

ANALYSIS OF OLDEST-OLD MORTALITY: LIFETABLES REVISITED¹

BY JANE-LING WANG, HANS-GEORG MÜLLER AND WILLIAM B. CAPRA

University of California, Davis

This paper provides a data analysis and some methodological advances which contribute to an ongoing scientific debate about the patterns of aging. One of the problems we address is how to estimate a hazard function when only aggregated information on the lifetimes in the form of a life table is available. This problem affects the estimation of oldest-old mortality which in turn plays an important role in the quantification of biological lifespan and longevity. We illustrate these issues with an analysis of mortality data obtained from cohorts of nematodes. The methods involve data transformation with the aim of bias reduction when estimating the hazard function. We provide rigorous asymptotic results for the smoothing of life tables and show that the transformation approach is supported by both asymptotic and simulation results. We also demonstrate how the information contained in many samples of life tables, as typically obtained in aging experiments, can be summarized in a two-dimensional hazard surface.

1. Introduction. Recent publications have suggested that, in contrast to earlier assumptions, the mortality of the very oldest individuals of a species shows distinct signs of deceleration. Suzman, Willis, and Manton (1992) suggested that many of these “oldest-old” individuals are actually healthier than one might have expected. They observed that the mortality of the elderly in the United States, as measured by the hazard function, decelerates somewhat after age 85. Such findings, if confirmed, will have to be taken into consideration when predicting future numbers of elderly and in particular of “oldest-old,” that is, those aged 85 or above. The consequences for the future social security, pension, and health care systems are potentially enormous. Formally, the hazard function, also referred to as force of mortality by demographers, is defined by

$$\lambda(t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} P(T \in [t, t + \Delta] | T > t),$$

where T is the random lifetime of an individual and P denotes probability.

Since human data are neither very precise nor obtained in a controlled environment, large-scale experiments with the aim of determining biological

Received September 1996; revised September 1997.

¹Research supported in part by NSF Grants DMS-93-12170, DMS-93-05484, DMS-94-04906 and DMS-96-25984.

AMS 1991 subject classifications. Primary 62G07, 62P10; secondary 62N05, 62P05.

Key words and phrases. Hazard function, hazard surface, smoothing, local polynomial fitting, discretization bias, transformation, samples of life tables, aging, nematodes, biodemography.

patterns of survival and longevity were conducted with medflies by Carey, Liedo, Orozco and Vaupel (1992), with *Drosophila* by Curtsinger, Fukui, Townsend and Vaupel (1992) and with nematodes by Brooks, Lithgow and Johnson (1994). A particular goal of these studies was to clarify the question whether a possible “deceleration of oldest-old mortality” exists in these populations.

In a study based on the survival of more than a million medflies which caught the attention of demographers and the scientific world at large, Carey, Liedo, Orozco and Vaupel (1992) observed a notable deceleration of mortality in a study of the long-term survival of cohorts of the Mediterranean fruit fly *Ceratitis capitata*. For this species, after a certain age, further aging apparently does not lead to further heightened risk of death, and the hazard rates were observed to decline for the “oldest-old” group. In particular, this rules out the classical Gompertz model assumption [Gompertz (1825)] for the survival distribution. The scientific debate on future trends in life expectancies and whether the Gompertz model is applicable or not is still unsettled [see Perls (1995), Kelner and Marx (1996) and Jazwinski (1996)]. In this paper, we address some statistical issues which are relevant to this debate.

The Gompertz model, which is still in common use to describe the trajectory of aging, implies that mortality increases exponentially with age; the logarithm of the hazard function $\lambda(t)$ is assumed to be a linear function of the age t . Thus,

$$(1.1) \quad \log(\lambda(t)) = \log(\beta_0) + \beta_1 t,$$

or equivalently, $\lambda(t) = \beta_0^{\beta_1 t}$. This is an extreme value distribution which corresponds to sharply increasing risk for the oldest-old. Oldest-old mortality is commonly expected to reflect high-risk situations. The plausibility of this expectation and the competing risks interpretation explains the popularity of the Gompertz distribution. Thus the results of Carey, Liedo, Orozco and Vaupel (1992), who showed that in fact a deceleration takes place, were quite unexpected.

One possible explanation for the deceleration of mortality is heterogeneity of the population. Models for composite populations, which take heterogeneity into account, are often referred to as frailty models and could provide an explanation for these findings [see, e.g., Vaupel and Carey (1993)]. Frailty models in demographic applications go back to Vaupel, Manton and Stallard (1979); see also Manton, Stallard, and Vaupel (1986). In such models it is assumed that individual hazard rates follow a given parametric model, for instance the Gompertz model, where the coefficients are assumed to be random and to correspond to varying degrees of individual “frailty.” Hazard rate deceleration at later ages could then be attributed to the fact that the more frail individuals die at earlier ages, so that individuals still alive at older ages are a selected group of more robust subjects with relatively lower mortality.

A consequence of this demographic selection effect is that if the population were in fact homogeneous, then hazard rates would not decelerate at later

ages. Thus any observed deceleration in hazard rates is an artifact of the heterogeneity in a population, and not of a slowing or deceleration of risk at the individual level.

A simple demonstration of the fact that frailty models can lead to an apparent nonmonotone population hazard function even though the hazard functions of all individuals are strictly monotone increasing is given in Figure 1. It is assumed here that the population consists of a mixture of two subpopulations, each occurring with probability $\frac{1}{2}$. The survival distribution of both subpopulations is Gompertz, for subpopulation 1 with parameters $(\beta_{01}, \beta_{11}) = (2, 0.025)$, and for subpopulation 2 with parameters $(\beta_{02}, \beta_{12}) = (1.5, 0.15)$. The fact that individual monotone hazard functions can lead to a nonmonotone population hazard function was also pointed out in Vaupel and Yashin (1985).

Other parametric models that have been advocated for mortality include the Weibull and logistic models [Congdon (1993), Juckett and Rosenberg (1993), Wilson (1994)]. Instead of dwelling on the various parametric models, we reanalyzed these data using nonparametric smoothing techniques, with the aim of “letting the data speak for themselves” and focusing on the oldest-old group. While carrying out this analysis, we encountered two issues of general interest: first, how to estimate hazard functions from lifetable data

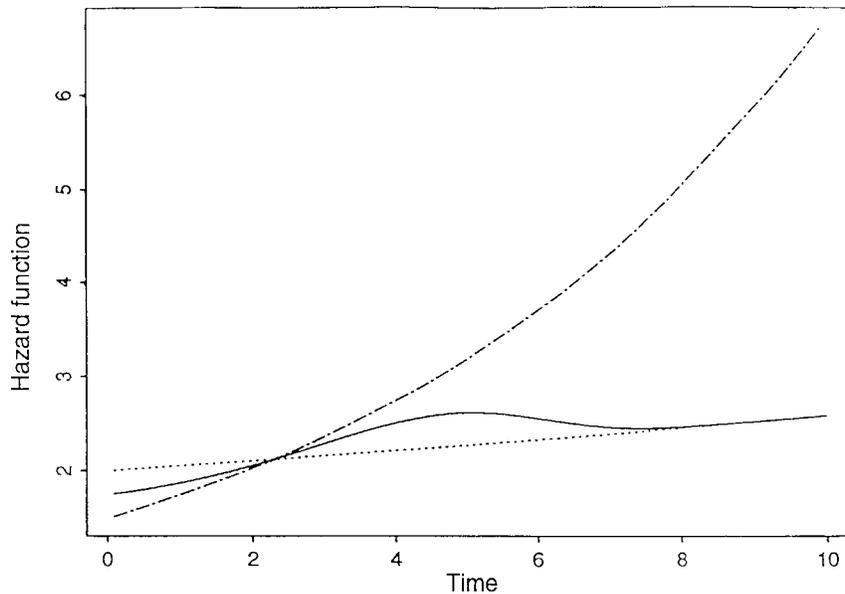


FIG. 1. Hazard functions for homogeneous and heterogeneous populations. Hazard function for Gompertz ($\beta_{01} = 2, \beta_{11} = 0.025$) is dotted, hazard function for Gompertz ($\beta_{02} = 1.5, \beta_{12} = 0.15$) is dash-dotted, and hazard function of mixture where each of the original models occurs with probability $\frac{1}{2}$ is solid.

which are often aggregated or discretized; and second, how to summarize the information contained in samples of lifetables.

The second issue arises since, in studies of oldest-old mortality, typically many cohorts are observed, each of which produces a set of lifetable data. The first issue has already been studied extensively both in the demographic and statistical literature. However, we discovered a bias effect due to the discretization of the lifetable data, which is caused by the aggregation of the lifetimes into intervals. This bias is particularly noticeable in the right tail when the hazard function is increasing in that region. Moreover, this bias may create the visual effect of a deceleration of the hazard function in the right tail. Therefore, this is an important issue when analyzing oldest-old mortality.

Our research was motivated by the analysis of data on the survival of 79 cohorts of genetically homogeneous nematodes (*Caenorhabditis elegans*) by Brooks, Lithgow and Johnson (1994), hereafter referred to as BLJ. These authors aimed to resurrect the universality of the Gompertz model. They attempted to demonstrate that for genetically homogeneous cohorts of nematodes, the Gompertz model fits well. The implication was that Carey, Liedo, Orozco and Vaupel's (1992) findings on mortality deceleration would be due to heterogeneity and would not be observed if homogeneous individuals were involved. However, the analysis of BLJ ignored the occurrence of censored data (40% were right censored) and, moreover, the Gompertz model fits presented in BLJ appear to provide at best a questionable fit to the data.

Nonparametric estimation of hazard functions from lifetables is an important alternative to parametric modeling, since the choice of a parametric model turns out to be rather difficult; moreover, nonparametric methods can be used to check informally the validity of a parametric model. Hoem (1976a) compared parametric modeling of mortality [so-called analytic graduation; see Hoem (1972)] with a nonparametric estimate [so-called linear graduation; see Hoem (1984)]. A number of commonly used smoothing methods like smoothing splines, kernel estimates and local polynomial fitting can be used to implement the basic step of smoothing lifetables. In the approach proposed here, local polynomial fitting methods are used as implementation of smoothing methods. Since nonparametric estimates do not impose restrictive model assumptions in contrast to the Gompertz model or other parametric approaches, they are especially suitable for exploratory data analysis.

As we show in the following, straightforward implementation of smoothing for lifetable data may result in sizeable biases of the resulting hazard function estimate, in particular for the oldest-old age range. This is due to the discretization bias mentioned above. The nematode data of BLJ provide a case in point. Here, 79 genetically homogeneous cohorts with 30 nematodes in each cohort were observed for 67 days, and a second experiment involved the observation of survival in a mass culture of 180,000 genetically homogeneous nematodes raised at a somewhat higher temperature. The discretization of these lifetime data occurs as dead nematodes were counted at the end of each day. The exact lifetimes are therefore unknown. Censoring also occurs, as some of the nematodes vanish from the observation area.

We use this example to point out the shortcomings of the directly smoothed lifetable estimate [Hoem (1976b, 1984)] and to investigate a modified estimate based on a data transformation which substantially reduces the bias encountered at the oldest lifetimes. Such a transformation is particularly beneficial for diagnostic assessment of the Gompertz model, as the nontransformed data lead to a bending of the hazard rate estimate near the right boundary. Thus even if the data were from a Gompertz distribution, the estimated hazard rate could create a false impression of deceleration at advanced ages. We also discuss how the data obtained from various cohorts, as customarily obtained in mortality studies, can be combined in a hazard surface plot.

Our paper is organized as follows: the problems associated with analyzing aggregated lifetime data are discussed in Section 2. The proposed transformation estimate is developed in Section 3. Details of the application to the problem of assessing oldest-old mortality for the BLJ nematode data are presented in Section 4. This section also contains a discussion of practical aspects such as bandwidth selection. The results of a simulation study evaluating the performance of the nonparametric estimates are reported in Section 5. Section 6 contains concluding remarks.

Asymptotic justifications for the proposed transformation method are provided in Sections 2 (Theorem 2.1) and 3 (Theorems 3.1, 3.2). Proofs and auxiliary results, some of which are of interest in their own right, are compiled in the Appendix. We note here that even for the untransformed smooth lifetable estimates, not much is known about their asymptotic properties. The bias and variance expressions in Theorem 3.1 are new, to the best of our knowledge.

2. Transforming the discretized hazard function. Associated with each of the n subjects enrolled in a study of mortality are the data pairs (T_i, C_i) , $i = 1, \dots, n$, where T_i is the lifetime and C_i is a censoring time. Usually these times are considered to be continuous random variables. If the subjects were followed continuously, then the observed data are (X_i, δ_i) , $i = 1, \dots, n$, under the usual random censorship model, where $X_i = \min(T_i, C_i)$, and $\delta_i = 1_{[T_i \leq C_i]}$ is the censoring indicator. Among many other authors, Anderson and Senthilselvan (1980), Gray (1990), and Müller and Wang (1994) discussed nonparametric estimation of the hazard function when the actual data (X_i, δ_i) , the observed time of death or failure, are observed. Theoretical results for this situation can be found in Ramlau-Hansen (1983), Tanner and Wong (1983), Yandell (1983) and Müller and Wang (1990). But usually there is some, possibly small, interval over which observed lifetimes are aggregated, since subjects typically cannot be monitored continuously. Aggregation effects were discussed previously in Tanner and Wong (1984).

In a typical lifetable or mortality study, data are not observed continuously or are rounded to the nearest unit. For example, when studying mortality of insects or nematodes, data are collected daily, and human mortality data are

usually recorded to the nearest year. Thus instead of actually observing (X_i, δ_i) for individual i , we observe data accumulated over a partition of p intervals, I_1, I_2, \dots, I_p , of the fixed interval $[0, L]$ for some constant $L > 0$.

