Can We Verify the Existence of a Human Mortality Plateau Today or in the Near Future?

László Németh[1], Maciej J. Dańko[1], and Trifon I. Missov[1,2]

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

[2]Mathematical Demography, University of Rostock, Ulmenstr. 69, 18057 Rostock, Germany

Preliminary Draft

Please do not cite without authors' permission

Abstract

The article studies three mixture (frailty) models that incorporate a Gompertz or Gompertz-Makeham baseline mortality: the proportional hazards (PH), the accelerated failure time (AFT), and the digamma-Gompertz model (DGG). Mortality plateaus are generated within the PH framework, while the force of mortality in the AFT and DGG settings has an inverse-bathtub shape. The article aims to answer if it is possible to verify the existence of a mortality plateau, i.e. if the goodness-of-fit of the PH model is better than the one of AFT and DGG, given current mortality data.

1 Introduction and research questions

Adult human-mortality is usually described by an increasing mortality risk. The most frequently used exponentially increasing hazard was suggested by Gompertz [7] and given as

$$\mu_G(x) = ae^{bx}\,.$$

Parameter a denotes the level of mortality at the starting adult age and b, equal to the relative derivative of $\mu\,(x)$, captures the rate of aging.

Vaupel et al. [21] introduced the concept of frailty to incorporate unobserved heterogeneity and resulting mortality models are known as relative risks models [21; 3] or proprotional hazards (PH). For a frailty value z, which acts on the starting level of mortality, the hazard is given by

$$\mu_{PH}(x \,|\, z) = zae^{bx}\,.$$

Vaupel and Yashin [20] showed that neglecting unobserved heterogeneity can lead to erroneous conclusions regarding the underlying mechanisms that generate the data. Frailer individuals tend to die first and more robust individuals survive longer, which leads to mortality deceleration and eventually to a mortality plateau at higher ages [21]. Empirical evidence by Gampe [5] also suggests that there might be a plateau after age 110.

A computationally more difficult alternative modeling framework assumes that the population is heterogeneous with respect to b, the rate of aging [9; 6]. In case of an accelerated failure time (AFT) framework the force of mortality for an invididual is described by

$$\mu_{AFT}(x \,|\, z) = zae^{zbx}\,.$$

The definition of the hazard suggests that frailty acts on the baseline hazard not only proportionally, but also on the time scale. It is possible that two separate and independent

frailty values modulate the starting level of mortality and the rate of ageing simultaneously. Such a mortality model is described by:

$$\mu_{D\Gamma G}(x \mid z_a, z_b) = z_b z_a a e^{z_b b x}\,. \tag{1}$$

For empirical and theoretical reasons frailty is usually considered to be gamma-distributed with an expected value of 1 and a squared coefficient of variation γ, i.e., $Z \sim \Gamma\left(1/\gamma, 1/\gamma\right)$. In this paper we assume that frailties, Z_a and Z_b are two independent gamma-distributed random variables and hence we are going to refer to (1) as the digamma-Gompertz model (DΓG).

Since frailty is unobserved, scientists estimate model parameters with the help of the hazards at the population level derived from the individual level risks. Figure 1 presents a set of logarithmic population hazard functions corresponding to the above discussed models. The logarithmic hazard generated by a PH model slows down from the linear increase to a constant level. Both AFT and DΓG models are characterized by a stronger curvature than a PH, and observed mortality decreases after a certain age.

Finkelstein and Esaulova [4] and Missov and Finkelstein [13] discussed that heterogeneity in the rate of ageing cannot produce a mortality plateau. The existence of the mortality plateau is yet to be proven. If the accelerated failure time framework can describe human mortality better than the proportional hazards, that could serve as an argument against the existence of a mortality plateau.

For this reason we check the behaviour of the previously mentioned models on simulated data whether it is possible to statistically identify each model. We simulate data generated by each of the frameworks, fit the three models to the same datasets and assess the arising bias in the estimated parameters. As a next step, we evaluate the identifiability performance with the help of Akaike's information criterion [1] and the bayesian information criterion for different sample sizes.

Looking for traces of accelerated failure time processes, we selected countries from the Human Mortality Database [8]. However, empirical data does not provide a clear answer which framework generated them, which indicates that the scarce data on supercentenarians currently available is not enough to decide.

In the next section we describe the model fitting method used in this paper and the simulation results. In section 3 we present the fitting conducted on empirical data.

