How Wrong Could Parameter Estimates Be?

Statistical Consequences of Fitting the Wrong Model to Human Mortality Data

Extended Abstract

Trifon I. Missov[1,2] and Laszlo Nemeth[1]

[1]Max Planck Institute for Demographic Research, Rostock, Germany
[2]Institute of Sociology and Demography, University of Rostock, Germany

November 15, 2013

Abstract

In this paper we check for and quantify the systematic bias regarding mortality parameter estimates when the fitted model is misspecified. We focus on the Gompertz, Gompertz-Makeham, gamma-Gompertz, gamma-Gompertz-Makeham models, which are commonly applied when studying adult human mortality. We simulate lifespans from each of these models, calculate the corresponding age-specific death counts and exposures, and fit the correct and a set of misspecified models to each resulting dataset. We check whether or not the known parameters are captured correctly by the estimation procedure and quantify the bias if it exists.

Introduction and Motivation

The Gompertz (G) model [1] is characterized by an exponentially increasing hazard function

$$\mu(x) = ae^{bx}\,. \tag{1}$$

In this model setting parameters a (the level of mortality at the starting age) and b (the rate of aging) describe the aging mechanism and are to be estimated. The Gompertz-Makeham (GM) model [2] contains, apart from a and b, a third parameter c, which captures extrinsic (aging-independent) mortality. Its force of mortality is given by

$$\mu(x) = ae^{bx} + c. \tag{2}$$

If populations are, in addition, stratified according to a measure of individual susceptibility (frailty), then we arrive at models for heterogeneous populations (see, for example, [3]). When frailty is gamma-distributed and acts multiplicatively on the baseline Gompertz hazard (see [4]), the resulting model is known as the gamma-Gompertz (ΓG) and has the following hazard:

$$\mu(x) = \frac{a\exp(bx)}{1+\gamma\frac{a}{b}\left(\exp(bx)-1\right)}\,. \tag{3}$$

The additional parameter γ measures the coefficient of variation of the frailty distribution. If frailty affects the aging process (captured by the Gompertz), but does not influence extrinsic mortality (Makeham's c), the model is known as the gamma-Gompertz-Makeham (ΓGM) characterized by a hazard function

$$\mu(x) = \frac{a\exp(bx)}{1+\gamma\frac{a}{b}\left(\exp(bx)-1\right)} + c\,. \tag{4}$$

To our knowledge models (1-4) have so far been estimated without applying any goodness-of-fit tests for them first. The design of such tests, in particular for the Gompertz model, has emerged just recently (see [5]). Our goal is to quantify the bias if a misspecified model is fit to a given dataset. E.g., how are estimates of the Gompertz parameters a and b affected if a Gompertz model is fit to a dataset with a non-zero Makeham term.

Methods

We simulate lifespans given a model with known hypothetical parameter values. 250 samples were simulated for chosen parameter values and each dataset contains 5000 individuals. From these data death counts and exposure times are also calculated. Then different models are fitted to each dataset to check if the known parameters are captured by the estimation procedure.

There are more possible methods to fit models to these datasets. First, maximizing the appropriate likelihood function fitted on the lifespan data. Second, maximizing a Poisson likelihood for the death count data. Third, if possible, to fit a linear regression.

Preliminary Results

For a GM-generated dataset, if we compare the fit of a G model with the fit of a GM one, we find that neglecting the Makeham term introduces a bias in the a and b estimates (Figure 1). If we estimate the parameters based on a GG-generated dataset, then there is no evidence that estimates from the GG and the GGM models are statistically distinguishable. However, when we fit a GG to a GGM-generated data, the estimated parameters are biased (Figure 2).