In order to simplify the presentation and avoid unnecessary technicalities, we present in the following the practically most common and relevant case, where the intervals I_i are of equal length. The nonequal length case leads to analogous results as long as certain regularity conditions like the existence of a design density for the end points of the intervals in the sense of Sacks and Ylvisaker (1970), are satisfied. In the equal length case, we assume that the length of these intervals is fixed at $\Delta > 0$. Thus Δ determines the degree of aggregation, that is, $\Delta = L/p$, and may range from one year, when studying human mortality from lifetables, to one day when studying mortality of insects or nematodes.

The midpoint of the j th interval I_j is then $t_j = \Delta(j - 1) + \Delta/2$, while $I_j = [\Delta(j - 1), \Delta j)$, $j = 1, \dots, p$. In mortality studies based on lifetables, one does not observe the actual lifetimes (X_i, δ_i) , but instead the aggregated data $(\mathbf{1}_{(X_i \in I_j, j=1, \dots, p)}, \delta_i)$, $i = 1, \dots, n$. Thus in lifetable studies, one does not observe the true time of death or censoring X_i , but only the fact that $X_i \in I_j$ for some j .

The data observed for each interval I_j can be summarized as (n_j, d_j) , $j = 1, \dots, p$, where d_j is the number of observed deaths in the interval I_j , and n_j is the number of subjects at risk at the beginning of the j th interval I_j :

$$d_j = \sum_{i=1}^n \mathbf{1}_{(X_i \in I_j, \delta_i=1)}, \quad n_j = \sum_{i=1}^n \mathbf{1}_{(X_i > \Delta(j-1))}.$$

A raw estimate of the hazard function based on lifetables consists of the data pairs $(t_j, \tilde{q}(t_j))$, $j = 1, \dots, p$. Here,

$$(2.1) \quad \tilde{q}(t_j) = \frac{d_j}{\Delta n_j}$$

is the proportion of subjects at risk at the start of the j th interval I_j who die during that interval, standardized by the interval length Δ . For $\Delta = 1$, demographers refer to $\tilde{q}(\cdot)$ as the (crude) death rate. Note, however, that $\tilde{q}(t)$ is an empirical estimate of the discretized version of the hazard function defined by

$$(2.2) \quad q(t) = \Delta^{-1} P \left(T \in \left(t - \frac{\Delta}{2}, t + \frac{\Delta}{2} \right] \middle| T > t - \frac{\Delta}{2} \right) \\ = \frac{\bar{F}_T(t - \Delta/2) - \bar{F}_T(t + \Delta/2)}{\Delta \bar{F}_T(t - \Delta/2)},$$

where $\bar{F}_T(t) = P(T > t)$ is the survival function. Demographers call this function q the (conditional) probability of dying at age t (when $\Delta = 1$). One expects heuristically that $E(\tilde{q}(t)) \approx q(t)$ and $q(t) \approx \lambda(t)$. Contrary to this

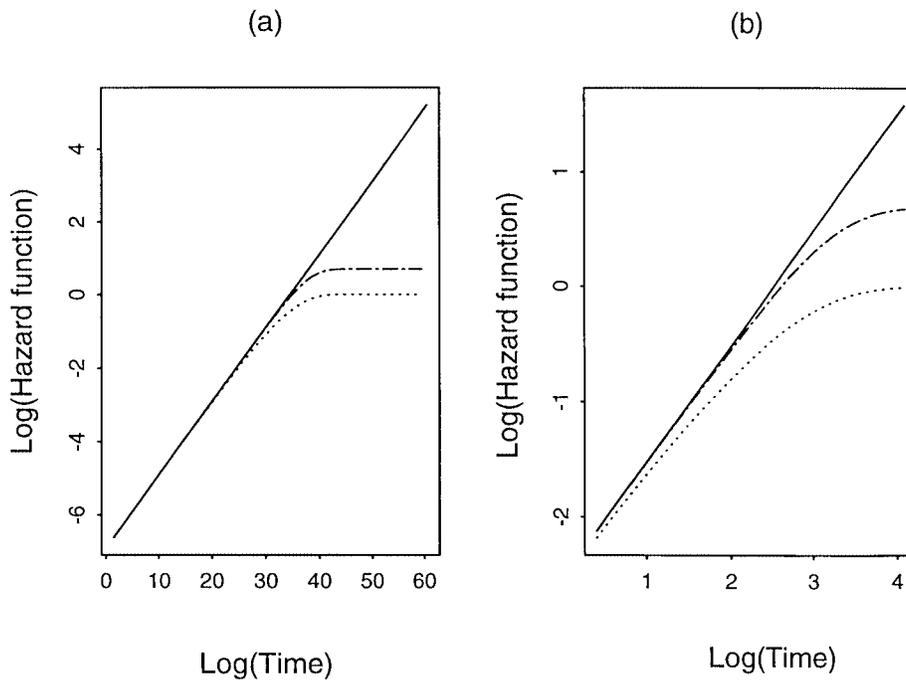


FIG. 2. (a) Comparison of true hazard function $\lambda(t)$ with $q(t)$, $q_c(t)$ and $\phi(q(t))$ for the Gompertz model on the log scale, $\log(\lambda(t)) = \log(0.001) + 0.2t$. Shown are $\log(\lambda(t))$ (solid), $\log(q(t))$ (2.2) (dotted), $\log(q_c(t))$ (2.4) (dash-dotted), and $\log(\phi(q(t)))$ (2.6) (dashed). Horizontal axis is t . Here the difference between the $\log(\lambda(t))$ and $\log(\phi(q(t)))$ is constant at 0.00167 and the differences between the two curves are thus not noticeable; that is, solid and dashed graphs coincide. (b) Same as (a) but for Weibull case with log hazard function $\log(\lambda(t)) = \log(0.08t)$. Horizontal axis is $\log(t)$. Here $\log(\lambda(t))$ is exactly the same as $\log(\phi(q(t)))$.

expectation, Figure 2a shows that $\log(q(t))$ differs significantly from $\log(\lambda(t))$ for the Gompertz distribution at later ages, where λ is the hazard function (1.1). We see that $\log(q(t))$ actually bends at the later ages whereas $\log(\lambda(t))$ is a linear function for the Gompertz model in accordance with (1.1).

Figure 2b shows a similar result for a Weibull distribution with increasing failure rate. Here, $\log(\lambda(t))$ is linear in $\log(t)$, yet we see that $\log(q(t))$ bends for large values of $\log(t)$. The reason is that it follows from (2.2) that $q(t) \leq 1/\Delta$ so that $\Delta \geq 1$ implies $q(t) \leq 1$. This means that any estimator like $\tilde{q}(t)$ or a simply smoothed version which targets $q(t)$ must incur a substantial bias whenever $\lambda(t) > 1$. These straightforward estimates are therefore unsuitable for the analysis of lifetable data as for instance those obtained in the BLJ study, where the focus is on oldest-old mortality, and high values of the hazard function $\lambda(t)$ are expected to occur.

Nonparametric methods with the aim of smoothing lifetable data $(t_j, \tilde{q}(t_j))$ have been discussed by various authors. Borgan (1979) and Hoem and Len-

nemann (1988) discussed moving average techniques in this context, Copas and Haberman (1983) and Bloomfield and Haberman (1987) used kernel smoothers, and Hoem (1976b, 1984) investigated the use of locally weighted least squares, which he referred to as the method of minimum modified chi-square. No matter which of these methods is used, the target function to which the various smoothed lifetable estimates converge is the function $q(t)$ rather than the hazard function $\lambda(t)$. As a consequence, they are all subject to the discretization bias problem.

Another hazard estimate, commonly adopted by practitioners, in particular demographers and actuarial statisticians, is the so-called “central death rate”:

$$(2.3) \quad \tilde{q}_c(t_j) = \frac{d_j}{(\Delta/2)(n_j + n_{j+1})},$$

which targets the function

$$(2.4) \quad \begin{aligned} q_c(t) &= \frac{P(T \in (t - \Delta/2, t + \Delta/2))}{(\Delta/2)(P(T > t - \Delta/2) + P(T > t + \Delta/2))} \\ &= \frac{\bar{F}_T(t - \Delta/2) - \bar{F}_T(t + \Delta/2)}{(\Delta/2)(\bar{F}_T(t - \Delta/2) + \bar{F}_T(t + \Delta/2))} \end{aligned}$$

rather than $q(t)$. This estimate is motivated by the assumption that deaths occur uniformly during the lifetable interval.

Figure 2 also displays the functions $\log(q_c(t))$ for the Gompertz and Weibull distributions, respectively. These figures show that although $\log(q_c)$ stays close to $\log(\lambda)$ for a longer period than $\log(q)$, at later ages it also suffers from a bias problem when $\lambda(t)$ gets large. This bias is simply a consequence of the observation that $q_c(t) \leq 2/\Delta$. Although the size of this bias problem is somewhat reduced as compared to the one for q , it still looms large for large values of $\lambda(t)$. We note that in many applications with modest values of the hazard function, as they typically occur in medical applications, the discretization bias incurred by q or q_c is not particularly prominent.

Since both q and q_c are found to be essentially unsatisfactory to deal with the discretization bias, we take a closer look at the relationship between $q(t)$ and $\lambda(t)$ with the aim of finding an improved transformation of the function $q(t)$. From the relation between hazard and survival functions,

$$\bar{F}_T(t) = \exp\left(-\int_0^t \lambda(u) du\right),$$

we find by simple algebra,

$$(2.5) \quad -\log(1 - \Delta q(t)) = \int_{t-\Delta/2}^{t+\Delta/2} \lambda(u) du.$$

This suggests the transformation

$$(2.6) \quad \phi(x) = -\log(1 - \Delta x)/\Delta.$$

Note that $\phi(q(t)) = (1/\Delta)\int_{t-\Delta/2}^{t+\Delta/2}\lambda(u) du$ is the slope of the cumulative hazard function over the time interval $(t - \Delta/2, t + \Delta/2)$. Some demographers would refer to $\phi(q(t))$ as “discrete-interval hazard function.”

An appealing property of $\phi(q(t))$ is as follows. For the Gompertz distribution,

$$\begin{aligned} \phi(q(t)) &= \frac{1}{\Delta} \int_{t-\Delta/2}^{t+\Delta/2} \beta_0 \exp(\beta_1 t) dt \\ &= \frac{\beta_0}{\Delta \beta_1} \exp(\beta_1 t) \exp(\beta_1 \Delta/2) (1 - \exp(-\beta_1 \Delta)), \end{aligned}$$

and therefore

$$\begin{aligned} \log(\phi(q(t))) &= \log(\beta_0) + \beta_1 t + \beta_1 \frac{\Delta}{2} + \log(1 - \exp(-\beta_1 \Delta)) - \log(\Delta \beta_1) \\ &= \log(\lambda(t)) + \beta_1 \frac{\Delta}{2} + \log(1 - \exp(-\beta_1 \Delta)) - \log(\Delta \beta_1). \end{aligned}$$

This demonstrates that $\log(\phi(q(t)))$ is linear in t with the same slope as $\log(\lambda(t))$, but with a different intercept. Estimation of $\log(\phi(q(t)))$ therefore provides a diagnostic tool to check the Gompertz model assumption. If one finds that $\log(\phi(q(t)))$ is not linear, then the Gompertz model is suspect. Also, $\log(\phi(q(t)))$ is expected to be close to $\log(\lambda(t))$ when Δ or β_1 is small. In Figure 2a, where $\Delta = 1$ and $\beta_1 = 0.2$, the difference between $\log(\phi(q(t)))$ and $\log(\lambda(t))$ disappeared from view as the difference in intercepts between the two lines is only 0.00167. For the particular Weibull distribution in Figure 2b there is no difference between $\log(\phi(q(t)))$ and $\log(\lambda(t))$.

A comparison of q, q_c and $\phi(q)$ in Figure 2 shows that $\log(\phi(q(t)))$ provides by far the best approximation of $\log(\lambda(t))$. Figure 3 provides the same comparison on the original scale. Here the discretization bias associated with $q(t)$ and $q_c(t)$ is even more apparent, while the bias associated with $\phi(q(t))$ still remains negligible and thus disappears from view.

The following result quantifies the asymptotic improvement achievable when one uses $\phi(q(t))$ rather than $q(t)$.