2 Model Fitting and Simulation

The fitting procedure of the models is based on the assumption that death counts $D(x)$ at age x are Poisson-distributed [2], i.e. $D(x) \sim$Poisson$(E(x)\mu(x))$, where $E(x)$ is the exposure at age x and $\mu(x)$ denotes the population hazard. As a result, we maximize a Poisson log-likelihood

$$\ln L = \sum_x \left[D\left(x\right) \ln \mu\left(x\right) - E\left(x\right) \mu\left(x\right) \right].$$

Other fitting procedures can also be applied, but only when their underlying assumptions hold. [For further reference, see [10], and [18], [19] for formal relationships.]

Maximization of the log-likelihood was carried out by applying differential evolution [17] using the `DEoptim` R-package [16].

Individual data are, in most of the cases, not available and only data aggregated at the population level can be aquired. However, the population hazard and survival do not bear an analytically closed form solution and must be numerically approximated for AFT and DΓG models.

We simulate individual lifespans from each model framework by inverting the corresponding survival functions. Death counts and exposures are then aggregated age-wise from individual lifespans. Parameter values for the simulation have been chosen in accordance with the range of estimates [see 12; 14; 15] obtained from fitting models from the Gompertz family to period mortality data from the Human Mortality Database [8].

We simulate 500 datasets from each model framework, then we fit every framework to each dataset. Figure 2 summarizes the parameter estimation results. Histograms represent the estimated a, the starting level of mortality, b, the rate of ageing, γ, the heterogeneity in a and δ, the heterogeneity in b, values (columns) for each model (rows).

Values of estimated DΓG parameters show the highest variance, mainly because the two frailties are in the model at the same time. Given a dataset, regardless of the underlying data-generating process, an AFT model always estimates higher a and lower b values than a PH model. This means that according to the AFT framework the standard baseline individual has a higher level of mortality at the starting age and a lower rate of aging in the heterogeneous group than the standard individual in a heterogeneous PH group. It might be counterintuitive that the population hazard has a stronger curvature in an AFT model than in a PH model but the standard individual has a lower rate of aging in the former. The stronger curvature is the consequence of having frailty modulating the rate of aging in the AFT framework.

Histograms in the first row in Figure 2 show the fitting results on simulated PH data. Fitting a PH model to this data captures the true parameter values of the underlying model. In the meanwhile, the AFT framework indicates heterogeneity in b, i.e. positive δ, which results in lower rate of aging and higher a values. Large part of estimated values for the DΓG model are capturing the true parameter values. However, the variation for the DΓG values is high, even showing estimated values similar to those of the AFT.

If the data-generating process is accelerated failure time (second row of Fig. 2), fitting an AFT model estimates the empirical parameter values correctly.

Estimation of a PH model bears bias regarding the estimated parameter values. Parameter a is underestimated, b and γ is overestimated. DΓG identification is again problematic, but the variation is smaller compared to an estimation on PH data. Estimated DΓG values tend to capture the true values, but unreliable if a is small.

Fitting a DΓG model to a DΓG generated dataset (third row of Figure 2) results in huge

variance by the estimation. AFT estimations have a overestimated, b underestimated and δ overestimated. The bias in δ is higher if γ is higher. Estimated PH parameters are also biased, a is underestimated, b is overestimated, γ is overestimated. The bias in γ is higher, when δ is higher.

Another batch of simulations is conducted to investigate how long we have to wait to be able to recognise with high certainty each discussed model framework by various sample sizes. We simulate data and fit models from each framework to data obtained by intervals starting from the highest age, in our case 110, backwards. Figures 3-5 present the results of model selection based on AIC and BIC.

Graphs suggest that more than 20 age groups of data is sufficient to correctly recognise PH or AFT models. Data generated by a DΓG process seems to behave quite differently (Figure 5). The same sample size does not provide the same threshold as it does for PH and AFT. In case of a small sample the dataset is captured either as a PH or an AFT and even increasing the sample size helps only to eliminate incorrect AFT estimations, but not the incorrectly recognised PH models.

3 Human Mortality Data

We selected a few countries from the Human Mortality Database to check if a framework is markedly present if we estimate on empirical data. We chose Japan, England and Wales, France, Italy and Finland, Norway and Sweden taken together. We start the model fitting to period data from age 65, because senescence related mortality starts to be more prevalent instead of the highly influental extrinsic middle-age mortality (Makeham's c [11]). We also fit models to the same data with different starting ages to explore the robustness of each framework. Figure 6 shows the parsimonious models for the above mentioned countries selected by AIC.

