Six new models for the expectation of life at old age: a comparative review

Dalkhat M. Ediev [1,2,3]

[1] North-Caucasian State Academy, Department of Applied Mathematics and Information Technologies, Stavropolskaya 36, Cherkessk, 369000, Russia.

[2] Lomonosov Moscow State University, Demography Chair (HSMSS), Leninskie Gory 51, r. 752, Moscow, 119992, Russia.

[3] Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Schlossplatz 1, 2361 Laxenburg, Austria.

E-mail: ediev@iiasa.ac.at; dalkhat@hotmail.com

The research leading to these results has received funding from the Russian Foundation for Basic Research under Grant 18-01-00289 "Mathematical models and methods of correcting the distortions of the age structure and mortality rates of elderly population".

Age exaggeration, other deficiencies of population statistics at old age as well as small population numbers preclude extending mortality analysis to advanced old age in many countries. Age heaping (Myers, 1993; Shryock & Siegel, 1973) – an indirect sign of problematic data on age recording – is prevalent world-wide. Surprisingly, it is observed even in such an advanced data collection as the Human Mortality Database (University of California & Max Planck Institute for Demographic Research (Rostock), 2019), where about two percent of entries (raw-data population distributions by age) show accumulation at ages ending with digit '0', see Figure 1 (Whipple (Shryock & Siegel, 1973) index K_{10} above 105, pane to the left). This proportion is, in fact, an underestimate of prevalence of age heaping in the database, because most HMD data refers to inter-censal years with heaping shifted to non-round ages. Assuming censuses are carried out once in about every decade, the true prevalence of heaping may be up to ten times higher. Indeed, Myers (1940) digit preference index exceeds two percent (a critical threshold roughly corresponding, in our data, to K_{10}=105) in nine percent of HMD data (Figure 1, pane to the right). Not surprisingly, HMD protocol assumes extensive data cleaning and adjustment at ages above age 80 years (Wilmoth, Andreev, Jdanov, & Glei, 2007). The age heaping is even more prevalent in data with lower quality, such as data for most African countries (Randall & Coast, 2016). 66 percent of census data for African countries (United Nations Statistics Division, 2019) are cases with heaping index K_{10} above 105, same proportion show digit preference index above two percent (Figure 2). Similar problems apply to many Latin American and Asian populations.

Age exaggeration is particularly problematic for survival analysis, because it leads to apparently elder population age composition, underestimated mortality at old age and overestimated life expectancy at all ages. A typical 'solution' to this problem is to close the life table at relatively young open age interval, so that most of the age exaggeration-related redistributions are confined within the open age. Such a truncation of the age scale is a misfortunate choice in view of rapidly expanding elderly populations (Robine & Caselli, 2005) and extending human lifespan (Ediev, 2011a; Oeppen & Vaupel, 2002; White, 2002).

Apart from being an obstacle in studying elderly, abridging the life table at oldest-old ages leads to another problem: the classical life table relation for the life expectancy in the open age interval

$$e_a = {M_{a+}}^{-1}, \tag{1}$$

is biased when the age composition of the population deviates from that of the (stationary) life table population (here, e_a is the remaining life expectancy at age a, M_{a+} is the aggregate death rate in the open age interval a+).

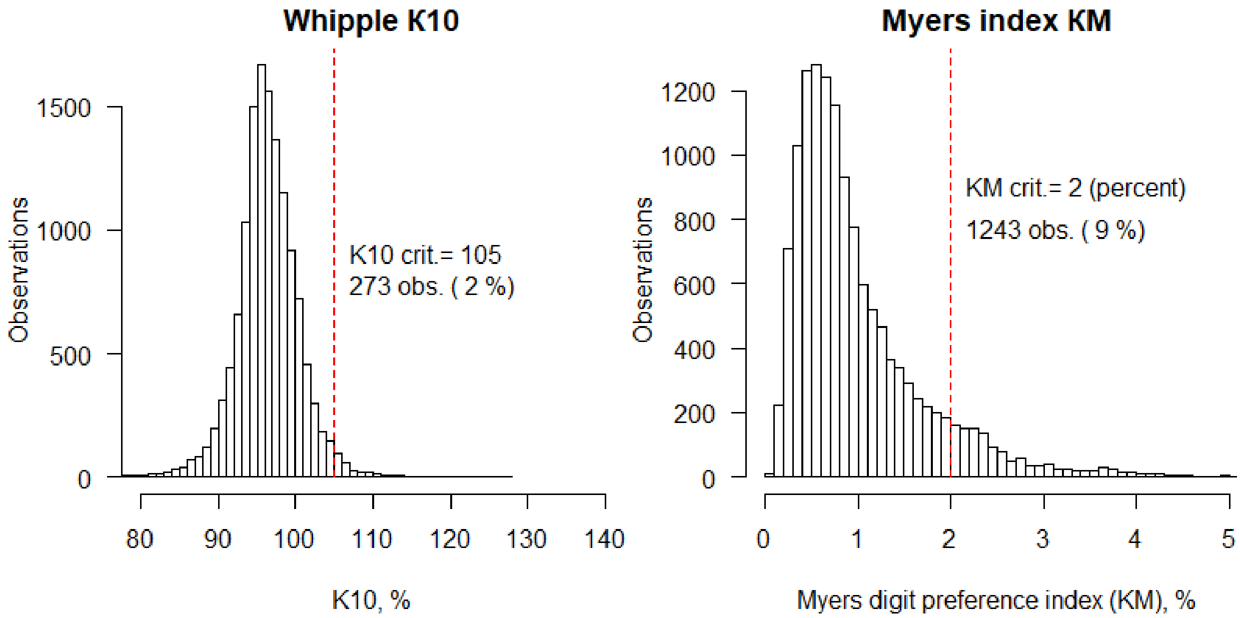


Fig. 1. Distributions of Whipple accumulation index K10 (the pane to the left) and Myers' digit preference index KM (the pane to the right) in HMD data (own calculations based on raw data from HMD (University of California & Max Planck Institute for Demographic Research (Rostock), 2019)).

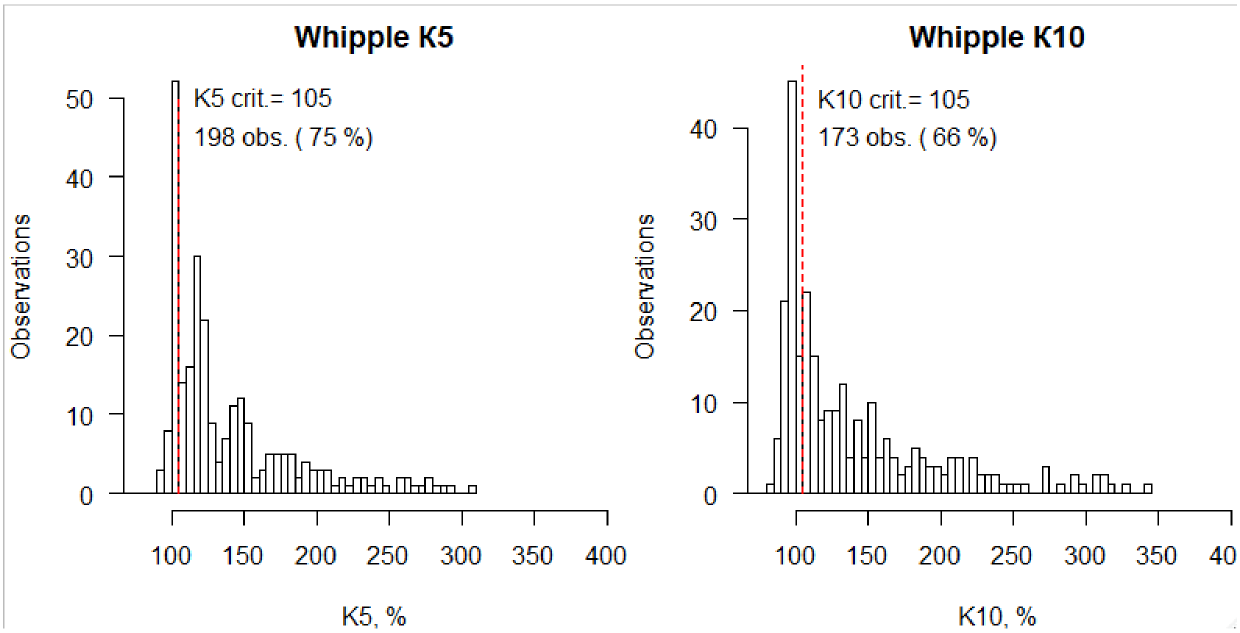


Fig. 2. Distributions of Whipple accumulation index K10 (the pane to the left) and Myers' digit preference index KM (the pane to the right) in Census data for African countries (own calculations based on UN data (United Nations Statistics Division, 2019)).

To mitigate the biasedness of the classical estimate (1), Horiuchi and Coale (Coale, 1985; Horiuchi & Coale, 1982) suggested an adjustment procedure that takes into account population growth and relies on stable population assumption:

$$e_a = M_{a+}{}^{-1} e^{-\beta_a r M_{a+}{}^{-\alpha_a}}, \qquad (2)$$

here, r is the annual growth rate of the population in the open age interval; α_a and β_a are the model parameters (for numerical values, see (Ediev, 2017; Horiuchi & Coale, 1982)). Mitra (1984, 1985) has developed an alternative approximation based on more accurate formal-mathematical account of stable population relations but also involving an additional parameter $\bar{x}$ that stands for the mean age of the population in the open age interval:

$$e_a = {M_{a+}}^{-1} e^{-r\left[{M_{a+}}^{-1} - \left(1 + r{M_{a+}}^{-1}\right)(\bar{x} - a)\right]}. \quad (3)$$

Although reliance on the empirical mean population age $\bar{x}$ in presence of age exaggeration may be questionable (Coale, 1985), this problem may be addressed by discarding the (possibly biased) empirical $\bar{x}$ and replacing it by the following regression approximation:

$$\bar{x} = C + k_1 {M_{a+}}^{-1} + k_2 r {M_{a+}}^{-1}, \quad (4)$$

see (Ediev, 2017) for regression (4) parameters and prediction errors.