THEOREM 2.1. *If assumptions (A.1)–(A.3) of the Appendix are satisfied for a fixed $t > 0$, then:*

- (i) $\phi(q(t)) = -\frac{\log(1 - \Delta q(t))}{\Delta} = \lambda(t) + \frac{\Delta^2}{24} \lambda''(t) + o(\Delta^2);$
- (ii) $q(t) = \lambda(t) - \frac{\Delta}{2} \lambda^2(t) + \frac{\Delta^2}{24} [\lambda''(t) + 4\lambda^3(t)] + o(\Delta^2).$

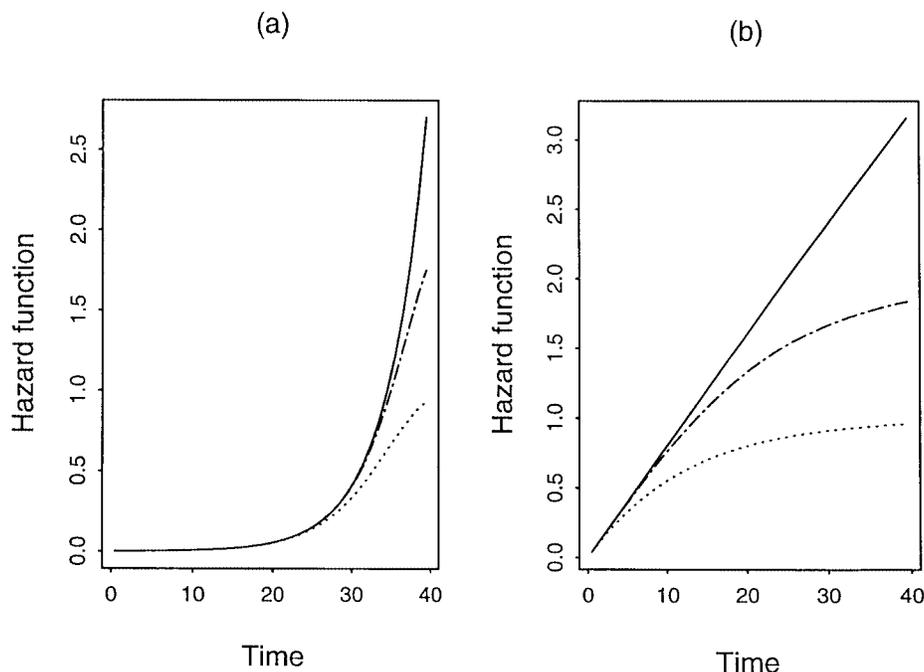


FIG. 3. (a) Comparison of true hazard function $\lambda(t)$ with $q(t)$, $q_c(t)$ and $\phi(q(t))$ for the Gompertz model on the original scale, $\lambda(t) = 0.001 \exp(0.2t)$, for the same functions as in Figure 2. Shown are $\lambda(t)$ (solid), $q(t)$ (2.2) (dotted), $q_c(t)$ (2.4) (dash-dotted), and $\phi(q(t))$ (2.6) (dashed). Horizontal axis is t . Again the differences between $\lambda(t)$ and $\phi(q(t))$ are so small that they are not noticeable. (b) Same as (a) but for Weibull case with hazard function $\lambda(t) = 0.08t$. Horizontal axis is t . Here $\lambda(t)$ is exactly the same as $\phi(q(t))$.

The proof is in the Appendix. This result shows that $\phi(q(t)) = \lambda(t) + O(\Delta^2)$, whereas $q(t) = \lambda(t) + O(\Delta)$, as $\Delta \rightarrow 0$, where Δ is the length of the aggregation intervals. Thus $\phi(q)$ provides a better approximation to the hazard function than q does, finitely as well as asymptotically.

3. Nonparametric estimation of the hazard function from aggregated data. In this section we describe the proposed methods to obtain smooth estimates of the functions $q(t)$ and $\phi(q(t))$. We first note that a variety of smoothing methods are available to estimate a smooth regression function $g(x) = E[Y|X = x]$, given scatterplot data (x_j, y_j) , $j = 1, \dots, p$. We place the estimation problem into a fixed design regression setting, since the lifetable data are aggregated over fixed intervals. The predictors will be the midpoints t_j of the interval I_j , and the corresponding observations are the raw estimates $\tilde{q}(t_j)$ in (2.1).

Let

$$(3.1) \quad S(x, b, (x_j, y_j, w_j)_{j=1, \dots, p}) = \sum_{j=1}^p W_j(x, b) y_j$$

be a general smoothing procedure acting on the scatterplot data (x_j, y_j) and employing case weights w_j . The smoother S is defined by specifying the smoothing weight functions $W_j(x, b)$, where b is a bandwidth or smoothing parameter. Specification of appropriate properties of these smoothing weight functions W_j then leads to consistency properties of S for the target, the regression function $E(Y|X = x)$.

Specific estimates for fixed design nonparametric regression which have this form are, for instance, convolution-type kernel estimators [Gasser and Müller (1979); for an overview, see Müller (1988)] or smooth spline estimators [for an introduction and review see Eubank (1988), Green and Silverman (1994)]. More recently, estimators based on local polynomial fitting, discussed previously by Cleveland (1979), Lejeune (1985) and Müller (1987) have become more popular, in particular after their superior behavior in random design nonparametric regression was established by Fan (1992); see also the monograph by Fan and Gijbels (1996).

It is of interest to note here that the local polynomial fitting method, also known as locally weighted least squares, was used already in 1879 by J. P. Gram. This famous Danish mathematician was the codiscoverer of the Gram-Schmidt orthogonalization and originator of Gram's determinant. Less well known is the fact that he was an early pioneer and perhaps the originator of the method of local least squares. Interestingly enough, he developed this method with the aim of smoothing lifetable data. His original dissertation was written in Danish [Gram (1879)] and the main results were later published in German [Gram (1883)]. The section on smoothing lifetable data was excluded from the 1883 publication and has eluded most statisticians. Compare Hoem (1983) for an interesting historical account of the early history of lifetable smoothing.

A version of Gram's smoothing procedure is what is nowadays known as locally weighted least squares, or local linear or polynomial fitting. The most common version corresponds to fitting linear lines to scatterplot data (x_j, y_j) , $j = 1, \dots, p$, within windows of the predictor variable around a fixed plot. This estimate is defined by

$$(3.2) \quad \begin{aligned} & S_L(t, b, (x_j, y_j, w_j)_{j=1, \dots, p}) \\ &= \arg \min_{a_0} \min_{a_1} \left[\sum_{j=1}^p w_j K\left(\frac{t - x_j}{b}\right) (y_j - a_0 - a_1(x_j - t))^2 \right], \end{aligned}$$

where w_j are case weights, that is, the curve estimate at t is the estimated intercept of the fitted line, which is centered at t . We choose the kernel

function

$$K(u) = \begin{cases} 1 - u^2, & \text{if } |u| \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

which is known to be optimal in the mean square sense [Müller (1987)]. While we use estimators S_L (3.2) in our data examples, the asymptotic results derived in the Appendix are valid for more general smoothers of type (3.1) under regularity conditions on the smoothing weights $W_j(x, b)$ in (3.1).

Case weights are expected to play an important role due to the high degree of heteroscedasticity in lifetable data: assuming that the number of subjects n_j at risk at the start of each interval I_j , $j = 1, \dots, p$, is fixed, then the distribution of $\Delta \tilde{q}(t_j)$, the proportion of deaths in the j th interval, is approximately binomial with variance $\Delta q(t_j)(1 - \Delta q(t_j))/n_j$. Thus the variance increases over time as the number of subjects at risk n_j decreases. Several simulation studies, the results of which are reported in Müller, Wang and Capra (1997), showed that the best weighting scheme is one of the simplest such schemes: choose case weights simply proportional to the number of subjects at risk, that is,

$$(3.3) \quad w_j = n_j.$$

This scheme was therefore implemented in the data applications and simulations reported in the remainder of this section and Section 5.

Our smoothed estimate $\hat{q}(t)$ of $q(t)$ is now

$$(3.4) \quad \hat{q}(t) = \hat{q}(t, b) = S_L\left(t, b, (t_j, \tilde{q}(t_j), w_j)_{j=1, \dots, p}\right),$$

where $\tilde{q}(t_j) = d_j/\Delta n_j$ as in (2.1), and for the hazard function $\lambda(t)$ we propose the estimate

$$(3.5) \quad \phi(\hat{q}(t)) = -\log(1 - \Delta \hat{q}(t))/\Delta.$$

The following asymptotic results, the proofs of which are in the Appendix, allow comparisons of the asymptotic bias and variance behavior of \hat{q} and $\phi(\hat{q})$, viewed as estimates of the hazard function λ . Exact expressions for the bias and variance of \tilde{q} can be found in Theorem A.1, and the following results are based on this core result.

THEOREM 3.1. *If assumptions (A.2)–(A.5) in the Appendix are satisfied for a fixed $t > 0$, then:*

$$(i) \quad E(\hat{q}(t)) = \lambda(t) - \frac{\Delta}{2} \lambda^2(t) + \frac{\Delta^2}{24} \lambda''(t) \\ + 4\lambda^3(t) + \frac{1}{2} b^2 \lambda''(t) \int u^2 K(u) du + o(\Delta^2 + b^2);$$

$$(ii) \quad \text{Var}[\hat{q}(t)] = \frac{\lambda(t)}{nb(1 - F(t))} \int K^2(u) du + o\left(\frac{1}{nb}\right).$$

THEOREM 3.2. *If assumptions (A.1)–(A.5) in the Appendix are satisfied for a fixed $t > 0$, then:*

$$(i) \quad E[\phi(\hat{q}(t))] = \lambda(t) + \frac{\Delta^2}{24} \lambda''(t) + \frac{1}{2} b^2 \lambda''(t) \int u^2 K(u) du + o(\Delta^2 + b^2);$$

$$(ii) \quad \text{Var}[\phi(\hat{q}(t))] = \frac{\lambda(t)}{nb(1 - F(t))} \int K^2(u) du + o\left(\frac{1}{nb}\right).$$

The mean square consistency and rate of convergence of $\hat{q}(t)$ and $\phi(\hat{q}(t))$ now follows immediately from Theorems 3.1 and 3.2. We note that the derivation of the mean squared error properties of the transformed estimators in Theorem 3.2 as given in the Appendix relies on specific properties of lifetable data.

We conclude that while the pointwise variances for these two estimates are the same, the bias behavior relative to the target $\lambda(t)$ differs markedly. While \hat{q} has leading bias terms of the order $\Delta + b^2$, $\phi(\hat{q})$ has leading bias terms of the order $\Delta^2 + b^2$ and is therefore less subject to discretization bias as $\Delta \rightarrow 0$. Moreover, $\phi(\hat{q})$ has the usual and desirable bias behavior for nonparametric curve estimates, in the sense that the bias is proportional to the second derivative $\lambda''(t)$ of the target function, $\lambda(t)$. This allows predicting local bias which will be small near flat parts of the target function λ , downwards near peaks where $\lambda''(t) < 0$ and upwards near troughs where $\lambda''(t) > 0$. In contrast, the bias behavior for \hat{q} is quite complicated and unwieldy for practical purposes.

4. Application to biological lifespan and oldest-old mortality for nematodes.

4.1. Hazard functions from nematode cohort lifetables. In an attempt to demonstrate that the Gompertz model is valid whenever there is only little heterogeneity in a population, BLJ conducted an experiment where they studied the survival of 79 cohorts of nematodes, each of which consisted of 30 nematodes which were genetically identical for each cohort. A large amount of right censoring which occurs in these data was ignored by BLJ in their analysis. In our own analysis of these data, we applied the new estimate $\phi(\hat{q})$, as well as \hat{q} , which both adjust for censoring. For each of these estimators, case weights proportional to the number at risk were used according to (3.3). In a second experiment, BLJ obtained data on the survival of 180,000 genetically homogeneous nematodes which were kept in a mass culture and were maintained at a higher temperature. We compared the estimators \hat{q} and $\phi(\hat{q})$ for these data as well.

Starting with the results for the second experiment, Figure 4 shows the estimates \hat{q} and $\phi(\hat{q})$ for the mass culture data. Until day 13, the sampling

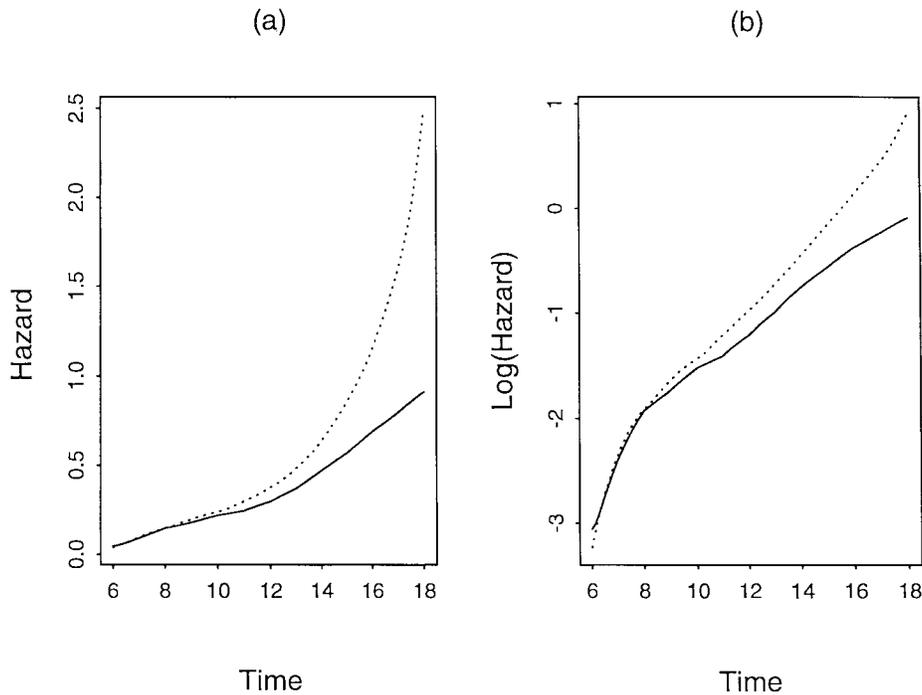


FIG. 4. Estimates \hat{q} (dotted) and $\phi(\hat{q})$ (solid) for the hazard functions of a mass culture of nematodes, for a cohort of 180,000 genetically homogeneous nematodes. Bandwidths \hat{b}_q, \hat{b}_ϕ were fixed at 3 days, respectively, 3.7 days. (a) Estimates shown on original scale. (b) Estimates shown on log scale. Clear differences emerge in the area of oldest-old mortality in the right tail of the hazard function.

procedure was different from that of an ordinary lifetable: on each day, 200 nematodes were sampled at random from the surviving nematodes at the beginning of that day, and the number of dead nematodes in this subsample was recorded. Starting from day 14, when 824 nematodes were still at risk, the full lifetable was observed. This design was devised by the biologists in order to make the experiment feasible.

We took this design into account when assigning the case weights, which were set equal to $w_i = \min(n_i, 200)$, $i = 1, 2, \dots, n_i$ denoting the number of nematodes at risk at the beginning of the i th day. This amounted to $w_i = 200$ for $i = 1, \dots, 14$ and $w_i = n_i$ for $i > 14$. We chose bandwidths $b = 3$ days for \hat{q} and $b = 3.7$ days for $\phi(\hat{q})$. The latter bandwidth was determined according to formula (4.4) below.

