Demographic Research a free, expedited, online journal
of peer-reviewed research and commentary
in the population sciences published by the
Max Planck Institute for Demographic Research
Doberaner Strasse 114 · D-18057 Rostock · GERMANY
www.demographic-research.org

DEMOGRAPHIC RESEARCH
VOLUME 3, ARTICLE 8
PUBLISHED 17 OCTOBER 2000
www.demographic-research.org/Volumes/Vol3/8/
DOI: 10.4054/DemRes.2000.3.8

Frailty Modelling for Adult and Old Age Mortality: The Application of a Modified DeMoivre Hazard Function to Sex Differentials in Mortality

Hans-Peter Kohler

Iliana Kohler

Frailty Modelling for Adult and Old Age Mortality: The Application of a Modified DeMoivre Hazard Function to Sex Differentials in Mortality

Hans-Peter Kohler[1]
Iliana Kohler[2]

Abstract

Unobserved differences in individual's susceptibility to death are an important aspect in the analysis of contemporary mortality patterns. However, observed mortality rates at adult ages, which are usually well-described by a Gompertz curve, are often perceived inconsistent with frailty models of mortality. We therefore propose a modified DeMoivre hazard function that is suitable for the application of frailty models to adult and old ages. The proposed hazard increases faster than exponential, and when combined with unobserved frailty it can capture a broad range of patterns encountered in the analysis of adult mortality. Our application to Bulgaria during 1992–93 suggests that the stronger selection process in the male population, caused by an overall higher level of mortality, may constitute a primary mechanism leading to the convergence of male and female mortality at higher ages. Hence, the convergence between male and female mortality is not necessarily caused by a differential process of aging across sexes, but is merely a consequence of the different levels of mortality at adult ages.

[1]Head of Research Group on Social Dynamics and Fertility, Max Planck Institute for Demographic Research, Doberaner Str. 114, 18057 Rostock, Germany. *Tel:* +49-381-2081-123, *Fax:* +49-381-2081-423, *Email:* kohler@demogr.mpg.de, *www:* http://user.demogr.mpg.de/kohler.

[2]Ph.D. student at the Max Planck Institute for Demographic Research and the Danish Center for Demographic Research at the University of Southern Denmark, Odense. *Contact address:* Max Planck Institute for Demographic Research, Doberaner Str. 114, 18057 Rostock, Germany. *Tel:* +49-381-2081-125, *Fax:* +49-381-2081-425, *Email:* iliana.kohler@demogr.mpg.de, *www:* http://www.demogr.mpg.de.

1 Introduction

Male mortality exceeds female mortality rates at adult ages in many populations. This female advantage in survival often diminishes with age, and male and female mortality rates converge at higher ages [Carey and Judge 2000, Hummer et al. 1998, Manton et al. 1995, Waldron 1985]. Similar patterns of convergence or mortality cross-overs are also observed between other populations that are subject to quite different mortality levels at adult, but not necessarily at old and oldest-old ages [Gavrilov and Gavrilova 1991, Vaupel and Yashin 1985]. Consider for instance the Bulgarian male and female mortality pattern during 1992–93 in Figure 1(a). The striking aspect of this mortality pattern is on one hand the substantially higher mortality level for males, especially during adult ages, and on the other hand the differential increase in the force of mortality by age [see also Kohler 2000a,b]. The mortality sex-ratio in Figure 1(b) shows that males in Bulgaria around age 40 experience a mortality level that is almost 200% higher than that of females. This male-female difference diminishes to 25% around age 80 and it virtually vanishes at ages above age 90. This convergence in male and female mortality levels occurs because, despite their mortality-advantage at adult ages, females in Bulgaria are subject to a substantially more rapid increase in the level of mortality by age. The life-table aging rate [Carey and Liedo 1995, Horiuchi and Coale 1990], i.e., the relative increase of the mortality hazard per additional year of age, depicted in Figure 1(b) shows that the relative increase of female mortality with age is above the relative male mortality increase at all ages above age 45. [Note 1] In addition to the difference in absolute level, the life-table aging rate (LAR) reflects some known sex-specific deviations from the fitted Gompertz model in the left graph of Figure 1. In particular, the female life-table aging rate is increasing between age 40–75, which has been attributed to a post-menopausal mortality increase that is due to lower evolutionary selection forces at post-reproductive ages [Horiuchi 1997]. After age 75, the relative increase of mortality by age is declining and the mortality increase by age is slowing down at these old and oldest-old ages. The female life-table aging rate is thus clearly bell-shaped, while the male life-table aging rate is increasing, with some minor fluctuations, up to age 80. Afterwards the increase of male mortality by age is slowing down similar to the female pattern. Ignoring these age specific patterns and averaging across the whole age range 40–100 years, female mortality rates increase by approximately 10.8% per year of age (based on the estimates of the Gompertz model in Figure 1a), while male mortality rates increase by only 8.25% per year of age. This differential increase in mortality by on average 2.5 percentage points implies the strong convergence between male and female mortality rates at higher ages in Figure 1(a).

The important question in this context is whether the differential increase in the force of mor-

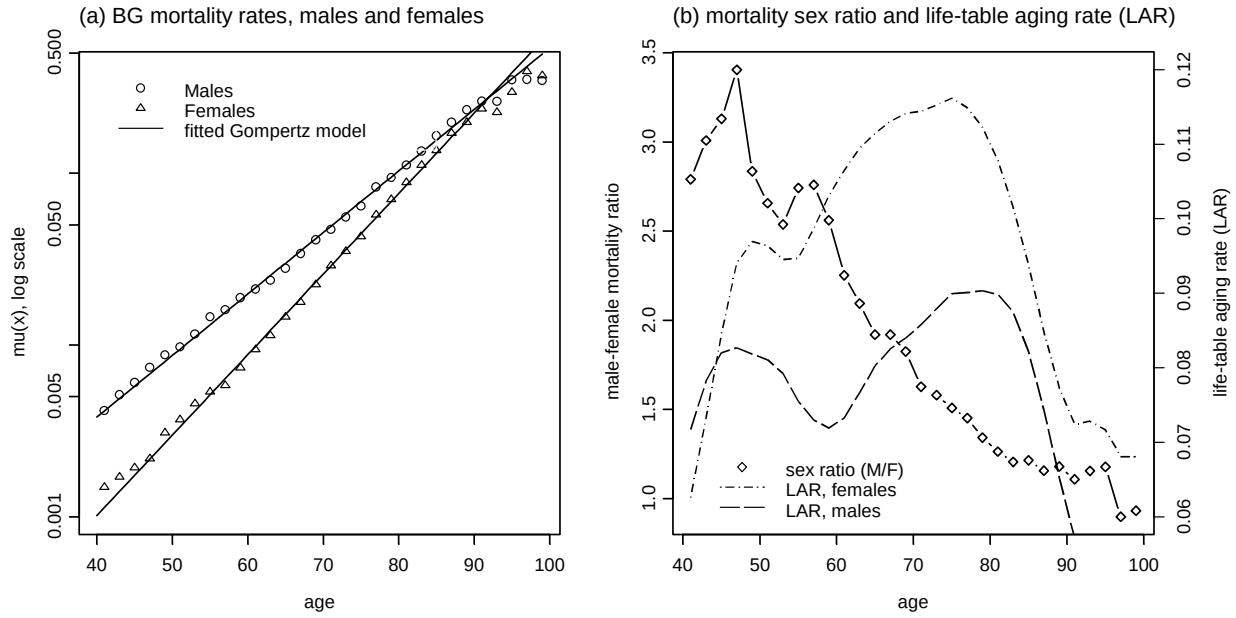

Figure 1: Left graph: Bulgarian mortality rates for males and females in 1992–93. Right graph: Male-female mortality ratio and life-table aging rate (LAR) for males and females in Bulgaria in 1992–93.

tality per year of age is due to a differential aging process between males and females, or whether this difference can be attributed to a stronger selection of the male population towards low-frailty individuals that is caused by the higher overall level of male mortality. The knowledge which of these two factors is primarily responsible for the above convergence pattern is essential for the development of appropriate theories of aging and mortality change [Carnes et al. 1996, Vaupel and Yashin 1985]. In the former case, the male-female convergence of mortality is attributed to factors such as a post-menopausal acceleration of mortality for females [Horiuchi 1997], sex-differences in metabolism, hormonal levels and other fundamental biological aspects [Hazzard 1986, Hazzard and Applebaum-Bowden 1989, Waldron 1985], genetic differences related to the female 'advantage' of having two X chromosomes [Christensen et al. 2000], potential systematic behavioral and psychological sex-differences in coping with stress and the aging process itself [Baltes et al. 1999], and age-related social and behavioral changes [House et al. 1990]. In the latter scenario, which emphasizes the process of differential selection, the higher mortality level for males — especially at adult ages — implies that the male population is more rapidly selected towards individuals with a relatively low risk of mortality. In a heterogeneous population, the differential strength of the

selection process between males and females then leads to a slower increase in the observed male mortality as compared to the female mortality [Vaupel et al. 1979, Vaupel and Yashin 1985]. A convergence in the observed male and female mortality pattern can thus occur even when both sexes are subject to a mortality curve that differs — conditional on a constant frailty composition — only by a factor of proportionality. That is, the mortality convergence occurs even if male and female mortality are characterized by an age pattern that, conditional on the frailty level, exhibits an identical life-table aging rate and an identical relative increase in the mortality rates with age. In this selection hypothesis, therefore, the main difference in the 'law of mortality' for males and females is in the level of the mortality risk, i.e., a proportionally higher level of male as compared to female mortality. Differences in the life-table aging rate and patterns of mortality convergence between sexes are thus primarily attributed to changes in the frailty composition of population, instead of fundamental differences in the process of aging itself.