A more traditional approach to estimating the death rates and lifespan at old age is extrapolating the death rates into the open age interval by a proper mortality model (e.g., Gompertz or Kannisto (Gompertz, 1825; Heligman & Pollard, 1980; Missov, Németh, & Dańko, 2016; Thatcher, Kannisto, & Vaupel, 1998)) based on their rate of increase at ages below the open age interval. Extrapolation might allow for both obtaining an estimate of the remaining life expectancy in the open age interval that is not affected by age exaggeration and extending the mortality profile into elder ages.

Comparative empirical analysis of accuracy of the traditional approaches shows that Horiuchi-Coale and Mitra models are by far more accurate than the classical life table estimate or the extrapolation method (Ediev, 2018). Mitra model is, on average, more accurate than the Horiuchi-Coale formula; yet, the Mitra model is subject to frequent outliers with large errors that limit its practical usability. Instability of the Mitra model originates from its sensitivity to inadequate estimates of the population growth parameter (Ediev, 2019a), a problem that will be addressed further down in the paper.

The extrapolation method, on the contrary, appeared in tests to be of as low accuracy as the classical estimate (1) with additional disadvantage of instable estimation errors. Indeed, in spite of using the mortality extrapolation to improve life expectancy estimates, it appears to be better to use estimates of remaining life expectancy in the open age interval to constrain and hence improve accuracy of the extrapolation itself (Ediev, 2017). Improved estimates of old-age remaining life expectancy together with the constrained extrapolation of the death rates to elder ages enables obtaining as accurate estimates of old-age mortality with open age interval 60+ as may typically be obtained by traditional unconstrained extrapolation with open age interval 80+. This presents a remarkable possibility of extending mortality statistics for countries, like many in Africa, Latin America and Asia, where data quality is already very low by age 80. The key ingredient of this possible extension of mortality statistics is bettering the estimates of remaining life expectancy.

The Mitra model (3) may be reformulated as:

$$e_a \approx {M_{a+}}^{-1} e^{-r\left({M_{a+}}^{-1} - (\bar{x} - a)\right) + r^2 {M_{a+}}^{-1}(\bar{x} - a)}, \quad (5)$$

which illuminates that the growth parameter *r*, entering the formula in a quadratic form, turns problematic when obtained with a bias or when the population is, in fact, non-stable. This particular representation of the growth parameter comes from approximations used when deriving the formula (Ediev, 2019a) and may not be addressed within the model itself, although some improvements may be achieved by combining the model with other, less accurate yet stable alternatives, such as the classical estimate (Ediev, 2018, 2019a).

Here, we suggest improving the estimates of life expectancy at old age by better accounting for population stability or dropping the stability assumption completely in a set of alternative models described further below.

In the *first two alternative models*, we suggest using the following (exact) relation applicable to a stable population:

$$e_a = {M_{a+}}^{-1} K, \quad (6)$$

where, K is the adjustment coefficient that equals, for a stable population:

$$K = \frac{\int_a^{\omega} d(x)e^{-rx}dx}{\int_a^{\omega} d(x)dx}\frac{\int_a^{\omega} l(x)dx}{\int_a^{\omega} l(x)e^{-rx}dx}, \quad (7)$$

where $d(x)$ and $l(x)$ are life table deaths and survival. These functions may be obtained from the extended life table that, in turn, may be computed using the death rates at advanced old ages extrapolated in the constrained extrapolation method (Ediev, 2017). This, essentially, makes the remaining life expectancy a (calculable) function of the growth rate and the remaining life expectancy itself:

$$e_a = K\big(r, f(e_a, M_{a-1})\big)/M_{a+}. \quad (8)$$

In our first alternative (*the iterative model with exogenous population growth*), we suggest resolving (8) for the remaining life expectancy given a growth rate estimate. This model is similar to models by Horiuhi-Coale and Mitra in relying on population stability and assuming and exogenously assessed growth parameter. Our model is free, however, of approximations made in deriving the traditional estimates by Horiuchi-Coale and Mitra. Instead, we use the constrained extrapolation of the death rates that is shown to yield reliable patterns of the death rates at old age, to better describe the age structure of the stable population.

The idea of the next model, *the iterative model with endogenous population growth,* is to supplement the mortality data (M_{a-1}, M_{a+}) by population data in order to estimate both the life expectancy and the growth rate. To this end, we suggest using the ratio of the population in the open age interval to population in an interval below the open age (both quantities, supposedly, are not affected by age exaggeration):

$$\frac{\boldsymbol{P_{a+}}}{{}_{\Delta}\boldsymbol{P_a}} = \frac{\int_a^{\omega} l(x)e^{-rx}dx}{\int_{a-\Delta}^{a} l(x)e^{-rx}dx} = g\Big(r, K\big(r, f(e_a, M_{a-1})\big)\Big). \quad (9)$$

Equations (8), (9) may numerically be resolved for unknown e_a and r.

The two alternatives have been tested on HMD data and appear to outperform both the classical methods and the Mitra model (Figure 3), the model with endogenous growth parameter performing marginally better.

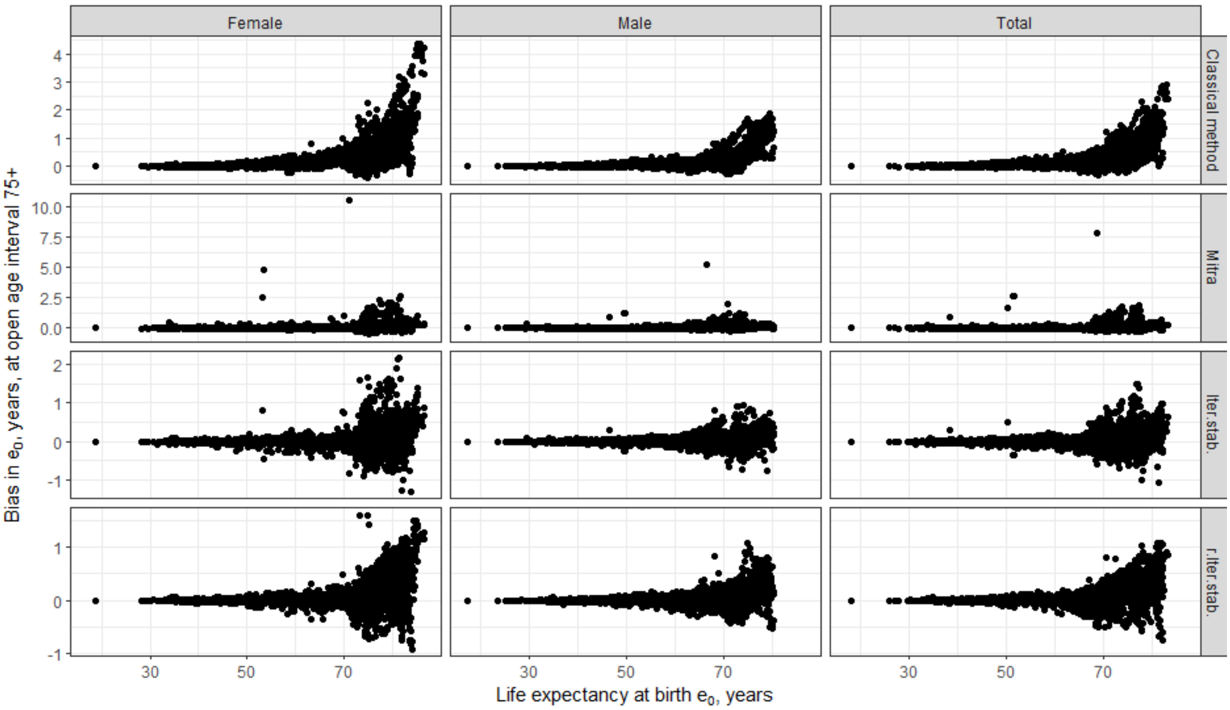


Fig. 3. Life expectancy estimation errors for selected models applied to the open age interval: the classical life table model (the first line), the Mitra model (the second line), the stable population model with exogeneouse population growth parameter based on relation (8) and empirical growth rates, and the stable population model with endogenouse population growth parameter based on Eqs. (8), (9). Open age interval 75+ in all cases.

Our next two alternative models drop the assumption of population stability and, instead, rely on sorts of backward population projection to obtain the population age composition in the open age interval. First, we use the *robust backward projection* (Ediev, 2011b) from cohorts in age groups below the open age interval to elder ages. The method (results not shown in the abstract) appears to be of accuracy similar to that of the stable population model (8)-(9).