Figure 1: Gompertz and Gompertz-Makeham estimation comparison (parameters: a=0.00002, b=0.14, c=0.001)

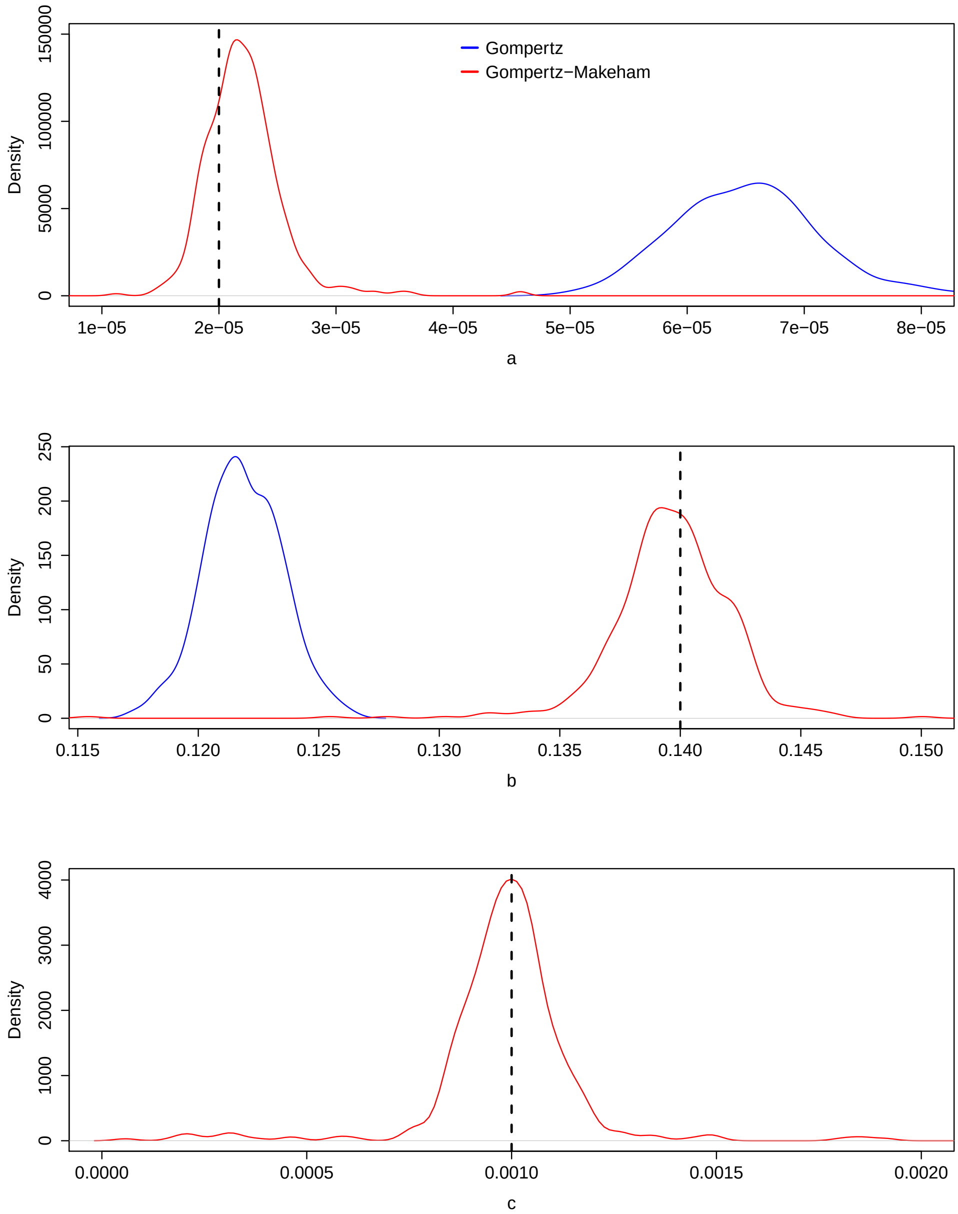
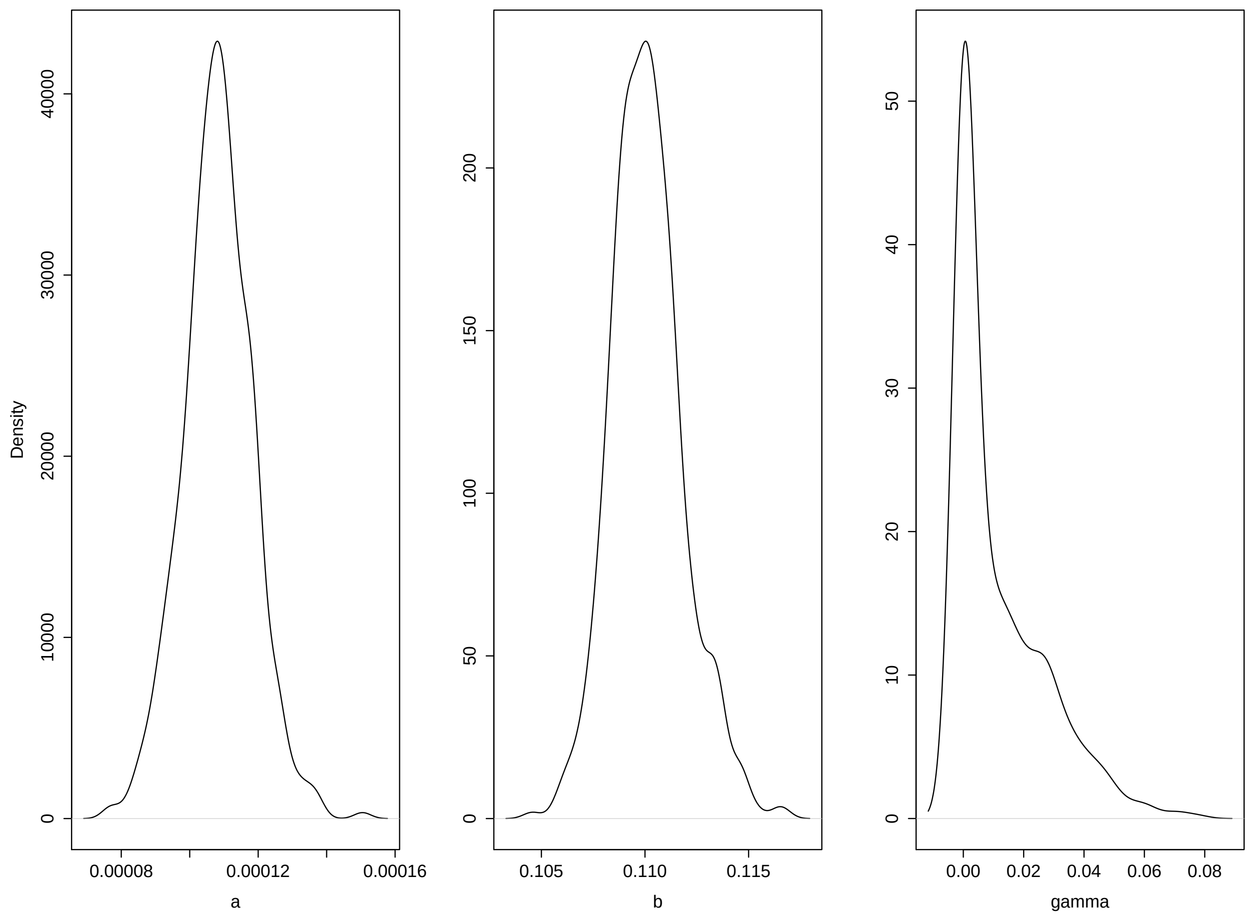


Figure 2: gamma-Gompertz fit (parameters: a=0.00002, b=0.14, gamma=0.02, c=0.001)



Preliminary conclusions

Based on our results we can conclude that the incorrect model specification could result in biased estimation. We need to be aware that omitting seemingly insignificant parts can influence our results in undesirable ways.

References

[1] Gompertz, B. (1825). On the nature of the function expresive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophal Transactions of the Royal Society of London* 115, 513–585.

[2] Makeham, W. M. (1860). On the law of mortality. *Journal of the institute of actuaries* 13, 283–287.

[3] Vaupel, J.W., Missov, T.I. (2013). Population Heterogeneity: A Primer. *Demographic Research* (forthcoming).

[4] Vaupel, J.W., Manton, K., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16, 855–860.

[5] Lenart, A., Missov, T.I. (2013). Goodness-of-fit tests for the Gompertz distribution. *Communications in Statistics - Theory and Methods* (forthcoming).