Statistical Evidence for the Preference of Frailty Distributions with Regularly-Varying-at-Zero Densities

(preliminary draft submitted to the 2015 PAA Annual Meeting, San Diego, USA, April 30-May 2)

Trifon I. Missov[1,2] and Jonas Schöley[1]

[1]Max Planck Institute for Demographic Research, Konrad-Zuse-Str. 1, 18057 Rostock, Germany

[2]Institute of Sociology and Demography, University of Rostock, Ulmenstr. 69, 18057 Rostock, Germany

Abstract

Missov and Finkelstein (2011) prove an Abelian and its corresponding Tauberian theorem regarding distributions for modeling unobserved heterogeneity in fixed-frailty mixture models. The main property of such distributions is the regular variation at zero of their densities. According to this criterion admissible distributions are, for example, the gamma, the beta, the truncated normal, the log-logistic and the Weibull, while distributions like the log-normal and the inverse Gaussian do not satisfy this condition. In this article we show that models with admissible frailty distributions and a Gompertz baseline provide a better fit to adult human mortality data than the corresponding models with non-admissible frailty distributions. We implement estimation procedures for mixture models with a Gompertz baseline and frailty that follows a gamma, beta, truncated normal, log-logistic, Weibull, log-normal, or inverse Gaussian distribution.

Keywords: unobserved heterogeneity; fixed frailty; regular variation at zero; AIC; Akaike weights; Gompertz distribution; adult human mortality

Introduction

Fixed-frailty models extend standard survival models by accounting for unobserved heterogeneity. In the absence of covariates the hazard function $\mu(x \,|\, Z)$ for individuals with frailty Z is defined as

$$\mu(x \,|\, Z) = Z\, \mu(x \,|\, 1)\,, \tag{1}$$

where $\mu(x \,|\, 1)$ is the baseline hazard and Z is a random variable that captures individual unobserved or unmeasurable susceptibility to the study event (Vaupel et al. 1979). Frailty acts multiplicatively on the baseline hazard $\mu(x)$ and each individual-specific realization $Z = z$ stays fixed throughout the observation period, i.e. each individual is characterized by an "assigned" unknown frailty number throughout his or her life. The multiplicative connection between frailty and baseline mortality in (1) can be justified in the context of human mortality by the shape of death rates at the oldest ages (Gampe 2010; Missov and Finkelstein 2011; Missov and Vaupel 2015). As Z is unobserved, the estimation of frailty models is carried out by specifying a frailty distribution and working with the resulting marginal distribution with a hazard function

$$\mu(x) = -\mu(x \,|\, 1) \cdot \frac{d}{ds}\, \mathcal{L}_Z(s)\Big|_{s=H(x\,|\,1)}\,, \tag{2}$$

where $\mathcal{L}_Z(\cdot)$ denotes the Laplace transform of Z and $H(x \,|\, 1) = \int_0^x \mu(t \,|\, 1)\, dt$ is the baseline cumulative hazard (for detailed discussion see Vaupel et al. 1979; Vaupel and Yashin 2006; Vaupel and Missov 2014).

Human mortality by senescent causes at adult ages is well captured by a Gompertz

distribution (Yashin et al. 2000), i.e. $\mu(x \mid 1) = ae^{bx}$, where a is the starting level of mortality at adult age $x = 0$ and b is the rate of aging. The choice of a frailty distribution, though, is often driven by convenience, i.e. distributions with a closed-form Laplace transform, e.g., the gamma and the inverse Gaussian, are preferable as the marginal distribution in (2) can be obtained explicitly. The general class of three-parameter distributions with closed-form Laplace transforms is derived in Hougaard (1984) and Aalen (1988, 1992). Another "convenient" framework arises when (1) is taken on a logarithmic scale:

$$\ln \mu(x \mid Z) = \ln \mu(x \mid 1) + \ln Z \,. \tag{3}$$

In this case (3) is viewed as a regression equation and $\ln Z$ as an error term in it. The standard normality assumption regarding the errors implies that frailty Z has a log-normal distribution.

Computational convenience – working with closed-form Laplace transforms in a maximum-likelihood setting or assuming a regression setting that provides closed-form least-squares solutions – is often decisive for choosing a frailty distribution. Generic properties of frailty, though, are little known. As it is a measure of *unobserved* heterogeneity, i.e. variation in the study population due to covariates we cannot measure or we do not have information on, it poses a difficult task to characterize the admissible distributions of frailty. By proving an Abelian and its corresponding Tauberian theorem for frailty distributions, Missov and Finkelstein (2011) derived a property they should comply with: the p.d.f. of frailty needs to be regularly varying at zero with power greater than -1 (for details, see Missov and Finkelstein 2011: p.66–67). One can easily show that this property is fulfilled for the gamma, beta, Weibull, log-logistic, and truncated (at zero) normal distributions, while popular frailty "candidates" like the log-normal and the inverse Gaussian are in this sense non-admissible.