Substantially more accurate results are obtained for another non-stable model that relies on population and mortality estimates for earlier years (if available) to project approximate age composition of the population in the open age interval in the given year. In essence, the model builds on the cohort-component projection of the elderly (open age interval) population based on the initial population distribution by age and on the dynamics of the population below the open age interval. For a single-year step, the method is illustrated in Figure 4. Because the death rates are calculated on the basis of mid-year populations, the cohort-component projection step will depend on death rates in both years:

$$\frac{P_{t,x}}{P_{t-1,x-1}} = \frac{L_x}{L_{x-1}} \sim LT\left(\frac{M_{t-1,x}+M_{t,x}}{2}\right) = LT\left(\frac{M_{t-1,x}+f(e_a,M_{a-1})}{2}\right), \tag{10}$$

here, $LT(m_x)$ stands for a function of the life-table constructed from the death rates m_x. All in all, population at time t will be determined by combination of population in the previous year and death rates of the previous and current years, where the latter will be a function of the expectation of life to be estimated:

$$\mathrm{P}(t, x \geq a) \sim \begin{cases} \mathrm{P}(t-1, x \geq a-1) \\ \mathrm{M}(t-1, x \geq a-1) \\ \mathrm{M}(t, x \geq a-1) \sim f(e_a, M_{a-1}) \end{cases} \tag{11}$$

Once the age composition is determined, it may be used to produce the correction coefficient in (6):

$$K = \frac{\int_a^\omega P(x)M(x)dx}{\int_a^\omega d(x)dx} \frac{\int_a^\omega l(x)dx}{\int_a^\omega P(x)dx} \tag{12}$$

that generalizes the stable-population relation (7) to an arbitrary non-stable case. Combining relations above, another recurrent relation may be developed that may be resolved for the remaining life expectancy:

$$e_a = K\big(\mathrm{P}(t-1,x), M(t-1,x), f(e_a, M_{a-1})\big)/\boldsymbol{M_{a+}}. \tag{13}$$

On HMD data, the cohort-component-based approach shows estimation errors of order of magnitude lower as compared to the Mitra or the stable-population models (Figure 5). The model, however, relies on availability of long time series of demographic data, a case not relevant to many countries with deficient demographic data.

The fifth alternative relates – in a regression model – the remaining life expectancy at given age a to the death rate at that age (Ediev, 2019b):

$$ln(e_a) = C + k_1 ln(M_a) + k_2 M_a + k_3 {M_a}^2 + k_4 a + k_5 a^2 + k_6 Sex + k_7 Period + \varepsilon, \tag{14}$$

here, 'Sex' and 'Period' stand for categorical variables representing sex and calendar year, $C, k_1, k_2, \dots, k_7$ are the model parameters, ε is the error term. The method is less accurate than the models suggested above; yet, it is robust and is of accuracy close to (yet, marginally worse than) that of methods by Horiuchi-Coale and Mitra (Figure 6). Hence, it may be recommended when the population stability assumption is violated substantially and, yet, data do not allow applying the more advanced backward-projection based estimation models. An important feature of the regression model is its statistical independence of other alternatives considered here (Ediev, 2019b).

Our last (*behavioral*) model builds upon explicit account for the mechanism of age exaggeration. In the behavioral model, we assume that a fraction α of people above age x_0, randomly chosen, exaggerate their true age by δ years each. Although this assumption simplifies the real-life age misreporting in many ways, the model does reflect the most important aspect of the age exaggeration; more complex patterns of age exaggeration may be represented as a mix of the elementary exaggerations we consider here. For brevity, we will call the introduced age exaggeration model as (x_0, α, δ) model. The model and population/mortality profiles generated by it has been studied formally (Appendix 1). Although it is hard to compare results for the behavioral model to those of other models proposed here, empirical results presented in the Appendix suggest that the behavioral model may be efficient in reconstructing the age exaggeration profiles and the true death rates distorted by the age exaggeration.

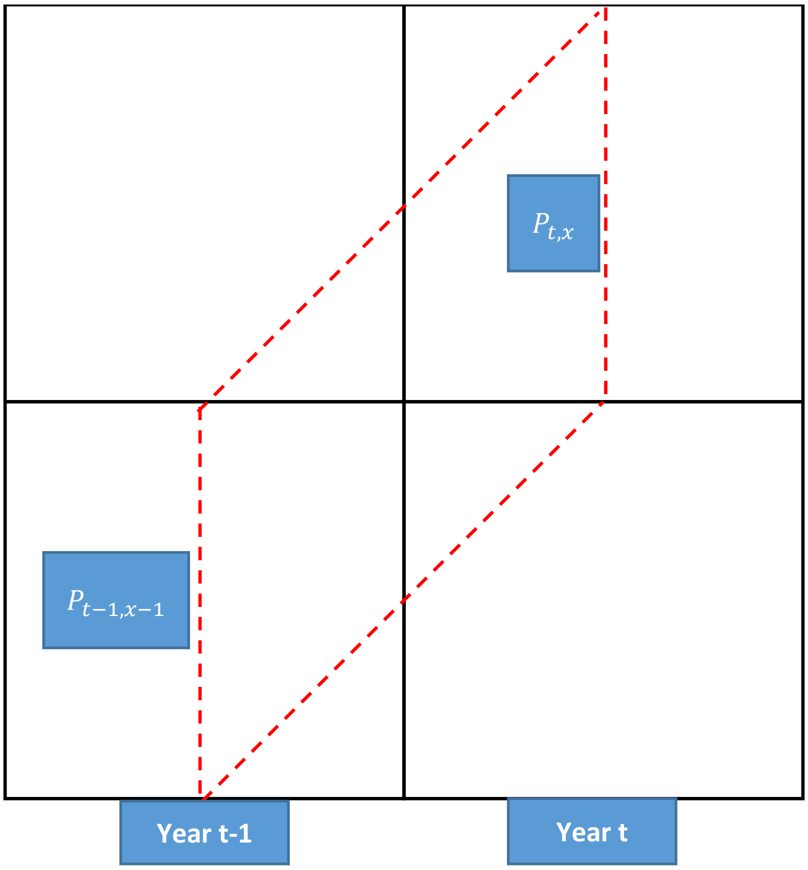


Fig. 4. Schematic illustration to the cohort-component-based method of estimating the life expectancy at old age.

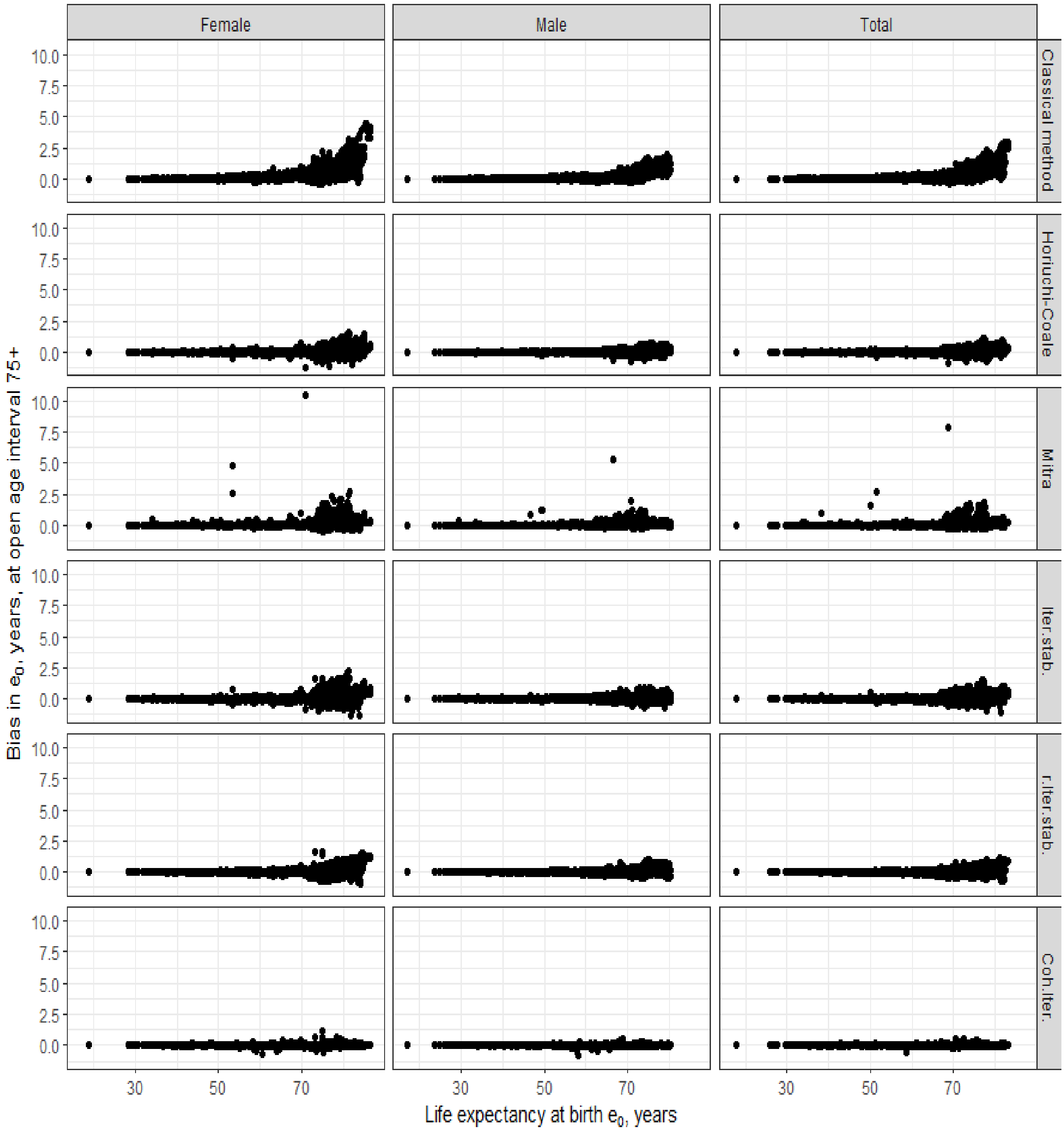


Fig. 5. Life expectancy estimation errors for selected models applied to the open age interval: the classical life table model (the first line), the Mitra model (the second line), the stable population model with exogeneouse population growth parameter based on relation (8) and empirical growth rates (line 4), the stable population model with endogenouse population growth parameter based on Eqs. (8), (9) (line 5), and the cohort-component-based approach (line 6). Open age interval 75+ in all cases.