We find notable differences in the right tail of these estimates. The observed differences for the transformed versus the untransformed estimates are due to the much larger discretization bias for the nontransformed estimate at high levels of the hazard function. It is clear that the differences

between the two estimates affect conclusions regarding the tail behavior and thus affect the assessment of goodness-of-fit of parametric models such as the Gompertz model.

In this example, both estimates show an initial phase of linear increase of the log hazard function, followed by a slowing of the increase. The estimate \hat{q} clearly shows signs of deceleration in the right tail, while this is not so for $\phi(\hat{q})$, which remains linear or even accelerates on the log scale. This is an instance where visual right tail inspection is heavily influenced by the choice of estimator. It is likely that \hat{q} is severely biased on the right tail, as by its very nature, $\hat{q} \leq 1/\text{day}$; in contrast, $\phi(\hat{q})$ indicates that values much above 1/day do occur here. The fact that the right tail is linear or even accelerated on the log scale probably is possibly due to the accelerated nature of this experiment which was conducted at 25.5°C rather than the normal temperature of 20°C and also at higher food concentrations [see Vaupel, Johnson and Lithgow (1994)]. The early concavity in both log hazard function estimates provides strong evidence against the validity of the Gompertz model for these data, in contrast to the claim made in BLJ.

Further evidence against this claim is found by evaluating the data from the first experiment, where complete lifetables were obtained for 79 cohorts consisting each of 30 genetically homogeneous nematodes. The transformed estimates $\phi(\hat{q})$ for all 79 cohorts, grouped into four quartiles according to mean lifetime of the cohort, are shown in Figure 5. The estimated hazard functions are shown in the log scale and the bandwidths for all curves were fixed at 6 days, in order not to magnify differences between individual hazard curves which may be driven by differences in individual bandwidth choices.

Viewed in their entirety, these nonparametric hazard function estimates demonstrate that the simple Gompertz model certainly is not the correct underlying parametric model for all homogeneous cohorts, contradicting the claim made by BLJ. We find that almost none of the individually estimated hazard functions for the 79 cohorts corresponds to roughly a straight line, as would be required by the Gompertz model. Instead, virtually all these estimates show concave behavior in the log scale. We note here that some or perhaps all authors of BLJ no longer contend that their cohorts fit a Gompertz model at extreme ages.

4.2. Bandwidth choice. The selection of the bandwidth b in the smoother is important for the practical behavior of estimates \hat{q} and $\phi(\hat{q})$. Assuming that the raw data d_i/n_i are approximately independent (this assumption will be justified in Theorem A.1 below), classical methods of bandwidth choice as have been developed for nonparametric regression can be used. We are not going to embark on a discussion of the pros and cons of the various bandwidth selectors. Instead, we focus on the problem how to choose bandwidths for the transformed estimates $\phi(\hat{q})$, once a bandwidth choice for \hat{q} is available. For the latter choice one could use cross-validation, due to its simplicity, or any other of a variety of bandwidth selection methods, including visual bandwidth choices.

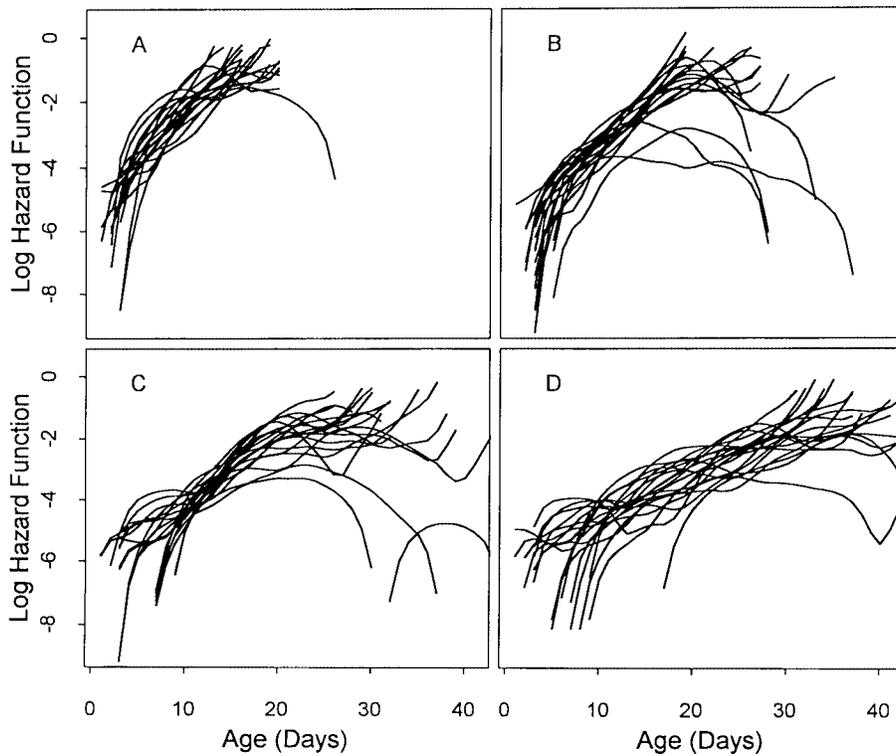


FIG. 5. All estimated hazard functions $\phi(\hat{q})$ for the 79 cohorts of 30 genetically homogeneous nematodes, arranged according to quartile of mean lifetime of the respective cohort. Bandwidths were fixed at 6 days, and hazard function estimates are shown on log scale. First quartile shown in panel (A), second quartile in panel (B), third quartile in panel (C) and fourth quartile in panel (D).

Let $\text{ISE}(\hat{\lambda}, b)$ denote the integrated squared error for an arbitrary smooth estimate $\hat{\lambda}(t, b)$ of $\lambda(t)$, obtained by involving at some step the smoothing procedure S_L with smoothing parameter b as in (3.1), that is,

$$(4.1) \quad \text{ISE}(\hat{\lambda}, b) = \int [\hat{\lambda}(t, b) - \lambda(t)]^2 dt.$$

This can be approximated by the summed squared error

$$(4.2) \quad \text{SSE}(\hat{\lambda}, b) = \frac{1}{p} \sum_{j=1}^p [\hat{\lambda}(t_j, b) - \lambda(t_j)]^2.$$

We define optimal bandwidths b_q and b_ϕ as minimizers of (4.2) when $\hat{\lambda}(t, b) = \hat{q}(t)$ or $\hat{\lambda}(t, b) = \phi(\hat{q}(t))$. These finitely optimal bandwidth choices require knowledge of $\lambda(t)$, and are only useful for simulation comparisons.

We describe now how cross-validation could be used as a method to determine b_q and b_ϕ from the data; compare for instance Rice (1984) for a discussion of cross-validation and related bandwidth selectors. Cross-validation is based on a measure of how well the smoothed estimates predict observed data. Optimal bandwidths b_q are estimated by

$$(4.3) \quad \hat{b}_q = \arg \min_b \sum_{j=1}^p w_j (\hat{q}^{-j}(t_j, b) - \tilde{q}(t_j))^2.$$

Here w_j are the case weights used in the smoothing step (3.2), and \hat{q}^{-j} is the smoothed estimate of $q(t)$, obtained from (3.2) by omitting the j th data pair (t_j, \tilde{q}_j) .

Estimation of optimal smoothing parameters b_ϕ used to obtain the transformed estimate $\phi(\hat{q}(t, b_q))$ requires additional considerations. Since this estimate is obtained by transforming another smooth estimate, cross-validation cannot be applied directly. By the delta method, one finds

$$\text{Var}(\phi(\hat{q})) \approx \text{Var}(\hat{q}) / (1 - \Delta\hat{q})^2.$$

This implies that if we used the same bandwidth for \hat{q} and $\phi(\hat{q})$, we would incur a larger variance for $\phi(\hat{q})$. We propose to choose \hat{b}_ϕ in such a way that the resulting estimate $\phi(\hat{q})$ has the same variance on average as the variance of \hat{q} when using \hat{b}_q .

Given a bandwidth estimate \hat{b}_q of b_q , which could be obtained by (4.3) or by any other reasonable bandwidth selector, this consideration leads to the bandwidth estimate

$$(4.4) \quad \hat{b}_\phi = \hat{b}_q \left[\frac{\sum_{j=1}^p \widehat{\text{Var}}(\tilde{q}(t_j)) / (1 - \Delta\hat{q}(t_j, \hat{b}_q))^2}{\sum_{j=1}^p \widehat{\text{Var}}(\tilde{q}(t_j))} \right]^{1/5}$$

for b_ϕ . Here $\widehat{\text{Var}}(\tilde{q}(t))$ is an estimate of the variance of $\tilde{q}(t)$, given by

$$(4.5) \quad \widehat{\text{Var}}(\tilde{q}(t)) = S_L \left[t, b, \left(t_j, \widehat{\text{Var}}(\tilde{q}(t_j)), w_j \right)_{j=2, \dots, p-1} \right].$$

This estimate of the variance is a smoothed version of the empirical variances $\widehat{\text{Var}}(\tilde{q}(t_j))$, $j = 2, \dots, p - 1$, obtained from the squared residual of the least squares estimate of the curve $q(t)$ at time t_j from the data at times $j - 1$ and $j + 1$. For equidistant t_j 's we obtain

$$\widehat{\text{Var}}(\tilde{q}(t_j)) = \frac{2}{3} \left[\frac{\tilde{q}(t_{j-1}) + \tilde{q}(t_{j+1})}{2} - \tilde{q}(t_j) \right]^2.$$

This residual variance has been proposed by Rice (1984); compare also Hall, Kay and Titterton (1990), Müller and Stadtmüller (1993), and Seifert, Gassser and Wolf (1993) for further developments. We choose the bandwidth $b = (t_p - t_1)/5$ in the smoothing step (4.5).

We note that the use of case weights w_i is only one way to address the heteroscedasticity in the lifetable data. Another possibility would be to adopt

a local (instead of global) bandwidth choice procedure [see Müller, Wang and Capra (1997)].

4.3. *Summarizing samples of lifetables by a hazard surface.* Since the individual sizes of the 79 cohorts of nematodes studied in the main experiment of BLJ are fairly small, it is of interest to combine the survival information contained in the 79 cohorts in one statistical graph. A hazard surface was used previously for this purpose in Wang, Müller, Capra and Carey (1994). We assume that not only age affects the value of the hazard function, but also the mean lifetime of the specific cohort considered, and that both influences are smooth. In particular, we take here the point of view that the hazard function observed for a particular cohort is a random function.

Denote observed mean lifetime (a cohort-specific random variable) by Ω , and age by t , and denote the (random) hazard function for a cohort with mean lifetime Ω at age t by $\lambda(t, \Omega)$. It is then natural to define—in analogy to nonparametric regression—the hazard surface as

$$(4.6) \quad \lambda(t, \omega) = E\{\lambda(t, \Omega) | \Omega = \omega\}.$$

Estimation of this hazard surface requires two smoothing steps, first smoothing in the age direction t , in order to estimate the random hazard functions associated with each cohort, and then smoothing in the mean lifetime direction ω . Using the transformed estimate $\phi(\hat{q}(t, b_\phi))_j$ in (3.5) with bandwidth b_ϕ , based exclusively on data from the j th cohort with observed mean lifetime Ω_j , one thus obtains

$$(4.7) \quad \hat{\lambda}(t, \omega) = S_L(\omega, b_\omega(t), (\Omega_i, \phi(\hat{q}(t, b_\phi))_i)_{i=1, \dots, 79}).$$

Here $b_\omega(t)$ is the bandwidth used for smoothing in the mean lifetime direction when age is t , and b_ϕ is the bandwidth used for smoothing in the age direction.

It turned out to be necessary to let b_ω depend on the age t . Variable bandwidth choice was necessary for this smoothing step, owing to the heteroscedasticity in the hazard function estimates, with higher variances toward the right tail, and also because of the different shapes of the functions $\lambda(t, \omega)$, viewed as functions in the argument ω , when t is varying. For the choice of b_ϕ , we explored two options: (1) fixing b_ϕ at the same value for all cohorts by an intuitively appealing value, for which we choose 4 days; and (2) cross-validation (4.4), carried out separately for each cohort. The choice of the function $b_\omega(t)$ is then done by cross-validating the smoother S_L in (4.7), given the data $(\Omega_j, \phi(\hat{q}(t, b_\phi))_j)$, $j = 1, \dots, 79$, for all fixed t equal to $0, 1, 2, \dots, 30$ days. The resulting cross-validation scores are shown as dots in Figure 6, for the cases where b_ϕ is fixed at 4 days and where b_ϕ is chosen by cross-validation. While these cross-validation bandwidths appear to follow a trend, their variability is quite high and there occur also outlying values. The two choices of b_ϕ are seen to lead to very similar results.