There are a very limited number of cases for which in a given year the same model is

chosen as the best one. Majority of the models starting from ages 75 and 80 indicates that the parsimonious model is PH and these are also seems to be more robust regarding the time period. However, huge proportion is spread between parsimonious AFT models, which are more dominant if the fitting is started from 65, 70, 85 or 90. According to the figure, females tend to be chosen as a process generated by PH rather than males. Akaike weights support both PH and AFT frameworks. Only a sporadic number of DΓG models were chosen as parsimonious, which is not surprising according to the simulation results.

4 Conclusion

Adult human mortality cannot be captured by a simple exponentially increasing hazard. We compared three alternative model frameworks incorporating frailty, the proportional hazards, the accelerated failure time and the digamma-Gompertz models. Given the fact that the AFT model structure is more complicated, it is hard to give a definitive answer to the question regarding the existence of the mortality plateau.

Another reason is that we haven't reached the point of the plateau yet, so we have to wait for more data to be available on supercentenarians. If a mortality plateau indeed exists, we might have to wait even longer to eventually reach it, since mortality shifts to older ages.

We showed on simulated data that PH and AFT models are identifiable if the data series is long enough and sample size amplifies the identifiability. On the other hand, capturing a DΓG process is very difficult and increasing sample size cannot improve the recognition. We assessed the bias of the parameter estimators arising when the wrong model is fitted. We explored empirical human mortality data from the Human Mortality Database, looking for evindence whether the mortality plateau exists and concluded that we have to wait for more data on supercentenarians in order to be able to give a definitive answer.

References

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest, 1973.

[2] D.R. Brillinger. The natural variability of vital rates and associated statistics. *Biometrics*, pages 693–734, 1986.

[3] L. Duchateau and P. Janssen. *The Frailty Model*. Springer, New York, 2008.

[4] M.S. Finkelstein and V. Esaulova. Asymptotic behavior of a general class of mixture failure rates. *Advances in Applied Probability*, 38:242–262, 2006.

[5] J. Gampe. Human mortality beyond age 110. In H. Maier, J. Gampe, B. Jeune, J.-M. Robine, and J.W. Vaupel, editors, *Supercentenarians*, number 7 in Demographic Research Monographs, chapter III, pages 219–230. Heidelberg [et al.]: Springer, 2010.

[6] Joshua R. Goldstein and Thomas Cassidy. How slowing senescence translates into longer life expectancy. *Population Studies*, 66(1):29–37, 2012. doi: 10.1080/00324728.2011.635215.

[7] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–583, 1825.

[8] HMD. The human mortality database. http://www.mortality.org/, 2014.

[9] John D. Kalbfleisch and Ross L. Prentice. *The statistical analysis of failure time data*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2002. ISBN 0-471-36357-X. doi: 10.1002/9781118032985. URL `http://dx.doi.org/10.1002/9781118032985`.

[10] A. Lenart. The moments of the Gompertz distribution and maximum likelihood estimation of its parameters. *Scandinavian Actuarial Journal*, 2014(3):255–277, 2014.

[11] W.M. Makeham. On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries*, 8:301–310, 1860.

[12] T.I. Missov. Gamma-Gompertz life expectancy at birth. *Demographic Research*, 28(9): 259–270, 2013.

[13] T.I. Missov and M. Finkelstein. Admissible mixing distributions for a general class of mixture survival models with known asymptotics. *Theoretical Population Biology*, 80 (1):64–70, 2011.

[14] T.I. Missov and A. Lenart. Gompertz-Makeham life expectancies: Expressions and applications. *Theoretical Population Biology*, 90:29–35, 2013.

[15] T.I. Missov, A. Lenart, L. Nemeth, V. Canudas-Romo, and J.W. Vaupel. The Gompertz force of mortality in terms of the modal age of death. *Demographic Research (forthcoming)*, 2015.

[16] K.M. Mullen, D.L. Gil D. Ardia, D. Windover, and J. Cline. DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6): 1–26, 2011.

[17] R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[18] J.W. Vaupel. Life expectancy at current rates vs current conditions: A reflexion stimulated by Bongaarts and Feeney's "How long do we live?". *Demographic Research*, 7: 365–378, 2002.

[19] J.W. Vaupel and T.I. Missov. Unobserved population heterogeneity: A review of formal relationships. *Demographic Research (in press)*, 2014.