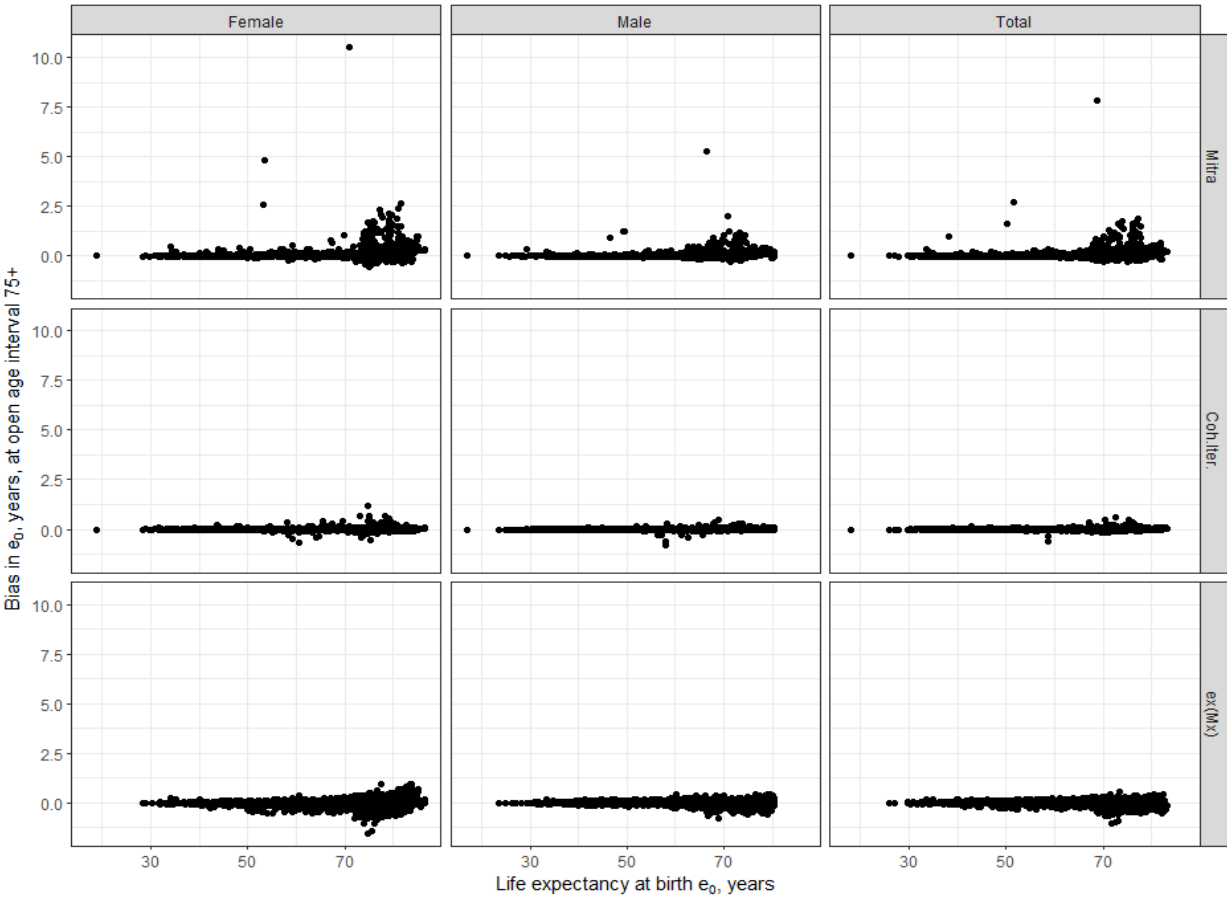


Fig. 6. Life expectancy estimation errors for selected models applied to the open age interval: the Mitra model (the first line), the cohort-component-based approach (the second line), and the regression model (14) (the third line). Open age interval 75+ in all cases.

Conclusions

Although the Horiuchi-Coale and Mitra models do provide a substantial improvement over the two classics (conventional life table and the extrapolation), their reliance on stable population assumption and approximations used pose limits to their practical applicability. The Mitra model, in particular, being the most accurate among traditional models, is prone to produce outliers, which preclude from a wider practical applications of the model. To resolve those problems, we suggested six alternative models: two relying on stable population assumption and on general analytical relations; two based on backward-projection ideas used to reconstruct the age structure of elderly, one regression model and one behavioral model accounting explicitly for the mechanism of age exaggeration. The models presented differ in their accuracy and data demand. The most accurate (assuming accurate and detailed input data exists) method is the one based on cohort-component method. The least accurate among the first five is the regression model that, however, does not rely on population stability assumption and is least data-demanding. Depending on data availability and on whether the stability assumption seems valid, one may select a proper model in a particular application case.

The behavioral model stands alone in our list of proposed models, as it produces the entire set of death rates rather than correcting the remaining life expectancy. The model could not have been tested in the same way as the other models, as it is based on detailed (even if biased) age patterns of population and death rates. Nonetheless, empirical approbations of the model point to its usefulness in reconstructing the non-distorted death rates and quantifying the magnitude of age exaggeration. The model may, apparently, be used as a supplement (an alternative independent assessment) to other models when the latter are used in combination with the constrained extrapolation to produce age details of mortality change at advanced old age.

References

Coale, A. J. (1985). Estimating the Expectation of Life at Old Ages: Comments on the Article by Mitra. *Population Studies*, *39*(3), 507–509. https://doi.org/10.1080/0032472031000141656

Ediev, D. M. (2011a). Life Expectancy in Developed Countries is Higher Than Conventionally Estimated. Implications from Improved Measurement of Human Longevity. *Journal of Population Ageing*, *4*(1), 5–32. https://doi.org/10.1007/s12062-011-9040-x

Ediev, D. M. (2011b). Robust backward population projections made possible. *International Journal of Forecasting*, *27*(4), 1241–1247.

Ediev, D. M. (2017). Constrained Mortality Extrapolation to Old Age: An Empirical Assessment. *European Journal of Population*. https://doi.org/10.1007/s10680-017-9434-4

Ediev, D. M. (2018). Expectation of life at old age: revisiting Horiuchi-Coale and reconciling with Mitra. *Genus*, *74*(1). https://doi.org/10.1186/s41118-018-0029-7

Ediev, D. M. (2019a). On the sources of instability of the Mitra model for years of life at old-age. *Communications in Statistics: Case Studies, Data Analysis and Applications, forthcomin*.

Ediev, D. M. (2019b). On the Statistical Dependence of Life Expectancy and Mortality Rate at a Given Age. *Voprosy Statistiki*, *26*(6), 39–46.

Gompertz, B. (1825). 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.

Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, *107*(01), 49–80. https://doi.org/10.1017/S0020268100040257

HLTDB. (2017). The Human Life-Table Database jointly developed by Max Planck Institute for Demographic Research (Rostock, Germany), Department of Demography at the University of California (Berkeley, USA), and Institut national d'études démographiques (Paris, France). Retrieved January 10, 2017, from http://www.lifetable.de/

Horiuchi, S., & Coale, A. J. (1982). A Simple Equation for Estimating the Expectation of Life at Old Ages. *Population Studies*, *36*(2), 317–326. https://doi.org/10.2307/2174203

Horiuchi, S., & Wilmoth, J. R. (1997). Age patterns of the life table aging rate for major causes of death in Japan, 1951-1990. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *52*(1), B67-77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9008660%5Cnhttp://biomedgerontology.oxfordjournals.org/cgi/content/abstract/52A/1/B67

Keyfitz, N., & Caswell, H. (2005). *Applied mathematical demography.* Springer.

Missov, T. I., Németh, L., & Dańko, M. J. (2016). How much can we trust life tables? Sensitivity of mortality measures to right-censoring treatment. *Palgrave Communications*, *2*, 15049. https://doi.org/10.1057/palcomms.2015.49

Mitra, S. (1984). Estimating the Expectation of Life at Older Ages. *Population Studies*, *38*(2), 313–319. https://doi.org/10.2307/2174079

Mitra, S. (1985). On Estimating the Expectation of Life at Old Ages: Reply to Professor Coale. *Population Studies*, *39*(3), 511–512. https://doi.org/10.1080/0032472031000141666

Myers, R. J. (1940). Errors and bias in the reporting of ages in census data. *Transactions of the Actuarial Society of America*, *41*(II), 395–415.

Myers, R. J. (1993). Errors and bias in the reporting of ages in census data. In *Donald J. Bogue, Eduardo E. Arriaga, Douglas L. Anderton, George W. Rumsey (Eds.) Readings in population research methodology. Volume 1. Basic tools* (pp. 4-23-4–29). Chicago, Illinois: Social Development Center. Retrieved from https://www.popline.org/node/335534

Oeppen, J., & Vaupel, J. W. (2002). Demography. Broken limits to life expectancy. *Science*

(New York, N.Y.), *296*(5570), 1029–1031. https://doi.org/10.1126/science.1069675
Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. https://doi.org/10.2307/1535065
Randall, S., & Coast, E. (2016). The quality of demographic data on older Africans. *Demographic Research*, *34*, 143–174. https://doi.org/10.4054/DemRes.2016.34.5
Robine, J.-M., & Caselli, G. (2005). An unprecedented increase in the number of centenarians. *Genus*, *61*(1), 57–82.
Shryock, H. S., & Siegel, J. S. (1973). *The Methods and Materials of Demography*. Washington D.C.: United States Bureau of the Census.
Thatcher, A. R., Kannisto, V., & Vaupel, J. W. (1998). *The Force of Mortality at Ages 80-120. Monographs on Population Aging*. Odense, Denmark: Odense University Press. Retrieved from http://www.demogr.mpg.de/Papers/Books/Monograph5/ForMort.htm
United Nations Statistics Division. (2019). Demographic and Social Statistics. Retrieved September 28, 2019, from https://unstats.un.org/unsd/demographic/products/dyb/dybcens.htm
University of California, B., & Max Planck Institute for Demographic Research (Rostock). (2019). Human Mortality Database. Online database sponsored by University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Retrieved May 15, 2018, from www.mortality.org
Vaupel, J. W., & Yashin, A. (1985). Heterogeneity's ruses: some surprising effects of selection on population dynamics. *The American Statistician*, *39*(3), 176–185. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12267300
White, K. M. (2002). Longevity Advances in High-Income Countries, 1955-96. *Population and Development Review*, *28*(1), 59–76. https://doi.org/10.1111/j.1728-4457.2002.00059.x
Wilmoth, J. R., Andreev, K., Jdanov, D., & Glei, D. A. (2007). Methods Protocol for the Human Mortality Database. Retrieved from http://www.mortality.org/Public/Docs/MethodsProtocol.pdf

