LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386

Rieder:

Estimation of Mortalities

Sonderforschungsbereich 386, Paper 28 (1996)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

MAX-PLANCK-GESELLSCHAFT

Estimation of Mortalities

Helmut Rieder
Department of Mathematics
University of Bayreuth *

May 23, 1996

Abstract

If a linear regression is fit to log-transformed mortalities and the estimate is back-transformed according to the formula $\mathrm{E}\, e^{Y} = e^{\,\mu+\sigma^2/2}$ a systematic bias occurs unless the error distribution is normal and the scale estimate is gauged to normal variance. This result is a consequence of the uniqueness theorem for the Laplace transform.

We determine the systematic bias of minimum–L_2 and minimum–L_1 estimation with sample variance and interquartile range of the residuals as scale estimates under a uniform and four contaminated normal error distributions. Already under innocent looking contaminations the true mortalities may be underestimated by 50% in the long run.

Moreover, the logarithmic transformation introduces an instability into the model that results in a large discrepancy between *rg_Huber* estimates as the tuning constant regulating the degree of robustness varies.

Contrary to the logarithm the square root stabilizes variance, diminishes the influence of outliers, automatically copes with observed zeros, allows the 'nonparametric' back-transformation formula $\mathrm{E}\, Y^2 = \mu^2 + \sigma^2$, and in the homoskedastic case avoids a systematic bias of minimum–L_2 estimation with sample variance.

For the company–specific table 3 of [Loeb94], in the age range of 20–65 years, we fit a parabola to root mortalities by minimum–L_2, minimum–L_1, and robust *rg_Huber* regression estimates, and a cubic and exponential by least squares. The fits thus obtained in the original model are excellent and practically indistinguishable by a χ^2 goodness-of-fit test.

Finally, dispensing with the transformation of observations, we employ a Poisson generalized linear model and fit an exponential and a cubic by maximum likelihood.

* Part of the work was done while the author was visiting the Statistics Department, University of Munich.

1 Introduction

The determination of true mortalities π_x at age x of a certain population on the basis of the observed mortalities q_x is a classical topic of insurance mathematics, which may be treated either numerically or statistically.

Numerical Smoothing vs. Statistical Estimation

The numerical method, on the one hand, employs some spline smoothing (Reinsch, Whittaker–Henderson), minimizing a combination of weighted least squares and a roughness penalty. Without requiring any model assumptions it produces smooth fits that may reflect several substructures.

The statistical method, on the other hand, needs to specify a suitable model for the function π_x and the error structure (the law of the deviations $q_x - \pi_x$), and then estimates the model parameters according to some statistical criteria. Examples include a weighted least squares estimation of a Gompertz–Makeham curve $-\pi_x \approx \log(1-\pi_x) = a + bc^x$ to $\log(1-q_x)$, by Knight and Hardy (1908/9), and the weighted minimum χ^2 fit of $\pi_x = a+bc^x$ to q_x by Cramèr and Wold (1935),

$$\sum_x L_x \frac{(q_x - \pi_x)^2}{q_x} = \min_{a,b,c}! \tag{1.1}$$

Here and subsequently, L_x denotes the size of the population at age x.

As for these (rather old) references in particular, and standard textbook accounts of the subject in general, we refer to [Wolff70] and [BePo82]. Judging from the more recent article [Loeb94], the model-free numerical method seems to prevail in German actuarial practice.

Company–Specific Small Populations

For small populations with scanty data and large variability one might expect the numerical smoothing in trouble and to produce wiggly curves. This is of course not necessarily the case since fit and smoothness can be determined by the choice of fit criterion and roughness penalty and can be balanced by their weights. In fact, the standard numerical fit employed in [Loeb94] turns out smooth for all three tables considered there.

The attractiveness of the statistical method consists in the possibility of obtaining a smooth and simply structured fit by specifying a suitable model and estimating but a few parameters. In this sense, the statistical estimation of mortalities appears particularly suited to small populations. Moreover, the graduation by reference to a basis table—the subject of [BePo82], Chapter 15—is recommended in such situations, but also aggravates the dependence on the model assumptions (to be satisfied by two populations) and adds the dependence on the basis table (whose structure is inherited).

By this technique the ratios $\pi_x/\pi_x^{\mathrm{bas}}$ of true mortalities are estimated using the observed ones, q_x/q_x^{bas}. One variant treated in [BePo82] employs LIDSTONE's transformation

$q \mapsto \log(1-q)$ in the model

$$\log(1-q_x) - \log(1-q_x^{\mathrm{bas}}) = f(x) + \mathrm{error} \tag{1.2}$$

and fits a cubic $f(x)$ by least squares with (squared) weights $L_x/\pi_x \approx L_x/q_x^{\mathrm{bas}}$.

Since the opening of the European market the problem seems to attract some new attention as reinsurance companies may wish to calculate premiums for each (smaller) life insurance company individually, taking into account the particular mortality structure.

Log-Linear Regression

In [OlMi96] a linear regression $bx+c$ for the log-mortalities $\log q_x$ (set constant on an initial section) is proposed, the two regression parameters and scale are estimated by weighted minimum–L_1 and the (standardized) interquartile range of the residuals, respectively. These estimates, to obtain an estimate $\hat{q}_x$ of the original mortality π_x, have to be transformed back suitably.

For the estimation of graduation ratios $\pi_x/\pi_x^{\mathrm{bas}}$, the two mortality data sets are not combined in one model by [OlMi96] but the log-linear regression is applied twice, once to the company–specific table and separately for the DAV 1994 T basis table (already numerically smoothed), and then $\pi_x/\pi_x^{\mathrm{bas}}$ are estimated by $\hat{q}_x/\hat{q}_x^{\mathrm{bas}}$.

Several issues are left unsettled by [OlMi96]: The exponential structure of the mortality curves, hence of the graduation ratios, is based on mere belief and not checked statistically. Observed zero mortalities $q_x = 0$, for which $\log q_x = -\infty$ would create difficulties, are arbitrarily modified. The problem of heteroskedasticity is ignored and homoskedasticity implicitly assumed. No distributional assumptions are made.

Without proof the authors of [OlMi96] claim in their paper that their estimation method be "robust" and yield "confirmed results" and, in the discussion to [OlKo96], that it "save ca. 10% premiums".

Safety Margins—a Separate Problem

One notable inconsistency in [OlMi96] concerns the treatment of safety margins. This topic is only secondary since the declared aim in [OlMi96], Section 0, p 2, is the estimation of raw mortalities, for which a basis table, however, seemed unavailable. It was apparently not observed that the corresponding data without any additional safety components can in fact be obtained from columns 2 and 3 of DAV 1994 T (tables 1 and 2 in [Loeb94]). Instead, the numerically smoothed mortalities $\bar{q}_x^{\alpha}$ from column 6 containing two additional safety components have been taken as a basis for the subsequent statistical fitting.

The entire Section 4 of [OlMi96] is devoted to safety margins but only employs a special variant—actually a coarsening—of the method of [Loeb94].

The first level α condition (4) of [Loeb94], Section 2, is adopted with the notational identification $s_x^{\alpha} = (c-1)q_x$ by [OlMi96] while the second level α^* condition (2) of [Loeb94] is omitted. Consequentially, uncontrolled subventions across different age groups must be accepted.

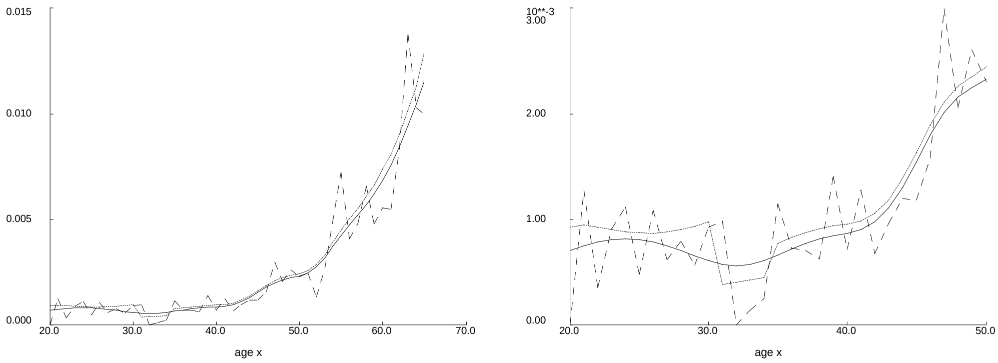

Figure 1: Raw mortalities t_x^u/L_x^u (– –), numerical smooth q_x^u (—), and the numerical smooth $q_x^{u\alpha}$ $(\cdots)$ containing one additional safety component, from table 3 of [Loeb94]

These statistical calculations concerning width and level of one-sided confidence intervals in a well-known way determine a minimum sample size. In [OlMi96], Section 0, p 2, this minimum sample size requirement is ascribed to the numerical smoothing technique.

Also, in [OlKo96], the column of numerically smoothed mortalities q_x^u in the company–specific table 3 of [Loeb94] was confounded with the following column of $q_x^{u\alpha}$ containing one additional safety component. The q_x^u are smooth but reflect substructures like the accident and heart–attack bumps; confer Figure 1. Over most of the range, the $q_x^{u\alpha}$ follow the q_x^u closely with a slight upward shift—except for the bump of $q_x^{u\alpha}$ at $x = 30, 31$. This artifact is unrelated to the numerical smoothing either.

In the said table of [Loeb94] we find $q_{30}^u = 0.613 \cdot 10^{-3}$, $q_{31}^u = 0.579 \cdot 10^{-3}$, and that $q_x^{u\alpha}$ drops from $0.977 \cdot 10^{-3}$ down to $0.388 \cdot 10^{-3}$, which is less than q_{31}^u. This artifact is caused by the alternating method I and II in [Loeb94], Section 5, of calculating safety margins, which at some instances (namely, for age groups with less than 5 deaths) switches between company and basis table. Moreover, the one-sided grouping of $x = 34, 33, 32$ with $x = 31$ seems doubtful. The method certainly needs to be improved.

Estimates Based on Root–Mortalities

Our paper concentrates on the statistical estimation of raw mortalities. The determination of additional safety components is considered a separate issue (to be treated elsewhere).

The log-linear model may be criticized from the viewpoint of destabilization of error variance, artificial generation of outliers, arbitrary treatment of zeros, a rather restricted back-transformation formula, and inevitable estimator bias.