[20] J.W. Vaupel and A.I. Yashin. Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician*, 39(3):176–185, 1985.

[21] J.W. Vaupel, K.G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–454, 1979.

Figures

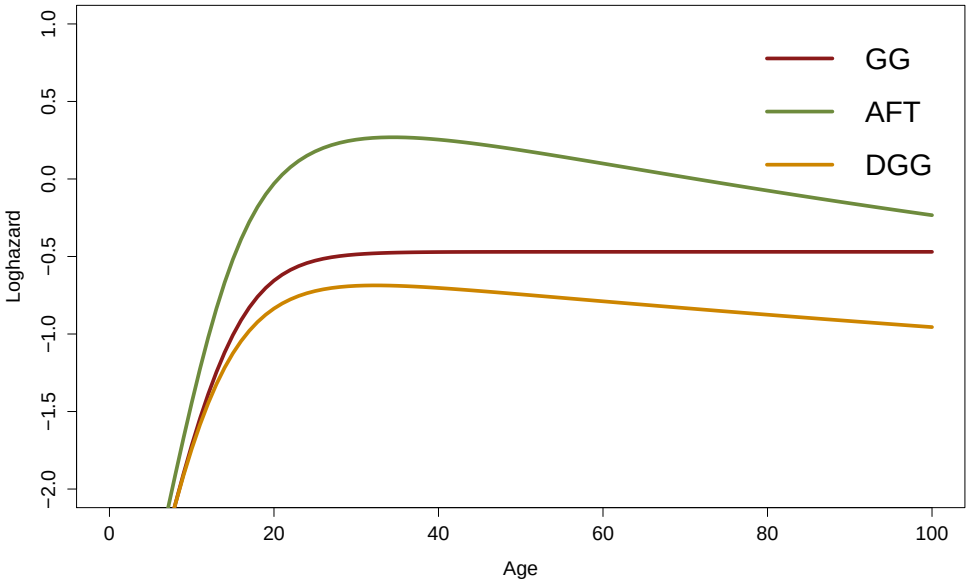

Figure 1: Population hazards generated by PH, AFT and DΓG models. Parameters corresponding to all models are: $a = 0.2$, $b = 0.25$, γ=0.4, $\sigma_b = 0.1$

