Reappraising Medfly Longevity: A Quantile Regression Survival Analysis

Roger KOENKER and Olga GELING

Chaudhuri, Doksum and Samarov (1997) have recently stressed the usefulness of the quantile regression formulation for survival analysis and for transformation models, more generally. In this paper, we explore the use of quantile regression in survival analysis by reanalysing a large experimental study by Carey, Liedo, Orozco, and Vaupel (1992), that monitored age-specific mortality in a population of roughly 1.2 million Mediterranean fruit flies (*Ceratitis Capitata*). The quantile regression approach appears useful in refining several of the conclusions drawn from this study including the apparent decline in mortality rates at advanced ages, and the gender cross-over effect in survival functions for medflies.

1. INTRODUCTION

The biology of aging has attained a robust adolescent stage as a scientific discipline and seems destined for a prolonged maturity. The enduring human fascination with "intimations of immortality," nurtured by modern developments in cell biology, provides a powerful impetus for the field. From a statistical vantage point, one of the most exciting recent developments in this emerging field involves large scale demographic experiments on lower animals designed to explore features of the survival distribution and determinants of longevity. The largest, and most influential of these is the work of Carey, Liedo, Orozco, and Vaupel (1992). The primary experiment described there consisted of monitoring age-specific mortality in a population of roughly 1.2 million Mediterranean fruit flies (*Ceratitis Capitata*). Several findings from this experiment challenged notions that might be regarded as central to the conventional wisdom of population biology, and demography more generally:

- Mortality rates actually *declined* at the oldest observed ages contradicting the view that aging is an ineluctable, monotone process of senescence. In the most extreme form of the traditional view the survival distribution is assumed to take the Gompertz form so the mortality rate (hazard) is log linear. This view is clearly contradicted by the medfly experiments.
- The right tail of the survival distribution was, at least by human standards, remarkably long. By 33 days 90% of the flies had died, by 50 days 99% were dead, but 120 (0.01%) lived to 86 days and 12 (0.001%) lived to 146 days. This finding casts doubt on the common practice of characterizing species-specific maximum life-spans, shifting the focus instead to the analysis of tail behavior of the survival distribution. See Thatcher (1999) for an extended discussion of related evidence for human populations.
- The experiment provided, really for the first time, strong evidence for a crossover in gender specific mortality rates for a non-human population. In medflies, Carey *et al* (1995) report that females have higher mortality rates than males up to roughly 20 days, while from 20 to 60 days males have higher rates than females, and after 60 days the rates are essentially indistinguishable. These results suggest a considerably more complicated view of adaptability of the sexes for survival at various stages of the life cycle than is provided by prior literature.

The statistical analyses employed in prior work on the medfly data are based largely on standard life table methods, as described, for example, in Carey (1993). Life table methods are well adapted for the study of the effects of gender and other discrete covariates on survival and mortality, however, they are less well suited to investigating the effect of continuous covariates like population density, a variable which has played a significant role in subsequent debate over the interpretation of the experimental results.

In contrast, parametric and semiparametric survival models while accommodating a broader class of covariates typically impose stringent conditions on how the covariates are permitted to influence survival prospects. For example, in the accelerated failure time model covariates are assumed to exert a pure location shift effect on log survival times, while in the Cox proportional hazard model the covariates exert a pure location shift on a transformation of the baseline survival probability evaluated at the random survival time. In both cases, the underlying assumption that the covariates only affect the location, not the shape of the distribution of transformed survival times, strongly restricts the nature of their impact. This is particularly unfortunate in the analysis of the medfly data because interest is focused so strongly on the upper tail of the survival distribution. It is highly undesirable that inferences about the effect of covariates on upper tail

Version: November 15, 2000. Roger Koenker is Professor, Departments of Economics and Statistics, University of Illinois, Champaign, IL 61820 (Email: rkoenker uiuc.edu). Olga Geling is graduate student, Department of Statistics, University of Illinois, Champaign, IL 61820 (Email: o-geling uiuc.edu). The authors would like to thank Professor James Carey for providing access to the medfly data. The authors wish to express their appreciation to participants of the Summer Institute on Social and Biological Determinants of Longevity at the Max-Planck Institute for Demographic Research, Rostock, and the research seminar of the Weierstrass Institute, Berlin for comments on earlier versions of this research. The research of Roger Koenker was partially supported by NSF grant SBR-9617206.