In this article we show that models with theoretically admissible frailty distributions provide a better fit to human mortality data than models with non-admissible frailties. We use high-quality mortality data for Denmark, France, Italy, Japan and Sweden from HMD

(2014). The frailty distributions we consider are the gamma, the beta, the Weibull, the truncated (at zero) normal, and the log-logistic (admissible); the inverse Gaussian and the log-normal (non-admissible). In all cases we assume a Gompertz baseline $\mu(x \,|\, 1) = ae^{bx}$ and start fitting each of the frailty models after age 80 to avoid the effect of possible non-negligible extrinsic mortality. For each year and cohort in each of the five selected countries we compare the model fits by AIC and calculate the associated Akaike weights (Burnham and Anderson 2004). We calculate in what proportion of the cases the best-fitting model has an admissible frailty distribution.

Fixed-Frailty Models with a Gompertz Baseline Hazard

The fixed-frailty frailty (conditional) model with a Gompertz baseline is given by

$$\mu(x \,|\, Z) = Z\, ae^{bx}\,. \tag{4}$$

The associated marginal model is given by (2) taking into account the Laplace transform for the distribution of Z.

The gamma frailty model

The gamma-Gompertz fixed-frailty model is given by

$$\mu(x) = \frac{ae^{bx}}{1 + \frac{a\gamma}{b}\,(e^{bx} - 1)}\,, \tag{5}$$

where Z follows a single-parameter gamma distribution with $\gamma = 1/k = 1/\lambda$ being the squared coefficient of variation. In this assumption frailty has a unit expectation which implies that the "average" individual is exposed to the baseline hazard.

Models with other admissible frailty distributions

The beta, Weibull, truncated (at zero) normal and log-logistic distributions do not have a closed-form Laplace transform, which makes working with them inconvenient. Table 1 provides an overview of the four distributions, including the integral or series representation of their Laplace transforms.

Distribution	Parameters	Density	Laplace Transform
Beta	α, β	$\frac{z^{\alpha-1}(1-z)^{\beta-1}}{\mathbb{B}(\alpha,\beta)}$	$1+\sum_{n=1}^{\infty}\left(\prod_{m=0}^{n-1}\frac{\alpha+m}{\alpha+\beta+m}\right)\frac{(-s)^n}{n!}$
Weibull	a, b	$\frac{a}{b}\left(\frac{z}{b}\right)^{a-1}\exp\left\{-(\frac{z}{b})^a\right\}$	$\sum_{n=0}^{\infty}\frac{(-s)^n b^n}{n!}\,\Gamma\left(1+\frac{n}{a}\right)$
Truncated N	μ, σ^2	$\frac{1}{\sqrt{2\pi}\sigma\left\{1-\Phi\left(-\frac{\mu}{\sigma}\right)\right\}}\exp\left\{\frac{-(z-\mu)^2}{2\sigma^2}\right\}$	$\frac{1-\Phi\left(-\frac{\mu}{\sigma}+\sigma s\right)}{1-\Phi\left(-\frac{\mu}{\sigma}\right)}\cdot e^{-\mu s+\frac{\sigma^2 s^2}{2}}$
Log-Logistic	α, β	$\frac{\frac{\beta}{\alpha}\left(\frac{z}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{z}{\alpha}\right)^{\beta}\right)^2}$	$\int_0^{\infty} e^{-sz}\,\frac{\frac{\beta}{\alpha}\left(\frac{z}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{z}{\alpha}\right)^{\beta}\right]^2}\,dz$

Table 1: An overview of the beta, Weibull, truncated normal, and log-logistic distributions – plausible frailty distributions according to the theoretical criterion in Missov and Finkelstein (2011).

The beta frailty model

The beta distribution has two positive shape parameters α and β. Its p.d.f. arises from the beta function

$$\mathbf{B}(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx\,,$$

which is closely linked to the gamma function

$$\Gamma(t)=\int_0^{\infty} x^{t-1}e^{-x}dx$$

by the relationship

$$\mathbf{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,.$$

As a result the beta-distributed frailty can be viewed as a "version" of the gamma-distributed frailty that is "rescaled" to $[0, 1]$. However, the Laplace transforms of the gamma and beta distributions are not linked to one another by a direct relationship. In fact the Laplace transform of the beta distribution is expressed by

$$\mathcal{L}_{\mathbf{B}}(s) = \int_0^1 e^{-sz} \frac{z^{\alpha-1}(1-z)^{\beta-1}}{\mathbb{B}(\alpha,\beta)} dz = {}_1F_1(\alpha; \alpha+\beta; -s)\,,$$

where ${}_1F_1(\alpha; \alpha+\beta; -s) = 1 + \sum_{n=1}^{\infty} \left(\prod_{m=0}^{n-1} \frac{\alpha+m}{\alpha+\beta+m} \right) \frac{(-s)^n}{n!}$ is the confluent hypergeometric function. Hypergeometric functions are series, and in practice one can use just several terms to get an accurate approximation (for examples of approximation formulae see Missov and Lenart 2013).