The investigation of whether the above 'selection hypothesis' can provide a plausible explanation for the male–female differences in the increase of mortality with age requires the estimation of mortality models with unobserved frailty. In a seminal analysis, Vaupel, Manton, and Stallard [1979] have introduced *relative frailty models* in which individuals in a population are heterogeneous with respect to their susceptibility to death. This relative risk of death, denoted by the frailty z, is unobserved on the individual level. Despite this unobservability, the mortality patterns can be adjusted for the distortions caused by the selection process due to differential mortality in heterogeneous populations. In particular, based on assumptions about the initial distribution of unobserved frailty in the population and its effect on the force of mortality, inferences can be made about the composition of the population with respect to frailty at some specific age x and the level of mortality that would have prevailed if there had been no changes in the frailty-composition of the population over time. Frailty models of mortality therefore allow the investigation of whether observed mortality patterns can be explained by selection processes within a population, or by differential selection across subpopulations. For instance, the analyses in this paper focus on the question of whether frailty models can provide a plausible explanation for the convergence between male and female mortality in Bulgaria and possibly other countries.

Relative frailty models assume that the mortality rate at age x of a person with frailty z equals $z\mu(x)$, where $\mu(x)$ is the mortality rate of individuals with $z = 1$. Individuals with $z > 1$ therefore experience a force of mortality that is proportionally higher than $\mu(x)$ at all ages, while individuals with $z < 1$ experience proportionally lower mortality rates. The composition of a cohort with respect to the frailty z changes as a cohort grows older because the most frail individuals tend to die earlier than the least frail individuals. The increase of the observed mortality rates with age

is therefore determined by two factors: (*a*) the age-increase in mortality holding frailty constant, which is reflected in $\frac{d}{dx}\mu(x)$; and (*b*) the extent to which the cohort at age x becomes selected towards low-frailty individuals, which is reflected in the distribution of the frailty z in the population conditional on survival up to age x.

Vaupel et al. [1979] assume a gamma-distributed frailty with mean one and variance σ^2, and show that the *observed* hazard and survival curve at age x, denoted $\bar{\mu}(x)$ and $\bar{s}(x)$, are equal to

$$\bar{\mu}(x) = \frac{\mu(x)}{1 - \sigma^2 \log s(x)} \tag{1}$$

and

$$\bar{s}(x) = (1 - \sigma^2 \log s(x))^{-1/\sigma^2}, \tag{2}$$

where $\mu(x)$ and $s(x)$ are the baseline hazard rate and the survival curve for individuals with a constant frailty of $z = 1$. Moreover, the mean frailty $\bar{z}(x)$ of the population who is alive at age x equals $\bar{z}(x) = \bar{s}(x)^{\sigma^2} = (1-\sigma^2 \log s(x))^{-1}$, which indicates that the mean frailty of the population who has survived to some age x decreases as the fraction of survivors to age x declines.

Equations (1) and (2) indicate that the selection process in heterogeneous populations drives a wedge between the baseline hazard rate $\mu(x)$, which pertains to individuals with constant frailty $z = 1$, and the observed hazard rate $\bar{\mu}(x)$. The extent to which these rates differ depends on two factors: (*a*) the variance σ^2, i.e., the variation in unobserved frailty in the population at the beginning of the observation period, and (*b*) the extent to which the population has already been selected, which is indicated by the survival function $s(x)$. Moreover, the derivative

$$\frac{d}{dx} \log \bar{\mu}(x) = \frac{\mu'(x)}{\mu(x)} - \frac{\sigma^2 \mu(x)}{1 - \sigma^2 \log s(x)} \tag{3}$$

shows that the slope of the observed mortality pattern is also affected by the selection towards low-frailty individuals in heterogeneous populations. On the one hand, the first term $\frac{\mu'(x)}{\mu(x)}$ indicates the relative increase in the force of mortality holding frailty constant at $z = 1$. On the other hand, the second term $\sigma^2\mu(x)/(1-\sigma^2 \log s(x))$ measures the strength of the selection process at age x. This strength depends on the mortality level $\mu(x)$, the overall variance of frailty in the population σ^2 and the mean frailty $\bar{z}(x) = (1 - \sigma^2 \log s(x))^{-1}$ of the population who has survived to age x. This selection process implies that the observed relative increase in the mortality rate x is slower than the relative increase of the baseline hazard $\mu(x)$. This difference between the observed and 'true'

increase of the mortality curve by age increases the larger is the final term in equation (3).

A typical pattern of observed mortality rates, which is implied by the above frailty model, is depicted in Figure 2(a). In this Figure we have assumed that $\mu(x)$ follows a Gompertz curve with $\log \mu(x) = ae^{bx}$, which implies a linear increase in the logarithm of the mortality hazard with age. The arrow in this graph reveals the wedge between the observed mortality rate $\bar{\mu}(x)$ and the underlying rate $\mu(x)$ for frailty $z = 1$. As long as mortality is relatively low, the two curves for $\bar{\mu}(x)$ and $\mu(x)$ trace each other closely. As soon as mortality has increased to moderate levels, however, the selection process towards low-frailty individuals in the population becomes significant and $\bar{\mu}(x)$ increasingly diverges from $\mu(x)$. In particular, while $\log \mu(x)$ increases linearly with age due to the Gompertz specification, the observed pattern $\log \bar{\mu}(x)$ is markedly flatter. The observed mortality increases slower than linear on the log scale, and the slope of the observed mortality pattern becomes increasingly less than the slope of Gompertz hazard. While the life-table aging rate in Figure 2(b) is constant across all ages in the Gompertz model, the introduction of the unobserved heterogeneity leads to a marked decline of the life-table aging rate at higher ages due to an increased selection of the population towards less frail individuals.

A potential problem in estimating the above frailty model becomes apparent when comparing the mortality pattern observed in Bulgaria (Figure 1) with the typical mortality pattern implied by a relative-frailty Gompertz model (Figure 2). First, a Gompertz model fits the Bulgarian mortality in Figure 1 relatively well, and to a first approximation a standard Gompertz hazard function provides a quite good description of the Bulgarian male and female adult and old age mortality pattern for the age range 40–100 years. While this good empirical fit of the Gompertz model may initially seem very desirable, it poses considerable problems in the context of frailty models. The problem arises because the characteristic feature of frailty models is the 'flattening' of the mortality curve and a decline in the life-table aging rate as shown in Figure 2. This decline of the life-table aging rate is due to the fact that the population alive at some age x becomes increasingly more selective towards 'healthy' individuals. The resulting flattening of the mortality curve should be most pronounced for the population that faces the highest level of mortality, that is, in our example the male population. The male empirical pattern in Figure 1, however, does *not* reflect such a flattening of the male mortality curves or a marked decline of the male life-table aging rate until relatively old ages. Instead, a Gompertz model with a linear increase of $\log \mu(x)$ provides a very good fit across all adult and old ages in Bulgaria, especially for males, and a divergence between the Gompertz model and the observed pattern — similar to the one depicted in Figure 2(b) — is absent.

The case for an alternative to the relative-frailty Gompertz model is furthermore strengthened

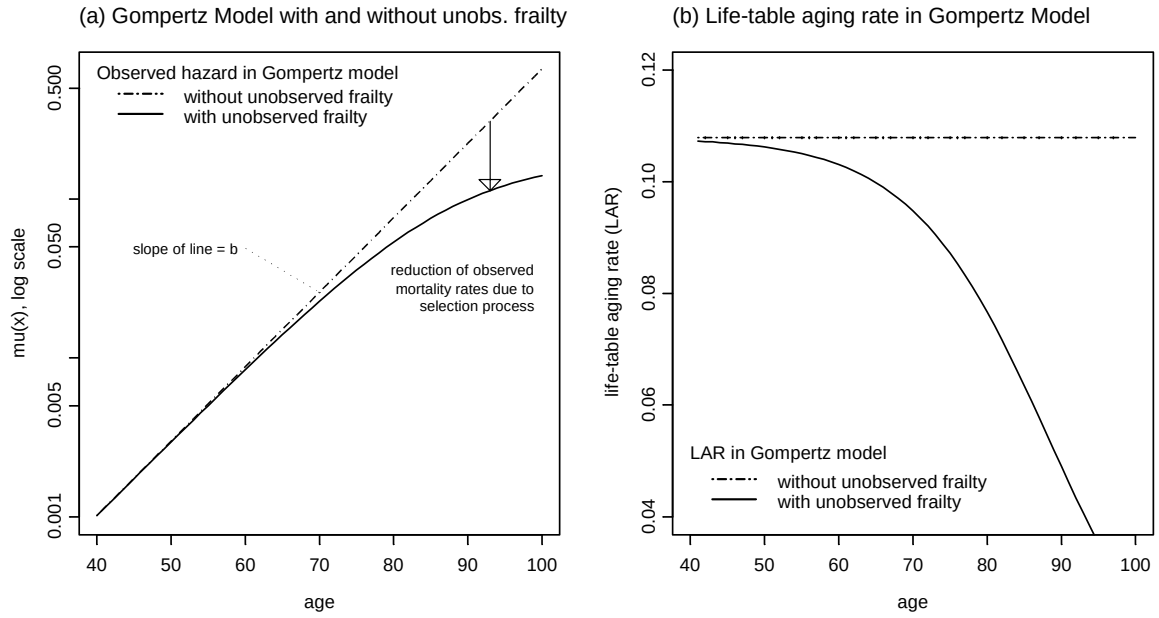

Figure 2: Left graph: Typical pattern of mortality rates implied by a Gompertz model with and without unobserved heterogeneity. Right graph: Life-table aging rate in a Gompertz model with and without unobserved heterogeneity.

by the deviations of the observed mortality pattern from the Gompertz model in Figure 1. As discussed above, the Bulgarian female mortality pattern exhibits a post-menopausal increase in the life-table aging rate and a decline at old and oldest-old ages, while the male pattern exhibits a modest increase in the life-table aging rate until about age 80. The observed life-table aging rate in the relative-frailty Gompertz model, however, exhibit a markedly different pattern. The life-table aging rate in this model attains it highest values at relatively young ages before mortality has affected the frailty composition of the population, and the life-table aging rate declines as the population becomes increasingly selected towards low-frailty individuals. The convex-concave pattern of mortality change observed in Figure 1, hence, is not implied by the relative-frailty Gompertz model in Figure 2 that exhibits a monotonously declining life-table aging rate.