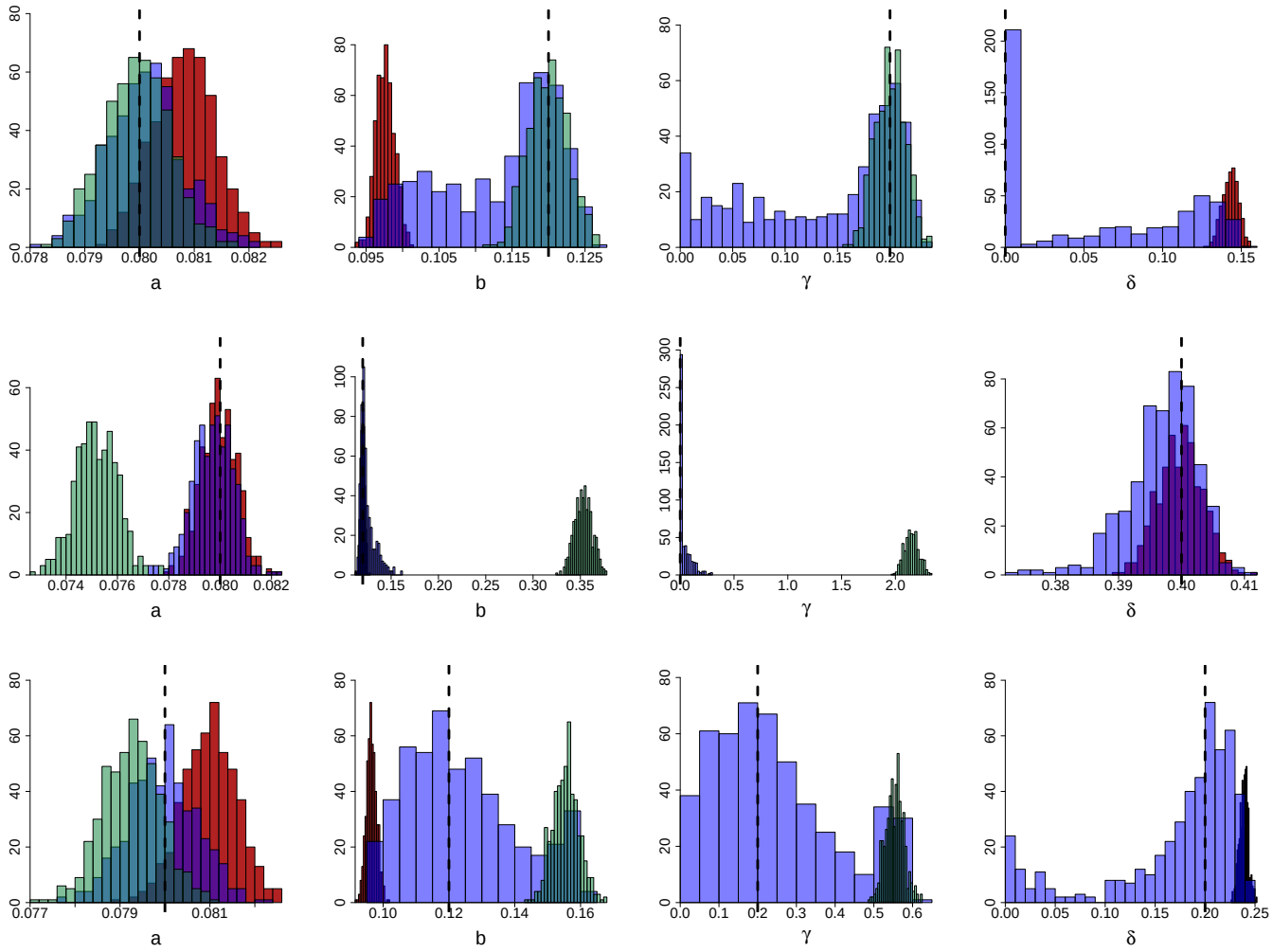

Figure 2: Histograms of estimated a, b and γ of the PH model (green bars), estimated a, b and δ of the AFT model (red bars) and estimated a, b, γ and δ of the DΓG model (blue bars). The PH, AFT, DΓG models were fitted to simulated data generated by a PH model (first row) with parameters $a = 0.08$, $b = 0.14$ and $\gamma = 0.2$; fitted to simulated data generated by an AFT model (second row) with parameters $a = 0.08$, $b = 0.12$ and $\delta = 0.4$; and fitted to simulated data generated by a DΓG model (third row) with parameters $a = 0.08$, $b = 0.12$, $\gamma = 0.2$ and $\delta = 0.2$. The true value of each parameter is designated by a dashed line.