As the support of the beta distribution is in $[0, 1]$, an assumption about unit average frailty does not make sense. Instead suppose the average frailty is 0.5, i.e., $\mathrm{E}Z = \alpha/(\alpha+\beta) = 0.5$ which is equivalent to $\alpha = \beta$. As a result, we can re-parameterize the beta's Laplace transform in terms of α only:

$$\mathcal{L}_{\mathbf{B}}(s) = {}_1F_1(\alpha; 2\alpha; -s) = 1 + \sum_{n=1}^{\infty} \left(\prod_{m=0}^{n-1} \frac{\alpha+m}{2\alpha+m} \right) \frac{(-s)^n}{n!}\,.$$

The frailty parameter to estimate is α. If we look for a correspondence with the gamma-frailty model, in which γ is the variance of frailty at the starting age, then the comparable characteristic of the beta distribution will be $\frac{1}{4(2\alpha+1)}$.

The Weibull frailty model

The Weibull distribution is the "piece of mystery" in the list of admissible frailties. It is a generalized extreme value distribution (GEV) and is used exclusively as a baseline mortality distribution (for further discussion, see Lenart and Missov 2014; Missov et al. 2014). However, it does satisfy the regular-variation-at-zero property (Missov and Finkelstein 2011) and is in this sense a plausible candidate for a frailty distribution, too. The Weibull is

described by a couple of positive parameters $a > 0$ (shape) and $b > 0$ (scale), and its Laplace transform is represented by a series:

$$\mathcal{L}_W(s) = \sum_{n=0}^{\infty} \frac{(-s)^n b^n}{n!} \Gamma\left(1 + \frac{n}{a}\right) .$$

Suppose the average frailty is 1. This means that $b\,\Gamma\left(1 + \frac{1}{a}\right) = 1$, i.e., $b = \frac{1}{\Gamma\left(1+\frac{1}{a}\right)}$. Thus, a (the shape of the Weibull) is the only frailty parameter to be estimated, and the variance of the Weibull distribution at the starting age (a characteristic comparable with γ in the gamma-frailty model) is equal to

$$\frac{\Gamma\left(1 + \frac{2}{a}\right)}{\left[\Gamma\left(1 + \frac{1}{a}\right)\right]^2} - 1 .$$

The Laplace transform of the single-parameter Weibull is given by

$$\mathcal{L}_W(s) = \sum_{n=0}^{\infty} \frac{(-s)^n}{n!} \frac{\Gamma\left(1 + \frac{n}{a}\right)}{\left[\Gamma\left(1 + \frac{1}{a}\right)\right]^n} .$$

The truncated normal frailty model

The truncated (at 0) normal distribution has the same set of parameters $\mu \in \mathbb{R}$ (location) and $\sigma^2 \geq 0$ (variation) as the generating (not truncated) normal distribution. The two functions that play an important role for the truncated normal are the p.d.f. of the standard normal distribution

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

and the corresponding c.d.f. of the standard normal

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}\, dy .$$

The Laplace transform of the truncated normal is given in an integral form by

$$\mathcal{L}_{tN}(s) = \frac{1 - \Phi\left(-\frac{\mu}{\sigma} + \sigma s\right)}{1 - \Phi\left(-\frac{\mu}{\sigma}\right)} \cdot e^{-\mu s + \frac{\sigma^2 s^2}{2}} .$$

The log-logistic frailty model

The log-logistic distribution has two parameters $\alpha > 0$ (scale) and $\beta > 0$ (shape). It is another "unexpected" frailty candidate as in survival analysis it is used exclusively as a baseline mortality distribution. Unlike the Gompertz and the Weibull, the log-logistic hazard is not necessarily monotonic and, thus, provides additional flexibility in the model. The Laplace transform of the log-logistic is given by the integral

$$\mathcal{L}_{LL}(s) = \int_0^\infty e^{-sz} \frac{\frac{\beta}{\alpha}\left(\frac{z}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{z}{\alpha}\right)^{\beta}\right]^2} dz .$$

Models with non-admissible frailty distributions

The inverse-Gaussian frailty model

The inverse Gaussian distribution with parameters $\mu > 0$ (mean) and $\lambda > 0$ (shape) has a p.d.f.

$$\pi^{(IG)}(z) = \left(\frac{\lambda}{2\pi z^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(z-\mu)^2}{2\mu^2 z}\right\} .$$

It becomes a single-parameter distribution $\mathrm{InvG}(\sigma^2)$ when we assume unit average frailty $\mathrm{E}Z = \mu = 1$ and denote the resulting variance $\mathrm{Var}Z = \mu^3/\lambda = 1/\lambda =: \sigma^2$. The inverse Gaussian distribution has a closed-form Laplace transform, which leads to the following marginal hazard for the inverse-Gaussian-Gompertz fixed-frailty model:

$$\mu(x) = \frac{ae^{bx}}{\sqrt{1 + \frac{2a\sigma^2}{b}\left(e^{bx} - 1\right)}} . \tag{6}$$

The log-normal frailty model

Log-normal frailty arises from 3: normally distributed random effect acts on a linear predictor, which suits generalized linear models (GLM) framework (McCullagh and Nelder 1989). The standard assumption for the error term $\ln Z$ is that on average it is 0, i.e. $\mathrm{E}Z = 1$. The log-normal distribution does not have a closed-form Laplace transform.