In the chosen descriptive framework of [OlMi96], stochastic reasoning suggests the root transformation over the logarithm: The square root stabilizes error variance, diminishes the influence of outliers, automatically copes with observed zeros, allows a 'nonparametric' back-transformation formula, and at least in the homoskedastic case avoids a systematic bias of least squares.

For the company–specific table 3 of [Loeb94], in the range of 20–65 years, we fit a parabola to the root observed mortalities by classical and robust estimation methods: minimum–L_2 (with sample variance), minimum–L_1 (with average absolute deviation), and by robust regression *rg_Huber* as implemented in ISP; confer [ISP95]. A cubic and exponential are fitted by least squares. Transformed back these estimators give excellent fits in the original model that are practically indistinguishable by a χ^2 goodness-of-fit test.

Generalized Linear Models

Heteroskedasticity inevitably accounts for bumps in the back-transformed estimates unless we fit rather complicated regression functions $f(x)$ to the transformed observations $\log q_x$ and $\sqrt{q_x}$. Instead, it seems more natural to assume a smooth functional form of the true mortalities π_x themselves and not to transform observations at all. Formulating a Poisson generalized linear model we fit a cubic and an exponential by maximum likelihood.

2 Transformation–Based Robust Estimators

2.1 The Huber Family

In the following, the observed mortalities q_x are transformed to g_x (logarithm or square root), and for the transformed observations g_x some regression model is assumed,

$$g_x = f_\theta(x) + \mathrm{error} \tag{2.1}$$

The parameter θ of the regression function has to be estimated using the transformed observations g_x .

M Estimates As estimators we shall employ regression M estimates in the sense of [Hubr81], but to account for heteroskedasticity we actually need weighted versions. Generalizing weighted least squares these estimators are defined by

$$\sum_{x=20}^{65} \varrho\big(w_x\, \hat\sigma^{-1}(\, g_x - f_{\hat\theta}(x))\big) = \min_\theta ! \tag{2.2}$$

or

$$\sum_{x=20}^{65} \psi\big(w_x\, \hat\sigma^{-1}(\, g_x - f_{\hat\theta}(x))\big)\, w_x\, df_{\hat\theta}(x) = 0 \tag{2.3}$$

where $\psi = \dot\varrho$ and $df = \partial f / \partial\theta$, and $\hat\sigma$ denotes a scale estimate (to be determined by a further equation). The weights w_x will be specified in Sections 3 and 6 depending on the transformation.

Huber Functions To obtain a representative class of such M estimators we employ the family of Huber functions $\varrho = \varrho_k$ and $\psi = \psi_k$,

$$\varrho_k(u) = \min\{k|u|, u^2\} - \tfrac{1}{2}\min\{k^2, u^2\}, \qquad \psi_k(u) = u\,\min\{1, k/|u|\} \tag{2.4}$$

with clipping constant $0 \le k \le \infty$ that determines the degree of robustness; the smaller k the more robust the estimator.

The choice $k = \infty$ defines the weighted minimum–L_2 (least squares) estimator, denoted by minL2, which is implemented in the ISP macro *regress*. The case $k = 0$ leads to the weighted minimum–L_1 estimator [in (2.2) and (2.3) divide $\varrho = \varrho_k$ and $\psi = \psi_k$ by k and let $k \to 0$]. It is implemented in the ISP macro *l1fit* and denoted by minL1.

The ISP macro *rg_Huber* determines k adaptively as a constant *kfac* times the median

$$k = \mathit{kfac} * \mathrm{MED}\{|u_x|\} \tag{2.5}$$

of the absolute value of weighted residuals,

$$u_x = w_x\big(\,g_x - f_{\hat\theta}(x)\big) \tag{2.6}$$

Scale Estimates With the regression estimates goes an estimate $\hat\sigma$ of scale.

For $0 < \mathit{kfac} < \infty$ essentially the sample variance $\hat\sigma$ of the Winsorized residuals $\psi_k(u_x)$ is used by the ISP macro *rg_Huber* and suitably gauged to normal variance; confer [Hubr81], Section 7.10, formulae (10.2), p 196, with correction factor $K = 1$.

For minL2 ($k = \infty$) the ISP macro *regress* supplies the sample variance (VAR) of the weighted residuals,

$$\hat\sigma^2 = \frac{1}{46 - \dim} \sum_{x=20}^{65} u_x^2 \tag{2.7}$$

where dim denotes the dimension of the parameter θ. The combined estimator will be denoted by minL2/VAR.

For minL1 ($k = 0$) the ISP macro *l1fit* supplies the average absolute deviation (AAD) of weighted residuals, which may be gauged to normal variance by division through $\sqrt{2/3.14}$,

$$\hat\sigma = \sqrt{1.57}\,\mathrm{AAD}\,, \qquad \mathrm{AAD} = \frac{1}{46 - \dim} \sum_{x=20}^{65} |u_x| \tag{2.8}$$

The combined estimator will be denoted by minL1/AAD.

Robustness Standard The Huber family defines some algorithmic robustness standard: If the estimates differ for various values of the tuning constant there is a robustness problem, if they practically agree there is none.

Backtransformation The estimates $\hat\theta$ based on g_x must be transformed back according to suitable formulae given in Subsections 3.3 and 6.3, so as to yield estimates $\hat q_x$ for the theoretical mortalities π_x referring to the original model.

Population–Size Invariance The weights $w_x = \sqrt{L_x \eta_x}$ and $w_x = \sqrt{L_x}$ specified below and the corresponding back-transformation formulae achieve that the back-transformed minL2/VAR and minL1/AAD estimators stay the same if the observed mortalities q_x are kept while the population sizes L_x are all rescaled by a factor *Lfac*. General Huber estimates, for $0 < k < \infty$, are not size-invariant in this sense.

2.2 Standard Data Set

For all the plots in this paper, the set of observed raw mortalities q_x will be taken from the company–specific table 3 of [Loeb94]; namely, as the ratios t_x^u / L_x^u of columns 2 and 3 in that table. The age range will be $x = 20, \ldots, 65$ years. The total size of the population in that range is 659115.

At some instances (calculating goodness-of-fit, ...) this rather large population will be scaled down by dividing each L_x^u through the factor $Lfac = 3, \ldots, 5$ supplying a population size comparable to the ones of DAV 1994T (200260), table 1 of [Loeb94], and [OlMi96] (120000).

2.3 Goodness-of-Fit

The goodness of the fit in the original model is measured by the following χ^2 criterion,

$$S^2 = \sum_{x=20}^{65} L_x \frac{(q_x - \hat{q}_x)^2}{\hat{q}_x(1 - \hat{q}_x)} \tag{2.9}$$

If the model is correct and no parameters were estimated, S^2 were in fact approximately central χ^2_{46} with 46 degrees of freedom [deMoivre–Laplace]. It is (crude) statistical practice, as d parameters are estimated, to simply subtract d degrees of freedom.[1] The p-value v_p is the corresponding χ^2 tail-probability,

$$v_p = \Pr(S^2, \infty) \tag{2.10}$$

If the model is correct, v_p is approximately uniformly distributed on $(0, 1)$. The larger v_p the better the fit.

Population–Size Dependence The goodness-of-fit statistic (2.9) obviously does depend on the population size. The dependence of S^2 is straight proportionality if the employed estimators are invariant under rescaling of L_x by the factor *Lfac*. Therefore, Figure 8 plots the p-value as a function of $Lfac = 3, \ldots, 5$.

The larger the underlying population the less likely the acceptance of the model (null hypothesis). More complex models ought to be fitted as the population size increases.

[1] The χ^2 distribution may not be realized if parameter estimates other than minimum–χ^2 are plugged in; confer Chernoff, H. and Lehmann, E.L. (1956), Ann. Math. Stat. **25** pp 579–586.

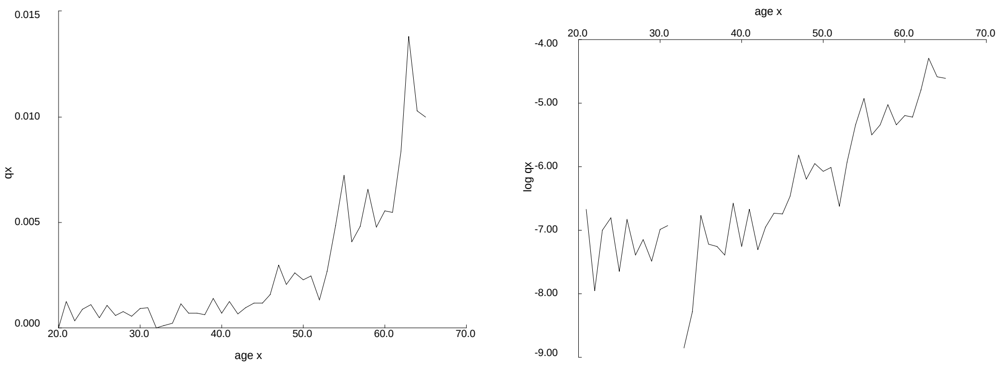

Figure 2: Raw mortalities t_x^u/L_x^u from table 3 of [Loeb94] and their logarithms, with $\log 0$ treated as missing value

3 Log-Linear Mortalities

3.1 Transformation and Preliminary Model

Here $g_x = \log q_x$. Without specifying the error structure, a linear regression

$$\log q_x = f_{b,c}(x) + \text{error} \tag{3.1}$$

is considered in [OlMi96] assuming a function $f_{b,c}(x)$ that is piecewise linear

$$f_{b,c}(x) = b\tilde{x} + c\,, \qquad \tilde{x} = \max\{31, x\} \tag{3.2}$$

A constant initial section is supposed to model the accident–bump in the range of 20–30 years, while other substructures, for example the heart–attack bump in the range of 45–50 years, are not taken into account.