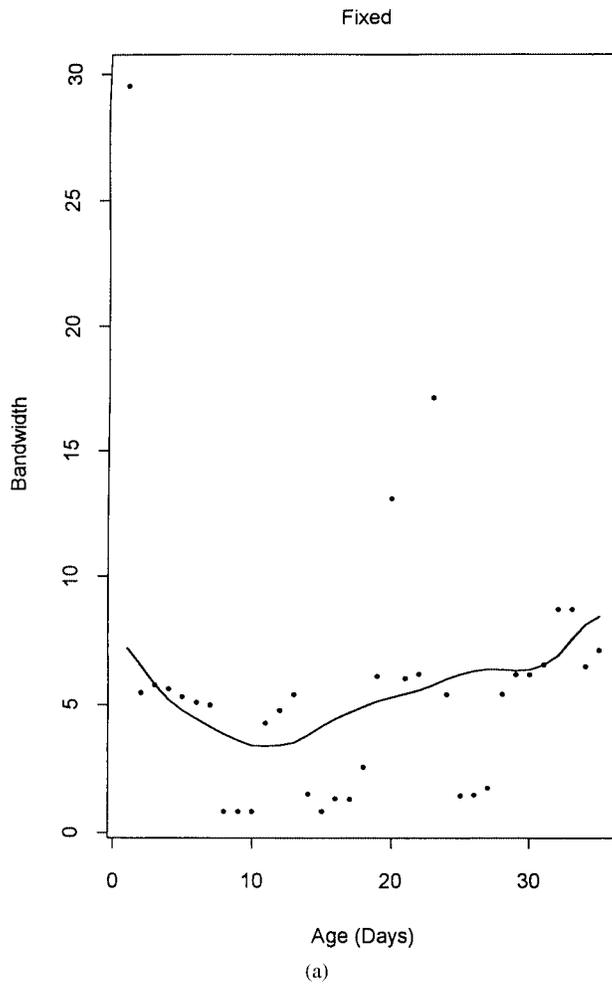


FIG. 6. Cross-validation score functions for choosing the bandwidth for b_ω , smoothing in the mean lifetime direction. (a) With bandwidth b_ϕ for age direction fixed at 4 days. (b) With cross-validation bandwidth choice for b_ϕ .

To arrive at the functions $b_\omega(t)$ actually used for the second smoothing step, we smoothed the scatterplots in Figure 6 by using the locally weighted least squares smoother S_L with bandwidth fixed at 5 days. This led to the estimated log hazard surfaces (4.7) shown in Figures 7 and 8 for the two cases where b_ϕ is fixed at 4 days and where b_ϕ is chosen by cross-validation. As expected, the hazard surface obtained for the cross-validation option is more variable, but overall the estimated hazard surfaces are quite similar.

The hazard surfaces as shown on the log scale in Figures 7 and 8 confirm strong initial bending and concavity for all cohorts, and some flattening

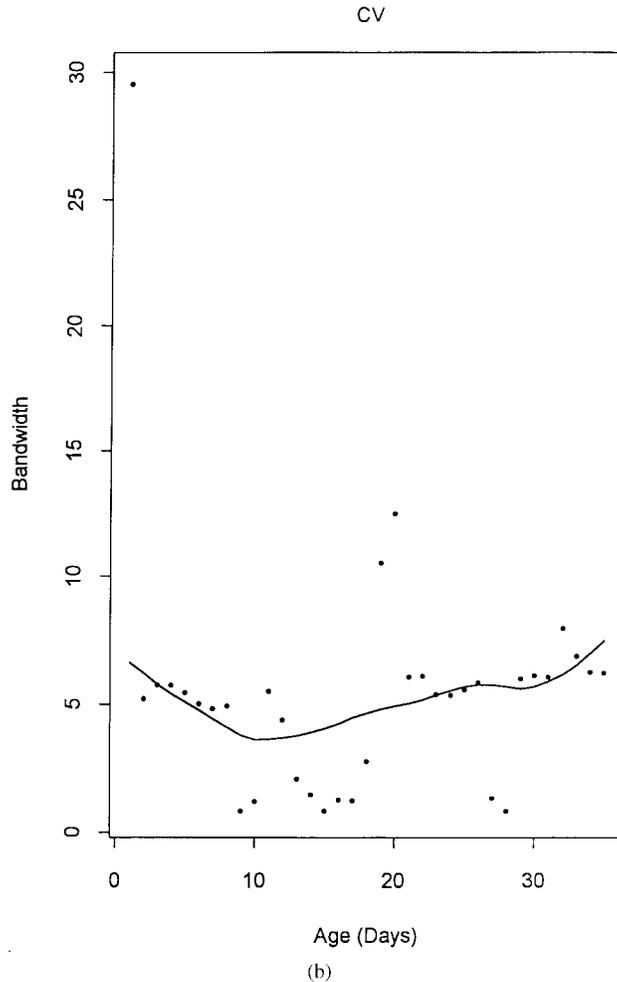


FIG. 6.—Continued

toward the right tail for the cohorts with relatively large mean lifetimes. It also appears that there is a consistent “valley” across cohorts at around age 10, followed by a “mountain ridge.” These features are not compatible with the Gompertz model, which would predict a “ruled” log hazard surface in the sense that cross-sections through the surface for any fixed mean lifetime are linear.

Summarizing the survival behavior observed for various cohorts in the form of a hazard surface plot is a simple and effective graphical tool. It can be applied whenever a covariate or characteristic can be associated with each cohort, and individual hazard functions depend in a smooth fashion on this covariate. If, as is mostly the case in practice, the individual cohort behavior

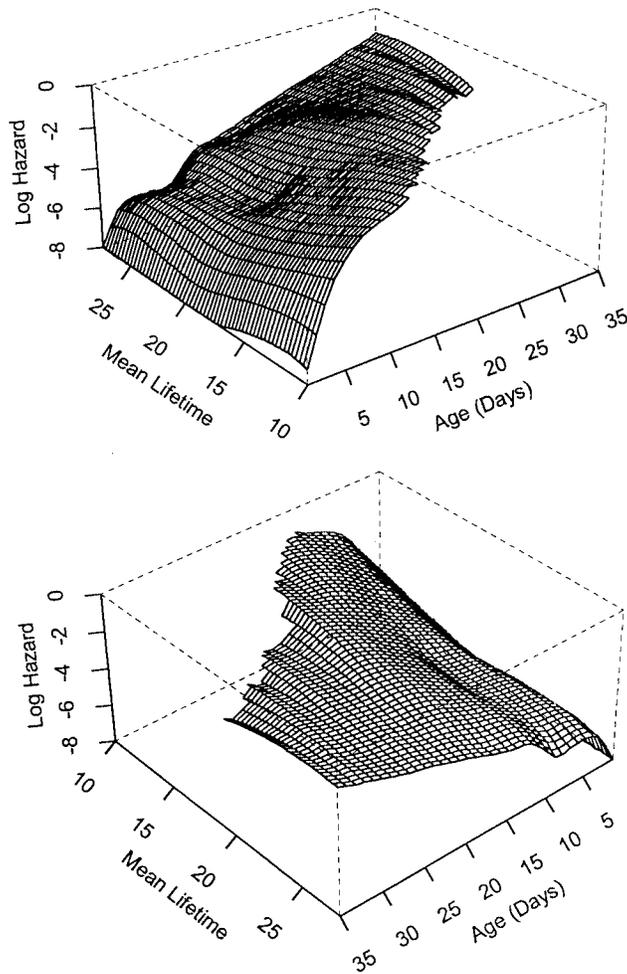


FIG. 7. Estimated hazard surface as a function depending on age and cohort mean lifetime; shown on the log scale from two perspectives, using fixed bandwidth $b_\phi = 4d$ in the age direction, and cross-validation for selecting b_ω , the bandwidth in the mean lifetime direction.

is observed in the form of a lifetable, the hazard surface plot needs to be combined with the transformation method to estimate cohort-specific hazard functions as described in the previous section. The combination of the two methods then leads to a practically useful graphical device for survival analysis in general and the analysis of oldest-old mortality in particular.

5. Simulation results. We conducted a simulation study to address the issues of how the estimators \hat{q} (3.4) and $\phi(\hat{q})$ (3.5) compare in estimating $\lambda(t)$. The performance criterion was the SSE (4.2) achieved by an estimator.

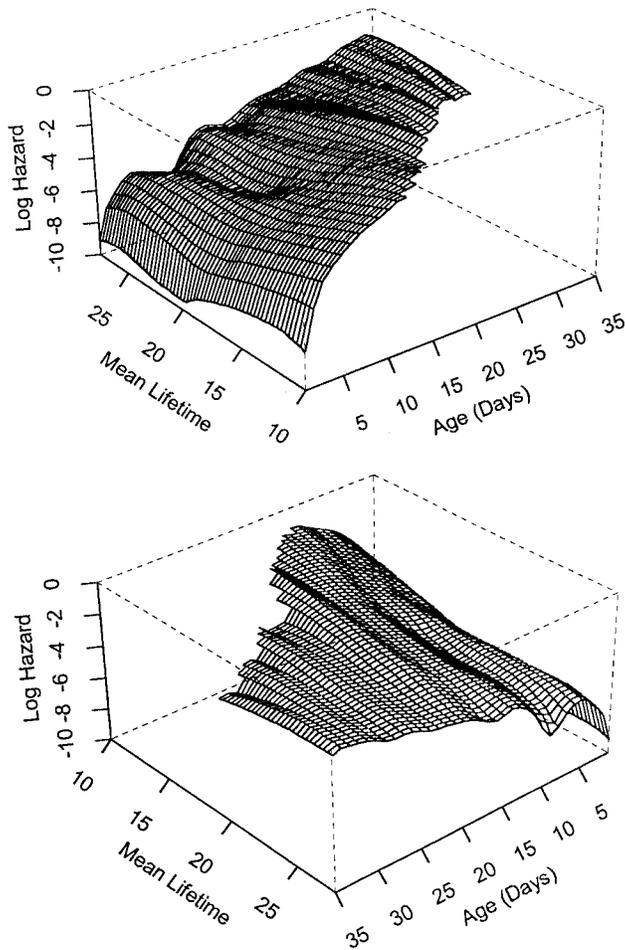


FIG. 8. Estimated hazard surface on the log scale as in Figure 7, but using cross-validation bandwidths for both b_{ϕ} and b_{ω} , the bandwidths for age and mean lifetime directions.

This comparison involved various sample sizes, $n = 30, 100, 1000, 10^4, 10^5$ and 10^6 . The larger the sample size, the more subjects will survive to high ages and therefore the larger the right end point of the range up to which the hazard function can be estimated. For each single Monte Carlo run, the right end point of this range was determined separately as the right end point where the number of subjects at risk dropped to 4. Since this right end point increases with sample size and the SSE is not adjusted for range, this means that SSE's for different sample sizes n , as well as for different models, are not comparable. The values tend to increase with sample size. The only meaningful comparison is between estimates at the same sample size and for the same model.

TABLE 1

Simulation results, comparing estimators \hat{q} (3.4) and $\phi(\hat{q})$ (3.5) for complete (uncensored) data

Model	Estimate	Sample Size					
		30	100	1000	10^4	10^5	10^6
Gompertz	\hat{q}	0.0314	0.0626	0.2941	0.9441	2.224	4.320
Gompertz	$\phi(\hat{q})$	0.0334	0.0329	0.0268	0.0241	0.0170	0.0402
Weibull	\hat{q}	0.0433	0.0850	0.2418	0.5279	0.9226	1.470
Weibull	$\phi(\hat{q})$	0.0571	0.0354	0.0246	0.0175	0.0186	0.0198

Case weights as in (3.3) and finitely optimal bandwidths, minimizing the SSE (4.2) are used. Gompertz model is $\lambda(t) = 0.001 \exp(0.2t)$, and Weibull model is $\lambda(t) = 0.08t$. Reported are SSE's over a range determined separately for each Monte Carlo run. The right end point of this range was the time at which the number of subjects at risk dropped to 4. Note that this point is increasing with sample size n , and so is SSE. For this reason, only comparisons of SSE's for the same sample size and the same model make sense.

In terms of models, we chose a Gompertz model with hazard function $\lambda(t) = 0.001 \exp(0.2t)$ and a Weibull model with hazard function $\lambda(t) = 0.08t$. From these models, lifetime data were simulated. For the Gompertz model, the case of censored data was also investigated by using an exponential censoring distribution. The mean of the censoring distribution was adjusted in such a way that approximately 50% of the data were censored. The originally continuous data were then aggregated into lifetables with aggregation interval length $\Delta = 1$ day.

Weighting (3.3) was used, and bandwidths were chosen in either one of two ways: for the results reported in Table 1, covering only uncensored data, bandwidths were chosen finitely optimal as minimizers of SSE (4.2). For the results reported in Table 2, which include the case of censored Gompertz data as well as uncensored data, data-based bandwidths \hat{b}_q (4.3) for \hat{q} and \hat{b}_ϕ (4.4) for $\phi(\hat{q})$ were used.

From the results in Table 1 it is immediate that, when using optimal bandwidths, it is advantageous to use the transformed estimate $\phi(\hat{q})$ rather than \hat{q} for the Gompertz and Weibull models considered for moderately large or large sample sizes. For sample sizes of 1000, the achievable gain in SSE is of the order of a factor 10, and for sample sizes of 10^6 it is of the order of a factor of 100.

Table 2 shows that these relations are more or less preserved when data-based bandwidth choices (4.3) and (4.4) are used. In this case, the necessary sample size for strong gains of the transformed estimate is still about 100 for the Weibull model and somewhat larger for the Gompertz model with uncensored or censored data. In all cases, the gains are again dramatic for sample sizes of 1000 and beyond.

These simulation results clearly demonstrate the superiority of the transformed estimates for moderate to large sample situations. The severe bias which afflicts the untransformed estimate is visualized in Figure 9, which

TABLE 2

Simulation results paralleling the setting of Table 1, but with estimated bandwidths \hat{b}_q (4.3) and \hat{b}_ϕ (4.4)

Model	Estimate	Sample Size					
		30	100	1000	10^4	10^5	10^6
Gompertz	\hat{q}	0.0523	0.0819	0.3260	1.006	2.311	4.474
Gompertz	$\phi(\hat{q})$	0.0823	0.0826	0.0850	0.0903	0.0656	0.2927
Weibull	\hat{q}	0.0559	0.1120	0.2832	0.6691	1.162	1.803
Weibull	$\phi(\hat{q})$	0.0859	0.0742	0.0851	0.1085	0.0762	0.0739
Gompertz, censored	\hat{q}	0.0406	0.0585	0.1856	0.6510	1.545	3.021
Gompertz, censored	$\phi(\hat{q})$	0.0502	0.0627	0.0653	0.0581	0.0388	0.0416

The case of censored data is included for the Gompertz model, with approximately 50% censoring. Reported are SSE's as in Table 1. Ranges of support and therefore SSE's differ for uncensored and censored data and are not comparable.