Model Fitting

We fit fixed-frailty models with a Gompertz baseline and gamma, beta, Weibull, truncated normal, log-logistic, inverse Gaussian or log-normal frailty by maximum likelihood.

Frailties with a closed-form Laplace transform

For the gamma and inverse Gaussian distributions we can express the marginal distribution explicitly. The hazard function is given by (5) and (6), while the survival function equals the Laplace transform calculated for the baseline cumulative hazard.

Frailties without a closed-form Laplace transform

For the beta, Weibull, truncated normal, log-logistic and log-normal models we use high accuracy approximations of the series or integrals which appear in the expressions for the Laplace transform.

Optimization Strategies

Three different optimization strategies are applied to maximize the log-likelihood functions of the aforementioned frailty models given the data. All strategies are tested on the same data and with the same objective functions. In this section we compare the three methods in terms of parameter estimates and overall model fit.

`optim` is a quasi-Newton algorithm for optimization (see Dalzell 2013) for an interview with the developer of `optim`). Method `L-BFGS-B` is used to be able to apply the box-constraints $[1e-8, 1]$ on all parameters. The initial parameters are $a = 0.01$, $b = 0.1$ and $\sigma^2 = 0.01$. Default values for the optimization options are used.

`goptim` Using the observed survival data, a Gompertz model is estimated as a Poisson regression. The resulting values for a and b constitute the baseline starting values for `optim`. On each data subset ten different models are computed with these baseline values, but different initial frailty values ($\sigma = \{0.01, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7\}$). The best model in terms of the AIC is chosen to be the final model.

`DEoptim` This *genetic algorithm* (see Mullen et al. 2011) evaluates the constrained parameter space (see optim for constraints) initially at multiple random positions (`NP` = 100). Through *differential evolution* the parameters estimates converge towards a global maximum. The number of iterations (generations) until final estimation is 200.

Computation speed

A total of 12 468 models are fitted across the dataset dimensions *Timeframe* (Period/Cohort), *Country* (Denmark, France, Italy, Japan, Sweden), *Sex* and *Year*. For each subset of the mortality data four models are estimated: the *gamma-Gompertz*, *inverse-Gaussian-Gompertz*, *log-normal-Gompertz* and the *zero-truncated-normal-Gompertz* frailty model. This constitutes a non-trivial computational task and calculation takes minutes to hours, depending on model, method and hardware.

The models are not dependent on one another and therefore parallel computation across the dataset dimensions is used to allow for faster results. The computations takes place on a remote 24 core 3.33 GHz Intel Xeon X5680 workstation.

Results: Standard `optim` is the fastest method, followed by `gomptim`. Despite calculating ten models instead of a single one in each run, the `goptim` method does not run ten times as

long. This might be because the precomputed starting values for a and b allow for a quick convergence. `DEoptim` is by far the slowest method.

All models except the log-normal-Gompertz are estimated in a matter of minutes up to half an hour. The log-normal-Gompertz model computations take hours. The objective function includes an integral which has to be numerically calculated in order to get the log-likelihood. This is a likely cause for the slow performance.

Convergence

`DEoptim` is the only algorithm which converges for all models; `optim` and `goptim` are not able to estimate the zero-truncated-normal-Gompertz model. `optim` gets stuck at the initial values while `goptim` changes only the a parameter (σ^2 gets stuck at the starting value and b stays at the upper bound).

All methods are able to converge for the gamma-Gompertz, the inverse-Gaussian-Gompertz, and the log-normal-Gompertz model. Most of the time the parameter estimates stay within the bounds. The frequency of the estimates being right at the bounds changes by method with `optim` achieving a low count of outliers and `DEoptim` hitting the wall more often.

Model Comparison

Differences in parameter estimates

For the gamma-Gompertz model there are no meaningful differences among parameter estimates by each method. Most differences are in the magnitude of 1E-04 or smaller (see Figure 1). For the inverse-Gaussian-Gompertz model the estimates for the parameter a are stable across methods. For b and the frailty parameter, however, a significant number of estimates differ by more than 0.01 units in different methods. The log-normal-Gompertz model yields the biggest estimation differences. For all three parameters a large amount of

estimates differs significantly across methods. Note that estimation differences for the zero-truncated-normal-Gompertz model could not be computed because only `DEoptim` converged.

Overall there are smaller differences in parameter estimates between `DEoptim` and `goptim` than between `optim` and the former two methods.