Remark 3.1 Often the logarithmic transformation is recommended when the variability of the data seems to increase with the absolute value of observations; for example, confer [SliSt95], p 18. In the present context, however, this phenomenon can be explained without assuming a multiplicative structure. Ideally, the observed mortalities are binomial: $L_x q_x \sim \text{Bin}(L_x, \pi_x)$ and $\operatorname{Var} q_x = L_x^{-1}\pi_x(1-\pi_x)$. For large x the L_x decrease and the $\pi_x \le 0.5$, hence $\operatorname{Var} q_x$, increase. ////

The Estimate minL1/IQR of [OlMi96]

In [OlMi96] the two parameter estimates $\hat{b}$ and $\hat{c}$ are weighted minimum–L_1 with weights $w_x = L_x$,

$$\sum_{x=20}^{65} L_x \big| \log q_x - f_{\hat{b},\hat{c}}(x) \big| = \min_{b,c}! \tag{3.3}$$

As scale estimate the interquartile range (IQR) of the unweighted minimum–L_1 residuals $\log q_x - f_{\hat{b},\hat{c}}(x)$ (with two zeros omitted) is used, standardized by 1.35,

$$\hat{\sigma} = \mathrm{IQR}/1.35 \tag{3.4}$$

Their estimate will be denoted by minL1/IQR; it is implemented in the ISP macros *l1fit* and *fivenum*.

Back-Transformation These estimates based on $\log q_x$ are in [OlMi96] transformed back according to the formula

$$\hat{q}_x = \exp\bigl(f_{\hat{b},\hat{c}}(x) + \tfrac{1}{2}\hat{\sigma}^2\bigr) \tag{3.5}$$

3.2 Heteroskedasticity

Variance Ideally, without further knowledge, the observed mortalities are truly binomial, $L_x q_x \sim \mathrm{Bin}(L_x, \pi_x)$ and $q_x \longrightarrow \pi_x$ as $L_x \longrightarrow \infty$ (law of large numbers). The Taylor expansion $\log q_x = \log \pi_x + \pi_x^{-1}(q_x - \pi_x) + \ldots$ suggests that approximately $\mathrm{E}(\log q_x) \approx \log \pi_x$ and

$$\mathrm{Var}(\log q_x) \approx \pi_x(1-\pi_x)\big/(L_x\pi_x^2) \approx 1\big/(L_x\pi_x) \tag{3.6}$$

The Δ-method, linking Taylor expansion and deMoivre–Laplace central limit theorem, even tells us that approximately, in distribution,

$$\sqrt{L_x}\,(\log q_x - \log \pi_x) \approx \mathcal{N}(0, \pi_x^{-1}) \tag{3.7}$$

Hence the variance σ_x^2 of $\log q_x$ at age x is approximately

$$\sigma_x^2 \approx 1\big/(L_x\pi_x) \tag{3.8}$$

and the errors are actually heteroskedastic.

Remark 3.2 In view of (3.8), as π_x is small ($0.5 \cdot 10^{-3} \le \pi_x \le 10^{-2}$ roughly), the logarithmic transformation destabilizes variance. ////

Implicit Homoskedasticity The fact that the scale estimate is not adjusted to x in the back-transformation formula (3.5) implies the assumption of homoskedastic errors in [OlMi96]. Thus, model equation (3.1) more precisely reads

$$\log q_x = f_{b,c}(x) + \sigma\,\varepsilon_x \tag{3.9}$$

with errors $\varepsilon_{20}, \ldots, \varepsilon_{65}$ stochastically independent, identically distributed according to some law F (i.i.d. $\sim F$) and σ some unknown scale.

3.3 Model Formulation

In view of (3.8), we replace model (3.9) by

$$\log q_x = f_{b,c}(x) + \frac{\sigma}{\sqrt{L_x \eta_x}}\, \varepsilon_x \tag{3.10}$$

with ε_x i.i.d. $\sim F$ and σ unknown. Following [BePo82], p 329, we have substituted the unknown π_x by some preliminary estimates η_x; for example, $\eta_x = \hat{q}_x^{\rm bas}$ (estimated) or $\eta_x = q_x^{\rm bas}$ (observed) from some basis table. Corresponding weights are

$$w_x = \sqrt{L_x \eta_x} \tag{3.11}$$

Interpretation of $\boldsymbol{f(x)}$ In view of (3.10) the regression function $f(x) = f_{b,c}(x)$ cannot simply be equated with $\log \pi_x$. Since

$$\pi_x = {\rm E}_F \exp(\log q_x) = \exp\bigl(f(x)\bigr) \int \exp(\sigma\varepsilon/w_x)\, F(d\varepsilon) \tag{3.12}$$

in fact

$$f(x) = \log \pi_x - \log \widetilde{F}(\sigma/w_x) \tag{3.13}$$

where $\widetilde{F}$ denotes the Laplace–transform of F. In this way regression and scale parameters are related with the estimand π_x. On the one hand, (3.13) might reflect an intricate dependence of mortalities and population sizes. On the other hand, if the original mortalities π_x are assumed to have a certain functional form (e.g., polynomial, exponential), that of the function $f(x)$ to be fitted in the transformed model is then prescribed by (3.13).

Back-Transformation The back-transformed estimate is

$$\hat{q}_x = \exp\bigl(f_{\hat{b},\hat{c}}(x) + \hat{\sigma}^2/(2w_x^2) \bigr) \tag{3.14}$$

According to (3.14), bumps of $L_x\eta_x$ inevitably create bumps of the estimate $\hat{q}_x$. Only under the assumption of homoskedasticity smoothness is preserved.

Special Case Homoskedasticity is the special case

$$L_x \pi_x \approx \text{constant} \tag{3.15}$$

The values of $L_x q_x$ for $x = 20, \ldots, 65$ of the table DAV 1994 T (males) in [Loeb94] and the company–specific table in [OlMi96] range from 5.214 to 65.03 and from 0.4872 to 11.63, respectively. Hence the corresponding σ_x range from 0.1240 to 0.4379 (with a median of 0.2993) and from 0.2932 to 1.433 (with a median of 0.4468), respectively. The corresponding values of table 3 in [Loeb94] with $Lfac = 5$ are 0.1386 and 1.502 (with a median of 0.2900). Thus, the value of σ in the homoskedastic case may be assumed between $1/4$ and $1/2$, but the value $\sigma = 1$ is also plausible.

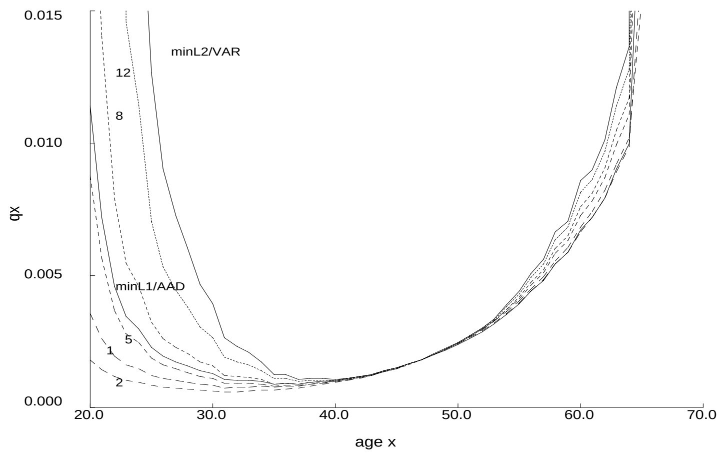

Figure 3: Back-transformed Huber estimates based on $\log q_x$ (zeros modified) for tuning constant $kfac = 0, 1, 2, 5, 8, 12, \infty$

3.4 Discrepancy of Huber Estimates

For the company–specific table 3 of [Loeb94] we evaluate a few representative members of the Huber family of robust estimators introduced in Section 2 with weights $w_x = \sqrt{L_x \eta_x}$ where $\eta_x = \hat{q}_x^{\text{bas}}$ denote the numerically smoothed mortalities from column 4 of the basis table DAV 1994 T (males), table 1 in [Loeb94]. In the computation of these estimates, observed numbers of deaths equal to zero have been replaced by the value 10^{-4}.

Figure 3 reveals a large discrepancy between the back-transformed Huber estimates if the tuning constant, which regulates the degree of robustness, varies. The same estimates based on root mortalities, with weights $w_x = \sqrt{L_x}$ and back-transformed, practically coincide; confer Figure 7. This phenomenon indicates an instability of the model introduced by the logarithmic transformation.

4 Implicit Log-Normality

Nonparametric? Robust? The omission of any distributional assumption in [OlMi96] suggests a method which is 'nonparametric', its performance being the same for all error distributions. Also, the minimum–L_1 criterion and interquartile range in the place of minimum–L_2 and sample variance, respectively, are supposed to guarantee estimates which are "robust", not requiring any strict distributional assumption (e.g., that of exact normality). The following argument, however, shows that the special form of formula (3.5) for the back-transformation entails exactly normal mean zero errors in the log-transformed model of [OlMi96], provided only the estimates are consistent for the true mortalities.

4.1 Asymptotic Setup

Model According to (3.9) this model is

$$\log q_x = f(x) + \sigma\,\varepsilon_x \tag{4.1}$$

with ε_x i.i.d. $\sim F$. The error distribution F may be arbitrary and the regression function $f(x)$ possibly more general than $f_{b,c}(x)$ defined by (3.2).

Estimand Of interest are the true mortalities π_x. Being probabilities they are the expection of the corresponding empirical frequencies,

$$\pi_x = \mathrm{E}_F\, q_x = \mathrm{E}_F \exp\bigl(f(x) + \sigma\,\varepsilon_x\bigr) = \exp\bigl(f(x)\bigr) \int e^{\,\sigma\,\varepsilon}\, F(d\varepsilon) \tag{4.2}$$

Product–Model Now suppose that k similar populations may be observed or that k years have passed (note that in [OlMi96] $k=4$ by the time of this writing). Then not only one observed mortality per age x but k values $q_{x,1},\dots,q_{x,k}$ are available. This situation is described by the product model,

$$\log q_{x,i} = f(x) + \sigma\,\varepsilon_{x,i} \tag{4.3}$$

with $\varepsilon_{x,i}$ i.i.d. $\sim F$ for $i = 1,\dots,k$ and $x = 20,\dots,65$.