A Gompertz model with relative frailty therefore does not provide a good explanation for the mortality pattern in Bulgaria. Two conflicting hypotheses can be considered in order to explain this apparent inability of the relative-frailty Gompertz model to replicate and explain the Bulgarian mortality pattern in Figure 1 and the convergence between male and female mortality levels:

(*a*) There are no unobserved differences in individuals' frailty in the Bulgarian population. A selection towards low-frailty individuals is thus absent and the male and female mortality pattern is characterized by systematically different mortality curves with different life-table aging rates. This implies that males and females in Bulgaria are characterized by a differential aging process. (*b*) The Bulgarian population is characterized by unobserved frailty, and the 'true' hazard $\mu(x)$ in equation (1), which applies to individuals with a constant frailty, is increasing faster than a Gompertz hazard so that the convergence between male and female mortality and the convex-concave pattern of the mortality hazard can result from the selection process in the population towards low-frailty individuals.

The argument given in (*a*), namely that the Bulgarian population is homogeneous with respect to frailty, seems rather unlikely and it contradicts most recent mortality research that points to important variations in the determinants of survival and longevity that are due to variation in genetic factors [Herskind et al. 1996, McGue et al. 1993] and early-life experiences [Barker 1992, Doblhammer 1999, Elo and Preston 1992, Horiuchi 1983]. The investigation of the second hypothesis (*b*) for Bulgaria, however, is difficult because the Gompertz hazard function, and most other commonly used hazard functions like Kannisto, Makeham, Logistic hazard functions [for a review of these models see Manton and Yashin 2000, Thatcher et al. 1998], do not allow a meaningful incorporation of unobserved frailty in order to explain the Bulgarian pattern in Figure 1(a). Moreover, the explanation of the male-female mortality convergence in Figure 1(a) using selection processes in heterogenous populations requires that the selective forces of mortality start operating already at adult ages. For instance, this relatively early onset of a selection of the population towards low-frailty individuals is supported by recent evidence from twin studies which suggest that unobserved heterogeneity is important for the estimation of mortality pattern at adult ages, and not only for mortality at old and oldest-old ages [Caselli et al. 2000, Iachine et al. 1998].

In this paper we therefore propose an alternative specification, a *modified DeMoivre hazard function*, that is suitable to investigate the hypothesis of whether the differential slopes of the mortality pattern between males and females in Bulgaria could merely be the result of a differentially strong selection process in heterogeneous populations.

2 A Modified DeMoivre Hazard Function

In an early attempt to describe mortality patterns with a mathematical formula, Abraham DeMoivre [DeMoivre 1725, p. 4, cited in Keyfitz and Smith 1977, p. 273] hypothesized that 'the number of lives existing at any age is proportional to the number of years intercepted between the age given

and the extremity of old age', i.e.,

$$l(x) = l_0(1 - \frac{x}{\omega}),$$

where ω is the maximum attainable age in the population. From this, the hazard rate, or the force of mortality at age x is defined as

$$\mu^D(x) = \frac{1}{\omega - x}, \tag{4}$$

where $\mu^D(x)$ denotes the DeMoivre hazard function. The hazard rate in this example increases towards infinity as x approaches the maximum attainable age ω. The existence of this maximum attainable age implies that the force of mortality increases faster than in the Gompertz model, especially as x approaches the maximum age ω [for a related discussion see Zelterman 1992].

For the application to contemporary mortality patterns the hazard function $\mu^D(x)$ in (4) is not sufficiently flexible. We therefore propose a modified DeMoivre hazard given by

$$\mu^{MD}(x) = a\left(1 - \frac{x}{\omega}\right)^{-b\omega}, \tag{5}$$

where $\mu^{MD}(x)$ is the *modified DeMoivre hazard function*. The survival curve $s^{MD}(x)$ corresponding to the above hazard function is given by

$$s^{MD}(x) = \exp\left[-\frac{a\omega}{b\omega - 1}\left((1 - \frac{x}{\omega})^{-(b\omega-1)} - 1\right)\right]. \tag{6}$$

The parameter b in (5) and (6) needs to satisfy $b > \frac{1}{\omega}$ in order that the hazard and survival curves are meaningful.

The hazard function $\mu^{MD}(x)$ defined in (5) has two limiting properties that render it a plausible and easily interpretable specification. First, the original DeMoivre hazard $\mu^D(x)$ emerges — up to a factor of proportionality — from the modified DeMoivre hazard $\mu^{MD}(x)$ when the product $b\omega$ in (5) approaches one. This occurs, for instance, when $b \to 1/\omega$ for a fixed maximum attainable age ω. Formally this property is represented as $\lim_{b\to 1/\omega} \mu^{MD}(x) = \lim_{b\omega\to 1} \mu^{MD}(x) = \frac{a}{b}\mu^D(x)$. Second, as the maximum attainable age ω becomes large, the modified DeMoivre hazard approaches the Gompertz hazard $\mu^G(x) = ae^{bx}$. That is, the hazard $\mu^{MD}(x)$ in (5) has the limit $\lim_{\omega\to\infty} \mu^{MD}(x) = \mu^G(x)$. The reason for this convergence to the Gompertz hazard is easily seen by noting that the limit $\lim_{\omega\to\infty}(1 - \frac{x}{\omega})^{\omega} = e^x$.

The interpretation of the parameters in the modified DeMoivre hazard is further facilitated by the fact that $\mu^{MD}(x)|_{x=0} = a$ and $\frac{d}{dx}\log\mu^{MD}(x)\big|_{x=0} = \frac{d}{dx}\log\mu^{G}(x)\big|_{x=0} = b$, which implies that the modified DeMoivre hazard agrees with a Gompertz hazard with equal parameters a and b at the age $x = 0$.

In heterogeneous populations with unobserved frailty the observed mortality rate differs from $\mu^{MD}(x)$ because the population becomes increasingly selected towards low-frailty individuals with age. If we assume a Gamma-distributed relative frailty model, then the observed hazard rate and survival curve implied by the modified DeMoivre model, denoted $\bar{\mu}^{MD}(x)$ and $\bar{s}^{MD}(x)$, follow directly from equations (1) and (2) as

$$\bar{\mu}^{MD}(x) = \frac{\mu^{MD}(x)}{1 - \sigma^2 \log s^{MD}(x)} \tag{7}$$

and

$$\bar{s}^{MD}(x) = (1 - \sigma^2 \log s^{MD}(x))^{-1/\sigma^2}, \tag{8}$$

where $\mu^{MD}(x)$ and $s^{MD}(x)$ are the hazard function and survival curve in equations (5) and (6) for individuals with a constant frailty $z = 1$.

Figure 3 plots the modified DeMoivre hazard function and the implied life-table aging rate for different values of σ^2 representing different degrees of unobserved heterogeneity in the population. In this figure we have used a maximum attainable age ω of 122.45 years, i.e., a ω that corresponds to Madame Jeanne Calment age at death, and we used the modified DeMoivre hazard function to model mortality during ages 40–100. [Note 2] The dashed-dotted line in both graphs represents the values for the modified DeMoivre hazard without unobserved frailty, or equivalently, for individuals with a constant frailty $z = 1$. Conditional on a constant frailty z the modified DeMoivre hazard increases faster than the exponential Gompertz hazard and the life-table aging rate is also an increasing function of age. As the age approaches the maximum attainable age ω, both the mortality hazard and the life-table aging rate approach infinity and the probability of surviving to ages larger than ω is zero.

The existence of such an upper limit to life-span is in contrast to the Gompertz model that does not imply a maximum attainable age and it is certainly controversial in view of the recent debate about the limits to the increase in life expectancy and particularly to the biological limits of life-span [Gavrilov and Gavrilova 1991, Manton and Stallard 1996, Vaupel et al. 1998, Wilmoth 1997, Wilmoth et al. 2000]. An emerging consensus in this debate seems to be that if upper limits to

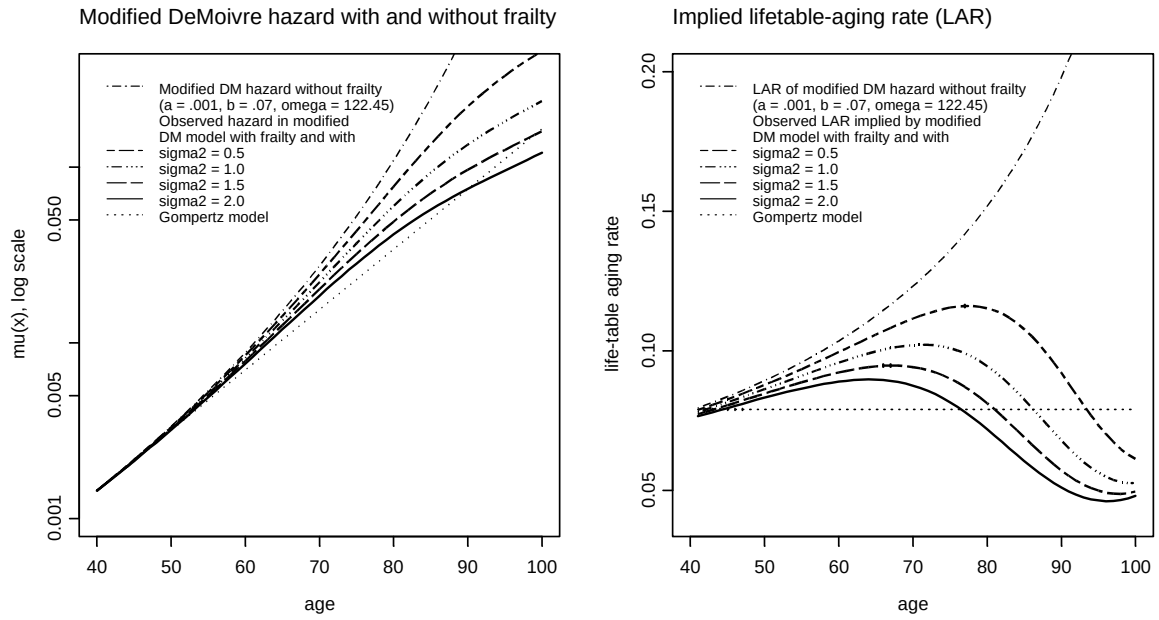

Figure 3: Observed mortality rates implied by the modified DeMoivre hazard function for different degrees of heterogeneity in the population

life-span exist, they need to be seen in a dynamic perspective [Carey and Judge 2000] and almost certainly do not represent immutable biological limits that are insurmountable by medical or social progress in survival to very old ages. At the same time, in any given socioeconomic environment the observed human life-span is finite and no human being has been documented to survive above Madame Jeanne Calment's age at death. Hence, while there probably does not exist an absolute biological limit to life-span, human life-span in any socioeconomic context may be limited and the changes of this upper limit to life in itself may constitute an interesting area of research [e.g., Carey and Liedo 1995, Wilmoth and Lundström 1996].