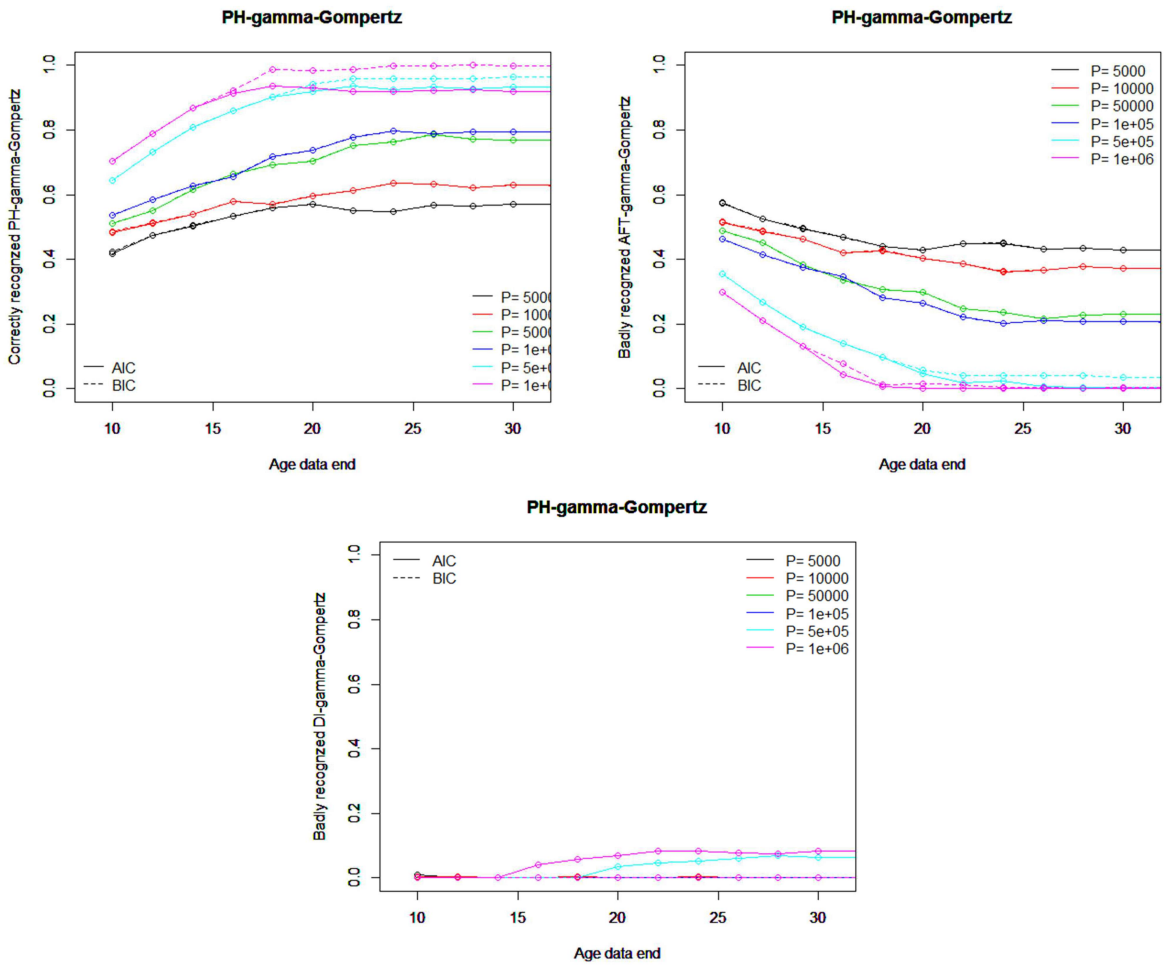

Figure 3: Proportion of correctly recognized PH and incorrectly recognized AFT and DΓG models as a function of data length. The PH, AFT, DΓG models were fitted to simulated data generated by a PH model with parameters $a = 0.08$, $b = 0.14$ and $\gamma = 0.2$. Model selection is based on AIC.