Remark 4.1 Although the product–model (4.3) is possibly not realized exactly, it is a reasonable framework in which results on the distribution (bias) of estimators can be derived to see what principally happens in the long run. ////

4.2 Consistent Estimability

As k grows larger, more and more information becomes available and it will be possible to find estimates $\hat{f}_k$ and $\hat{\sigma}_k$ using all the observations $\log q_{x,i}$ that estimate $f(x)$ and σ consistently, in the sense that for each x and all σ,

$$\hat{f}_k(x) \longrightarrow f(x)\,, \qquad \hat{\sigma}_k \longrightarrow \sigma\, s_F \tag{4.4}$$

in probability or even almost surely, as $k \to \infty$. A factor $s_F \neq 1$ means that the scale estimate is not gauged to the given F. In fact, requirement (4.4) of consistent estimability is the only restriction we impose on the regression function $f(x)$ and the error distribution F. Verification amounts to proving a suitable law of large numbers.

Consistency of minL2 and minL1 for $f_{b,c}(x)$

Condition (4.4) for $f_{b,c}(x)$ given by (3.2) can be verified using minL2/VAR provided the error distribution satisfies

$$\int \varepsilon\, F(d\varepsilon) = 0\,, \qquad \int \varepsilon^2\, F(d\varepsilon) < \infty \tag{4.5}$$

and then $s_F = 1$ holds.

Using minL1/AAD it suffices to assume an error distribution with unique median such that

$$\mathrm{MED}(F) = 0\,, \qquad \int |\varepsilon|\, F(d\varepsilon) < \infty \tag{4.6}$$

For minL1/IQR in addition the two quartiles $Q_{.25}(F)$ and $Q_{.75}(F)$ must be unique; confer [BaKo78] and [BlSt83]. The division of AAD and IQR by $\sqrt{2/3.14}$ and 1.35, respectively, achieves that $s_F = 1$ for normal F.

Consistency of Huber estimates with $0 < kfac < \infty$ holds if the equation

$$\int \psi_k(\varepsilon - m)\, F(d\varepsilon) = 0 \tag{4.7}$$

has the unique solution $m = 0$. We omit the details of the proof.

Consistent Back-Transformation In general, the estimates $\hat{f}_k(x)$ and $\hat{\sigma}_k$ are transformed back according to formula (3.5) to yield the following estimate of π_x,

$$\hat{q}_k(x) = \exp\big(\hat{f}_k(x) + \tfrac{1}{2}\hat{\sigma}_k^2\big) \tag{4.8}$$

By continuity, consistency (4.4) carries over so that, on the one hand,

$$\hat{q}_k(x) \longrightarrow \exp\big(f(x) + \tfrac{1}{2}\sigma^2 s_F^2\big) \tag{4.9}$$

On the other hand, we want to insist that the estimates $\hat{q}_k(x)$ aim at the right target value; that is,

$$\hat{q}_k(x) \longrightarrow \pi_x = \mathrm{E}_F\, q_x \tag{4.10}$$

4.3 Unique Laplace Transform

Equation (4.2) and the convergences (4.9) and (4.10) imply that

$$\exp\big(f(x)\big) \int e^{\sigma\varepsilon}\, F(d\varepsilon) = \exp\big(f(x) + \tfrac{1}{2}\sigma^2 s_F^2\big) \tag{4.11}$$

and, upon cancellation of the term $\exp f$, the following identity,

$$\int e^{\sigma\varepsilon}\, F(d\varepsilon) = \exp(\tfrac{1}{2}\sigma^2 s_F^2) \tag{4.12}$$

On the RHS, we recognize the Laplace transform of the $\mathcal{N}(0, s_F^2)$ distribution, on the LHS that of F. Identity (4.12) is valid for all σ which may appear in model equation (4.1). We assume that the set of such σ-values includes some nondegenerate interval $(0, \delta)$.

Then the uniqueness theorem for the Laplace transform is in force and tells us that necessarily

$$F = \mathcal{N}(0, s_F^2) \tag{4.13}$$

4.4 Normality or Systematic Bias

In other words, if F is not exactly normal mean zero or the scale estimate is not gauged to normal variance the consistency (4.10) is violated,

$$\hat{q}_k(x) \not\longrightarrow \pi_x \tag{4.14}$$

that is, the estimates $\hat{q}_k(x)$ have a systematic bias relative to the mortalities π_x.

Remark 4.2 This result should be distinguished from the discussion of bias in [OlMi96]. With reference to [Millr84] these authors only mention the fact that in the case $F = \mathcal{N}(0,1)$ the back-transformed estimates $\hat{q}_x$ given by (3.5) may have expectation unequal to π_x. This is the case since the (theoretical) back-transformation formula holding for normal $Y \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathrm{E}\, e^Y = e^{\,\mu+\sigma^2/2} \tag{4.15}$$

does no longer obtain if mean μ and variance σ^2 are substituted by estimates. After all, even in the normal case only consistency (4.9) can be achieved. Exact unbiasedness would require a more sophisticated back-transformation formula, as in [Hoyl75] and [Vera91].

Our argument on the contrary shows that, as soon as (4.13) is violated—in particular, if F is not normal—the estimates $\hat{q}_k(x)$ do not approach the true mortalities π_x. This systematic bias goes unnoticed in [OlMi96]. ////

Even under ideal conditions, (4.13) is not fulfilled exactly because of the unbounded support of nondegenerate normals and $\log q_x \le 0$.

5 Systematic Bias

Huber's Neighborhood In robust statistics—confer [Hubr81]—the following full neighborhood of the standard normal is considered: the set of all probabilities F of the form

$$F = (1-r)\mathcal{N}(0,1) + rM \tag{5.1}$$

where $r \in (0,1)$ denotes some fixed radius and the contamination M may be an arbitrary probability measure. This neighborhood models the situation that $(1-r)100\%$ of the observations are generated by the standard normal and $r100\%$ by an arbitrary probability, or that 100% observations may come from a slightly and arbitrarily deformed normal law. Condition (4.13) restricts this neighborhood rather severely, namely to the subset of normals

$$F = \mathcal{N}(0, \tau^2)\,, \qquad 1 \le \tau \le (1-r)^{-1} \tag{5.2}$$

But pure normal distributions do not model outliers. The log-linear approach of [OlMi96], therefore, cannot claim the label "robust" unless systematic bias due to nonnormality, or only approximate normality, is taken into account.

Estimators Bias will be evaluated for the estimate minL1/IQR of [OlMi96] and, for comparison, minL2/VAR as defined in Subsections 2.1 and 3.1.

5.1 Relative Bias—Symmetric Case

For the investigation of systematic bias we relate quantity (4.9), which is actually estimated, to the target value π_x. Thus, under consistent estimability (4.4) the relative bias $\gamma_F(\sigma)$ is

$$\gamma_F(\sigma) = \pi_x^{-1} \exp\bigl(f(x) + \tfrac{1}{2}\sigma^2 s_F^2 \bigr) = \frac{\exp(\frac{1}{2}\sigma^2 s_F^2)}{\int e^{\sigma\varepsilon}\, F(d\varepsilon)} \tag{5.3}$$

which turns out the same for all x.

Symmetric *F*

As alternatives to the standard normal, we first consider error distributions of form (5.1) with contaminating M symmetric Dirac (generating outliers $\pm z$) or M symmetric normal (with larger variance z^2),

$$\textsf{(DS)} \quad M = \tfrac{1}{2}(\delta_{-z} + \delta_z), \qquad \textsf{(NS)} \quad M = \mathcal{N}(0, z^2) \tag{5.4}$$

with $z \ge 1$ such that the variance of M is greater than 1. For approximate normality the contamination radius is chosen $r = 10\%$. As a nonnormal F, the uniform on the interval $(-z, z)$

$$\textsf{(U)} \quad F = U(-z, z) \tag{5.5}$$

is considered, with $z \ge \sqrt{3}$ such that the variance of M is greater than 1.

The expression (5.3) may be evaluated analytically in more detail and then numerically. Via s_F it only depends on the scale estimate. The true scale matters: The larger σ the larger the bias. For $\sigma = 0.375$ and $\sigma = 1$ we have plotted $\gamma_F(\sigma)$ as a function of $z \le 6$. Only this range is relevant since the true $\log \pi_x$ are roughly between $\log 0.5 \cdot 10^{-3} \approx -7.6$ and $\log 10^{-2} \approx -4.6$, and $\log q_x \le 0$ always.

Relative Bias—Asymmetric Case

Outliers may also occur in asymmetric ways. In [OlMi96], p 5, on the one hand, only outliers to the left, for which q_x is small or even 0, are mentioned at all; these are the observations with large $|\log q_x|$. Actually, it is the logarithmic transform itself that generates such outliers.

Remark 5.1 The logarithm introduces a skewness to the left: Relative distances $|\log \rho\pi_x - \log \pi_x| = |\log \rho|$ to the right are decreased, since $|\log \rho| < |\rho - 1|$ for $\rho > 1$, whereas to the left, since $|\log \rho| > |\rho - 1|$ for $\rho < 1$, they are increased.

For example, let us consider an age group of size $L_x = 2000$ and mortality $\pi_x = 0.6 \cdot 10^{-2}$, such that $\sigma_x \approx 0.3$. Then a fluctuation of the observed about