The mortality hazard in the modified DeMoivre model, conditional on a constant frailty z, increases faster than exponential and implies an increasing life-table aging rate with age. Despite this fact, the observed mortality pattern in a heterogeneous population can reflect a substantially different pattern. In Figure 3 we have therefore included the observed mortality hazard and life-table aging rate that is implied by the modified DeMoivre hazard with different degrees of unobserved heterogeneity in the population. Most strikingly, the observed mortality pattern does not necessarily represent the mortality pattern conditional on the frailty z [for an influential related discussion

see Vaupel and Yashin 1985], but reflects properties that are characteristic of the Bulgarian mortality pattern in Figure 1 and also that of other countries. For instance, the observed life-table aging rate between ages 40–100 can be initially increasing, and the extent of this increase in the life-table aging rate depends on the degree of heterogeneity and the level of mortality. At the same time, the relative increase in mortality by age decreases at old and oldest-old ages. Once age approaches the maximum attainable life-span ω, this decline in the life-table aging rate reverses again and the observed mortality rate and the life-table aging rate increase and ultimately approach infinity at ω. While this last implication for extremely old ages is controversial in view of the discussion on the limits to human longevity, the modified DeMoivre hazard in combination with unobserved heterogeneity can represent observed age-patterns of mortality for the quite wide age-range from adult to old and oldest-old ages during which most deaths in humans occur.

In the age range below age 100, therefore, the modified DeMoivre hazard function $\mu^{MD}(x)$ has several properties that make it a plausible choice for estimating frailty models in mortality. For instance, nonparametric estimations of the hazard function, which are feasible in bivariate frailty model applied to twin data [Yashin et al. 1995], have revealed that the hazard for individuals with constant frailty $z = 1$ during adult ages is increasing substantially faster than the Gompertz hazard, while the observed mortality rates can be approximated by a Gompertz or Logistic hazard function. In addition, detailed multivariate follow-up studies document a differential selection in male and female cohorts with respect to physiological characteristics and functional abilities that indicate an important role of selection processes for understanding male-female differences in mortality [Manton et al. 1995]. On the basis of these and related findings, Caselli et al. [2000, p. 8] have concluded that the 'correction for unobserved heterogeneity in demographic life tables may be needed not only for the oldest-old but also for the traditional interval of aging between 35 and 85 years of age, for which the observed trajectory of mortality appears to be well-described by a Gompertz curve.'

The modified DeMoivre hazard function in Figure 3 provides a possibility to estimate frailty models in the above situation. In particular, the observed hazard $\bar{\mu}^{MD}(x)$ as well as the baseline hazard $\mu^{MD}(x)$ in this Figure agree highly with the nonparametric estimates reported in Caselli et al. [2000] and Yashin et al. [1995]. Since nonparametric estimation is only feasible with special data, as for instance data on the mortality of twins or data with proportional-hazard covariates, the modified DeMoivre hazard introduced in this paper provides a suitable alternative for frailty modeling with vital statistics data. The hazard function introduced in this paper, therefore, allows the investigation of the 'selectivity hypotheses' for the convergence of male and female mortality in Bulgaria, and possibly also other countries, where relative-frailty Gompertz models with

unobserved heterogeneity do not yield an accurate description of the observed mortality dynamics.

Two basic approaches exist for the estimation of the unknown parameters a, b, ω and σ^2. First, we can choose a plausible value for ω, such as Madame Jeanne Calment's age at death. The remaining parameters a, b, σ^2 can then be estimated via a maximum likelihood using $\omega = 122.45$. This approach should be taken if mortality data at very old ages, say above age 110 that could shed light on the value of ω, are not available or reliable. In our experience, this procedure yields a quite robust and plausible estimation, and the results of the parameters a and b are not very sensitive to the choice of ω as long as it is chosen within a plausible age range, say 110 – 150 years.

The second possibility is to estimate *all four* parameters a, b, ω and σ^2 using maximum likelihood estimation. Since ω determines the convexity of the hazard function, and σ^2 influences the extent to which the increase in $\bar{\mu}(x)$ is flattened due to the selection process towards low-frailty individuals, the joint estimation of all four parameters is not feasible when only one mortality pattern is observed. It is, however, feasible if mortality patterns of several subpopulations, e.g., by sex or educational attainment, are analyzed. An example for this estimation is given in Section 3. Quite naturally, an effective estimation of the life-span ω requires reliable data at very high ages that may not be available in many countries. If the focus of the investigation is on adult and old-age mortality, e.g., as in our Bulgarian example on the age range 40–100, an estimation strategy that assumes a specific value for ω and then conducts a sensitivity analysis may be preferable to the estimation of ω from data that are censored at some upper age limit. If ω is nevertheless estimated in these cases, our experience suggests that the resulting 'best-fitting mortality curve' is often characterized by a too low maximum age ω, and a too high level of unobserved heterogeneity.

Independent of which approach regarding the specification of ω is chosen, the parameters a, b, and σ^2 can be functions of characteristics of individuals or subpopulations. In particular, we implement the estimation of the modified DeMoivre mortality model with

$$a(y_a) = \exp(\alpha_0 + \alpha_1 y_{a,1} + \alpha_2 y_{a,2} + \dots) \tag{9}$$

$$b(y_b) = \beta_0 + \beta_1 y_{b,1} + \beta_2 y_{b,2} + \dots \tag{10}$$

$$\sigma^2(y_\sigma) = \exp(\gamma_0 + \gamma_1 y_{\sigma,1} + \gamma_2 y_{\sigma,2} + \dots), \tag{11}$$

where y_a, y_b, y_σ are vectors of covariates that influence the level of the parameters a, b, and σ^2. For instance, in the next Section we will estimate a model where some of the parameters can differ by sex. This dependence of the parameters on sex is incorporated by including a dummy for females in y_a, y_b and/or y_σ. While this specific analysis is a relatively simple dependence of the parameters a, b, and σ^2 on covariates, considerably more complex specifications are also feasible.

3 Application of the Modified DeMoivre Method to Bulgaria

3.1 Estimation of the modified DeMoivre hazard

The Bulgarian mortality pattern for males and females during 1992–93 has already been depicted in Figure 1. In this Section we use the modified DeMoivre hazard function in combination with unobserved heterogeneity in order to evaluate whether the differential increase in mortality by age between males and females in Bulgaria can plausibly be explained by a selection process in which the higher male mortality level leads to a quicker selection of the male population towards low-frailty individuals.

Our analyses are based on an unique database for Bulgaria which is based on a linkage between the death records for the period 5th of December 1992 – 31st of December 1993 and the census on 4th of December 1992. This data-set is the first comprehensive population-based data on mortality in Bulgaria that includes a broad array of socioeconomic information for individuals who are at risk of death. The linkage of the death certificates to the census records is carried out using a personal identification number (PID) included in all official records of an individual in Bulgaria. The linkage between the census and the death registration is of high quality, and in total 92.67 per cent of all death certificates are linked to the census records. Among the linked deaths, 95.07 per cent are based on the PID number. Only 4.93 per cent of the deaths are linked using other identification variables (e.g., place and region of residence, birth day, sex, education, marital status) because the PID number is missing or incomplete. A relatively small fraction of 7.33 per cent of all deaths could not be linked to the census records. [Note 3]

The subsequent analyses include all individuals in Bulgaria who are at least 40 years old at census and are below age 100 either at death or on 31st December 1993. Our data thus comprise 1.87 million males of whom 57,221 die during the observation period, and 2.09 million females of whom 45,831 die during the 13 months after the census.

We first estimate a standard piecewise-constant hazard model (with constant hazards in two-year age intervals) separately for males and females in order to obtain a nonparametric estimate for the observed mortality pattern. The respective estimates have already been depicted in Figure 1(a), and they will also be included in subsequent Figures for comparison with our parametric estimates.

In Table 1 we report the results of different parametric specifications. Model 0 is a standard Gompertz model without unobserved heterogeneity that allows for male-female differences in both parameters a and b. These estimates have been used for the fitted Gompertz hazard curve in Figure 1(a). The estimates reveal that the 'level-parameter' a for females is only a fraction of about 0.27

of the respective parameter for males, while the 'slope-parameter' b for females exceeds that of males by 0.025. The latter difference implies that the relative increase of mortality rates by each year of age is 2.5 percentage points higher for females than for males and it leads to the — already discussed — strong convergence between the male and female mortality rates in Bulgaria.

Models 1–4 in Table 1 include a Gamma-distributed frailty and are based on a modified DeMoivre hazard function. Except for Model 4, where the maximum attainable age ω is estimated from the data, these models assume a ω which is equal to Madame Calment's age at death. The parameters a, b, and σ^2 are specified as in equations (9–11) using sex as the only covariate.