Appendix 1. The behavioral model

1. Formal results

Under the (x_0, α, δ) model, death rates at ages $x \geq x_0 + \delta$ will be distorted because younger, in case $\delta > 0$, or older, in case $\delta < 0$, people will be erroneously added to the observed population thereby distorting the observed death rates to the weighted average of the true rates at ages x and $x - \delta$:

$$\widetilde{M}(x) = \frac{\widetilde{D}(x)}{\widetilde{P}(x)} = \frac{(1-\alpha)P(x)M(x)+\alpha P(x-\delta)M(x-\delta)}{(1-\alpha)P(x)+\alpha P(x-\delta)} = M(x)\left(1 - \frac{1-\frac{M(x-\delta)}{M(x)}}{1+\frac{1-\alpha}{\alpha}\frac{P(x)}{P(x-\delta)}}\right), x \geq x_0 + \delta, \quad (1)$$

here, $M(x)$ is the death rate at age x, $P(x)$ is the population of age x, and $D(x) \stackrel{\text{def}}{=} M(x)P(x)$ is the deaths' intensity at age x, and the tilde mark denotes observed (possibly biased due to the age exaggerations) quantities as opposed to the true, non-observed, ones that are not marked. If, as typical at old age, mortality increases by age and $\alpha, \delta > 0$, then the observed death rate is an underestimation of the true rate, $\widetilde{M}(x) < M(x), x \geq x_0 + \delta$. If, on the contrary, elderly were to reduce their true age ($\alpha > 0, \delta < 0$) the observed death rates would be overestimated as compared to the true rates: $\widetilde{M}(x) > M(x), x \geq x_0 + \delta$.

To examine qualitatively the patterns of distortions to the mortality curve in the age exaggeration model (1), we use Gompertz (Gompertz, 1825; Thatcher et al., 1998) mortality model that implies, for any two ages x and y:

$$\frac{M(y)}{M(x)} = e^{b(y-x)}, \quad (2)$$

and the stable population model (Preston, Heuveline, & Guillot, 2001) where:

$$\frac{P(y)}{P(x)} = e^{-r(y-x)}\frac{l(y)}{l(x)} = e^{-r(y-x)}e^{-\int_x^y M(z)dz} = e^{-r(y-x)}e^{-\frac{1}{b}M(x)\left(e^{b(y-x)}-1\right)}. \quad (3)$$

Here, $l(x)$ is the survival function, i.e., proportion surviving from birth to age 0.

Combining (1)-(3), we get the analytical expression for distortion of the observed death rates:

$$\frac{\widetilde{M}(x)}{M(x)} = 1 - \frac{1-e^{-b\delta}}{1+\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}}, x \geq x_0 + \delta. \quad (4)$$

As it follows from (4), the higher the death rate (i.e., the older the age x), the stronger the proportionate bias of the death rate. Two limit cases of (4) deserve closer attention. First, consider younger ages where the death rate is low as compared to the population growth rate, $M(x) \ll r$. In that case, (4) reduces to:

$$\frac{\widetilde{M}(x)}{M(x)} \approx 1 - \frac{1-e^{-b\delta}}{1+\frac{1-\alpha}{\alpha}e^{-r\delta}}, x \geq x_0 + \delta. \quad (5)$$

This means, *at ages where the growth rate dominates the death rate in shaping the population age structure, proportionate distortions to the observed death rates are age-independent*. When, additionally, $|r\delta| \ll 1$ and $|b\delta| \ll 1$, (5) reduces to simply $\frac{\widetilde{M}(x)}{M(x)} \approx 1 - \alpha\delta b \approx e^{-\alpha\delta b}$ and does not depend on the growth rate r. In other words, at young ages and at moderate age exaggeration, the proportionate distortion of the death rates is determined by the amount of increase of the death rate over the mean magnitude ($\alpha\delta$) of the age exaggeration and not on the growth rate.

In the opposite limit case of advanced old ages where $M(x) \gg |r|$,

$$\frac{\widetilde{M}(x)}{M(x)} \approx 1 - \frac{1-e^{-b\delta}}{1+\frac{1-\alpha}{\alpha}e^{-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}}. \quad (6)$$

Assuming $|b\delta| \ll 1$, $e^{-M(x)\delta} \ll \alpha$, (6) reduces to $\frac{\widetilde{M}(x)}{M(x)} \approx e^{-b\delta}$. Hence, at advanced old age with high mortality, the observed death rates are also distorted in an age-independent way. Furthermore, distortion of the death rates at old ages may be independent of the age exaggeration parameter α.

The two limit cases suggest that the population growth rate (and therefore, the departure of the population age structure from stationary) might be less important for the distortions of the mortality curve as compared to other factors such as the age exaggeration parameters and the level and the slope of the mortality curve.

Patterns of the distortions in the observed population composition by age fit the following relation:

$$\frac{\tilde{P}(x)}{P(x)} = 1 + \alpha\left(e^{r\delta + \frac{1}{b}M(x)(1-e^{-b\delta})} - 1\right), x \ge x_0 + \delta. \tag{7}$$

At young age, population distortions (7) are sensitive to the growth rate, $\frac{\tilde{P}(x)}{P(x)} \approx 1 + \alpha r\delta$, unlike in the death rates' distortions. At advanced old age, the population distortions explode remaining α-specific: $\frac{\tilde{P}(x)}{P(x)} \sim \alpha e^{M(x)\delta}$.

Because the distortions to the death rates due to the age exaggeration vary with age, they also affect the pace at which the death rates change. In particular, from (4), the Life-table Ageing Rate (LAR) (Horiuchi & Wilmoth, 1997) is subject to the following alterations:

$$\frac{dln\tilde{M}(x)}{dx} = \frac{dlnM(x)}{dx} + \frac{d}{dx}ln\left(1 - \frac{1-e^{-b\delta}}{1+\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}}\right) = b + \frac{1}{1-\frac{1-e^{-b\delta}}{1+\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}}}\frac{1-e^{-b\delta}}{\left(1+\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}\right)^2}\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)(1-e^{-b\delta})}\left(-\frac{1}{b}\right)(1-e^{-b\delta})bM(x) = b - \frac{\left(1-e^{-b\delta}\right)^2\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}}{\left(e^{-b\delta}+\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}\right)\left(1+\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}\right)}M(x). \tag{8}$$

Next, we infer analytically, if the classical relation (3) between the stable population growth, mortality and age composition may still hold if the population numbers and the death rates are subject to distortions due to the age exaggeration. In the stable population, as it follows from (3), there must be a relation between the age structure, mortality and growth rate:

$$\frac{dlnP(x)}{dx} = -r - M(x). \tag{9}$$

When observing the biased age structure (7) and the death rates (4) at ages $x > x_0 + \delta$,

$$\frac{dln\tilde{P}(x)}{dx} = \frac{dlnP(x)}{dx} + \frac{dln\left(1+\alpha\left(e^{r\delta+\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}-1\right)\right)}{dx} = -r - M(x) + \frac{\alpha e^{r\delta+\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}\frac{1}{b}\frac{dM(x)}{dx}\left(1-e^{-b\delta}\right)}{1+\alpha\left(e^{r\delta+\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}-1\right)} = -r - M(x)\left(1 - \frac{\alpha e^{r\delta+\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}\left(1-e^{-b\delta}\right)}{1+\alpha\left(e^{r\delta+\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}-1\right)}\right) = -r - M(x)\left(1 - \frac{1-e^{-b\delta}}{1+\frac{1-\alpha}{\alpha}e^{-r\delta-\frac{1}{b}M(x)\left(1-e^{-b\delta}\right)}}\right) = -r - \tilde{M}(x). \tag{10}$$

In other words, *at ages $x > x_0 + \delta$, the stable-population relation between the observed age composition, death rates and the growth rate holds despite the distortions caused by the age misstatement*. This, unfortunately, implies it will typically be hard to detect the age exaggeration from observed population numbers and death rates. Note, however, that the stable-population relation between the population growth, age composition and mortality will be broken at ages at ages x_0 to $x_0 + \delta$ where the death rates will not be affected by age exaggeration while the population size will.

More generally, the usual population balance (Keyfitz & Caswell, 2005):

$$\frac{\partial}{\partial t}P(x,t) + \frac{\partial}{\partial x}P(x,t) = -D(x,t), \tag{11}$$

where the second variable in all functions indicates time, will not be violated by age- and time-independent age misstatements in any (including non-stable) migration-closed population *at ages $x > x_0 + \delta$*:

$$\frac{\partial}{\partial t}\tilde{P}(x,t) + \frac{\partial}{\partial x}\tilde{P}(x,t) = \frac{\partial}{\partial t}\big((1-\alpha)P(x,t) + \alpha P(x-\delta,t)\big) + \frac{\partial}{\partial x}\big((1-\alpha)P(x,t) + \alpha P(x-\delta,t)\big) = (1-\alpha)\left(\frac{\partial}{\partial t}P(x,t) + \frac{\partial}{\partial x}P(x,t)\right) + \alpha\left(\frac{\partial}{\partial t}P(x-\delta,t) + \frac{\partial}{\partial x}P(x-\delta,t)\right) = -(1-\alpha)D(x,t) - \alpha D(x-\delta,t) = -\tilde{D}(x,t). \tag{12}$$

The population balance will be broken, however, at ages x_0 and $x_0 + \delta$ where population numbers will also be subject to jumps caused by the onset of age exaggeration process.