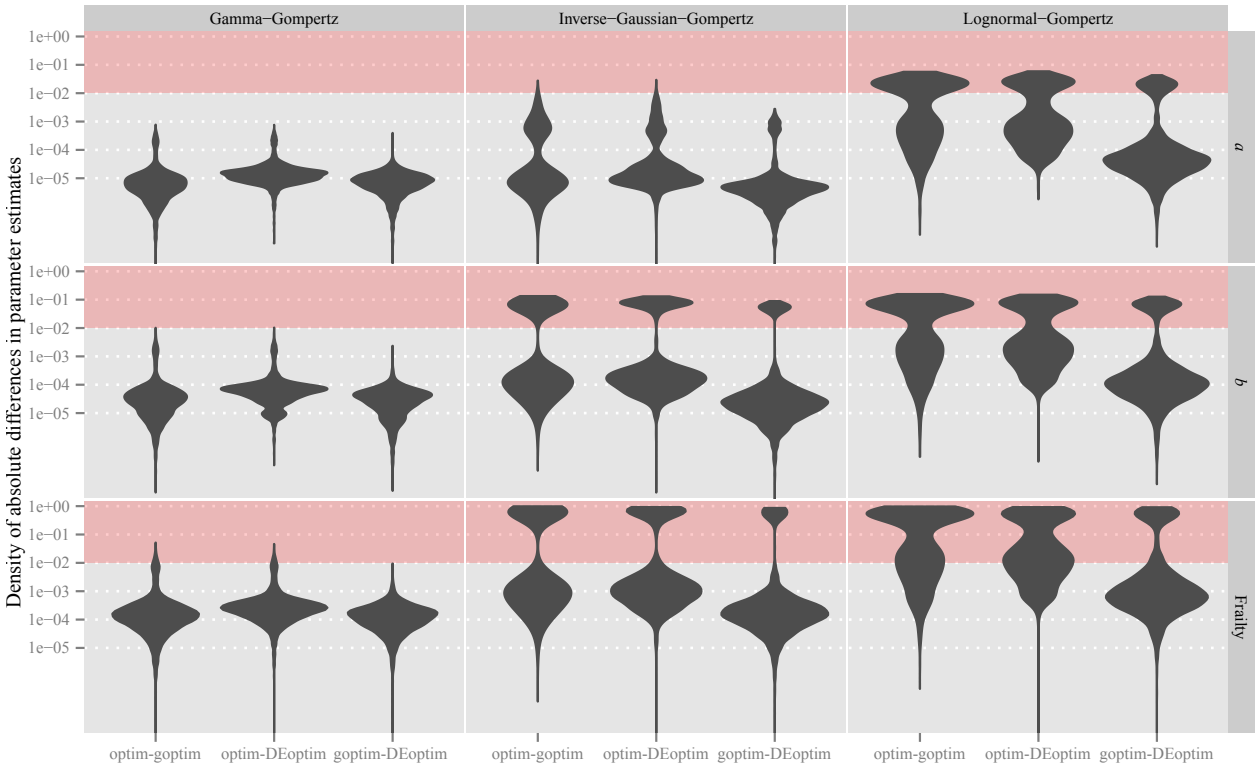

Figure 1: Density of absolute differences in parameter estimates between estimation methods

Comparing the distributions of the frailty parameter estimates by model shows a tendency of the `DEoptim` and `goptim` methods to produce higher frailty estimates compared with `optim` (see Figure 2). For the inverse-Gaussian-Gompertz and log-normal-Gompertz models not only is the median of the frailty estimates higher for these methods, but also a significant share of estimates for the frailty parameter are close to the upper boundary.

Different parameter estimates imply differences in the log-likelihood's maximum that the optimization methods found. To answer the question which estimation method to *trust*, we compare the differences in model AIC values for different methods: a smaller AIC implies a better model fit. Figure 3 shows the densities of the model AIC differences by method.