Journal of the American Statistical Association
???

behavior should be unduly influenced by global features of the model specification.

In this paper we describe a new approach to the analysis of survival data of this type using the quantile regression methods introduced in Koenker and Bassett (1978). Quantile regression offers a flexible semiparametric approach to survival analysis. Rather than making global assumptions about how covariates influence transformed survival times, quantile regression enables us to focus on the estimation of particular local features of the conditional survival distribution. Thus, we may explore the effect of covariates on just the upper tail of the survivals without risk that we may be unduly prejudicing the results by modeling assumptions about the rest of the conditional distribution.

In the next Section we provide a brief introduction to quantile regression methods. Section 3 describes the link between the quantile regression model and some common survival analysis models. Section 4 introduces the data to be analyzed. Section 5 presents the analysis. And some concluding remarks are collected in Section 6.

2. INTRODUCTION TO QUANTILE REGRESSION

Quantile regression is gradually evolving into a comprehensive approach to the statistical analysis of linear and nonlinear response models for conditional quantile functions. Just as classical linear regression methods based on minimizing sums of squared residuals enable one to estimate a general class of models for conditional mean functions, quantile regression methods offer a mechanism for estimating models for the conditional median function, and the full range of other conditional quantile functions. By supplementing least squares estimation of conditional mean functions with techniques for estimating an entire family of conditional quantile functions, quantile regression is capable of providing a more complete statistical analysis of the stochastic relationships among random variables.

2.1 Quantiles via Optimization

One exceedingly simple idea underlies the extension of the sample quantiles to the more general contexts of linear, and nonlinear, regression. This idea is just the observation that we can replace the notions of sorting, ordering, and ranking that appear to be inherent in the definition of the ordinary sample quantiles by an elementary optimization problem. Suppose that we have a single sample $\mathcal{S}_n = \{y_1, ..., y_n\}$ and we minimize,

$$R(\xi) = \sum_{i=1}^{n} \rho_\tau(y_i - \xi)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the piecewise linear "check function" of Koenker and Bassett (1978), then the solution $\hat{\xi}(\tau)$ is a τth sample quantile of $\mathcal{S}_n$. The median case where $\rho_{1/2}(u) = \frac{1}{2}|u|$ is well known, but the general case is easily seen by considering the directional derivatives of the function R. Now, just as we can extend the notion of estimation of the unconditional sample mean defined as $\hat{\mu} = \text{argmin} \sum(\text{y}_\text{i} - \mu)^2$ to the estimation of linear *conditional* mean functions by solving, $\hat{\beta} = \text{argmin} \sum(\text{y}_\text{i} - \text{x}_\text{i}'\beta)^2$ we can extend the notion of univariate unconditional quantiles to the estimation of linear conditional quantile functions by solving,

$$\hat{\beta}(\tau) = \text{argmin} \sum \rho_\tau(\text{y}_\text{i} - \text{x}_\text{i}'\beta). \qquad (1)$$

The problem 1 can be easily formulated as a linear program and simplex based methods provide efficient algorithms for most applications. See Koenker and d'Orey (1987,1993) for further details. In large problems recent development of interior point methods offer substantial computational advantages. Portnoy and Koenker (1998) demonstrate that quantile regression computational speed is now comparable to least squares computational speed throughout the range of problem dimensions observed in applications.

2.2 Quantile Regression Treatment Effects

The simplest formulation of quantile regression is the two-sample treatment-control model where as suggested by Lehmann (1974, p. 68) a general notion of the quantile treatment effect can be defined as,

> "Suppose the treatment adds the amount $\Delta(x)$ when the response of the untreated subject would be x. Then the distribution G of the treatment responses is that of the random variable $X + \Delta(X)$ where X is distributed according to F."