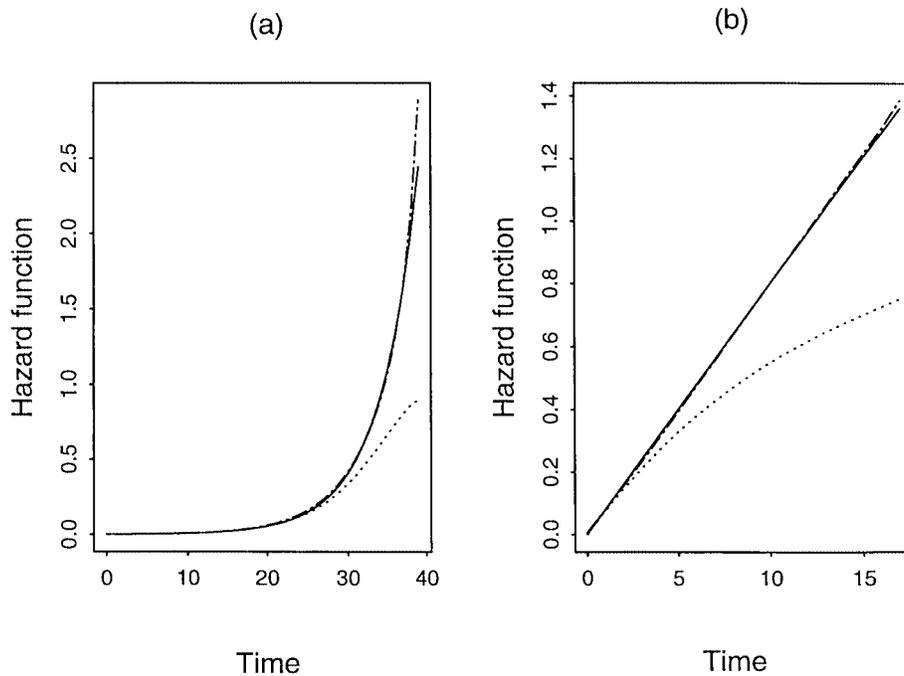


FIG. 9. (a) Mean hazard function estimates based on 500 simulations from the Gompertz distribution with $\lambda(t) = 0.001 \exp(0.2t)$, with sample size $n = 10^6$ for each simulation. Shown are $\lambda(t)$ (solid line), and the averages of estimates $\hat{q}(t)$ (3.4) (dotted line), and $\phi(\hat{q}(t))$ (3.5) (dash-dotted line). Cross-validation bandwidths \hat{b}_q and \hat{b}_ϕ were used for \hat{q} and $\phi(\hat{q})$, respectively. Horizontal axis is time (in days). (b) Same as (a) but for Weibull distribution with $\lambda(t) = 0.08t$. Horizontal axis is time (in days).

shows the means of the hazard function estimates $\hat{q}(t)$ and $\phi(\hat{q}(t))$, along with $\lambda(t)$, for 500 estimated functions, one per Monte Carlo run. The estimates use cross-validation bandwidths. These comparisons are shown for uncensored data, for $n = 10^6$, and for both Gompertz (Figure 9a) Weibull (Figure 9b) cases.

6. Concluding remarks. Although smoothing techniques have been applied extensively for hazard estimation, properties of the smoothed lifetable estimate \hat{q} remained mostly unexplored. In this paper we derive the mean and variance expression for the smoothed lifetable estimate of the hazard function. Such expressions can be utilized for bandwidth choices. They also demonstrate that a transformation of the smoothed lifetable hazard function estimate is necessary in many cases to reduce the bias resulting from the aggregation of the data in the lifetable. Both theoretical findings and simulations illustrate the advantage of such a transformation.

The transformation is particularly relevant for the study of oldest-old mortality in large sample experiments. Our analysis of nematode survival experiments casts doubts on the universality of the Gompertz model, even for populations which are genetically homogeneous. This becomes particularly obvious when we combine the sample of 79 cohorts of genetically homogeneous nematodes to construct a hazard surface. More work will be needed on the important issues of combining information from samples of lifetables and the inclusion of covariates.

APPENDIX

Auxiliary results and proofs. In this section we compile the basic assumptions for the asymptotic results, the proofs and a key auxiliary result (Theorem A.1) which is of interest in its own right. We use the following notation: let T be the random lifetime and C the random censoring time for a subject; T and C are assumed to be independent. Let $X = \min(T, C)$ be the observed time of failure of a subject and $\delta = 1_{\{X=T\}}$ be the corresponding censoring indicator. We define the following functions which are of relevance for lifetable data:

$F_T(t) = P(T \leq t)$, the distribution function of the lifetimes T ,

which is assumed to have a density $f_T(t) = F_T'(t)$;

$\lambda(t) = f_T(t)/(1 - F_T(t))$, the hazard function;

$F_C(t) = P(C \leq t)$, the distribution function of the censoring times C ;

$F(t) = P(X \leq t)$, the distribution function of the observed data;

$\bar{F}(t) = 1 - F(t) = [1 - F_T(t)][1 - F_C(t)]$;

$p(t) = P(X \in (t - \Delta/2, t + \Delta/2], \delta = 1)$;

and

$$q(t) = \Delta^{-1}P(T \in (t - \Delta/2, t + \Delta/2] | T > t - \Delta/2).$$

Let τ be a point such that $F(\tau) < 1$. We make the following assumptions for a fixed point t at which the estimators are to be compared:

$$(A.1) \quad t < \tau, \text{ and hence } \bar{F}(t) \geq \bar{F}(\tau) = \alpha \text{ for some } \alpha > 0;$$

$$(A.2) \quad \lambda \text{ is twice continuously differentiable in a neighborhood of } t.$$

A basic assumption for the asymptotic analysis is that the lengths of the aggregation intervals converge to 0 as $n \rightarrow \infty$;

$$(A.3) \quad \Delta \rightarrow 0 \text{ as } n \rightarrow \infty.$$

In addition, we assume that the kernel function K and the bandwidth b used in the smoothing procedure S_L (3.2) satisfy the conditions

$$K \text{ has compact support on } [-1, 1], K(u) \geq 0, \sup K(u) < \infty,$$

$$(A.4) \quad \int K(u) du = 1, \int uK(u) du = 0, \int u^2K(u) du < \infty,$$

$$\int K^2(u) du < \infty;$$

and

$$(A.5) \quad b \rightarrow 0, \quad nb \rightarrow \infty \quad \text{and} \quad \frac{\Delta}{b}(\log n)^3 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We note that

$$(A.6) \quad p(t_j) = \Delta q(t_j) \bar{F}(\Delta(j-1)).$$

PROOF OF THEOREM 2.1. (i) Using (2.5), (2.6) and a Taylor expansion of λ , we have

$$\begin{aligned} \phi(q(t)) &= \frac{1}{\Delta} \int_{t-\Delta/2}^{t+\Delta/2} \lambda(s) ds \\ &= \frac{1}{\Delta} \int_{t-\Delta/2}^{t+\Delta/2} \left[\lambda(t) + \lambda'(t)(s-t) + \lambda''(t)(s-t)^2/2 + o(s-t)^2 \right] ds \\ &= \lambda(t) + \frac{\Delta^2}{24} \lambda''(t) + o(\Delta^2). \end{aligned}$$

(ii) From (2.6), using (i), one finds

$$\begin{aligned}
q(t) &= \Delta^{-1} [1 - \exp(-\Delta\phi(q(t)))] \\
&= \phi(q(t)) - \frac{\Delta}{2} [\phi(q(t))]^2 + \frac{\Delta^2}{6} [\phi(q(t))]^3 + O(\Delta^3) \\
&= \lambda(t) + \frac{\Delta^2}{24} \lambda''(t) - \frac{\Delta}{2} \left[\lambda^2(t) + \frac{\Delta^2}{12} \lambda(t) \lambda''(t) \right] \\
&\quad + \frac{\Delta^2}{6} \left[\lambda^3(t) + \frac{\Delta^2}{8} \lambda^2(t) \lambda''(t) + o(\Delta^2) \right] + o(\Delta^2),
\end{aligned}$$

when the result follows. \square

Noting that $t_j = \Delta(j-1) + \Delta/2$ and $I_j = [\Delta(j-1), \Delta j)$, we state the following result regarding the moment structure of $\tilde{q}(t_j) = d_j/\Delta n_j$ in (2.1), which provides the key step for the proofs of Theorems 3.1, 3.2. Define, for any β in $[0, 1)$,

$$\begin{aligned}
I_n(\beta) &= \sum_{k=1}^n \frac{1}{k} \binom{n}{k} \beta^{n-k} (1-\beta)^k \\
\text{(A.7)} \quad &= \frac{1}{n} + \frac{\beta}{n-1} + \cdots + \frac{\beta^{n-1}}{1} - \beta^n \sum_{k=1}^n \frac{1}{k} \\
&= \int_0^{1-\beta} \frac{(\beta+x)^n - \beta^n}{x} dx.
\end{aligned}$$

THEOREM A.1. *Under assumptions (A.1)–(A.5), it holds for $\tilde{q}(t_i)$ and $\tilde{q}(t_j)$ in (2.1), $1 \leq i < j \leq p$, that:*

- (i) $E[\tilde{q}(t_j)] = [1 - F^n(\Delta(j-1))]q(t_j)$
 $= [1 - F^n(\Delta(j-1))]$
 $\times \left[\lambda(t_j) - \frac{\Delta}{2} \lambda^2(t_j) + \frac{\Delta^2}{24} (\lambda''(t_j) + 4\lambda^3(t_j)) + o(\Delta^2) \right];$
- (ii) $\text{Var}(\tilde{q}(t_j)) = \Delta^{-1} q(t_j) (1 - \Delta q(t_j)) I_n(F(\Delta(j-1)))$
 $+ q(t_j) (1 - q(t_j)) F^n(\Delta(j-1))$
 $+ q(t_j) F^n(\Delta(j-1)) [1 - q(t_j) F^n(\Delta(j-1))]$
 $= \frac{\lambda(t_j)}{n\Delta [1 - F(\Delta(j-1))]} \left[1 - \frac{3\Delta}{2} \lambda(t_j) + O(\Delta^2) \right];$
- (iii) $\text{Cov}(\tilde{q}(t_i), \tilde{q}(t_j)) = O((1-\alpha)^n)$ for $i \neq j$.

PROOF. Note that uniformly in $0 \leq \beta < 1$,

$$(A.8) \quad nI_n(\beta) \rightarrow \frac{1}{1-\beta}.$$

Note further that (A.1) and (A.8) imply

$$(A.9) \quad I_n(F(t_{j-1})) = (n\bar{F}(t_{j-1}))^{-1}(1 + o(1)).$$

Let $B(n, p)$ denote a random variable with a binomial distribution with parameters n and p . We note the following three facts:

$$(A.10) \quad \begin{aligned} E\left[\frac{1}{1+B(n-1, \bar{F}(t))}\right] &= \sum_{k=1}^n \frac{1}{k} \binom{n-1}{k-1} \bar{F}^{k-1}(t) F^{n-k}(t) \\ &= \frac{1}{n\bar{F}(t)} \sum_{k=1}^n \binom{n}{k} \bar{F}^k(t) F^{n-k}(t) \\ &= \frac{1}{n\bar{F}(t)} [1 - F^n(t)]; \end{aligned}$$

$$(A.11) \quad \begin{aligned} E\left([1+B(n-1, \bar{F}(t))]^{-2}\right) &= \sum_{k=1}^n \frac{1}{k^2} \binom{n-1}{k-1} \bar{F}^{k-1}(t) F^{n-k}(t) \\ &= (n\bar{F}(t))^{-1} \sum_{k=1}^n \frac{1}{k} \binom{n}{k} \bar{F}^k(t) F^{n-k}(t) \\ &= (n\bar{F}(t))^{-1} I_n(F(t)); \end{aligned}$$

and

$$(A.12) \quad \begin{aligned} E\left([2+B(n-2, \bar{F}(t))]^{-2}\right) &= \sum_{k=2}^n \frac{1}{k^2} \binom{n-2}{k-2} \bar{F}^{k-2}(t) F^{n-k}(t) \\ &= \frac{1}{n(n-1)\bar{F}^2(t)} \sum_{k=2}^n \frac{k-1}{k} \binom{n}{k} \bar{F}^k(t) F^{n-k}(t) \\ &= \frac{1}{n(n-1)\bar{F}^2(t)} [1 - F^n(t) - I_n(F(t))]. \end{aligned}$$

Defining

$$m(y) = E[\delta = 1 | X = y],$$

we find

$$\begin{aligned}
E[\hat{q}(t_j)] &= \frac{n}{\Delta} E\left[\frac{\mathbf{1}_{(X_1 \in I_j, \delta_1=1)}}{n_j}\right] \\
&= \frac{n}{\Delta} E\left[E(\mathbf{1}_{(X_1 \in I_j, \delta_1=1)} | X_1)\right. \\
&\quad \left. \times E\left(\frac{1}{\mathbf{1}_{(X_1 > \Delta(j-1))} + \sum_{k=2}^n \mathbf{1}_{(X_k > \Delta(j-1))}} | X_1\right)\right] \\
&= \frac{n}{\Delta} \int_{\Delta(j-1)}^{\Delta j} m(y) E\left\{\left[1 + B(n-1, \bar{F}(\Delta(j-1)))\right]^{-1}\right\} dF(y) \\
\text{(A.13)} \quad &= \frac{n}{\Delta} \int_{\Delta(j-1)}^{\Delta j} m(y) [1 - F^n(\Delta(j-1))] [n\bar{F}(\Delta(j-1))]^{-1} dF(y) \\
&= [1 - F^n(\Delta(j-1))] [\Delta\bar{F}(\Delta(j-1))]^{-1} \int_{\Delta(j-1)}^{\Delta j} m(y) dF(y) \\
&= [1 - F^n(\Delta(j-1))] [\Delta\bar{F}(\Delta(j-1))]^{-1} p(t_j) \\
&= [1 - F^n(\Delta(j-1))] q(t_j) \\
&= [1 - F^n(\Delta(j-1))] \\
&\quad \times \left[\lambda(t_j) - \frac{\Delta}{2} \lambda^2(t_j) + \frac{\Delta^2}{24} (\lambda''(t_j) + 4\lambda^3(t_j)) + o(\Delta^2)\right],
\end{aligned}$$

using Theorem 2.1(ii). This implies (i).