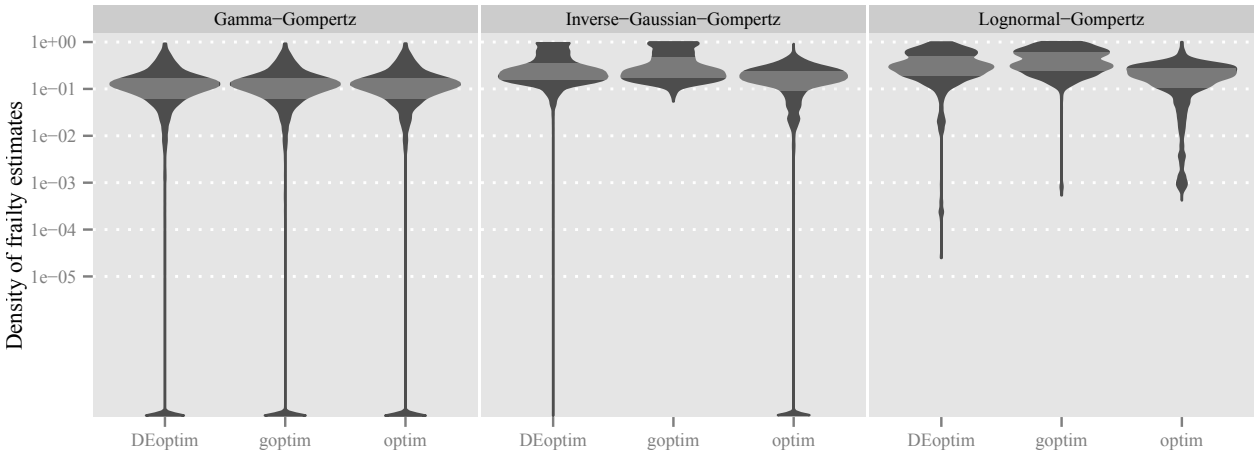

Figure 2: Density of frailty estimates by estimation method

The AICs of the estimated gamma-Gompertz models do not differ substantially depending on the optimization method. The differences in model fit follow a normal distribution with a median at a magnitude of about 1E-04. As previously, things look different when considering the mathematically more complicated inverse-Gaussian-Gompertz and especially the log-normal-Gompertz model. Here we see larger differences between estimation methods. While a large amount of differences still happens around the magnitude of 1E-04 the distribution exhibits a second local peak at around 1E+00. This *compound distribution* of differences in AIC between method reveals that for the inverse-Gaussian and log-normal-Gompertz models, different methods produce either normally distributed small differences or normally distributed large differences in AIC – no method produces vastly better fits in all cases. Regardless of the model the `DEoptim` method produces the smallest AIC in nearly all cases while the `optim` method produces the largest AICs.

Among the three applied estimation methods `DEoptim` was the only one able to handle the zero-truncated-normal-Gompertz model. It also consistently found the highest log-likelihood value. From a theoretical standpoint `DEoptim` is especially suited to maximize likelihood functions which are not smooth or might have multiple local maxima (see Mullen et al. 2011). While the gamma-Gompertz model does not produce such a likelihood function, the log-normal-Gompertz and the zero-truncated-normal-Gompertz models might do. On the

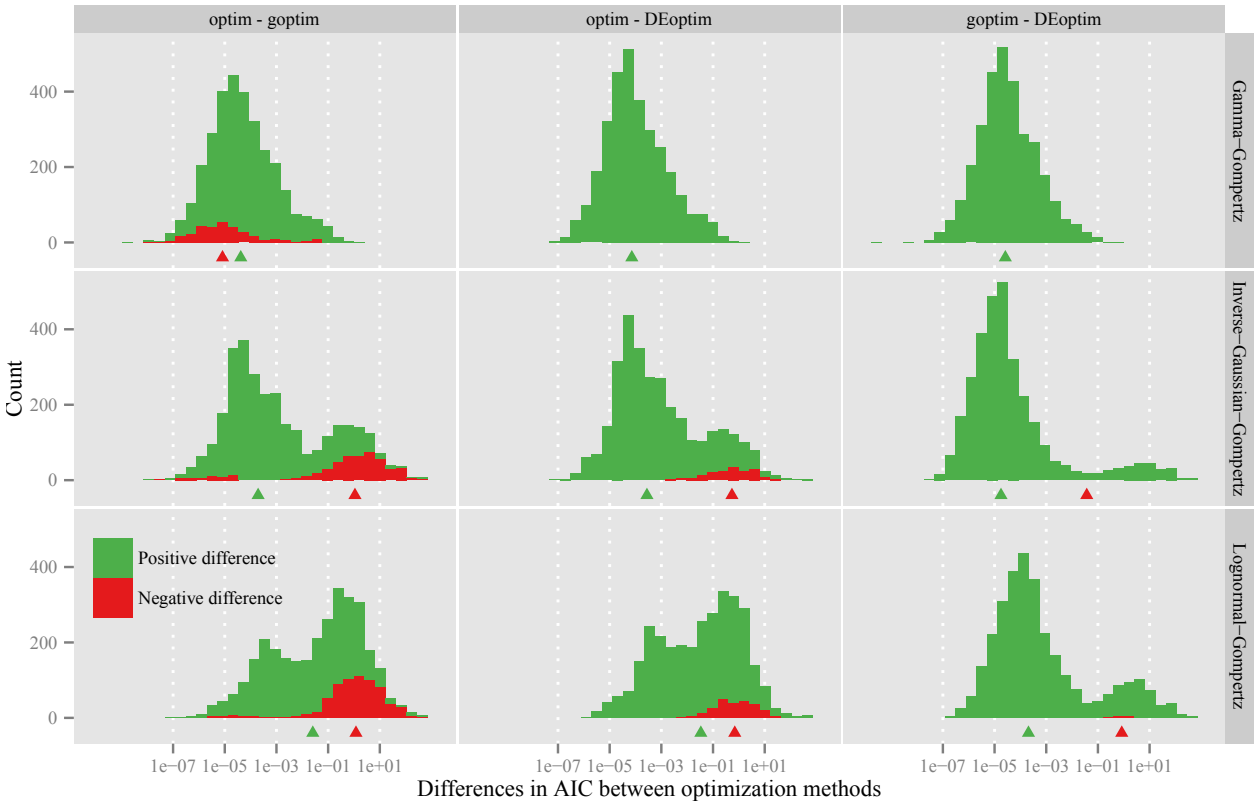

Figure 3: Density of differences in model AIC between estimation methods

other hand the `DEoptim` method produces wildly varying frailty estimates over time – a characteristic which is not so well detected by `optim`. Keeping that in mind, `DEoptim` is the only option if one seeks to optimize complicated likelihood functions.