The simplest Model 1 estimates the male and female mortality pattern using identical parameter values b and σ^2 for both sexes, while the parameter a is allowed to vary between males and females in order to capture the different mortality levels. That is, the model allows for different levels of mortality by sex, but it assumes an equal 'slope-parameter' b across sex. The model therefore attempts to explain the differential increase in mortality with age merely by the selection hypothesis, i.e., the fact that the male population faces a more rapid selection towards low-frailty individuals due to the higher overall male mortality level. Most importantly, the model yields an estimate of $\hat{\sigma}^2 = \exp(-.6186) = 0.54$, indicating a quite substantial heterogeneity in the population. According to this estimate, about 28 per cent of the population at age 40 have a frailty of $z \leq .5$ and 9.7 per cent have a frailty of $z \geq 2$.

Figure 4(a) shows that this model traces the convergence between the observed male and female mortality rates with increasing age quite well (full lines), despite the fact that the mortality rates for a constant frailty $z = 1$ increase in a parallel fashion (dashed-dotted line). Hence, Model 1 contributes a substantial part of the observed convergence between male and female mortality rates to the differential strength of the selection process in the male and female population (a formal measurement of the fit of this model and a comparison with alternative Gompertz specifications are provided in the sensitivity analysis in Section 3.3 below).

Model 2 in Table 1 provides an extension of the above model and incorporates a potentially different degree of heterogeneity between the male and female populations. Possible reasons for such a differential variance in unobserved frailty could be a greater variation in life styles (such as smoking habits or other risky behaviors) among the male as compared to the female population. Indeed, the estimates in Table 1 suggest that the male variance of frailty is $\sigma^2_{male} = \exp(-.4876) = 0.61$, while the corresponding variance for females is 35% smaller ($\sigma^2_{female} = 0.40$). This implies that about 30 per cent of males, but only 22 per cent of females, have a low frailty of $z \leq .5$, and more than 10 per cent of males, but about 8 per cent of females, have a high frailty of $z \geq 2$ at age 40.

Table 1: Estimates of a Gompertz model without frailty and modified DeMoivre hazard models with Gamma-distributed frailty (D_{female} denotes a dummy variable for females)

Data	**Male and Female Mortality in Bulgaria, 1992–93 Age 40–95**				
Method	**Gompertz hazard, no frailty**	**Modified DeMoivre hazard function with Gamma-distributed relative frailty**			
	Model 0	*Model 1*	*Model 2*	*Model 3*	*Model 4*
Specification for $a = \exp(\alpha_0 + \alpha_1 D_{female})$					
α_0 (Constant)	-5.5713 (0.0117)*	-5.6350 (0.0110)*	-5.5686 (0.0114)*	-5.4365 (0.0129)*	-5.5249 (0.0116)*
α_1 (Female)	-1.3205 (0.0198)*	-0.7302 (0.0100)*	-0.8385 (0.0117)*	-1.1086 (0.0182)*	-0.9163 (0.0139)*
Specification for $b = \beta_0 + \beta_1 D_{female}$					
β_0 (Constant)	0.0825 (0.0004)*	0.0755 (0.0004)*	0.0749 (0.0004)*	0.0687 (0.0005)*	0.0740 (0.0005)*
β_1 (Female)	0.0254 (0.0006)*	–	–	0.0103 (0.0004)*	–
Specification for $\sigma^2 = \exp(\gamma_0 + \gamma_1 D_{female})$					
γ_0 (Constant)	–	-0.6186 (0.0239)*	-0.4876 (0.0234)*	-0.7411 (0.0263)*	0.0960 (0.0281)*
γ_1 (Female)	–	–	-0.4285 (0.0255)*	–	–
Specification for ω					
ω	–	set to 122.45	set to 122.45	set to 122.45	104.2420 (0.3475)*

Notes: Standard errors in parentheses. *p-values:* $^{+}$ $p < 0.01$; * $p < 0.001$. Age has been rescaled in the estimation so that the parameter a reflects the mortality level of the population at age 40, and σ^2 indicates the variance of unobserved frailty at this age.

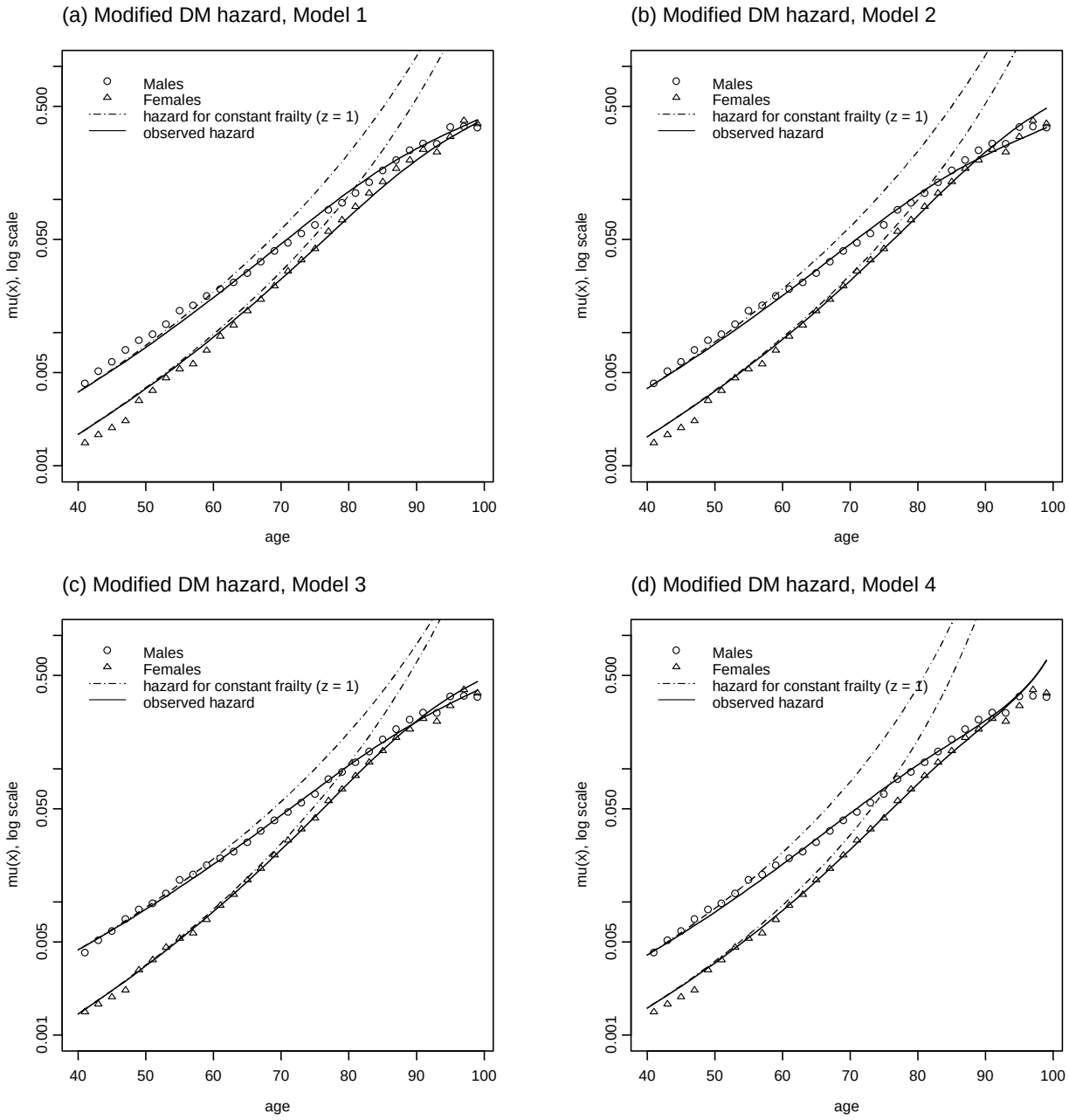

Figure 4: Application of modified DeMoivre hazard functions with Gamma distributed relative frailty to Bulgarian mortality pattern

The resulting fit of the Model 2 is depicted in Figure 4(b). Since this model allows for an additional parameter, it fits the Bulgarian mortality pattern slightly better than our earlier model. Differences in the observed slope of the mortality pattern in Model 2 are again only due to differences in the overall mortality level and in subsequent differences in the selection process in a heterogeneous population. The present model therefore suggests that the male population may be more heterogeneous with respect to various biological or socioeconomic determinants of mortality. The specific investigation of this issue is beyond the scope of the present paper, but it is feasible on the basis of our data that includes a broad range of covariates about individuals. Most important in the present context is that Model 2 further confirms our argument that differential selection process provides a plausible explanation for the male-female difference in the increase of mortality with age.

An alternative generalization of our initial estimation is provided in Model 3 in Table 1, where the parameter b, instead of the variance σ^2, is allowed to vary across sexes. The coefficients show that the relative difference in the 'slope-parameter' b between males and females is substantially reduced by incorporating unobserved heterogeneity as compared to the Gompertz model without any frailty considerations. This finding is again consistent with an important male-female difference in the strength of selection towards low-frailty individuals in the population. The fit of this model is given in Figure 4(c), where this model performs slightly better than the two earlier models. However, this improvement is not surprising since the specification of a separate slope-parameter b for males and females provides a direct modelling of the differential mortality increase between sexes.