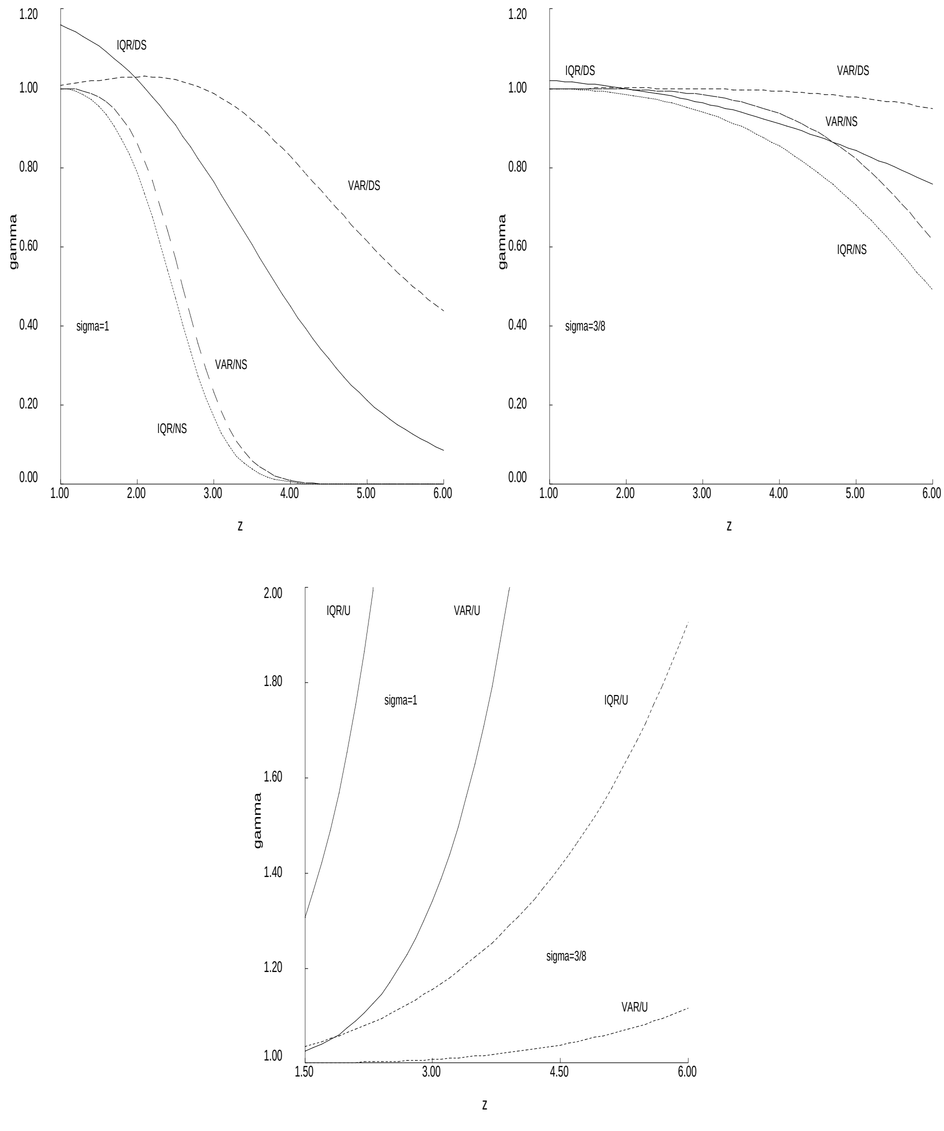

Figure 4: Systematic bias of minL2/VAR and minL1/IQR, for $\sigma = 0.375$ and $\sigma = 1$ under symmetric error distributions DS, NS, and the uniform $U(-z, z)$ (3rd plot)

the expected number (12) of deaths by ± 6 (± 8, ± 11, ± 12) translates into a fluctuation of $\log q_x$ about $\log \pi_x = -5.1$ by $\log 2 = 0.70$ (1.1, 2.5, ∞) to the left and to the right by only $\log 1.5 = 0.40$ (0.51, 0.65, 0.70, respectively). Thus, the logarithmic transform generates outliers to the left that are not in the data. ////

On the other hand it may be argued that, starting from a certain background mortality, outliers mostly occur to the right (accidental crashs of full carloads, ...), and so the real distribution would actually be asymmetric to the right. Both effects may result in a symmetric contamination, but should of course be formulated and investigated separately.

Asymmetric *F*

Therefore, we secondly consider error distributions $F = (1-r)\mathcal{N}(0,1) + rM$ of radius $r = 10\%$ and with the contaminating M asymmetric Dirac and shifted normal, respectively, both generating outliers z (exactly or on the average),

$$\text{(DA)} \quad M = \delta_z\,, \qquad \text{(NA)} \quad M = \mathcal{N}(z,1) \tag{5.6}$$

Relative Bias—General Case

To determine the systematic bias under asymmetry, we assume the regression function $f_{b,c}(x)$ of type (3.2) and rewrite model equation (4.1) in the following form,

$$\log q_x = f_{b,c+\sigma m}(x) + \sigma(\varepsilon_x - m) \tag{5.7}$$

Depending on the estimate of the regression coefficients—minL1 or minL2—the constant m denotes the mean and median of F, respectively, and the zero m of $\int \psi_k(\varepsilon - m)\,F(d\varepsilon) = 0$ for general Huber estimates. Then the consistency conditions (4.5) and (4.6) are fulfilled. Consequentially, (4.2) and (4.9) are replaced by

$$\hat{q}_k(x) \longrightarrow \exp\bigl(f_{b,c+\sigma m}(x) + \tfrac{1}{2}\sigma^2 s_F^2\bigr) \tag{5.8}$$

and

$$\pi_x = \exp\bigl(f_{b,c+\sigma m}(x)\bigr) \int e^{\,\sigma(\varepsilon - m)}\,F(d\varepsilon) \tag{5.9}$$

As the $\exp f$ term cancels again, the systematic bias is

$$\gamma_F(\sigma) = \frac{\exp(\sigma m + \frac{1}{2}\sigma^2 s_F^2)}{\int e^{\,\sigma\varepsilon}\,F(d\varepsilon)} \tag{5.10}$$

where m expresses the influence of the regression estimate under asymmetry.

6 Root-Transformation

The following stochastic reasoning suggests the square root transformation instead of the logarithm.

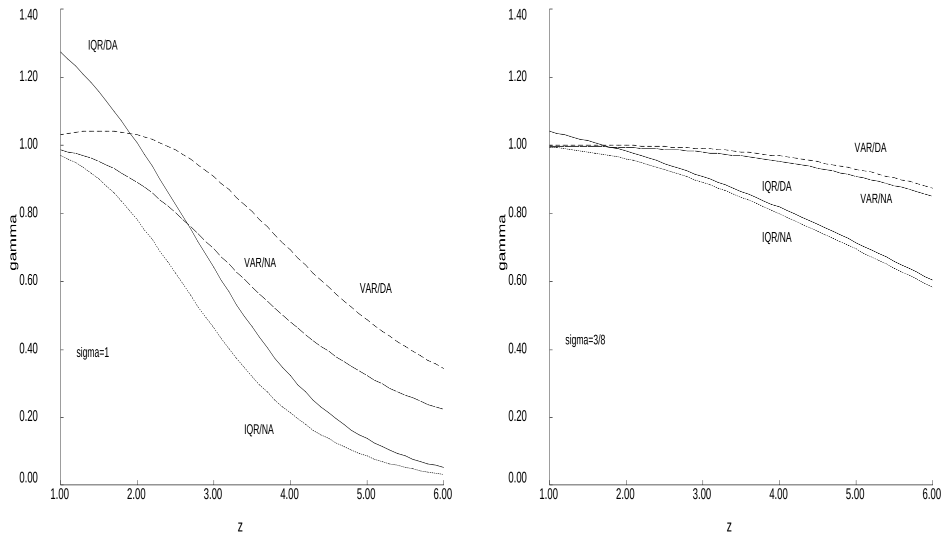

Figure 5: Systematic bias of VAR: minL2/VAR, and IQR: minL1/IQR, for $\sigma = 0.375$ and $\sigma = 1$, under asymmetric error distributions DA and NA

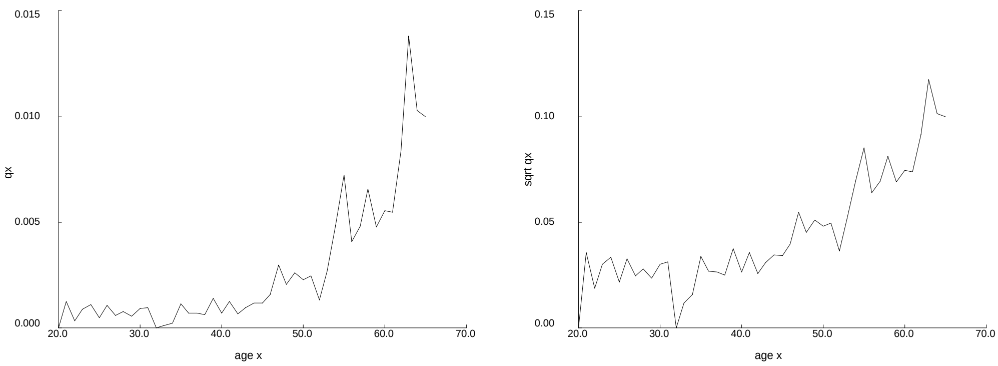

Figure 6: Raw mortalities t_x^u/L_x^u from table 3 of [Loeb94] and their square roots

6.1 Variance Considerations

Heteroskedasticity Without further knowledge, the observed mortalities are binomial, $L_x q_x \sim \mathrm{Bin}(L_x, \pi_x)$, and $q_x \longrightarrow \pi_x$ as $L_x \longrightarrow \infty$ (law of large numbers). The Taylor expansion $\sqrt{q_x} = \sqrt{\pi_x} + (2\sqrt{\pi_x})^{-1}(q_x - \pi_x) + \ldots$ suggests that approximately $\mathrm{E}(\sqrt{q_x}) \approx \sqrt{\pi_x}$ and

$$\mathrm{Var}(\sqrt{q_x}) \approx \pi_x(1-\pi_x)\big/(4L_x\pi_x) \approx 1\big/(4L_x) \tag{6.1}$$

The Δ-method, linking Taylor expansion and deMoivre–Laplace central limit theorem, even tells us that approximately, in distribution,

$$\sqrt{L_x}\,(\sqrt{q_x} - \sqrt{\pi_x}) \approx \mathcal{N}(0, \tfrac{1}{4}) \tag{6.2}$$

Therefore, the variance σ_x^2 of $\sqrt{q_x}$ at age x is approximately

$$\sigma_x^2 \approx 1\big/(4L_x) \tag{6.3}$$

Stabilization In comparison with (3.8), the square root has stabilized the variance by removing the dependence on π_x.

Homoskedasticity Homoskedasticity amounts to the condition that

$$L_x \approx \text{constant} \tag{6.4}$$

The values of L_x for $x = 20, \ldots, 65$ in DAV 1994 T (males), in the special company table of [OlMi96], and in table 3 of [Loeb94] scaled down by $Lfac = 6$ have a median of 4405, 3095, and 1430, respectively. Hence the corresponding σ_x-values have medians in the range from $0.75 \cdot 10^{-2}$, $0.9 \cdot 10^{-2}$, to $1.32 \cdot 10^{-2}$. Thus, in the homoskedastic case, the value of $\sigma = 10^{-2}$ is plausible and corresponds to age groups of size $L_x \approx 2500$.