Further, for the proof of (ii), we find

$$\begin{aligned}
E[\tilde{q}^2(t_j)] &= \frac{n}{\Delta^2} E\left[\mathbf{1}_{(X_1 \in I_j, \delta_1=1)} / n_i^2\right] \\
\text{(A.14)} \quad &+ \frac{n(n-1)}{\Delta^2} E\left[\mathbf{1}_{(\Delta(j-1) < X_1, X_2 \leq \Delta j, \delta_1=\delta_2=1)} / n_i^2\right].
\end{aligned}$$

By (A.6) and (A.9),

$$\begin{aligned}
&\frac{n}{\Delta^2} E\left[\mathbf{1}_{(X_1 \in I_j, \delta_1=1)} / n_j^2\right] \\
&= \frac{n}{\Delta^2} \int_{\Delta(j-1)}^{\Delta j} m(y) \\
\text{(A.15)} \quad &\quad \times E\left(\left[1 + B(n-1, \bar{F}(\Delta(j-1)))\right]^{-2}\right) dF(y) \\
&= I_n(F(\Delta(j-1))) [\Delta^2 \bar{F}(\Delta(j-1))]^{-1} p(t_j) \\
&= I_n(F(\Delta(j-1))) q(t_j) / \Delta.
\end{aligned}$$

Furthermore, using (A.6) and (A.12),

$$\begin{aligned}
& E\left[\mathbf{1}_{(\Delta(j-1) < X_1, X_2 \leq \Delta j, \delta_1 = \delta_2 = 1)}/n_j^2\right] \\
&= E\left[\mathbf{E}\left(\mathbf{1}_{(\Delta(j-1) < X_1, X_2 \leq \Delta j, \delta_1 = \delta_2 = 1)} \mid X_1, X_2\right)\right. \\
&\quad \times E\left(\left[\mathbf{1}_{(X_1 > \Delta(j-1))} + \mathbf{1}_{(X_2 > \Delta(j-1))}\right.\right. \\
&\quad \left.\left. + \sum_{k=3}^n \mathbf{1}_{(X_k > \Delta(j-1))}\right]^{-2} \mid X_1, X_2\right)\left. \right] \\
&= \int_{\Delta(j-1)}^{\Delta j} \int_{\Delta(j-1)}^{\Delta j} m(y)m(z) \\
&\quad \times E\left(\left[2 + B(n-2, \bar{F}(\Delta(j-1)))\right]^{-2}\right) dF(y) dF(z) \\
&= \frac{1 - F^n(\Delta(j-1)) - I_n(F(\Delta(j-1)))}{n(n-1)\bar{F}^2(\Delta(j-1))} p^2(t_j) \\
&= \frac{1 - F^n(\Delta(j-1)) - I_n(F(\Delta(j-1)))}{n(n-1)} \Delta^2 q^2(t_j).
\end{aligned}
\tag{A.16}$$

Thus, (A.14)–(A.16) imply

$$\begin{aligned}
E[\tilde{q}^2(t_j)] &= I_n(F(\Delta(j-1)))q(t_j)/\Delta + q^2(t_j) \\
&\quad - q^2(t_j)F^n(\Delta(j-1)) - q^2(t_j)I_n(F(\Delta(j-1))).
\end{aligned}$$

From (A.1), (A.8) and Theorem 2.1(ii), we obtain

$$\begin{aligned}
\text{Var}(\tilde{q}(t_j)) &= E[\tilde{q}^2(t_j)] - [E(\tilde{q}(t_j))]^2 \\
&= \Delta^{-1}q(t_j)(1 - \Delta q(t_j))I_n(F(\Delta(j-1))) \\
&\quad + q(t_j)(1 - q(t_j))F^n(\Delta(j-1)) \\
&\quad + q(t_j)F^n(\Delta(j-1))[1 - q(t_j)F^n(\Delta(j-1))] \\
&= \frac{\Delta q(t_j)(1 - \Delta q(t_j))}{n \Delta^2 \bar{F}(\Delta(j-1))} (1 + o(1)) \\
&= \frac{[1 - \Delta \lambda(t_j) + (\Delta^2/2)\lambda^2(t_j) + O(\Delta^3)]}{n \Delta^2 \bar{F}(\Delta(j-1))} \\
&\quad \times \left[\Delta \lambda(t_j) - \frac{\Delta^2}{2}\lambda^2(t_j) + O(\Delta^3)\right] (1 + o(1)) \\
&= \frac{\lambda(t_j)}{n \Delta \bar{F}(\Delta(j-1))} \left(1 - \frac{3\Delta}{2}\lambda(t_j) + O(\Delta^2)\right) (1 + o(1)),
\end{aligned}$$

which implies (ii). For the proof of (iii), we observe that for $i \neq j$,

$$\begin{aligned}
& E[\tilde{q}(t_i)\tilde{q}(t_j)] \\
&= E\left[\frac{\sum_{h=1}^n \mathbf{1}_{(X_h \in I_i, \delta_h=1)}}{\Delta n_i} \frac{\sum_{k=1}^n \mathbf{1}_{(X_k \in I_j, \delta_k=1)}}{\Delta n_j}\right] \\
&= \frac{n(n-1)}{\Delta^2} E\left[\mathbf{1}_{(X_1 \in I_i, X_2 \in I_j, \delta_1=\delta_2=1)} / (n_i n_j)\right] \\
&= \frac{n(n-1)}{\Delta^2} E\left\{E\left(\mathbf{1}_{(X_1 \in I_i, X_2 \in I_j, \delta_1=\delta_2=1)} \mid X_1, X_2\right)\right. \\
&\quad \times E\left(\left[\mathbf{1}_{(X_1 > \Delta(i-1))} + \mathbf{1}_{(X_2 > \Delta(i-1))} + \sum_{m=3}^n \mathbf{1}_{(X_m > \Delta(i-1))}\right]^{-1}\right. \\
&\quad \times \left[\mathbf{1}_{(X_1 > \Delta(j-1))} + \mathbf{1}_{(X_2 > \Delta(j-1))}\right. \\
&\quad \left. \left. \left. + \sum_{m=3}^n \mathbf{1}_{(X_m > \Delta(j-1))}\right]^{-1} \mid X_1, X_2\right)\right\} \\
&= \frac{n(n-1)}{\Delta^2} \int_{\Delta(i-1)}^{\Delta i} \int_{\Delta(j-1)}^{\Delta j} m(y)m(z) \\
&\quad \times E\left[(2 + M_1 + M_2)^{-1}(1 + M_2)^{-1}\right] dF(y) dF(z) \\
&= \frac{n(n-1)}{\Delta^2} p(t_i)p(t_j) E\left[(2 + M_1 + M_2)^{-1}(1 + M_2)^{-1}\right].
\end{aligned}$$

Here, (M_0, M_1, M_2) has a multinomial distribution with parameters $n - 2, \pi_0, \pi_1$ and π_2 , where

$$\begin{aligned}
M_0 &= \sum_{m=3}^n \mathbf{1}_{(X_m \leq \Delta(i-1))}, \\
M_1 &= \sum_{m=3}^n \mathbf{1}_{(X_m \in (\Delta(i-1), \Delta(j-1)])}, \\
M_2 &= \sum_{m=3}^n \mathbf{1}_{(X_m > \Delta(j-1))}
\end{aligned}$$

and $\pi_0 = F(\Delta(i-1))$, $\pi_1 = F(\Delta(j-1)) - F(\Delta(i-1))$, $\pi_2 = \bar{F}(\Delta(j-1))$. Using $\int_0^1 x^a dx = (1+a)^{-1}$, we find

$$\begin{aligned}
& E\left[(2 + M_1 + M_2)^{-1}(1 + M_2)^{-1}\right] \\
&= \int_0^1 \int_0^1 E(x^{1+M_1+M_2} y^{M_2} dx dy) \\
&= \int_0^1 \int_0^1 \sum_{m_0, m_1, m_2} x^{1+m_1+m_2} y^{m_2} \binom{n-2}{m_0, m_1, m_2} \pi_0^{m_0} \pi_1^{m_1} \pi_2^{m_2} dx dy
\end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \int_0^1 x(\pi_0 + \pi_1 x + \pi_2 xy)^{n-2} dy dx \\
&= \int_0^1 \frac{(\pi_0 + \pi_1 x + \pi_2 x)^{n-1} - (\pi_0 + \pi_1 x)^{n-1}}{(n-1)\pi_2} dx \\
&= \frac{1 - F^n(\Delta(i-1))}{n(n-1)} [\bar{F}(\Delta(i-1))\bar{F}(\Delta(j-1))]^{-1} \\
&\quad - \frac{(\pi_0 + \pi_1)^n - \pi_0^n}{n(n-1)\pi_1\pi_2}.
\end{aligned}$$

Now use (A.1) and (A.6) to obtain

$$E[\tilde{q}(t_i)\tilde{q}(t_j)] = q(t_i)q(t_j)[1 - F^n(\Delta(i-1))] + O(F^n(\Delta(j-1))),$$

which implies

$$\begin{aligned}
\text{Cov}(\tilde{q}(t_i), \tilde{q}(t_j)) &= q(t_i)q(t_j)[1 - F^n(\Delta(i-1))] \\
&\quad - q(t_i)q(t_j)[1 - F^n(\Delta(i-1))][1 - F^n(\Delta(j-1))] \\
&\quad + O(F^n(\Delta(j-1))) \\
&= q(t_i)q(t_j)[1 - F^n(\Delta(i-1))]F^n(\Delta(j-1)) \\
&\quad + O(F^n(\Delta(j-1))) \\
&= O(F^n(\Delta(j-1))) = O((1-\alpha)^n),
\end{aligned}$$

from which (iii) follows. \square

This result is now applied to the proofs of Theorems 3.1 and 3.2.

PROOF OF THEOREM 3.1. (i) By (3.1), the locally weighted least squares estimate $\hat{q}(t)$ (3.2) is defined as

$$\hat{q}(t) = S_L[t, b, (t_j, \tilde{q}(t_j), w_j), j = 1, \dots, p] = \sum_{j=1}^p W_j(t, b)\tilde{q}(t_j),$$

where the $W_j(t, b)$ are weight functions. According to Müller (1987), there exists a kernel function K satisfying (A.4) and $\lim_{p \rightarrow \infty} \sup_{1 \leq j \leq p} |W_j(t, b)/W_{K,j}(t, b) - 1| = 0$ (defining $0/0 = 1$), where the weight functions for the kernel estimate are given by

$$W_{K,j}(t, b) = \frac{1}{b} \int_{\Delta(j-1)}^{\Delta j} K\left(\frac{t-u}{b}\right) du.$$

The locally weighted least squares estimate therefore can be written as

$$\begin{aligned}
 \hat{q}(t) &= \sum_{j=1}^p W_{K,j}(t, b) \tilde{q}(t_j) \\
 &+ \sum_{j=1}^p W_{K,j}(t, b) \left(\frac{W_j(t, b)}{W_{K,j}(t, b)} - 1 \right) \tilde{q}(t_j) \\
 &= \sum_{j=1}^p W_{K,j}(t, b) \tilde{q}(t_j) (1 + o(1)),
 \end{aligned}
 \tag{A.17}$$

noting that $K(u) \geq 0$. Applying Theorem A.1(i) and (A.17),

$$\begin{aligned}
 E[\hat{q}(t)] &= \sum_{j=1}^p W_{K,j}(t, b) E[\tilde{q}(t_j)] (1 + o(1)) \\
 &= \sum_{j=1}^p W_{K,j}(t, b) \\
 &\times \left[(1 - F^n(\Delta(j-1))) \right. \\
 &\quad \left. \times \left(\lambda(t_j) - \frac{\Delta}{2} \lambda^2(t_j) + \frac{\Delta^2}{24} (\lambda''(t_j) + 4\lambda^3(t_j)) + o(\Delta^2) \right) \right] \\
 &\times (1 + o(1)).
 \end{aligned}
 \tag{A.18}$$

By the mean value theorem for integrals,

$$\begin{aligned}
 \sum_{j=1}^p W_{K,j}^2(t, b) &= \frac{\Delta}{b^2} \sum_{j=1}^p \mathbf{1}_{(|t-\Delta j| \leq 2b)} \\
 &\times \int_{\Delta(j-1)}^{\Delta j} K^2\left(\frac{t-u}{b}\right) du \left(1 + O\left(\frac{\Delta}{b}\right)\right) \\
 &= \frac{\Delta}{b} \left[\int K^2(u) du + O(\Delta) \right].
 \end{aligned}
 \tag{A.19}$$

From (A.19) and an application of the Cauchy–Schwarz inequality,

$$\begin{aligned}
 &\sum_{j=1}^p W_{K,j}(t, b) F^n(\Delta(j-1)) \lambda(t_j) \\
 &\leq \left(\sum_{j=1}^p W_{K,j}^2(t, b) \right)^{1/2} \left(\sum_{j=1}^p \mathbf{1}_{(W_{K,j}(t, b) \neq 0)} \lambda^2(t_j) F^{2n}(\Delta(j-1)) \right)^{1/2} \\
 &= O((1 - \alpha)^n).
 \end{aligned}$$

Next, using the moment conditions on the kernel function, and a Taylor expansion, it follows by a well-known argument that

$$(A.20) \quad \sum_{j=1}^p W_{K,j}(t, b) \lambda(t_j) = \lambda(t) + \frac{b^2}{2} \lambda''(t) \int u^2 K(u) du + o(b^2),$$

and similarly,

$$(A.21) \quad \sum_{j=1}^p W_{K,j}(t, b) \lambda^2(t_j) = \lambda^2(t) + O(b^2),$$

$$(A.22) \quad \sum_{j=1}^p W_{K,j}(t, b) \lambda''(t_j) = \lambda''(t) + o(1),$$

$$(A.23) \quad \sum_{j=1}^p W_{K,j}(t, b) \lambda^3(t_j) = \lambda^3(t) + O(b^2).$$

We note that $nb \rightarrow \infty$ according to (A.5). Therefore, $(1 - \alpha)^n / b^2 \rightarrow 0$, so that $(1 - \alpha)^n = o(b^2)$. The result follows from (A.18)–(A.23).