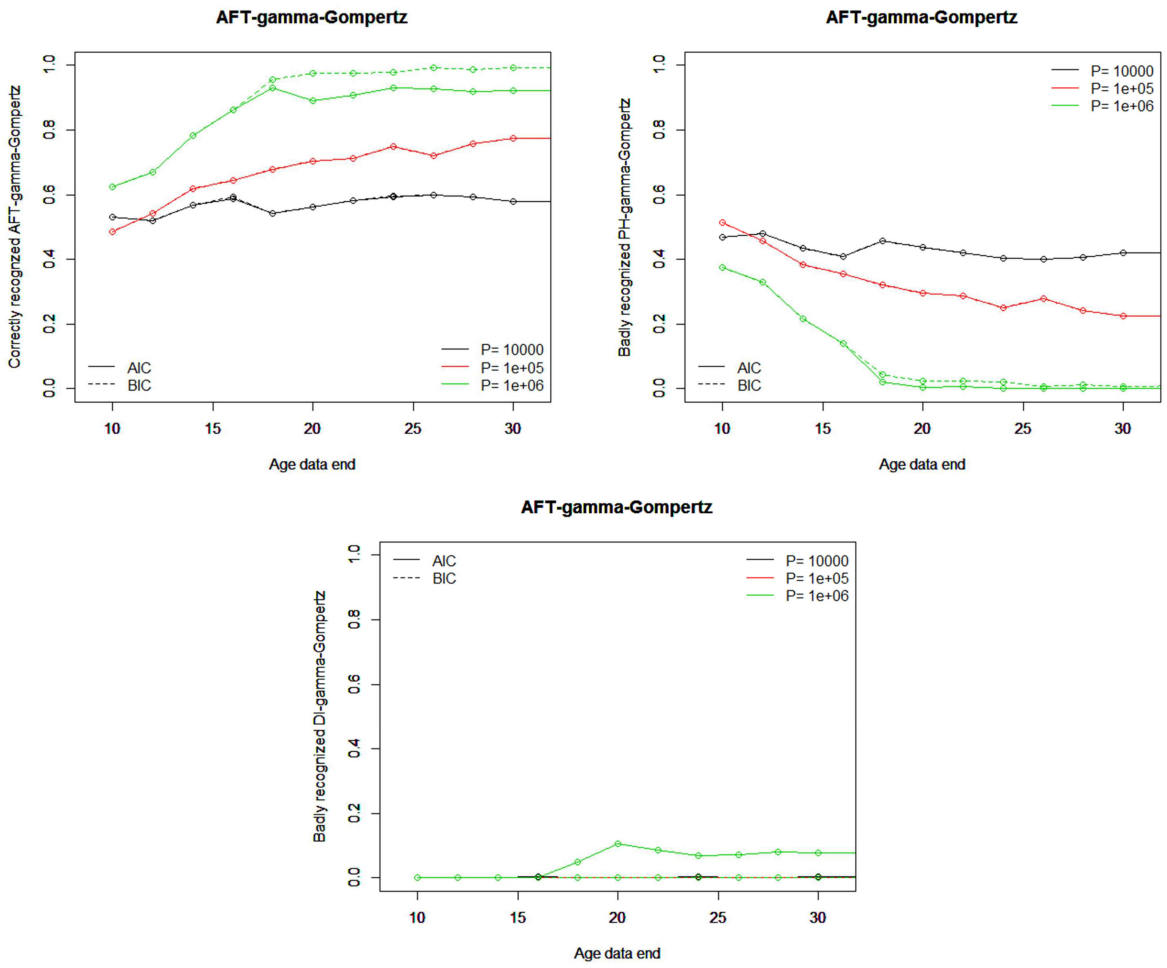

Figure 4: Proportion of correctly recognized PH and incorrectly recognized AFT and DΓG models as a function of data length. The PH, AFT, DΓG models were fitted to simulated data generated by a PH model with parameters $a = 0.08$, $b = 0.115$ and $\delta = 0.13$. Model selection is based on AIC.