The last two formal results show that detecting the age exaggeration from the data alone, without specific assumptions as regards, for example, the shape of the mortality curve, might be a very tough task.

We conclude this section by studying how the age exaggeration may affect the life expectancy estimates in the case of stationary population. The stationary population case is convenient to study because our earlier results suggest the age structure and distribution of deaths at ages $x > x_0 + \delta$ will be consistent with each other despite the distortions caused by the age exaggeration. In particular, the remaining life expectancy at age $x_0 + \delta$ in stationary population is the mean age of the observed distribution of deaths. At ages older than $x_0 + \delta$, the observed distribution of deaths is a mix of the true distribution taken with weight $(1-\alpha)l(x_0+\delta)$, where $l(x)$ is the survival curve at age x, i.e., stationary-population number of deaths at ages older than x, and of shifted distribution of deaths at ages $x > x_0$ with weight $\alpha l(x_0)$. Therefore, the observed remaining life expectancy at age $x_0 + \delta$ in the stationary population equals:

$$\tilde{e}(x_0+\delta) = \frac{(1-\alpha)l(x_0+\delta)e(x_0+\delta)+\alpha l(x_0)e(x_0)}{(1-\alpha)l(x_0+\delta)+\alpha l(x_0)} \quad (13)$$

here, $e(x)$ is the remaining life expectancy at age x. When mortality at ages x_0 to $x_0 + \delta$ falls to levels where $l(x_0+\delta) \approx l(x_0)$ and $e(x_0) \approx \delta + e(x_0+\delta)$, Eq. (13) leads to the limit

$$\tilde{e}(x_0) - e(x_0) \approx \tilde{e}(x_0+\delta) - e(x_0+\delta) \approx \alpha\delta. \quad (14)$$

this will be an upper limit to the remaining life expectancy bias at all ages up to $x \le x_0$, because, at those ages, using (13) and inequalities $e(x_0) \le \delta + e(x_0+\delta)$, $l(x_0+\delta) \le l(x)$, it follows:

$$\tilde{e}(x) - e(x) = \frac{l(x_0+\delta)}{l(x)}[\tilde{e}(x_0+\delta) - e(x_0+\delta)] = \frac{l(x_0+\delta)}{l(x)}\left[\frac{(1-\alpha)l(x_0+\delta)e(x_0+\delta)+\alpha l(x_0)e(x_0)}{(1-\alpha)l(x_0+\delta)+\alpha l(x_0)} - e(x_0+\delta)\right] = \frac{l(x_0+\delta)}{l(x)}\frac{\alpha l(x_0)[e(x_0)-e(x_0+\delta)]}{(1-\alpha)l(x_0+\delta)+\alpha l(x_0)} \le \frac{l(x_0+\delta)}{l(x)}\frac{\alpha l(x_0)\delta}{(1-\alpha)l(x_0+\delta)+\alpha l(x_0)} = \frac{l(x_0)}{l(x)}\frac{l(x_0+\delta)}{(1-\alpha)l(x_0+\delta)+\alpha l(x_0)}\alpha\delta \le \alpha\delta. \quad (15)$$

Hence, in the stationary population, at ages younger than the first age where the age exaggeration begins, distortion to the remaining life expectancy will be bounded by the upper limit simply equal to the mean magnitude of the age exaggeration $\alpha\delta$. That upper limit twill only be reached when mortality below age $x_0 + \delta$ falls to zero. For a non-stationary population, however, the stationary upper limit may be exceeded, as, for example, in stable growing populations.

2. Numerical simulations

To have a better idea of distortions due to age miss-statements in the (x_0, α, δ) model, we ran numerical simulations using (4), (7) and (8) (Figures A1-A4).

Distortions to population structure by age substantially depend on the population growth rate that may be taken as a proxy for deviations of real-life populations from the stationary population (Figure A1: in this and following figures, the horizontal axis is the true death rates in log-scale that, in Gompertzian mortality assumed here, represents age). At large age exaggerations δ, distortion patterns for various combinations of the growth rate and α overlap with each other that makes it problematic to reconstruct the age exaggeration parameters from the observed population age composition.

As expected from the results for the limit cases (5) and (6), distortions of the death rates (Figure A2) are not much sensitive to the population growth rate as compared to the age exaggeration parameters. This suggests it might be possible to reconstruct the exaggeration model parameters as well as the true death rates from the observed deviations of the death rates' patterns from assumed model mortality curves. The underestimation of the death rates at advanced ages is much stronger than at younger ages. In extreme cases, the observed death rates may go down to as low as only dozens of percent of the true rate. Such a strong biases at old age, both absolute and relative, in comparison to younger ages, make it possible for the observed death rates to follow some very unrealistic non-monotone trajectories at old age (Figure A3)[1].

Consequently, the observed Life-Table Ageing Rate (LAR) may profoundly mislead because of the age exaggeration (Figure A4). Somewhat counterintuitively, the age at observed maximum mortality (at which LAR equals zero) is younger for less-prevalent (but strong) age exaggerations than for the more-prevalent ones.

[1] Notably, such non-monotone patterns of old-age mortality exist, although not common, even in the high-quality data collection of the Human Mortality Database.

This is an important observation pointing to possibility to distinguish the cases of age exaggeration by a small percent of population from cases of heterogeneous population based on the observed mortality patterns. As shown by Vaupel and Yashin (1985), population heterogeneity in frailty may well produce sorts of distorted age patterns of the death rates that we demonstrate here for the age exaggeration case. Indeed, the two cases are identical mathematically, as a proportion of people who mis-placed themselves into a wrong age category may formally be interpreted as a sub-population of the reported age with different frailty. Yet, one might distinguish the two scenarios if the age misreporting affects a small fraction of the population who, nonetheless exaggerate their age by a large amount while heterogeneity leads to more substantial proportions of the population showing moderate differences in their frailty. Old-age mortality peaking at some age seems to be a sign of few people strongly exaggerating their age rather than more substantial parts of population differing in their frailty.

Figure A1. Distortions to the observed population numbers in relevant age groups as compared to the true population by age: by δ ("Delta", in years, columns), b (rows), α ("Alpha", in percent, colors), r (line types). Horizontal axis: the true death rate M(x), log-scale.
Note: values below 1 indicate underestimation; values above 1 indicate overestimation.

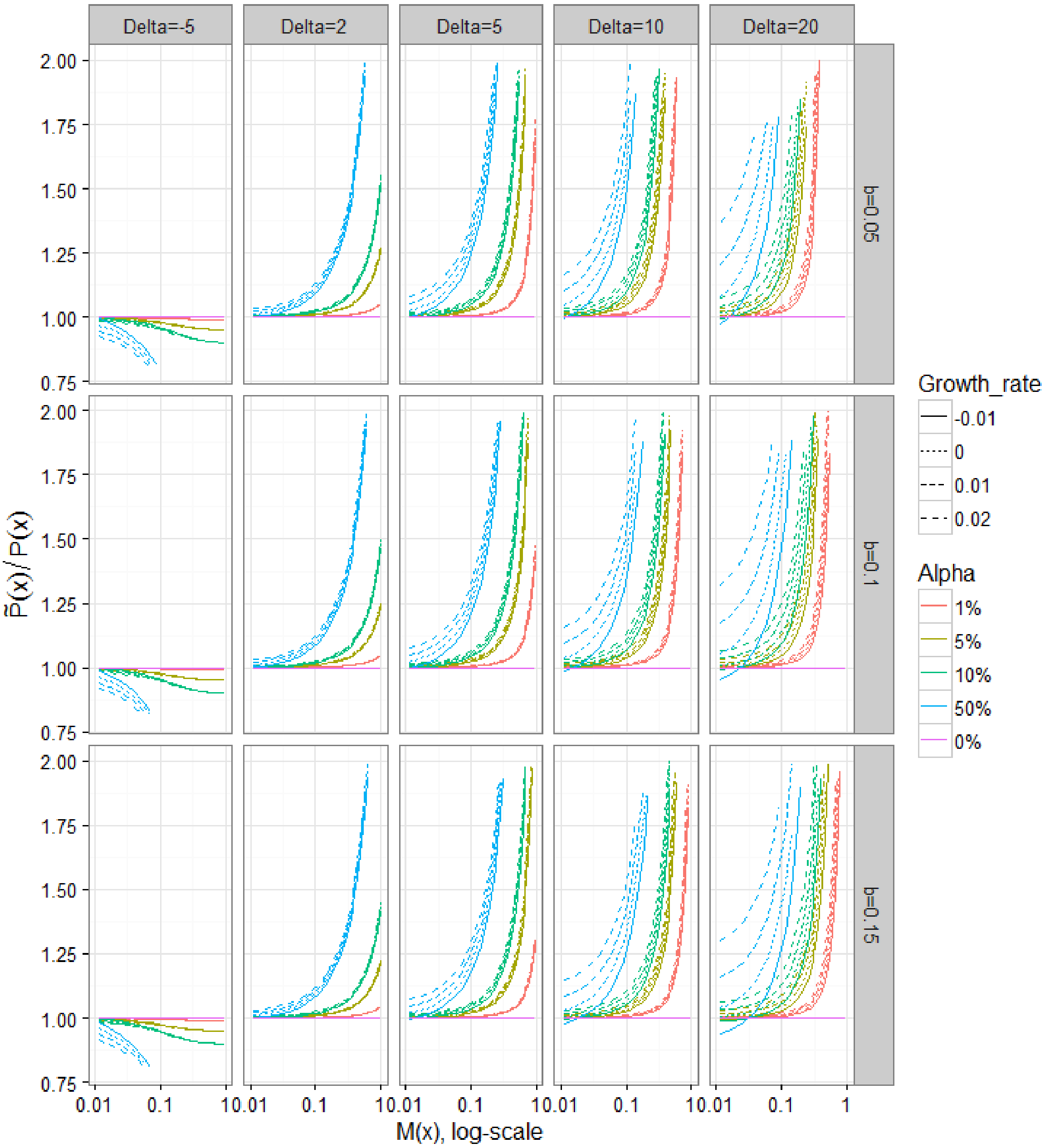


Figure A2. Distortions to the observed death rates as compared to the true rates: by δ ("Delta", in years, columns), b (rows), α ("Alpha", in percent, colors), r (line types). Horizontal axis: the true death rate M(x), log-scale.