In this view, elaborated by Doksum (1974), the quantile treatment effect is simply,

$$\delta(\tau) = G^{-1}(\tau) - F^{-1}(\tau),$$

that is, the difference between the quantiles of the marginal distributions of the treatment and control responses. This is precisely what is estimated in the two sample quantile regression model,

$$Q_{Y_i}(\tau|D_i) = \alpha(\tau) + \delta(\tau)D_i \qquad (2)$$

where D_i denotes the treatment indicator, with $D_i = 1$ indicating treatment, $D_i = 0$, control.

When the treatment variable takes more than two values, this interpretation requires only minor adaptation. In the case of p distinct treatments, we can write

$$Q_{Y_i}(\tau|D_{ij}) = \alpha(\tau) + \sum_{j=1}^{p} \delta_j(\tau)D_{ij}$$

where $D_{ij} = 1$ if observation i received the jth treatment and $D_{ij} = 0$ otherwise. Here $\delta_j(\tau)$ constitutes the quantile treatment effect connecting the distribution of control responses to the responses of subjects under treatment j. If the treatment is continuous as, for example, in dose-response studies, then it is natural to consider the assumption that the effect is linear, and write,

$$Q_{Y_i}(\tau|x_i) = \alpha(\tau) + \beta(\tau)x_i.$$

We assume thereby that the treatment effect, $\beta(\tau)$, of changing x from x_0 to $x_0 + 1$ is the same as the treatment effect of an alteration of x from x_1 to x_1+1. Interpreted in this fashion the quantile treatment effect offers a natural extension to continuously varying treatments of the Lehmann-Doksum formulation for the discrete case.

2.3 Transformation Equivariance of Quantile Regression

In the quantile regression model, for any monotone function, $h(\cdot)$, we have,

$$Q_{h(T)}(\tau|x) = h(Q_T(\tau|x)).$$

This equivariance to monotone transformations of the conditional quantile function is a crucial feature, allowing us to decouple the potentially conflicting objectives of transformations. This equivariance property is in direct contrast to the inherent conflicts in estimating transformation models for conditional mean relationships. Since, in general, $E(h(T)|x) \neq h(E(T|x))$ the transformation alters in a fundamental way what is being estimated in ordinary least squares regression.

2.4 Robustness

Robustness to distributional assumptions is an important consideration throughout survival analysis, so it is worth emphasizing that quantile regression inherits the robustness of the ordinary sample quantiles. The estimates and the associated inference apparatus have an inherent distribution-free character since quantile estimation is influenced only by the local behavior of the conditional distribution of the response near the specified quantile. Other semi-parametric methods like the Cox model are also distribution free in certain respects. But it is typical that these models assume that there is some distributional feature like the baseline hazard function in the Cox model that is common, up to a location/scale shift for all values of the covariates. In quantile regression there is considerably more flexibility in the sense that the shape of the conditional density may change with the covariates.

2.5 Inference in Quantile Regression

The asymptotic behavior of the quantile regression process $\{\hat\beta(\tau) : \tau \in (0,1)\}$ closely parallels the theory of ordinary sample quantiles in the one sample problem. Koenker and Bassett (1978) show that in the classical linear model,

$$y_i = x_i\beta + u_i$$

with u_i iid from dfF, with density $f(u) > 0$ on its support $\{u|0 < F(u) < 1\}$, the joint distribution of $\sqrt{n}(\hat\beta_n(\tau_i) - \beta(\tau_i))_{i-1}^m$ is asymptotically normal with mean 0 and covariance matrix $\Omega \otimes D^{-1}$. Here $\beta(\tau) = \beta + F_u^{-1}(\tau)e_1, e_1 = (1,0,\ldots,0)', x_{1i} \equiv 1, n^{-1}\sum x_i x_i' \to D$, a positive definite matrix, and

$$\Omega = (\omega_{ij} = (\tau_i \wedge \tau_j - \tau_i\tau_j)/(f(F^{-1}(\tau_i))f(F^{-1}(\tau_j))).$$