Conclusion

Missov and Finkelstein (2011) derive a formal criterion to check whether a frailty (or, what is equivalent, *mixing*) distribution is plausible (admissible) in a fixed-frailty model. The gamma distribution, a popular mixing distribution in frailty models, as well as the zero-truncated-normal are among the *admissible distributions*. The inverse Gaussian and the log-normal are not admissible within this framework. We compare admissible and non-admissible frailty distributions by fitting the corresponding frailty models on mortality data from HMD (2014) and comparing the fit of the different models. The relative likelihood

derived from the AIC is suited for comparing the fit of non-nested models: the AICs of all models are compared to the lowest AIC of all models, and a probability that model i minimizes the estimated information loss can be derived. The resulting statistic is called *Akaike weight* (see Burnham and Anderson 2004).

For the `optim`-based optimization methods the Gamma-Gompertz frailty model produces the best model fit in roughly $^2/_3$ of all cases (see table 2). However, including the zero-truncated-normal-Gompertz model in the "competition" by applying the `DEoptim` optimization algorithm changes the picture. In 42 % of all cases estimated with `DEoptim`, the zero-truncated-normal-Gompertz model produced the smallest AIC, leaving all other competing models behind with gamma-Gompertz coming in second. Regardless of the estimation method, the inverse-Gaussian-Gompertz model is the least likely to have the best model fit.

The zero-truncated-normal-Gompertz model performs especially well on cohort data whereas the gamma-Gompertz model has the strongest explanatory power when applied to period data. There are strong geographic effects on the relative differences of the model fit. In Sweden and Denmark the best models are for the most part only marginally better than all other competing models. This might be an artifact of the relatively small population sizes of these countries. In France, Italy and Japan the Akaike weights of the winning models are more equally distributed across the range of possible values. While the gamma-Gompertz model is not the overall winner, it does produce the most Akaike weights near the value of 1, so if a model wins by a high margin, it is most likely the gamma-Gompertz.

Our preliminary results show that fixed-frailty models with admissible (according to the criterion by Missov and Finkelstein 2011) frailty distributions (and a Gompertz baseline) are the ones that provide the best model fit for the majority (between 60.5% and 74%) of the high-quality human-mortality datasets we study.

	DEoptim		goptim		optim	
Gamma-Gompertz	987	*31.7*	1886	*60.5*	2137	*68.6*
Inverse-Gaussian-Gompertz	224	*7.2*	533	*17.1*	473	*15.2*
Lognormal-Gompertz	587	*18.8*	698	*22.4*	507	*16.3*
Zero-truncated-normal-Gompertz	1319	*42.3*	.	.	.	.

Absolute counts and column percent.
The zero-truncated-normal model was not estimated with the `goptim` and `optim` `methods`.

Table 2: Count of minimal AIC among competitive models estimated with `DEoptim`

Acknowledgements

This research is funded by the Max Planck Society.

References

Aalen, O.O. 1988. "Heterogeneity in Survival Analysis." *Statistics in Medicine* 7:1121–1137.

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

Burnham, K.P. and D.R. Anderson. 2004. "Multimodel inference: Understanding AIC and BIC in model selection." *Sociol. Meth. Res.* 33:261–304.

Dalzell, Catherine. 2013. "Optimization in R. A conversation with John Nash about optim and optimx." http://www.ibm.com/developerworks/library/ba-optimR-john-nash/index.html.

Gampe, J. 2010. "Human Mortality Beyond Age 110." In *Supercentenarians*, edited by H. Maier, J. Gampe, B. Jeune, J.-M. Robine, and J.W. Vaupel, number 7 in Demographic Research Monographs, chapter III, pp. 219–230. Heidelberg [et al.]: Springer.

HMD. 2014. "The Human Mortality Database." http://www.mortality.org/.

Hougaard, P. 1984. "Life table methods for heterogeneous populations: Distributions describing the heterogeneity." *Biometrika* 71:75–83.

Lenart, A. and T.I. Missov. 2014. "Goodness-of-fit tests for the Gompertz distribution." *Communications in Statistics - Theory and Methods (in press)* .

McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Model*. Monographs on Statistics Applied Probability. London: Chapman & Hall, 2nd edition.

Missov, T.I. and M. Finkelstein. 2011. "Admissible Mixing Distributions for a General Class of Mixture Survival Models with Known Asymptotics." *Theoretical Population Biology* 80:64–70.

Missov, T.I. and A. Lenart. 2013. "Gompertz-Makeham Life Expectancies: Expressions and Applications." *Theoretical Population Biology* 90:29–35.

Missov, T.I., A. Lenart, L. Nemeth, V. Canudas-Romo, and J.W. Vaupel. 2014. "The Gompertz Force of Mortality in Terms of the Modal Age of Death." *Demographic Research (under review)* .