Note: values below 1 indicate underestimation; values above 1 indicate overestimation.

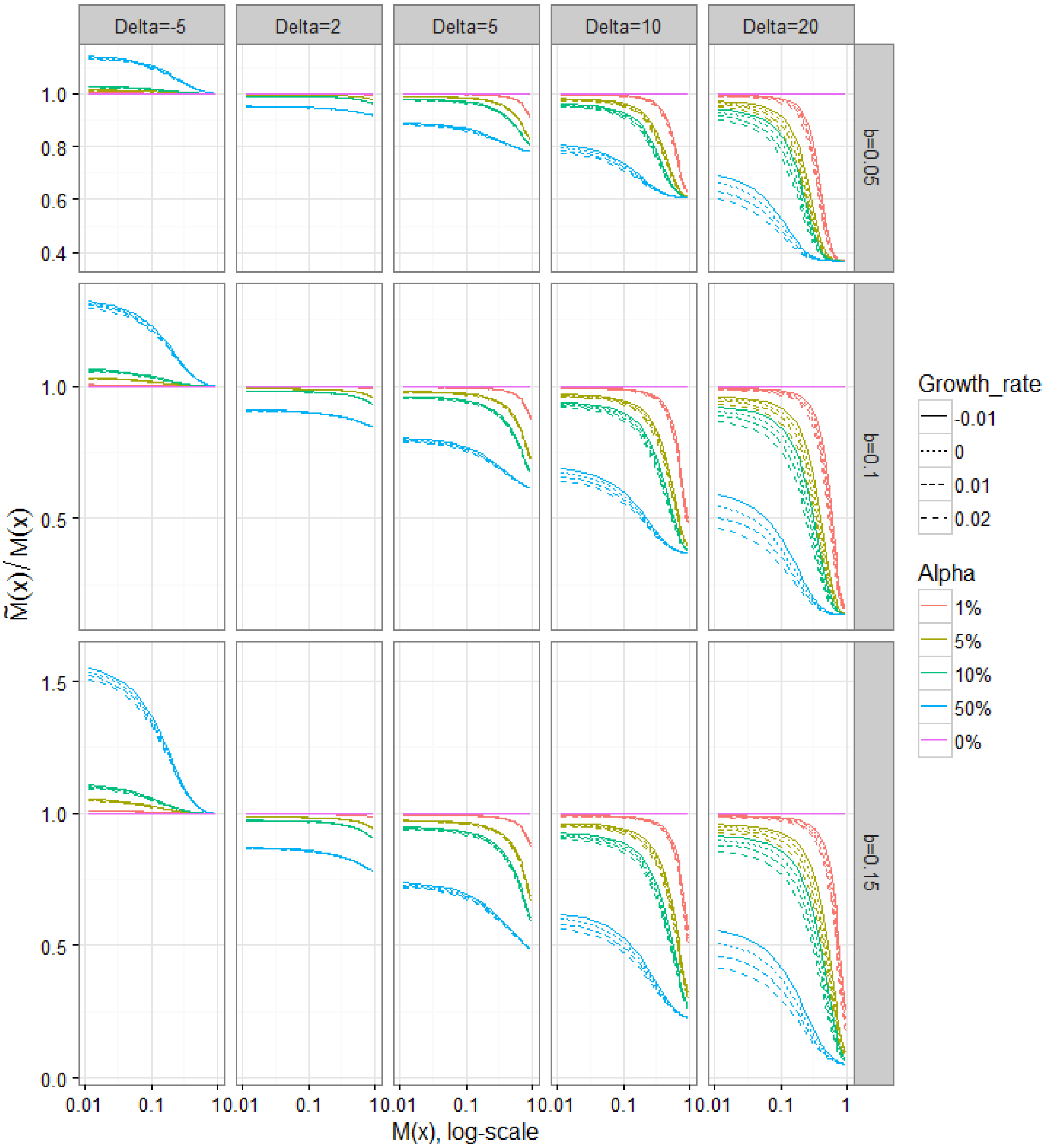


Figure A3. Observed death rates (vertical axis) vs true death rates (horizontal axis, log-scale): by δ ("Delta", in years, columns), b (rows), α ("Alpha", in percent, colors), r (line types).

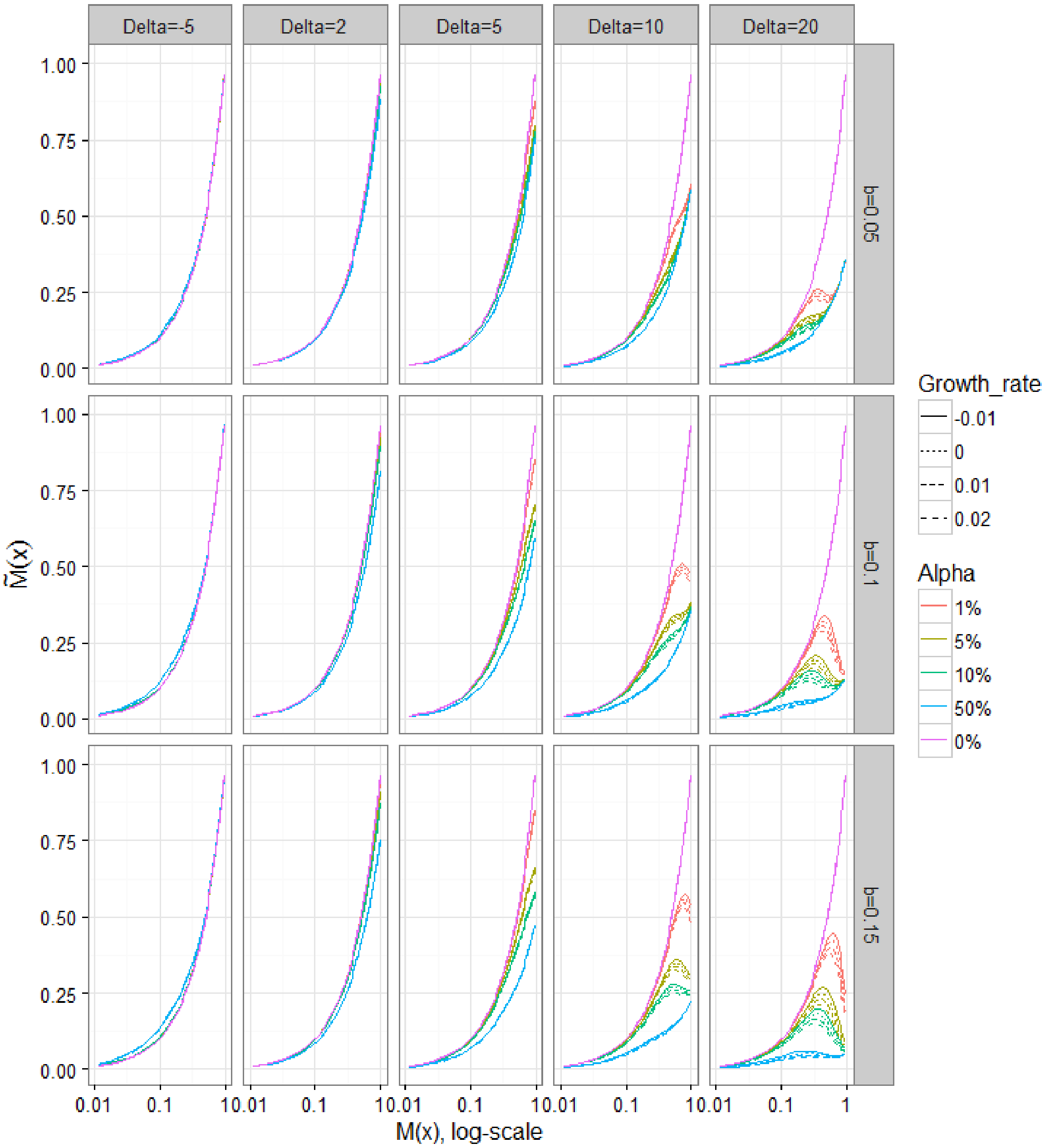
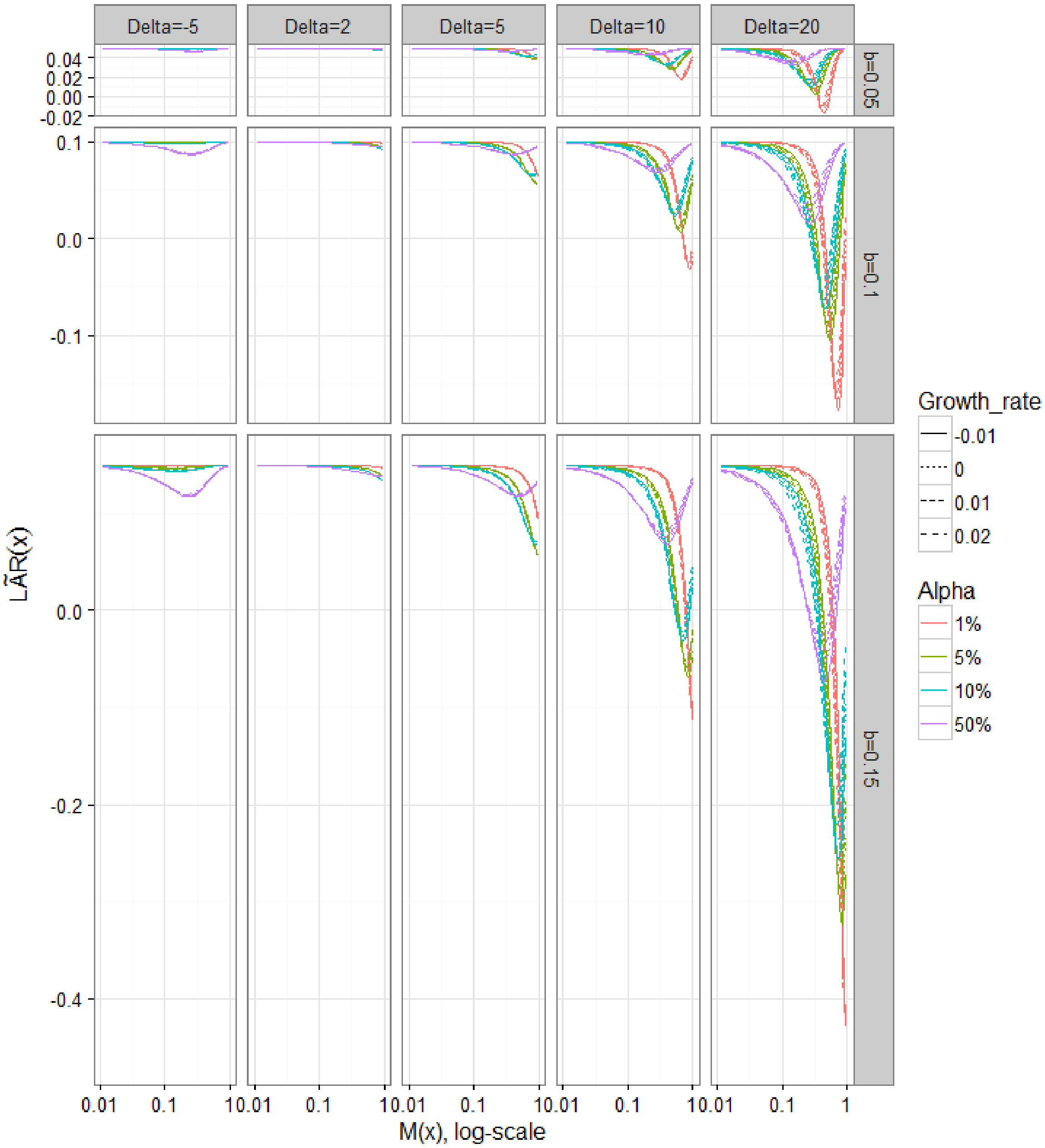