(ii) We find from Theorem A.1(ii), (iii) that

$$\begin{aligned} \text{Var}(\hat{q}(t)) &= \left[\sum_{j=1}^p W_{K,j}^2(t, b) \text{Var}(\tilde{q}(t_j)) \right. \\ &\quad \left. + \sum_{\substack{i=1 \\ i \neq j}}^p \sum_{j=1}^p W_{K,i}(t, b) W_{K,j}(t, b) \text{Cov}(\tilde{q}(t_i), \tilde{q}(t_j)) \right] (1 + o(1)) \\ &= \left[\sum_{j=1}^p W_{K,j}^2(t, b) \left[\lambda(t_j) [n \Delta \bar{F}(\Delta(j-1))]^{-1} (1 + o(1)) \right]^2 \right. \\ &\quad \left. - \sum_{\substack{i=1 \\ i \neq j}}^p \sum_{j=1}^p W_{K,i}(t, b) W_{K,j}(t, b) O((1 - \alpha)^n) \right] [1 + o(1)]. \end{aligned}$$

Now (A.19) and $K(u) \geq 0$ imply

$$\begin{aligned} &\sum_{\substack{i=1 \\ i \neq j}}^p \sum_{j=1}^p W_{K,i}(t, b) W_{K,j}(t, b) O((1 - \alpha)^n) \\ &= \left[\left(\sum_{j=1}^p W_{K,j}(t, b) \right)^2 - \sum_{j=1}^p W_{K,j}^2(t, b) \right] O((1 - \alpha)^n) \\ &= \left(1 - O\left(\frac{\Delta}{b}\right) \right) O((1 - \alpha)^n) = O((1 - \alpha)^n) + O\left(\frac{\Delta}{b} (1 - \alpha)^n\right), \end{aligned}$$

and

$$\sum_{j=1}^p W_{K,j}^2(t, b) o\left(\frac{1}{n\Delta}\right) = o\left(\frac{1}{nb}\right).$$

Here and in other places, we use the fact that the $O(\cdot)$ terms are uniform in $j = 1, \dots, p$. Furthermore, the smoothness assumptions on λ and F imply by an argument analogous to Gasser and Müller (1979), using (A.19),

$$\sum_{j=1}^p W_{K,j}^2(t, b) \frac{\lambda(t_j)}{n \Delta \bar{F}(\Delta(j-1))} = \frac{1}{nb} \frac{\lambda(t)}{\bar{F}(t)} \int K^2(u) du (1 + o(1)).$$

The result follows. \square

PROOF OF THEOREM 3.2. Applying a Taylor expansion of $\phi(x) = -\log(1 - \Delta x)/\Delta$ (2.5) at the point $x = q(t)$, we obtain for a mean value ξ between $q(t)$ and $\hat{q}(t)$,

$$(A.24) \quad \phi(\hat{q}(t)) = \phi(q(t)) + \frac{1}{1 - \Delta\xi} (\hat{q}(t) - q(t)).$$

We see from the proof of Theorem 3.1(i) and (A.5) that the weights $W_j(t, b)$ employed in the smoothing procedure S_L satisfy

$$(A.25) \quad \max_j |W_j(t, b)| \leq \sup |K(u)| \frac{\Delta}{b} (1 + \gamma)$$

for a constant $\gamma > 0$. We also note that (A.4) and the design assumptions imply that

$$(A.26) \quad \sum_{i=1}^p \mathbf{1}_{\{W_i(t, b) > 0\}} \leq \frac{2b}{\Delta}.$$

Letting $\Gamma = (1 + \gamma) \sup K(u)$, where γ is as in (A.25), we note that

$$(A.27) \quad \hat{q}(t) = \frac{1}{\Delta} \sum_{i=1}^p W_i(t, b) \frac{d_i}{n_i} \leq \frac{\Gamma}{b} \sum_{i=1}^p \frac{d_i}{n_i} \mathbf{1}_{\{W_i(t, b) > 0\}}.$$

We next observe that the number of summands on the r.h.s. of (A.27) for which one has

$$(A.28) \quad \frac{d_i}{n_i} \geq \frac{1}{\Gamma \log n},$$

among the indices $i = 1, \dots, p$, is of the order $O([\log n]^2)$. To see this, note that $d_i/n_i \geq 1/\Gamma \log n$ requires $n_{i+1} \leq n_i(1 - (\Gamma \log n)^{-1})$. Denote the count of indices i where (A.28) happens by ν . A conservative upper bound for ν is obtained by assuming that (A.28) happens for all indices $1, \dots, \nu$, that is, a subset of the index set $\{1, 2, 3, \dots, p\}$ composed of a string of initial indices. For this situation, an upper bound for ν is obtained from the requirement

$n(1 - (\Gamma \log n)^{-1})^\nu \leq 1$. This leads to the bound

$$(A.29) \quad \nu \leq -\frac{\log n}{\log(1 - (\Gamma \log n)^{-1})} = O([\log n]^2).$$

We infer from (A.26)–(A.29) that

$$\hat{q}(t) \leq \frac{\Gamma}{b} \left\{ O([\log n]^2 + \frac{b}{\Delta} \frac{1}{\Gamma \log n}) \right\} = O\left(\frac{1}{\Delta \log n}\right)$$

for n large enough and fixed t , observing (A.5). Since according to Theorem 2.1, $q(t) = \lambda(t) + O(\Delta)$, we conclude that for the mean value $\xi = \xi_n$, $\Delta \xi \rightarrow 0$, so that on the r.h.s. of (A.24),

$$(1 - \Delta \xi)^{-1} \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Therefore,

$$\phi(\hat{q}(t)) = \phi(q(t)) + (\hat{q}(t) - q(t))(1 + o(1)).$$

Now (i) follows immediately from Theorem 2.1(i) and Theorem 3.1(i), while (ii) follows from Theorem 3.1(ii). \square

Acknowledgments. We thank James Carey for directing our attention to the problem of oldest-old mortality, Jeng-Min Chiou for help with some of the graphs and T. Johnson for providing us with the original nematode data. Comments of a referee which led to improvements in the paper are also appreciated.

REFERENCES

- ANDERSON, J. and SETHILSELVAN, A. (1980). Smooth estimates for the hazard function. *J. Roy. Statist. Soc. Ser. B* **42** 322–327.
- BLOOMFIELD, D. and HABERMAN, S. (1987). Graduation: some experiments with kernel methods. *J. Inst. Actuaries* **114** 339–369.
- BORGAN, O. (1979). On the theory of moving average graduation. *Scand. Actuar. J.* **3** 83–105.
- BROOKS, A., LITHGOW, G. and JOHNSON, T. (1994). Rates of mortality in populations of *Caenorhabditis Elegans*. *Science* **263** 668–670.
- CAREY, J. R., LIEDO, P., OROZCO, D. and VAUPEL, J. W. (1992). Slowing of mortality rates at older ages in large medfly cohorts. *Science* **258** 457–461.
- CLEVELAND, W. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.
- CONGDON, P. (1993). Statistical graduation in local demographic analysis and projection. *J. Roy. Statist. Soc. Ser. A* **156** 237–270.
- COPAS, J. and HABERMAN, S. (1983). Nonparametric graduation using kernel methods. *J. Inst. Actuaries* **110** 135–156.
- CURTSINGER, J. W., FUKUI, H. H., TOWNSEND, D. R. and VAUPEL, J. W. (1992). Demography of genotypes—failure of the limited life-span paradigm in *Drosophila Melanogaster*. *Science* **258** 461–463.
- EUBANK, R. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York.
- FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.
- FAN, J. and GJIBELS, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London.

- GASSER, T. and MÜLLER, H. G. (1979). Kernel estimation of regression functions. *Smoothing Techniques in Curve Estimation. Lecture Notes in Math.* **757** 23–68. Springer, New York.
- GOMPERTZ, B. (1825). On the nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society* **2**.
- GRAM, J. P. (1879). *Om Raekkeviklinger, bestemte ved Hjaelp af de mindste Kradraters Methode*. A. F. Høst & Son, Copenhagen.
- GRAM, J. P. (1883). Ueber Entwicklung reeller Functionen in Reihen mittelst der Methode der Kleinsten Quadrate. *J. Math.* **94** 41–73.
- GRAY, R. (1990). Some diagnostic methods for Cox regression models through hazard smoothing. *Biometrics* **46** 93–102.
- GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London.
- HALL, P., KAY, J. and TITTERINGTON, D. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528.
- HOEM, J. (1972). On the statistical theory of analytic graduation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 569–600. Univ. California Press, Berkeley.
- HOEM, J. (1976a). The statistical theory of demographic rates. *Scand. J. Statist.* **3** 169–185.
- HOEM, J. (1976b). On the optimality of modified minimum chi-square analytic graduation. *Scand. J. Statist.* **3** 89–92.
- HOEM, J. (1983). The reticent trio: some little-known discoveries in life insurance mathematics by L. H. F. Opperman, T. N. Thiele, and J. P. Gram. *Internat. Statist. Rev.* **51** 213–221.
- HOEM, J. (1984). A contribution to the statistical theory of linear graduation. *Insurance Math. and Econ.* **3** 1–17.
- HOEM, J. and LENNEMANN, P. (1988). The tails in moving average graduation. *Scand. Actuarial J.* **11** 193–229.
- JAZWINSKI, S. M. (1996). Longevity, genes, and aging. *Science* **273** 54–59.
- JUCKETT, D. and ROSENBERG, B. (1993). Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. *Mechanisms of Ageing and Development* **69** 1–31.
- KELNER, K. L. and MARX, J. (1996). Patterns of aging. *Science* **273** 41–41.
- LEJEUNE, M. (1985). Estimation non-paramétrique par noyaux: régression polynomiale mobile. *Rev. Statist. Appl.* **33** 43–67.
- MANTON, K. G., STALLARD, E. and VAUPEL, J. W. (1986). Alternative models for the heterogeneity of mortality risks among the ages. *J. Amer. Statist. Assoc.* **81** 635–644.
- MÜLLER, H. G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.* **82** 231–238.
- MÜLLER, H. G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Springer, New York.
- MÜLLER, H. G. and STADTMÜLLER, U. (1993). On variance function estimation with quadratic forms. *J. Statist. Plann. Inference* **35** 213–231.
- MÜLLER, H. G. and WANG, J. L. (1990). Locally adaptive hazard smoothing. *Probab. Theory Related Fields* **85** 523–538.
- MÜLLER, H. G. and WANG, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50** 61–76.
- MÜLLER, H. G., WANG, J. L. and CAPRA, W. B. (1997). Estimating hazard functions from lifetables: the transformation approach. *Biometrika*. To appear.
- PERLS, T. T. (1995). The Oldest-Old. *Scientific American* **January** 70–75.
- RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.
- RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.
- SACKS, J. and YLVIKAKER, D. (1970). Designs for regression problems with correlated errors III. *Ann. Math. Statist.* **41** 2057–2074.

- SEIFERT, B., GASSER, T. and WOLF, A. (1993). Nonparametric estimation of residual variance revisited. *Biometrika* **80** 373–383.
- SUZMAN, R. M., WILLIS, D. and MANTON, K. (1992). *The Oldest Old*. Oxford Univ. Press.
- TANNER, M. and WONG, W. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* **11** 989–993.
- TANNER, M. and WONG, W. (1984). Data-based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis. *J. Amer. Statist. Assoc.* **79** 174–182.
- VAUPEL, J. W., JOHNSON, T. E. and LITHGOW, G. J. (1994). Rates of mortality in populations of *Caenorhabditis Elegans*. *Science* **266** 826.
- VAUPEL, J. W. and CAREY, J. R. (1993). Compositional interpretations of medfly mortality. *Science* **260** 1666–1667.
- VAUPEL, J. W., MANTON, K. G. and STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16** 439–454.
- VAUPEL, J. W. and YASHIN, A. I. (1985). Heterogeneity's ruses: some surprising effects of selection on population dynamics. *Amer. Statist.* **39** 176–185.
- WANG, J. L., MÜLLER, H. G., CAPRA, W. B. and CAREY, J. R. (1994). Rates of mortality in populations of *Caenorhabditis Elegans*. *Science* **266** 827–828.
- WILSON, D. (1994). The analysis of survival (mortality) data: fitting Gompertz, Weibull, and logistic functions. *Mechanisms of Ageing and Development* **74** 15–33.
- YANDELL, B. S. (1983). Nonparametric inference for rates with censored survival data. *Ann. Statist.* **11** 989–993.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616