6.2 Model Formulation

Thus, we consider the following model for the observed root-mortalities,

$$\sqrt{q_x} = f(x) + \frac{\sigma}{\sqrt{L_x}}\,\varepsilon_x \tag{6.5}$$

with the errors ε_x i.i.d. $\sim F$ and the scale σ unknown. Corresponding weights are

$$w_x = \sqrt{L_x} \tag{6.6}$$

Regression Function As regression function $f(x)$ we employ a parabola and a cubic,

$$f(x) = a + bx + cx^2\ (+\,dx^3) \tag{6.7}$$

and an exponential with additive constant,

$$f(x) = c + \exp(a + bx) \tag{6.8}$$

The regression functions $f(x)$ may be set constant for $x = 20, \ldots, 31$ by simply inserting $\tilde{x} = \max\{31, x\}$ for x.

Interpretation of $f(x)$ Again, one cannot simply equate $f(x)$ with $\sqrt{\pi_x}$. From (6.5) and $\mathrm{E}_F\, q_x = \pi_x$ we obtain

$$\pi_x = \big(f(x) + \sigma\, w_x^{-1} m\big)^2 + \sigma^2 w_x^{-2}\, \mathrm{Var}_F\, \varepsilon \tag{6.9}$$

where $m = \mathrm{E}_F\, \varepsilon$, and hence actually

$$f(x) = \sqrt{\pi_x - \sigma^2 w_x^{-2}\, \mathrm{Var}_F\, \varepsilon}\; - \sigma\, w_x^{-1}\, \mathrm{E}_F\, \varepsilon \tag{6.10}$$

In this way regression and scale parameters are related with the estimand π_x. On the one hand, this indicates an intricate dependence of mortalities and population sizes, which undoubtedly exists for an entire population. On the other hand, if the original mortalities π_x are assumed to obey a certain functional form, the $f(x)$ to be fitted in the transformed model is necessarily prescribed by (6.10).

6.3 Back-Transformation

The (theoretical) back-transformation formula is

$$\mathrm{E}\, Y^2 = \mu^2 + \sigma^2 \tag{6.11}$$

which is valid for any random variabe Y with mean μ and variance σ^2. In this sense, and contrary to (4.15), formula (6.11) is in fact 'nonparametric'.

Remark 6.1 Without setting $1 - \pi_x \approx 1$, the variance-stabilizing transformation would be $\arcsin \sqrt{q_x}$. The corresponding back-transformation formula is

$$\mathrm{E}(\sin Y)^2 = \tfrac{1}{2} - \tfrac{1}{4}\big(\varphi^Y(2) + \varphi^Y(-2)\big) \tag{6.12}$$

and involves the Fourier transform φ^Y of Y. Thus, instead of mean and variance, the Fourier transform at ± 2 would have to be estimated. ////

The back-transformation formula for estimates $\hat\theta$ of the unknown parameter θ of the regression function $f(x) = f_\theta(x)$ that are based on $\sqrt{q_x}$ is

$$\hat q_x = \hat f(x)^2 + \hat\sigma^2 / L_x \tag{6.13}$$

where $\hat f = f_{\hat\theta}$. Bumps of L_x inevitably create bumps of the estimate $\hat q_x$. Only in the homoskedastic case this problem does not appear.

6.4 The Huber Family

Estimates $\hat\theta$ of the regression parameter θ that are based on $\sqrt{q_x}$ may be defined according to (2.2)–(2.8) employing the weights $w_x = \sqrt{L_x}$. In particular, the Huber family of estimators minL2/VAR, minL1/AAD, and *rg_Huber* with tuning constant *kfac* is available.

6.5 Least Squares Unbiased

In the homoskedastic case, model (6.5) attains the form

$$\sqrt{q_x} = f(x) + \sigma\,\varepsilon_x \tag{6.14}$$

with ε_x i.i.d. $\sim F$ and σ unknown, and the back-transformation is given by

$$\hat{q}_x = \hat{f}(x)^2 + \hat{\sigma}^2 \tag{6.15}$$

Consistency of minL2/VAR Assuming an error distribution of finite variance $\mathrm{Var}_F\,\varepsilon < \infty$ and setting $m = \mathrm{E}_F\,\varepsilon$, we rewrite the homoskedastic model (6.14) in the following form,

$$\sqrt{q_x} = f(x) + \sigma\, m + \sigma\,(\varepsilon_x - m) \tag{6.16}$$

In the product–model, the minL2/VAR estimate can be proved to be consistent,

$$\hat{f}_k(x) \longrightarrow f(x) + \sigma\, m\,, \qquad \hat{\sigma}_k^2 \longrightarrow \sigma^2\,\mathrm{Var}_F\,\varepsilon \tag{6.17}$$

Thus,

$$\hat{q}_k(x) = \hat{f}_k(x)^2 + \hat{\sigma}_k^2 \longrightarrow \bigl(\,f(x) + \sigma\, m\,\bigr)^2 + \sigma^2\,\mathrm{Var}_F\,\varepsilon \tag{6.18}$$

where the limiting value is in fact the target since, in view of (6.16) with $m = \mathrm{E}_F\,\varepsilon$,

$$\pi_x = \mathrm{E}_F\, q_x = \bigl(\,f(x) + \sigma\, m\,\bigr)^2 + \sigma^2\,\mathrm{Var}_F\,\varepsilon \tag{6.19}$$

In this sense, minL2/VAR is always unbiased. The existence of such an estimator speaks in favor of the model.

6.6 Outliers

The problem of robustness will not be as severe as with the logarithm since the root-transform does not generate outliers. In particular, as $\sqrt{0} = 0$, observed mortalities $q_x = 0$ do not need an extra treatment.

Remark 6.2 In the numerical example given in Remark 5.1 the roots $\sqrt{q_x}$ fluctuate about $\sqrt{\pi_x} = 7.75\%$, with $\sigma_x \approx 1.12\%$, attaining the following values (observed number of deaths in brackets) 0% (0), 3.16% (1), 4.48% (4), 5.48% (6), 7.75% (12), 9.48% (18), 10% (20), 10.7% (23), and 10.95% (24).

Relative distances towards both sides $\bigl|\sqrt{\rho\,\pi_x} - \sqrt{\pi_x}\bigr| = |\sqrt{\rho} - 1|\sqrt{\pi_x}$, since $|\sqrt{\rho} - 1| \le |\rho - 1|$ for $\rho < 1$ and $\rho > 1$, are decreased. ////

6.7 Nonnegativity

Without an explicit nonnegativity condition, the fits to $\sqrt{q_x}$ have turned out so good for the data considered that $\hat{f}(x) \ge 0$ automatically.

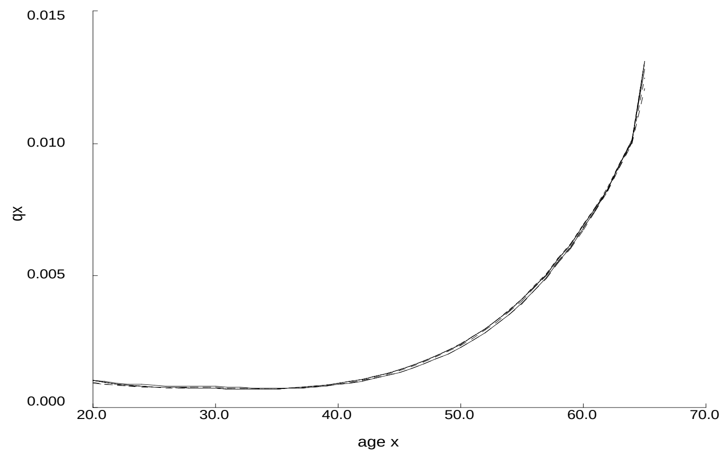

Figure 7: Back-transformed $\sqrt{q_x}$ based Huber estimates of a parabola with the previous choice of tuning factor $\mathit{kfac} = 0, 1, 2, 5, 8, 12, \infty$

7 Estimates Based on Root-Mortalities

7.1 Estimates and Functions in Comparison

Comparison of Huber Estimates

First, a parabola $f(\tilde{x})$ with constant initial section is fitted to $\sqrt{q_x}$ of our standard data set from table 3 of [Loeb94]. The different estimates of the Huber family (minL2/VAR, minL1/AAD, *rg_Huber*) are evaluated for $\sqrt{q_x}$, transformed back, and compared.

Contrary to $\log q_x$ based estimation—recall Figure 3—the different estimates plotted in Figure 7 practically agree [2], which indicates stability of the model. Thus, one may stay with minL2/VAR.

Comparison of Parabola, Cubic, and Exponential

Second, the fit by different regression functions is compared. By minL2/VAR we determine the $\sqrt{q_x}$ based fit of parabola, cubic, and exponential given by (6.7) and (6.8) with constant initial section; confer the left-hand plot of Figure 8.

Third, the right-hand plot of Figure 8 shows the goodness-of-fit measured by the p-value as a function of the population rescaling factor *Lfac*; confer Subsection 2.3.

The three estimates and corresponding p-values achieved are practically indistinguishable. Thus one may stay with the parabola.

[2] We have also tried the Hampel–Krasker family (computed by iterative weights) and obtained similar results.

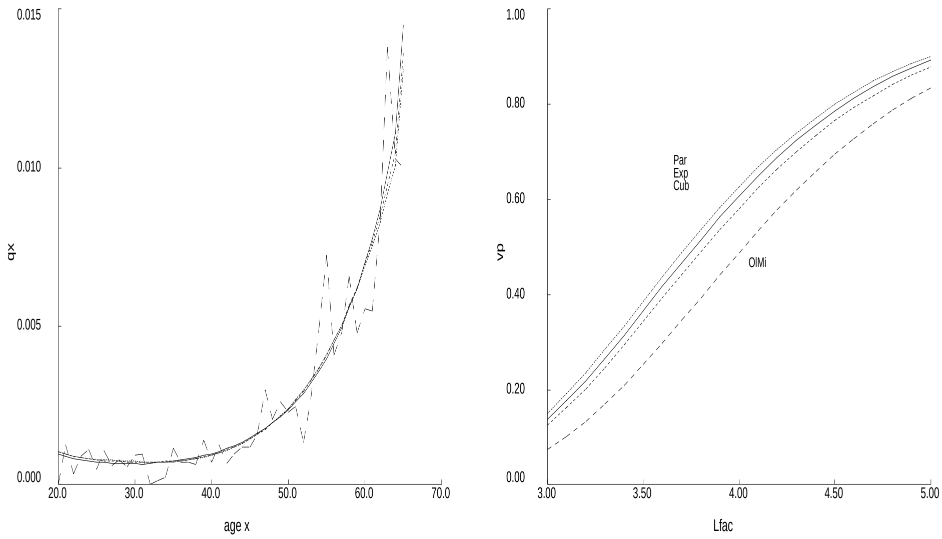

Figure 8: $\sqrt{q_x}$ based minL2/VAR fit of a parabola ($\cdots$), cubic (- - -), and exponential (—), with corresponding p-values as a function of the population rescaling factor *Lfac*; the fourth curve (– –) belongs to the $\log q_x$ based minL1/IQR fit (3.3)–(3.5).