Missov, T.I. and J.W. Vaupel. 2015. "Mortality Implications of Mortality Plateaus." *SIAM Review* 57.

Mullen, K.M., D. Ardia, D.L. Gil, D. Windover, and J. Cline. 2011. "DEoptim: An R Package for Global Optimization by Differential Evolution." *Journal of Statistical Software* 40:1–26.

Vaupel, J.W., K.G. Manton, and E. Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16:439–454.

Vaupel, J.W. and T.I. Missov. 2014. "Unobserved Population Heterogeneity: A Review of Formal Relationships." *Demographic Research* 31:659–686.

Vaupel, J.W. and A.I. Yashin. 2006. "Unobserved Population Heterogeneity." In *Demography: analysis and synthesis; a treatise in population studies, volume 1*, edited by G. Caselli, J. Vallin, and G. Wunsch, chapter 21, pp. 271–278. London: Academic Press.

Yashin, A.I., A.I. Iachine, and A.S. Begun. 2000. "Mortality Modelling: A Review." *Mathematical Population Studies* 8:305–332.

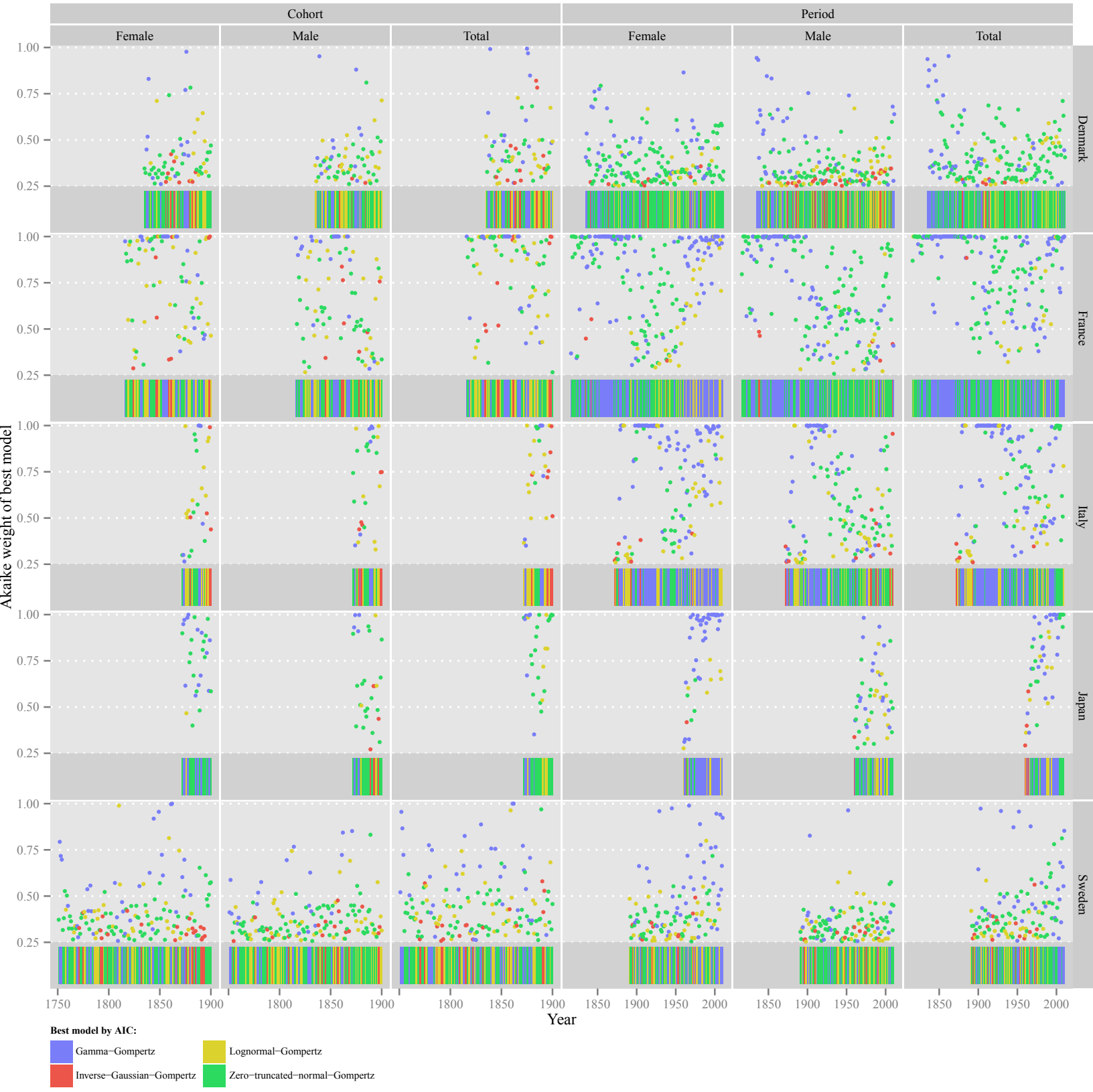

Figure 4: Akaike weights of winning models estimated with `DEoptim`