Figure A4. Observed life table ageing rate (LAR) at relevant age: by δ ("Delta", in years, columns), b (rows), α ("Alpha", in percent, colors), r (line types). Horizontal axis: the true death rate M(x), log-scale.



3. Calibrating the age exaggeration model

In order to assess possible values of the parameters if the behavioral model fit to empirical data, we estimated the optimal model parameters that would reproduce some selected empirical patterns of the death rates. The selected patterns were those from the Human Life Table Database (2017) that show signs of age exaggeration (such as levelling off and reversing age trends in old-age mortality). Results are presented in Table A2 and Figures A5 (produced assuming Gompertz model) and A6 (produced assuming Kannisto model).

Table 2. Reconstructed age exaggeration model parameters and observed and corrected life table indicators. Selected populations of the Human Life Table Database (2017).

Country	Sex	Period	b	x_0	α	δ	e(0)	e(0) cor-rected	e(0) bias	e(x_0)	e(x_0) cor-rected	e(x_0) bias
China rural	M	1981	0.087	73	4.1%	17.0	65.9	65.9	-0.02	8.1	8.2	-0.05
China rural	F	1981	0.098	56	1.0%	23.0	68.7	68.6	0.11	21.2	21.1	0.14
Cuba	M	2005_07	0.091	75	6.5%	13.9	76.1	75.9	0.11	10.9	10.7	0.18
Cuba	F	2005_07	0.103	77	9.5%	12.6	80.1	79.9	0.20	11.2	10.9	0.30
Mexico	M	1983_85	0.072	76	3.0%	17.0	66.2	66.1	0.02	10.1	10.0	0.06
Mexico	F	1983_85	0.084	76	4.4%	17.0	72.2	72.2	0.02	10.8	10.8	0.03
Turkey	M	2013_14	0.096	63	0.9%	21.0	75.3	75.3	0.03	17.6	17.6	0.03
Turkey	F	2013_14	0.117	71	6.7%	13.0	80.7	80.6	0.18	14.7	14.4	0.21
Uruguay	M	2004	0.099	80	4.5%	8.0	71.7	71.6	0.04	6.6	6.4	0.12
Uruguay	F	2004	0.124	68	21.3%	5.4	79.0	78.3	0.76	16.8	15.9	0.92

Source: Own elaboration on (HLTDB, 2017).

Figure A5. Empirical (circles) and reconstructed biased (red lines) and corrected unbiased death rates (red lines) obtained by fitting the behavioral model combined with the Gompertz model of old-age mortality rates. Selected populations of the Human Life Table Database (2017).

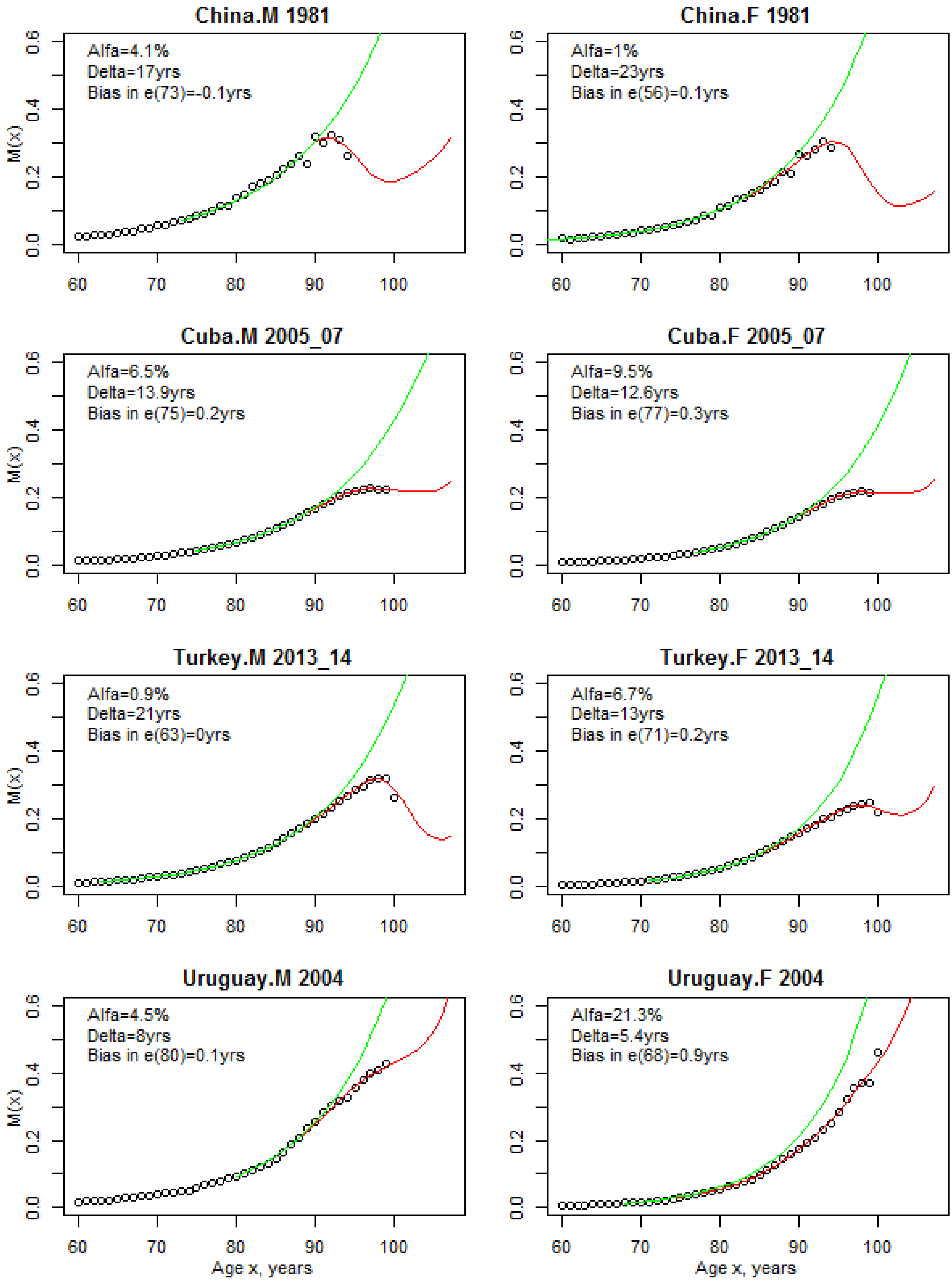


Figure A5. Empirical (circles) and reconstructed biased (red lines) and corrected unbiased death rates (red lines) obtained by fitting the behavioral model combined with the Kannisto model of old-age mortality rates. Selected populations of the Human Life Table Database (2017).