Finally, model 4 in Table 1 estimates the maximum attainable age ω in addition to the remaining parameters a, b, and σ^2. While our earlier models were based on a predetermined ω of 122.45 years, the present estimate reveals a ω of slightly above 104 years. Moreover, the variance of the unobserved frailty has increased to $\sigma^2 = 1.10$ (with $\sigma^2 = 1.10$, 11 per cent of the population have a frailty of $z \leq .1$, 41 per cent have a frailty of $z \leq .5$, and 14 per cent have a frailty of $z \geq 2$). While these parameter estimates yield a 'best-fitting model' in Figure 4(d), the estimate for ω is not plausible and it depends strongly on the age at which the data are censored. For instance, if the data include individuals who survive to age 105 or 110, then the respective estimates for ω increase respectively to 108 and 111.5. and that for σ^2 decline to 0.91 and 0.79. The estimates for the parameters a and b are relatively insensitive to changes in the age range above 100. Hence, while Model 4 provides the best fit of the Bulgarian mortality data using a modified DeMoivre hazard with no sex-differences in the slope parameter b or the variance of frailty σ^2, the estimates of this model about ω cannot be interpreted in terms of a maximum life-span because the estimation was deliberately censored at age 100 in order to focus on the convergence of male and female mortality

in the age-range 40–100. Without survivors to very old ages, however, an estimation that assumes a plausible value for ω as in Models 1–3 seems preferable to the direct estimation of ω from the data. It is beyond the scope of this paper to apply the modified DeMoivre hazard explicitly to reliable mortality data at ages 100+, but these future applications provide a possibility to estimate interpretable and realistic values for the maximum attainable age ω and the changes of this limit to life-span over time.

3.2 Comparison with piecewise-constant hazard with unobserved frailty

In order to assess the empirical plausibility of the modified DeMoivre hazard model we compare the above results with a nonparametric estimation of the hazard curve. This nonparametric alternative is feasible if the mortality hazards for males and females, conditional on the frailty z, differ only by a factor of proportionality. In this case it is possible to combine Gamma-distributed relative frailty with a piecewise-constant hazard function $\mu^{PW}(x)$ that does not impose parametric restrictions on the shape of the mortality pattern [Note 4]. This nonparametric estimation of the baseline-hazard $\mu^{PW}(x)$ can then be compared to the modified DeMoivre hazard $\mu^{MD}(x)$ in order to assess the implications of the parametric assumptions used in the previous section.

The left graph in Figure 5 shows the observed male and female mortality level in Bulgaria along with the estimated baseline hazard $\mu^{PW}(x)$ for individuals with a constant frailty $z = 1$ and the corresponding observed hazard $\bar{\mu}^{PW}(x)$ obtained from the piecewise-constant estimation. The model fits the observed male and female mortality pattern relatively well and the fit is comparable to our earlier Model 4 in Figure 4.

The primary question regarding the estimation of this piecewise-constant frailty model in Figure 5(a) is whether the respective estimates for the variance of the frailty distribution and the increase of the mortality rates for a constant frailty $z = 1$ are consistent with our knowledge about human mortality. In order to investigate this issue, we compare in Figure 5(b) the estimated baseline hazards $\mu(x)$ obtained from the piecewise-constant and the modified DeMoivre model.

The graph reveals that the piecewise-constant specification yields the fastest increasing baseline hazard across all estimated models, and it also yields the highest variance σ^2 for unobserved frailty ($\sigma^2 = 1.57$). The differences are largest between the piecewise-constant model and the DeMoivre models with $\omega = 122.45$ (Models 1 and 2 in Figures 4 and 5b), while it is only modest when compared to the DeMoivre model where ω is estimated from the data (Model 4 in Figures 4 and 5b). Since the last model already implied an implausible value of only 104 years for the highest attainable age ω, we also consider the increase of the baseline hazard obtained from the piecewise-

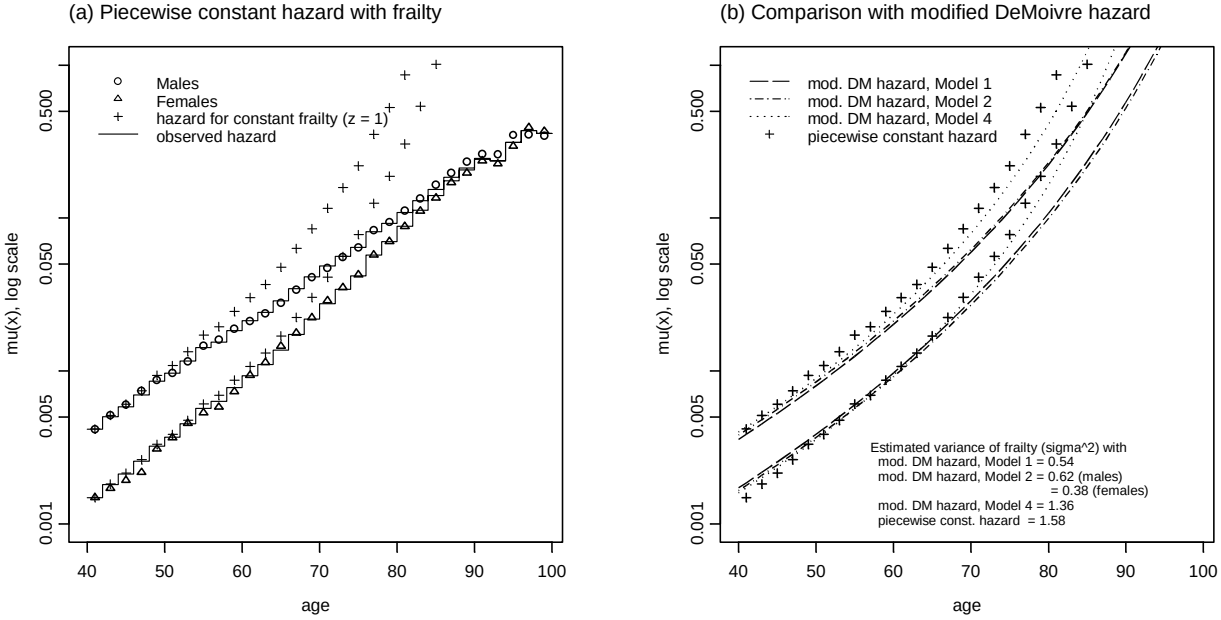

Figure 5: Comparison of a piecewise-constant hazard model with unobserved frailty with the modified DeMoivre model with unobserved frailty

constant estimation as too steep. Similarly, the variance of the frailty distribution in the piecewise-constant model seems unrealistically high since it suggests that about 19 per cent of the initial population have a frailty of $z \leq 0.1$ and about 15 per cent have a frailty of $z \geq 2$. However, Iachine et al. [1998] obtained similarly large values for the variance of unobserved frailty from twin data.

While the nonparametric estimation of the baseline hazard with a piecewise-constant model certainly has its virtues, it can also lead to estimates of the baseline-hazard that are implausible. The modified DeMoivre model with ω set to 122.5 years provides a possibility to restrict the increase of the baseline hazard with age to values that are consistent with observed survival to very old ages. Moreover, the differences between the observed and estimated mortality pattern in the simplest DeMoivre model in Figure 4(a) suggest alternative specifications that are not feasible with the piecewise-constant estimation. For instance, in Models 2 and 3 in Figure 4 we allow the variance of the frailty distribution or the slope-parameter b to be different between males and females. Both extensions substantially improve the fit of the model. In our opinion these extensions provide a more plausible, and probably also more accurate description of the mortality dynamics than the

piecewise-constant analysis. The modified DeMoivre function therefore provides a very suitable hazard function for the application of frailty models to adult ages, and it allows specifications of the parameters that are not available with nonparametric estimations.

3.3 Sensitivity analysis of estimated parameters

In this Section we provide a sensitivity analysis in order to investigate the extent to which the coefficients obtained from the modified DeMoivre hazard function depend on the choice of the maximum attainable age ω. Our earlier estimates of Models 1–3 in Table 1 and Figure 4 were based on $\omega = 122.45$, i.e., Madame Calment's age at death. In order to evaluate the sensitivity of our estimates with respect to the choice of ω, we reestimate Models 1 and 2 in Table 1 using a ω that ranges from 105 to 150 years (we do not report the sensitivity analysis for Model 3 since it leads to similar results).

The top-left graph in Figure 6 shows the relative deviation of the estimated coefficients a_{male}, a_{female}, b, and σ^2 as a function of ω. The top-right graph in this Figure shows the goodness-of-fit of the Model 1 as a function of ω. The goodness-of-fit is calculated as $1 - RSS/SST$, where RSS is the residual sum of squares on the logarithmic scale and SST is the total sum of squared deviations from the mean on the logarithmic scale.

The Figure shows that the parameters a and b are relatively insensitive to the specification of ω, while the estimated variance of frailty depends quite strongly on the choice of ω. The latter is not very surprising since ω determines the convexity of the hazard function $\mu^{MD}(x)$ in equation (5), i.e., the hazard for individuals with a constant frailty $z = 1$. A low ω implies a quite strongly increasing hazard $\mu^{MD}(x)$ with age. This subsequently results in a higher estimate for σ^2 in order for the model to fit to the observed mortality rate.

The top-right graph in Figure 6 shows a goodness-of-fit analysis of Model 1, i.e., a model that allows only for level-differences in mortality across sex but no differences in the 'slope-parameter' b The dash-dotted lines in this graph reveal on one hand the fit of the Gompertz model with separate parameters a and b for males and females (i.e., Model 0 in Table 1), and on the other hand the Gompertz model with b constrained equal across sexes. The fit of Model 1 is between these two benchmarks, and it tends to decrease the higher is the value for ω. For low values of ω, Model 1 fits almost as good as Model 0 with no parameter restrictions across sex. With increasing ω this goodness-of-fit declines. Ultimately it approaches the lower dash-dotted line because a rising ω renders the modified DeMoivre hazard more and more like a Gompertz model. However, for values of ω below 140 years the goodness-of-fit of Model 1 is closer to the Gompertz model with

separate slope-parameters for males and females than that of the Gompertz model with equal b for both sexes.

The sensitivity of the estimates for σ^2, therefore, is not as severe as the top-left graph in Figure 6 may suggest. First, the model with $\omega = 122.45$, i.e., Madame Calment's age at death, provides the best-fitting which is based on a maximum attainable age ω that is at least as high as the highest age lived by any person so far. Second, for a quite broad range of plausible choices for ω, say, between 115 and 130 years, the main conclusion of Model 1 remains unaltered: a substantial part of the male-female differences in the slope of the observed mortality pattern can be explained by a differential strength of the selection process towards low-frailty individuals that is caused by differences in the overall level of mortality between males and females.