When the response is conditionally independent over i, but not identically distributed, the asymptotic covariance matrix of $\xi(\tau) = \sqrt{n}(\hat\beta(\tau) - \beta(\tau))$ is somewhat more complicated. Let

$$\xi_i(\tau) = x_i\beta(\tau)$$

denote the conditional quantile function of y given x_i, and $f_i(\cdot)$ the corresponding conditional density, and define,

$$J_n(\tau_1,\tau_2) = (\tau_1 \wedge \tau_2 - \tau_1\tau_2)n^{-1}\sum_{i=1}^n x_i x_i',$$

and

$$H_n(\tau) = n^{-1}\sum x_i x_i' f_i(\xi_i(\tau)).$$

Under mild regularity conditions on the $\{f_i\}$'s and $\{x_i\}$'s, we have joint asymptotic normality for vectors $(\xi(\tau_i),\ldots,\xi(\tau_m))$ with mean zero and covariance matrix

$$V_n = (H_n(\tau_i)^{-1}J_n(\tau_i,\tau_j)H_n(\tau_j)^{-1})_{i=1}^m.$$

This "Huber sandwich" is the quantile regression version of the Eicker-White heteroscedasticity consistent covariance matrix for the least squares estimator.

In the present application we will estimate $f_i(\xi_i(\tau))$ using

$$\hat f_i(\hat\xi_i(\tau)) = \max\{0, 2h_n/(x_i'(\hat\beta(\tau+h_n) - \hat\beta(t-h_n)) - \varepsilon)$$

where $h_n = n^{-1/3}\Phi^{-1}(1-\alpha/2)^{2/3}((3/2\phi^2(0))/(2\Phi^{-1}(\tau)^2 + 1))^{1/3}$ is a bandwidth selected in accordance with the theory developed in Hall and Sheather (1989).

There are several alternative schemes for conducting inference for quantile regression. Rank based methods of inference for quantile regression are surveyed in Koenker (1996), and various approaches to inference based on resampling methods are discussed in Parzen, Wei and Ying (1994), Horowitz (1999), Buchinsky (1998) and Hahn (1995). Koenker and Machado (1999) discuss general goodness of fit measures and related inference methods based on the entire quantile regression process.

3. QUANTILE REGRESSION MODELS FOR SURVIVAL ANALYSIS

A wide variety of survival analysis models, as noted by Doksum and Gasko (1990), may be written as

$$h(T_i) = x_i'\beta + u_i$$

where h is a monotone transformation, T_i is an observed survival time, x_i is a vector of covariates, β is an unknown parameter vector in $\mathbb{R}^p$ and $\{u_i\}$ are assumed to be iid with distribution function F. The quantile regression approach to survival analysis can be seen as a natural extension of the transformation model formulation.

3.1 Three Examples

We illustrate this transformation model formulation of survival models with three leading examples.

Example 1. For the Cox (1972) proportional hazard model with

$$\log\lambda(t|x) = \log\lambda_0(t) - x'\beta$$

we can express the conditional survival function in terms of the integrated baseline hazard $\Lambda(t) = \int_0^t \lambda_0(s)ds$ as,

$$\log(-\log(S(t|x))) = \log\Lambda_0(t) - x'\beta$$

and thus write the model as

$$\log \Lambda_0(T) = x'\beta + u$$

for u_i iid with extreme value distribution $F(u) = 1 - e^{-e^u}$.

Example 2. For the Bennett (1983) proportional odds model where the conditional odds of death $\Gamma(t|x) = F(t|x)/(1 - F(t|x))$ are written as,

$$\log \Gamma(t|x) = \log \Gamma_0(t) - x'\beta,$$

Doksum and Gasko (1990) show that,

$$\log \Gamma_0(T) = x'\beta + u$$

for u iid logistic with $F(u) = (1 + e^{-u})^{-1}$.

Example 3 In the accelerated failure time model we have

$$\log(T_i) = x_i'\beta + u_i$$

with the distribution of u_i unspecified. A special case is the Cox model with Weibull baseline hazard, but in general we have

$$P(T > t) = P(e^u > te^{-x\beta}) = 1 - F(te^{-x\beta})$$

where $F(\cdot)$ denotes the distribution function of e^u, and therefore, in this model,

$$\lambda(t|x) = \lambda_0(te^{-x\beta})e^{-x\beta}$$

where $\lambda_0(\cdot)$ denotes the hazard function corresponding to F. In effect, the covariates act to rescale time in the baseline hazard.

