big)\Big(\mu_2(B)-\mu_2(D)+\mu_2(D)-\mu_2(C)\Big)\\
&+ \Big(\alpha_1(A)-\alpha_1(D)\Big)^3 + \Big(\mu_3(A)-\mu_3(D)\Big)\\
&- \Big(\alpha_1(B)-\alpha_1(D)\Big)^3 - \Big(\mu_3(B)-\mu_3(D)\Big)\\
&= 3\Big(\big(\alpha_1(A)-\alpha_1(D)\big)\Big)\Big(\big(\alpha_1(D)-\alpha_1(C)\big)^2 + \big(\mu_2(D)-\mu_2(C)\big)\Big)\\
&- 3\Big(\big(\alpha_1(B)-\alpha_1(D)\big)\Big)\Big(\big(\alpha_1(D)-\alpha_1(C)\big)^2 + \big(\mu_2(D)-\mu_2(C)\big)\Big)\\
&+ 3\Big(\big(\alpha_1(A)-\alpha_1(D)\big)^2 + \big(\mu_2(A)-\mu_2(D)\big)\Big)\Big(\alpha_1(D)-\alpha_1(C)\Big)\\
&- 3\Big(\big(\alpha_1(B)-\alpha_1(D)\big)^2 + \big(\mu_2(B)-\mu_2(D)\big)\Big)\Big(\alpha_1(D)-\alpha_1(C)\Big)\\
&+ 3\Big(\alpha_1(A)-\alpha_1(D)\Big)\Big(\mu_2(A)-\mu_2(D)\Big) - 3\Big(\alpha_1(B)-\alpha_1(D)\Big)\Big(\mu_2(B)-\mu_2(D)\Big)\\
&+ \Big(\alpha_1(A)-\alpha_1(D)\Big)^3 + \Big(\mu_3(A)-\mu_3(D)\Big)\\
&- \Big(\alpha_1(B)-\alpha_1(D)\Big)^3 - \Big(\mu_3(B)-\mu_3(D)\Big)\\
&= 3\Big(H_1(A,D)-H_1(B,D)\Big)H_2(D,C) + 3\Big(H_2(A,D)-H_2(B,D)\Big)H_1(D,C)\\
&+ \Big(H_3(A,D)-H_3(B,D)\Big).
\end{aligned}
$$

■