The bottom-left and bottom-right graph in Figure 6 show the corresponding sensitivity analysis for Model 2 in Table 1. Similar to Model 1, the estimates for a and b are not very sensitive with respect to the choice of ω, while the estimates for σ^2_{male} and σ^2_{female} change substantially with ω. The analysis in the bottom-right graph, however, reveals that these changing estimates for σ^2-parameters leave the goodness-of-fit of the model almost unaffected.

The choice of ω in Model 2 is thus not essential for the main conclusion of the analysis: A modified DeMoivre hazard model with only one 'slope-parameter' b for both males and females provides a very good description of the Bulgarian mortality pattern. Moreover, this model attributes a substantial part of the differential male-female increase in mortality by age to a differential strength of the selection process which is caused by the higher overall level of male mortality.

Our preferred choice for ω in analyses with the modified DeMoivre hazard is $\omega = 122.45$ years based on Madame Calment's age at death. While the specific estimates for σ^2 are sensitive to this choice, the primary conclusion resulting from the incorporation of frailty in the analysis of Bulgarian mortality is very robust with respect to this specification.

4 Conclusions

The incorporation of unobserved frailty in the analysis of mortality at adult ages is of considerable theoretical and empirical interest. For instance, the estimation of relative frailty models sheds light on questions of whether a convergence in the mortality rates of different subpopulations, or a 'flattening' of the mortality curve with age, can be explained by a selection process towards low-frailty individuals in heterogeneous populations.

While frailty models have been primarily applied to old and oldest-old mortality, recent evidence from twin studies suggests that the consideration of unobserved heterogeneity among in-

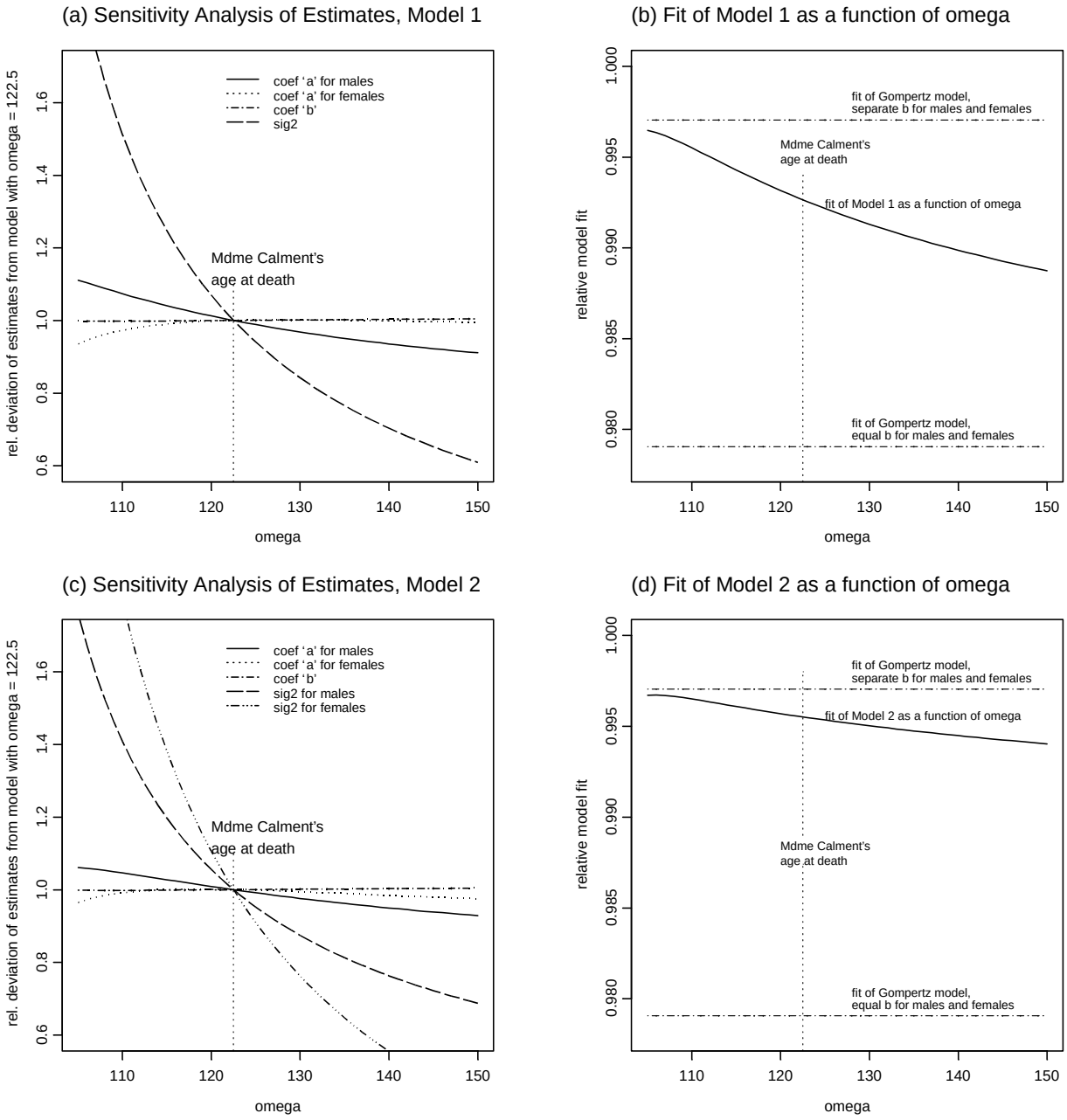

Figure 6: Sensitivity analysis of the parameter estimates obtained for Models 1 and 2 in Table 1 with respect to different values of ω

dividuals in their susceptibility to death is also relevant for the traditional age range from, say, 40–100 years. The application of relative frailty models to these ages, however, is often hampered by the fact that the observed mortality pattern across adult ages is well-described by a Gompertz curve. In this case the estimation of the usual frailty models, which are based on Gompertz or similar hazard functions, fails because the observed pattern lacks the typical 'flattening' of the mortality curve that results from the selection towards low-frailty individuals in a heterogeneous population.

In this paper we propose a *modified DeMoivre hazard function* which is suitable for the estimation of frailty models of mortality for adult ages, e.g., between 40–95 years. The hazard conditional on a constant frailty in this specification increases faster than exponential. The observed hazard, on the other hand, can capture a broad range of mortality patterns that are commonly encountered in the analysis of mortality at adult ages.

We apply the Gamma-distributed relative frailty model with a modified DeMoivre hazard to male and female mortality in Bulgaria during 1992–93. The two characteristic features of this mortality pattern are a substantial difference in the level of adult mortality between males and females, and a considerably steeper increase of mortality with age for females than for males. Our analyses show that a substantial part of this differential increase of mortality can be explained by a differential selection process in the male and female population. Since the level of mortality is higher for males than for females, the male population faces an earlier and stronger selection towards low-frailty individuals, and the resulting differences in the frailty composition of the population at older ages can explain the convergence between male and female mortality.

This finding of our analyses is robust across different specifications for the hazard function. Moreover, sensitivity analyses and a comparison with nonparametrically estimated hazard functions show that the modified DeMoivre hazard function leads to plausible and relatively robust estimates.

In summary, this paper provides a new substantive and methodological approach to the understanding of mortality at adult and old ages. First, we propose a new hazard function that implies, conditional on frailty, a faster than exponential increase of mortality with age. This model is therefore suitable for the estimation of frailty models at ages between 40–100 years, i.e., the age range where Gompertz or logistic models often yield a reasonable description of the observed mortality pattern. Second, we argue that unobserved heterogeneity provides a plausible explanation of the adult mortality pattern in Bulgaria — and possible also other countries — during the early 1990s. Our estimations using the modified DeMoivre hazard function suggest that the stronger selection process towards low-frailty individuals in the male population, caused by an overall higher level

of mortality, may constitute a primary mechanism leading to the convergence of male and female mortality at higher ages. This finding implies that the convergence is not necessarily caused by a differential process of aging across sexes, but is merely a consequence of the different levels of mortality, and a subsequently different selection process, for males and females.

Acknowledgments: We are indebted to James W. Vaupel, Francesco Billari, Tan Qihua, Heiner Maier, the members of the Laboratory of Survival and Longevity at the Max Planck Institute for Demographic Research and two anonymous referees for many helpful comments and suggestion that have greatly improved this paper. We gratefully acknowledge the support received for this research by the Max Planck Institute for Demographic Research and the National Statistical Institute (NSI) in Bulgaria. The Bulgarian mortality data has been made available by the National Statistical Institute of Bulgaria, and we are greatly indebted to Ekaterina Arnaudova, Ivan Balev and Mariana Dimova for their help and support in establishing this database.

Notes

1. The estimates for the life-table aging ratio have been smoothed using Kalman-Filter techniques.

2. Madame Jeanne Calment was born on the 21st of February 1875 and died on the 4th of August 1997 at the age of 122 years, 5 months and 14 days. Madame Jeanne Calment age at death is currently the highest verified age at death of a person.

3. In the period 5th of December 1992 – 31st of December 1993 $116,611$ deaths occured in Bulgaria. A relatively small fraction of $8,541$ deaths $(7.33\ \%)$, could not be linked to the census records (4,142 (48.50%) of the unlinked deaths are males, and 4,399 (51.50%) are females). Although the death certificate in Bulgaria contains some limited information on the socio-economic status of dead persons (e.g., education at death, marital status at death, place of residence at death), we omit the unlinked deaths from our analysis mainly for two reasons: (*a*) We do not know whether the persons corresponding to unlinked death certificates have participated the census, and their personal data were wrongly coded in the census records, or whether they are subject to census undercount. This latter possibility is supported by the fact that some relevant minorities in Bulgaria are more likely to be a subject of undercount, which is indicated by the fact that most of the unlinked deaths involve persons with low education (76.56% of the unlinked male, and 86.63% of the unlinked female deaths). (*b*) The unlinked deaths can also correspond to return migrants whose prevalence has increased after the first waves of rapid emigration in the early 1990s. We believe that census undercount and return migration are the most important factors leading to unlinked death certificates. If the persons corresponding to the unlinked deaths have not participated in the census due to the above reasons, then excluding them from the analyses is appropriate. If they have participated in the census, but the death cannot be linked due to incomplete identification numbers, our estimates about the mortality level are biased downward. However, we have no reasons to believe that this bias is highly sex-specific.