Numerical, minL2/VAR Parabola, and Log-Linear minL1/IQR Fit

As a reference curve, we appeal to the numerical fit q_x^u taken from column 4 of table 3 in [Loeb94]. Also, the $\log q_x$ based minL1/IQR estimate ${}^{\mathrm{r}}q_x^{\mathrm{roh}}$ defined by (3.3)–(3.5) and its p-value are determined. These estimates, the parabola $\hat{q}_x^{\mathrm{par}}$, and the observed mortalities are plotted on the left-hand of Figure 9 and tabulated in Table 1.

The right-hand plot of Figure 9 shows the differences between the various fits and the numerical smooth (reference curve).

7.2 Unresolved Problems

Not only bias but also variance of estimators should be treated.

Based on an analysis of likelihoods of the transformed model, the asymptotic distribution of regression estimators centered at the regression parameters may be derived. Likewise, robustness of Huber and other estimates has been defined and proved in terms of bias and variance, in the context of estimating the regression parameters of the transformed model.

How do these properties carry over to the estimation of π_x?

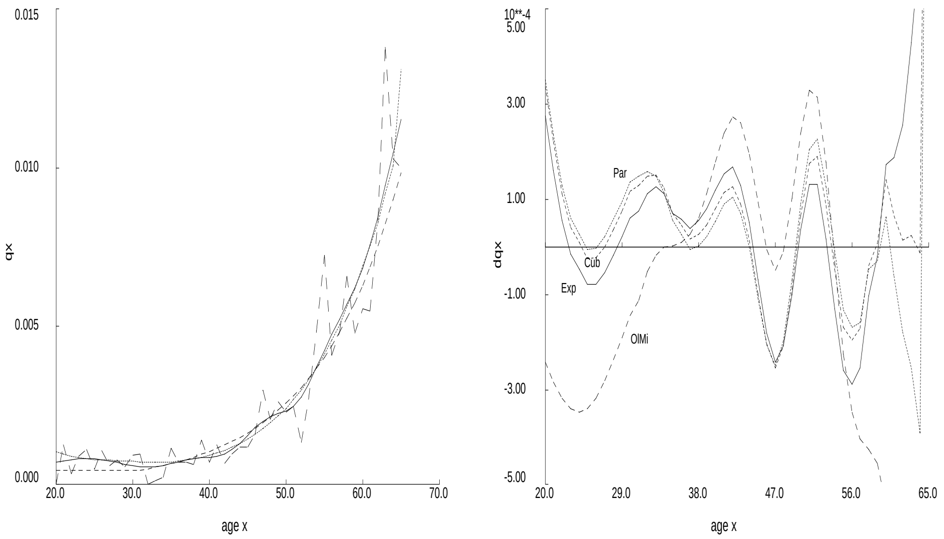

Figure 9: back-transformed $\sqrt{q_x}$ based minL2/VAR fit of a parabola $(\cdots)$, the numerical smooth (—), and $\log q_x$ based minL1/IQR fit (3.3)–(3.5) (– –), on the left-hand plot; on the right-hand plot, differences between the various fits and the numerical smooth

8 MLE in a Poisson Generalized Linear Model

8.1 Poisson GLM

Probability Model The number of deaths T_x observed in the group of age x may instead of binomial well be supposed Poisson,

$$\Pr(T_x = t_x) = e^{-\lambda_x} \frac{\lambda_x^{t_x}}{t_x!}\,, \qquad t_x = 0, 1, \ldots \tag{8.1}$$

with expected number of deaths

$$\lambda_x = L_x \pi_x \tag{8.2}$$

Assuming stochastic independence the log-likelihood function is

$$\log \Pr\bigl(T_x = t_x\,,\ x = 20, \ldots, 65\bigr) = \sum_{x=20}^{65} \bigl(-\lambda_x + t_x \log \lambda_x - \log(t_x!)\bigr) \tag{8.3}$$

Regression Functions We model the function of true mortalities π_x by an exponential and a cubic in $\tilde{x} = \max\{31, x\}$,

$$\pi_x = \exp\bigl(a + b\tilde{x}\bigr) \tag{8.4}$$

respectively,

$$\pi_x = a + b\tilde{x} + c\tilde{x}^2 + d\tilde{x}^3 \tag{8.5}$$

8.2 Maximum Likelihood Estimate (MLE)

The log-likelihood (8.3) with either choice (8.4) or (8.5) is differentiable with respect to the parameters and yields the maximum likelihood equations by setting the derivative equal to zero.

Likelihood Equations

In case (8.4) the likelihood equations to be solved for $\hat{a}$ and $\hat{b}$ are

$$\frac{\sum_x \tilde{x} L_x e^{\hat{b}\tilde{x}}}{\sum_x L_x e^{\hat{b}\tilde{x}}} = \frac{\sum_x \tilde{x} t_x}{\sum_x t_x}, \qquad e^{\hat{a}} = \frac{\sum_x t_x}{\sum_x L_x e^{\hat{b}\tilde{x}}} \tag{8.6}$$

In case (8.5) the likelihood equations to be solved for $\hat{a}$, $\hat{b}$, $\hat{c}$, and $\hat{d}$ are

$$\sum_{x=20}^{65} L_x X = \sum_{x=20}^{65} \frac{t_x}{\hat{a} + \hat{b}\tilde{x} + \hat{c}\tilde{x}^2 + \hat{d}\tilde{x}^3} X \tag{8.7}$$

where $X^\top = (1, \tilde{x}, \tilde{x}^2, \tilde{x}^3)$.

Iterative Solutions These equations are numerically well-behaved and can be solved by a dozen iterations of the Newton–Raphson algorithm. (As starting values for the cubic we took the first derivatives from the exponential fit at $x = 40$, for the exponential fit we chose the initial value $b_0 = 0$.)

8.3 Exponential and Cubic

The first plot of Figure 10 shows the exponential and cubic fits together with the numerical smooth and raw mortalities from the standard data set (confer Subsection 2.2). Second, the p-values are plotted as a function of the population rescaling factor *Lfac* (confer Subsection 2.3). The third plot shows the differences between these fits and the numerical smooth.

The numerical values $\tilde{q}_x^{\rm exp}$ and $\tilde{q}_x^{\rm cub}$ are tabulated in the last two columns of Table 1.

8.4 Further Developments

Asymptotics and Robustness

As model (8.1)–(8.5) is smoothly parametrized, it should be possible to bring modern asymptotic statistics and infinitesimal robustness to bear on the (non-i.i.d.) estimation problem. Thus, optimally robust alternatives to the MLE may eventually be derived—confer [Ridr94] and [Slatr94].

Dynamic Aspects

As mortalities may change over time a dynamic modelling seems to be required. The models of [FaTu94], Section 8.2, may turn out useful.

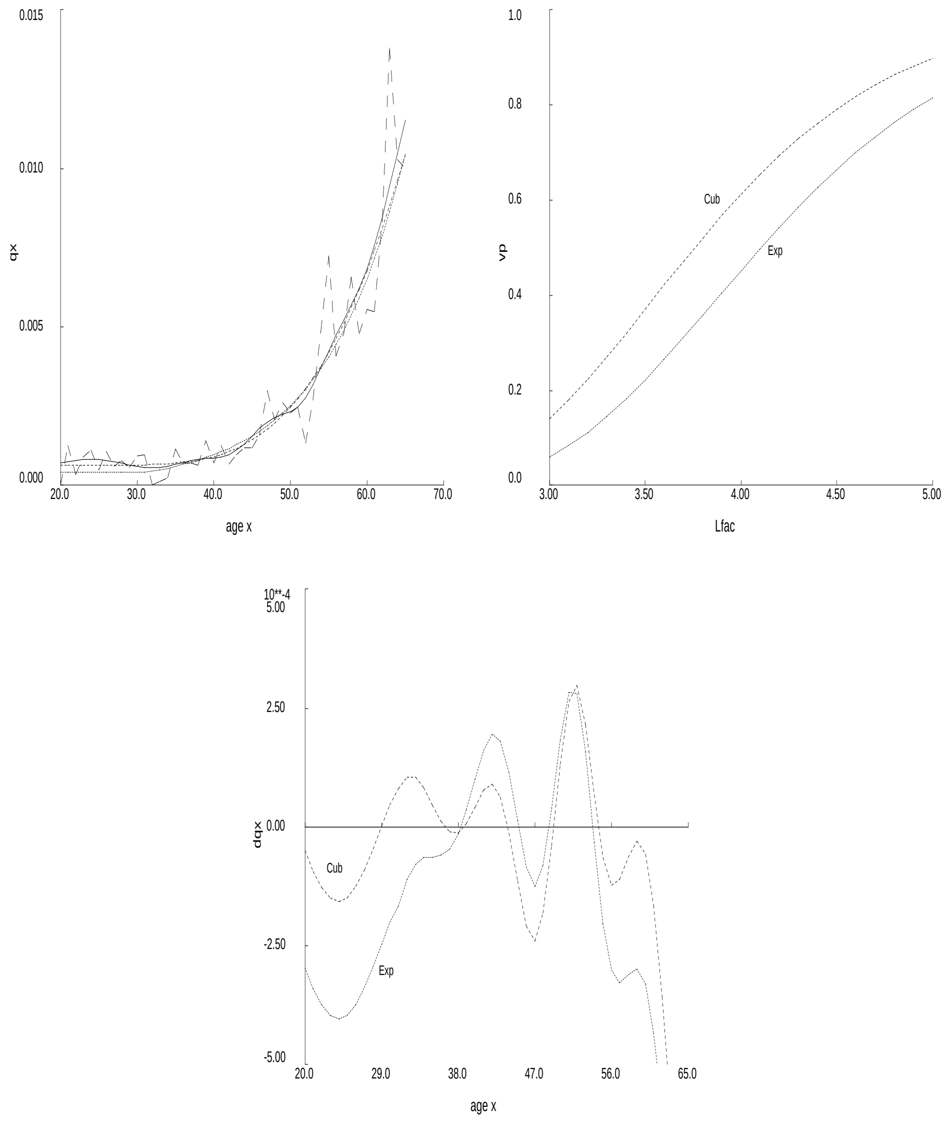

Figure 10: MLE of exponential ($\cdots$) and cubic (- - -) in GLM, with the numerical smooth (—) and corresponding p-values as a function of the population rescaling factor $Lfac$; differences with respect to the numerical smooth