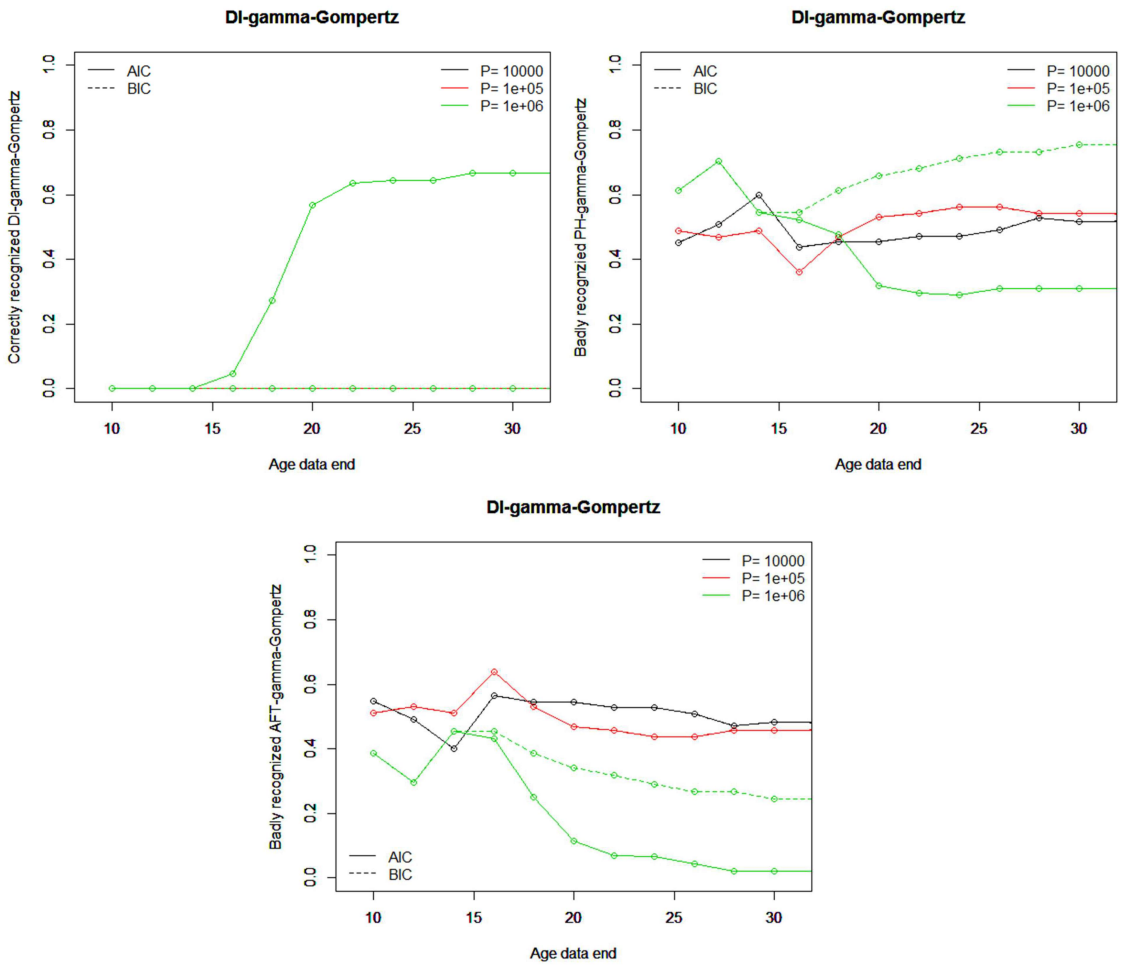

Figure 5: Proportion of correctly recognized PH and incorrectly recognized AFT and DΓG models as a function of data length. The PH, AFT, DΓG models were fitted to simulated data generated by a DΓG model with parameters $a = 0.08$, $b = 0.128$, $\gamma = 0.1$ and $\delta = 0.13$. Model selection is based on AIC.

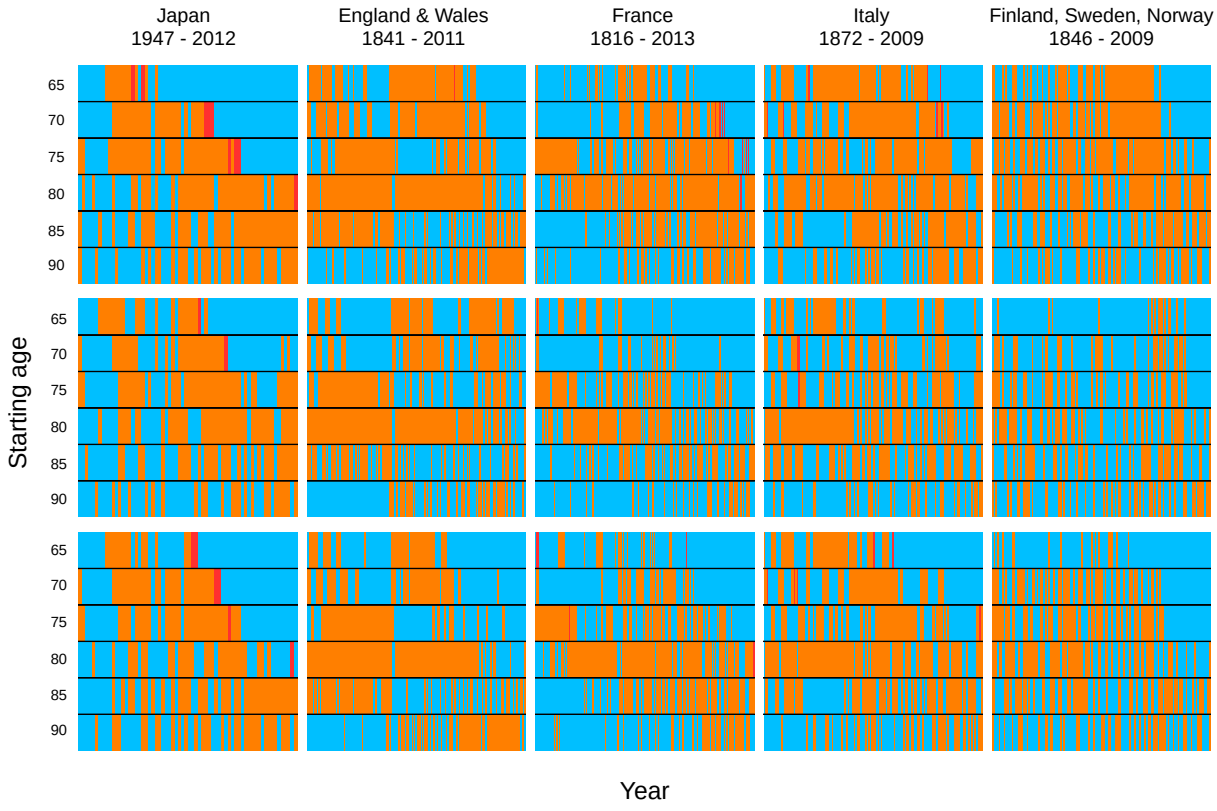

Figure 6: Parsimonious model selection according to AIC. PH model is designated by orange, AFT by blue, DΓG by red. Period data from HMD, first row for females, second for males and third for total population.