Therefore, the Bulgarian age and sex-specific patterns of mortality presented in this paper cannot be attributed to a bias caused by unlinked deaths because the overall numbers of unlinked deaths is quite small and is evenly distributed across sexes. Moreover, the age distribution of the linked and unlinked deaths is relatively similar for males, while it is slightly shifted to higher ages for females. For instance, the exact mean age at death of unlinked male deaths is 67.2 years, which is slightly below that of linked male deaths of 67.8 years, and female unlinked deaths have a mean

age of 76.7, which is above that of linked deaths with a mean age of 73.2.

Age reporting is of relatively high quality in the Bulgarian census. The age of a person is coded in the personal identification number (PID), which was assigned to the individuals either during 1976–78 for those alive at this time, or at birth for those born afterwards. The age in the census form is taken from the age coded in the PID number. Since the old and oldest-old in 1992 have received their PID about 15 years earlier, when the propensity to age-misreporting may have been lower due to younger ages and careful administrative checking, the age information in our census data should be of high quality even for the old population.

4. The piecewise-constant proportional hazard model with unobserved frailty is specified as follows. Consider the age-intervals $(c_0, c_1], \ldots, (c_{j-1}, c_j], \ldots, (c_{K-1}, c_K]$ that separate the observed age range into K disjoint intervals. Then assume that the mortality hazard, conditional on a frailty $z = 1$ and the observed covariates y_a, is constant within each of these age intervals and equals $a(y_a)\mu_j$ for $x \in (c_{j-1}, c_j]$. In this specification μ_j is the mortality hazard prevailing in the age interval $(c_{j-1}, c_j]$, $j = 1, \ldots, K$, and $a(y_a)$ is the factor of proportionality for individuals with characteristics y_a. Denote as $s^{PW}(x)$ the corresponding survival function at age x. The observed hazard at age x in a heterogeneous population with Gamma-distributed relative frailty then equals

$$\bar{\mu}^{PW}(x) = \frac{a(y_a)\mu_j}{(1 - a(y_a)\sigma^2 \log s^{PW}(x))} \qquad \text{for } x \in (c_{j-1}, c_j]. \tag{12}$$

Because of the numerical difficulties in estimating this piecewise-constant hazard function via maximum likelihood in the presence of many age-intervals and large data, we implement a slight approximation to the hazard function (12). In particular, the difficulties in the estimation arise because the hazard $\bar{\mu}^{PW}(x)$ is not constant within age intervals. This results from the fact that the value of the survival function $s^{PW}(x)$ in the denominator declines with age x. For sufficiently small age-intervals, however, the effect of this changing value of the survival function on the observed hazard $\bar{\mu}^{PW}(x)$ within an age interval is small. The piecewise-constant hazard function with relative frailty can therefore be approximated by replacing the value of the survival function $s^{PW}(x)$ in (12) with the value of the survival function at the mid-point of each age-interval. With this approximation, the observed hazard $\bar{\mu}^{PW}(x)$ is constant within age intervals and the MLE estimation is substantially simplified. We estimate this piecewise-constant hazard model with a constant mortality risk within two-year age intervals.

References

Baltes, M. M., A. M. Freund, and A. L. Horgas (1999). Men and women in the Berlin aging study. In P. B. Baltes and K. U. Mayer (Eds.), *The Berlin Aging Study. Aging from 70 to 100*, pp. 259–281. Cambridge: Cambridge University Press.

Barker, D. J. P. (1992). *Fetal and Infant Origins of Adult Disease*. London: Britisch Medical Journal.

Carey, J. R. and D. S. Judge (2000). Mortality dynamics of aging. *Generations - Journal of the American Society on Aging 24*(1), 19–24.

Carey, J. R. and P. Liedo (1995). Sex-specific life table aging rates in large medfly cohorts. *Experimental Gerontology 30*(3/4), 315–325.

Carnes, B. A., S. J. Olshansky, and D. Grahn (1996). Continuing the search for a law of mortality. *Population and Development Review 22*(2), 231–264.

Caselli, G., J. W. Vaupel, and A. I. Yashin (2000). Longevity, heterogeneity and selection. Paper presented at the XL Riunione Scientifica della Societa Italianà di Statistica, Florence, April 26–28.

Christensen, K., M. Kristiansen, H. Hagen-Larsen, A. Skytthe, L. Bathum, B. Jeune, K. Andersen-Ranberg, J. W. Vaupel, and K. H. Orstavik (2000). X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. *Blood 95*(7), 2449–2451.

DeMoivre, A. (1725). *Annuities on Lives: Or, the Valuation of Annuities Upon Any Number of Lives; as also of Reversions*. London.

Doblhammer, G. (1999). Longevity and month of birth: Evidence from Austria and Denmark. *Demographic Research [Online] Available http://www.demographic-research.org/Volumes/Vol1/3 1*(3).

Elo, I. T. and S. H. Preston (1992). Effects of early-life conditions on adult mortality: A review. *Population Index 58*(2), 186–212.

Gavrilov, L. A. and N. S. Gavrilova (1991). *The Biology of Life Span. A Quantitative Approach*. London: Harwood Academic Publishers.

Hazzard, W. R. (1986). Biological basis of the sex differential in longevity. *Journal of the American Geriatrics Society 34*, 455–471.

Hazzard, W. R. and D. Applebaum-Bowden (1989). Why women live longer than men: The biologic mechanism of the differential in longevity. *Transactions of the American Clinical and Climatological Association 101*, 168–189.

Herskind, A. M., M. McGue, N. V. Holm, T. I. A. Sorensen, B. Harvald, and J. W. Vaupel (1996). The heritability of human longevity: A population based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics 97*(3), 319–323.

Horiuchi, S. (1983). The long-term impact of war on mortality: Old-age mortality of the First World War survivors in the Federal Republic of Germany. *Population Bulletin of the United Nations 15*, 80–92.

Horiuchi, S. (1997). Postmenopausal acceleration of age-related mortality increase. *Journal of Gerontology: Biological Sciences 52A*(1), B78–B92.

Horiuchi, S. and A. J. Coale (1990). Age patterns of mortality for older women: An analysis using the age-specific rate of mortality change with age. *Mathematical Population Studies 2*(4), 245–267.

House, J. S., R. C. Kessler, and A. R. e. Herzog (1990). Age, socioeconomic status, and health. *The Milbank Quarterly 68*(3), 383–411.

Hummer, R. A., R. G. Rogers, and I. W. Eberstein (1998). Sociodemographic differentials in adult mortality: A review of analytic approaches. *Population and Development Review 24*(3), 553–578.

Iachine, I., N. V. Holm, J. R. Harris, A. Z. Begun, M. K. Iachina, M. Laitinen, J. Kaprio, and A. I. Yashin (1998). How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins. *Twin Research 1*(4), 196–205.

Keyfitz, N. and D. W. E. Smith (1977). *Mathematical Demography: Selected Papers.* Berlin: Springer Verlag.

Kohler, I. (2000a). Mortality dynamics in Bulgaria and Russia: Age, period and cohort approach. Paper presented at the annual meeting of the Population Association of America, Los Angeles, CA, March 23–25.

Kohler, I. (2000b). Structural changes in Bulgarian mortality and their socioeconomic determinants. Paper presented at the annual meeting of the Population Association of America, Los Angeles, CA, March 23–25.

Manton, K. G. and E. Stallard (1996). Longevity in the United States: Age and sex-specific evidence on life span limits from mortality patterns 1960-1990. *Journals of Gerontology Series A - Biological Sciences and Medical Sciences 51*(5), B362–B375.

Manton, K. G., M. A. Woodbury, and E. Stallard (1995). Sex differences in human mortality and aging at late ages: the effect of mortality selection and state dynamics. *The Gerontologist 35*(5), 597–608.

Manton, K. G. and A. I. Yashin (2000). *Mechanisms of Aging and Mortality: The Search for New Paradigms*. Odense: Monographs on Population Aging 7, Odense University Press.

McGue, M., J. W. Vaupel, N. V. Holm, and B. Harvald (1993). Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *Journal of Gerontology, Biological Sciences 48*(6), 237–244.

Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998). *The Force of Mortality at Ages 80 to 120*. Monographs on Population Aging 5, Odense University Press.

Vaupel, J. W., J. R. Carey, K. Christensen, T. E. Johnson, A. I. Yashin, N. V. Holm, I. A. Iachin, V. Kannisto, A. A. Khazaeli, P. Liedo, V. D. Longo, Y. Zeng, K. G. Manton, and J. Curtsinger (1998). Biodemographic trajectories of longevity. *Science 280*, 855–860.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography 16*(3), 439–454.

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

Waldron, I. (1985). What do we know about causes of sex differences in mortality? A review of the literature. *Population Bulletin U.N.* (18), 59–76.

Wilmoth, J. R. (1997). In search of limits. In K. W. Wachter and C. E. Finch (Eds.), *Between Zeus and the Salmon: The Biodemography of Longevity*, pp. 38–64. Washington, D.C.: National Academy Press.

Wilmoth, J. R., L. J. Deegan, H. Lundström, and S. Horiuchi (2000). Increase of maximum life-span in Sweden, 1861-1999. *Science 289*(5488), 2366–2368.

Wilmoth, J. R. and H. Lundström (1996). Extreme longevity in five countries. *European Journal of Population 12*, 63–93.

Yashin, A. I., J. W. Vaupel, and I. A. Iachine (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies 5*(2), 145–59.

Zelterman, D. (1992). A statistical distribution with an unbounded hazard function and its application to a theory from demography. *Biometrics 48*(3), 807–818.