x	t_x^u	L_x^u	t_x^u/L_x^u	q_x^u	${}^{\mathrm{r}}q_x^{\mathrm{roh}}$	$\hat{q}_x^{\mathrm{par}}$	$\tilde{q}_x^{\mathrm{exp}}$	$\tilde{q}_x^{\mathrm{cub}}$
20	0	2011	0.000	0.709	0.468	1.060	0.412	0.660
21	3	2350	1.277	0.752	0.468	0.989	0.412	0.660
22	1	2836	0.353	0.786	0.468	0.916	0.412	0.660
23	3	3287	0.913	0.808	0.468	0.868	0.412	0.660
24	4	3583	1.116	0.816	0.468	0.843	0.412	0.660
25	2	4184	0.478	0.808	0.468	0.804	0.412	0.660
26	5	4602	1.086	0.785	0.468	0.782	0.412	0.660
27	3	4852	0.618	0.750	0.468	0.771	0.412	0.660
28	4	5025	0.796	0.706	0.468	0.764	0.412	0.660
29	3	5300	0.566	0.658	0.468	0.754	0.412	0.660
30	5	5410	0.924	0.613	0.468	0.750	0.412	0.660
31	6	6092	0.985	0.579	0.468	0.729	0.412	0.660
32	0	6549	0.000	0.564	0.512	0.723	0.453	0.670
33	1	6935	0.144	0.576	0.560	0.726	0.498	0.681
34	2	7858	0.255	0.612	0.612	0.728	0.548	0.694
35	12	10428	1.151	0.666	0.669	0.722	0.603	0.712
36	8	10926	0.732	0.723	0.732	0.750	0.663	0.735
37	10	14121	0.708	0.775	0.801	0.770	0.729	0.765
38	9	14433	0.624	0.816	0.876	0.818	0.802	0.803
39	24	16988	1.413	0.845	0.958	0.866	0.882	0.851
40	15	21000	0.714	0.869	1.048	0.924	0.970	0.911
41	32	24955	1.282	0.906	1.146	0.996	1.067	0.984
42	19	28066	0.677	0.979	1.254	1.084	1.174	1.071
43	34	35233	0.965	1.110	1.371	1.182	1.291	1.173
44	37	30860	1.199	1.305	1.500	1.306	1.419	1.294
45	41	34627	1.184	1.546	1.641	1.439	1.561	1.432
46	52	32944	1.578	1.799	1.795	1.594	1.717	1.592
47	112	37359	2.998	2.013	1.963	1.763	1.888	1.773
48	70	33997	2.059	2.157	2.147	1.960	2.077	1.977
49	87	33374	2.607	2.245	2.348	2.177	2.284	2.206
50	73	31602	2.310	2.329	2.569	2.419	2.512	2.461
51	80	32373	2.471	2.480	2.810	2.686	2.763	2.744
52	42	31366	1.339	2.757	3.073	2.984	3.038	3.056
53	85	31737	2.678	3.183	3.361	3.311	3.342	3.399
54	83	17127	4.846	3.706	3.677	3.699	3.675	3.773
55	96	13225	7.259	4.246	4.022	4.114	4.042	4.182
56	39	9517	4.098	4.747	4.399	4.579	4.445	4.625
57	45	9303	4.837	5.216	4.811	5.058	4.889	5.105
58	39	5915	6.593	5.688	5.263	5.640	5.377	5.624
59	34	7091	4.795	6.211	5.756	6.181	5.914	6.181
60	23	4147	5.546	6.835	6.296	6.898	6.504	6.780
61	28	5118	5.471	7.586	6.887	7.524	7.153	7.421
62	37	4412	8.386	8.462	7.533	8.283	7.867	8.106
63	41	2970	13.805	9.434	8.239	9.180	8.652	8.837
64	27	2627	10.278	10.465	9.012	10.074	9.515	9.614
65	4	400	10.000	11.526	9.857	13.097	10.465	10.440

Table 1: Number of deaths t_x^u, size L_x^u of group at age x, raw mortality $q_x^u = t_x^u/L_x^u$, and numerical smooth q_x^u taken from table 3 of [Loeb94]; the $\log q_x$ based minL1/IQR fit ${}^{\mathrm{r}}q_x^{\mathrm{roh}}$ computed according to (3.3)–(3.5) and $\sqrt{q_x}$ based minL2/VAR fit $\hat{q}_x^{\mathrm{par}}$ of a parabola; $\tilde{q}_x^{\mathrm{exp}}$ denotes the exponential and $\tilde{q}_x^{\mathrm{cub}}$ the cubic MLE fit in the Poisson GLM. The unit of probabilities is 10^{-3}.

Acknowledgement I thank Simone Fischer for programming support during her assistantship at the University of Munich.

References

[BaKo78] Basset, G. and Koenker, R. (1978): Asymptotic Theory of Least Absolute Error Regression. *J. Amer. Statist. Assoc.* **73** 618–622.

[BePo82] Benjamin, B. and Pollard, J.H. (1982): *The Analysis of Mortality and Other Actuarial Statistics*, 2nd edition. Heinemann, London.

[BlSt83] Bloomfield, P. and Steiger, W.L. (1983): *Least Absolute Deviations—Theory, Applications, and Algorithms*. Birkhäuser, Boston.

[FaTu94] Fahrmeir, L. and Tutz, G. (1994): *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Verlag, New York.

[Hoyl75] Hoyle, M.H. (1975): Estimating generating functions. *Ann. Statist.* **3** 1361–1363.

[Hubr81] Huber, P.J. (1981): *Robust Statistics*. Wiley, New York.

[ISP95] PC-ISP (1995): *Interactive Scientific Processor*, Version 3.1. Datavision AG, Klosters, Switzerland.

[Loeb94] Loebus, H. (1994): Bestimmung einer angemessenen Sterbetafel für Lebensversicherungen mit Todesfallcharakter. *Bl. d. DGVM* **21** 497–524.

[Millr84] Miller, D. (1984): Reducing Transformation Bias in Curve Fitting. *The American Statistician* **38** 124–126.

[OlMi96] Olbricht, W. und Miller, K. (1996): Entwicklung unternehmenseigener Sterbetafeln. *Bl. d. DGVM* **23** 501–514.

[OlKo96] Olbricht, W. (1996): *Zur Entwicklung unternehmenseigener Sterbetafeln.* Math. Kolloquium, Univ. Bayreuth, 8 February 1996.

[Ridr94] Rieder, H. (1994): *Robust Asymptotic Statistics*. Springer, New York.

[Slatr94] Schlather, M. (1994): *Glattheit von Generalisierten Linearen Modellen und statistische Folgerungen*. Diplomarbeit, Universität Bayreuth.

[SliSt95] Schlittgen, R. and Streitberg, B. (1995): *Zeitreihenanalyse*, 6th ed. Oldenbourg Verlag, München.

[Vera91] Verrall, R.J. (1991): On the estimation of reserves from log-linear models. *Insurance: Mathematics and Economics* **10** 75–80.

[Wolff70] Wolff, K.–H. (1970): *Versicherungsmathematik*. Springer, Berlin.

Zusammenfassung

Bei der Schätzung von Sterblichkeiten mittels log-linearer Regression und Rücktransformation gemäß der Formel $\mathrm{E}\, e^{Y} = e^{\mu+\sigma^2/2}$ entsteht ein systematischer Bias, es sei denn, die Fehlerverteilung ist exakt normal und der Skalenschätzer schätzt die Varianz. Dies folgt aus dem Eindeutigkeitssatz für die Laplace–Transformierte.

Unter uniformer Fehlerverteilung und vier kontaminierten Normalverteilungen bestimmen wir den Bias für Minimum–L_2 und $-L_1$ Schätzungen mit Stichprobenvarianz und Quartilsabstand der Residuen als Skalenschätzer. Schon bei unscheinbarer Kontamination können die wahren Sterblichkeiten im statistischen Mittel systematisch um 50% unterschätzt werden.

Überdies führt die logarithmische Transformation zu einer Instabilität des Modells, welche sich in einer großen Diskrepanz der Schätzer vom Typ *rg_Huber* bei sich ändernder tuning–Konstante, die den Grad der Robustheit steuert, äußert.

Im Unterschied zum Logarithmus stabilisiert die Wurzel–Transformation die Varianz, sie dämpft den Einfluß von Ausreißern, beobachtete Null–Häufigkeiten verursachen keine Probleme, sie führt auf die 'nichtparametrische' Rücktransformationsformel $\mathrm{E}\, Y^2 = \mu^2 + \sigma^2$ und verhindert im homoskedastischen Fall einen systematischen Bias der Minimum–L_2 Schätzung mit Stichprobenvarianz.

Für die unternehmensspezifische Tafel 3 in [Loeb94] passen wir im Altersbereich 20–65 Jahre eine Parabel an die Wurzeln der Sterblichkeiten an und zwar mittels Minimum–L_2, $-L_1$ und robusten *rg_Huber* Schätzungen, sowie ein kubisches Polynom und eine Exponentialfunktion mittels Kleinste–Quadrate. Die damit im Originalmodell erzielten Anpassungen sind hervorragend und praktisch mit einem χ^2–Anpassungstest nicht zu unterscheiden.

Schließlich verwenden wir ein Poissonsches generalisiertes lineares Modell und schätzen eine Exponentialfunktion und ein kubisches Polynom nach der Maximum–Likelihood Methode ohne jegliche Transformation von Beobachtungen.

Prof. Dr. Helmut Rieder
Mathematik VII
Universität Bayreuth
D-95440 Bayreuth