A common feature of all of the foregoing models is the iid error assumption which implies that for some appropriate choice of $h(\cdot)$ we can express the transformed survival times as a pure location shift model. Thus if we were to formulate a family of conditional quantile models for $h(T)$ we would have a family of parallel conditional quantile functions,

$$Q_{h(T)}(\tau|x) = x'\beta + F_u^{-1}(\tau).$$

for $\tau \in (0, 1)$. The covariate effect $x'\beta$ shifts the location of the conditional density of $h(T)$, but the covariates have no effect on the shape of the conditional density, or even on its dispersion. This is obviously highly restrictive, and it seems only prudent to explore alternative models which relax this strict form of the location shift model.

3.2 Transformation Models and Quantile Regression

As in the ordinary linear regression model, one avenue of exploration is to consider a *family* of linear models for the conditional quantile functions of the transformed survival time $h(T)$,

$$Q_{h(T)}(\tau|x) = x'\beta(\tau)$$

where, potentially, all of the parameters composing the vector $\beta(\tau)$ may depend upon the quantile, $\tau \in (0, 1)$, of interest. The prior models constitute special cases in which all the dependence on τ is concentrated in the intercept coefficient, leaving the slope parameters independent of τ. As emphasized by Chaudhuri, Doksum and Samarov (1997), the quantile regression vector $\beta(\tau)$ is "a unifying concept that represents the coefficient vectors in the standard linear model, the Cox model, the proportional odds models, the accelerated failure time model and so on."

By allowing the slope coefficients of $\beta(\tau)$ to depend upon τ, we can introduce various forms of heterogeneity in the conditional distribution of $h(T)$ over the space of covariates. A particularly simple, yet important, case is the family of linear location-scale models

$$h(T_i) = x_i'\beta + (x_i'\gamma)u_i$$

for u_i iid from F; in this model we have the family of quantile regression models,

$$Q_{h(T)}(\tau|x) = x'\beta + x'\gamma F_u^{-1}(\tau) = x\beta(\tau)$$

where $\beta(\tau) = \beta + \gamma F^{-1}(\tau)$. In this case all the coordinates of $\beta(\tau)$ depend upon τ in the same way, up to a location and scale shift. This model captures a variety of plausible models of heteroscedasticity. More general forms of τ-dependence are obviously possible and reflect, as we shall see, more complicated notions of how the covariates influence the conditional distribution of the survival times. Particular applications may suggest reasons for focusing attention on restricted domains for $\beta(\tau)$. For example, in clinical trials one may be especially interested in treatment effects on long-term survival and thus wish to focus only on the upper tail of $\beta(\tau)$. The quantile regression framework permits this without worry that some global aspect of the model specification is unduly prejudicing the results in the region of particular interest.

An important consideration in the formulation of parametric quantile regression models is to find specifications for which the function

$$Q_T(\tau|x) = x'\beta(\tau)$$

can be assured to be monotone in τ over the full support of the design. In the simplest bivariate case where, $x = (1, z)'$,

$$Q_T(\tau|x) = \beta_0(\tau) + z\beta_1(\tau),$$

if z has support on the entire real line, then unless $\beta_1(\tau)$ is constant, i.e. independent of τ, we will have crossing of the conditional quantile functions and thus violations of the monotonicity requirement. Of course, in practice, we have bounded support for the $\{x_i\}$'s so this may be avoidable. At the centroid of the design $\bar{x} = n^{-1}\sum x_i$, it can be shown, see Theorem 2.1 of Bassett and Koenker (1982)), that

$$\hat{Q}_T(\tau|\bar{x}) = \bar{x}'\hat{\beta}(\tau)$$

must be monotonic in τ, and therefore $\hat{Q}_T(\tau|x)$ must be monotone in τ in a neighborhood of $x = \bar{x}$. Thus, even though quantile regression functions are estimated independently over τ's, incidents of crossing generally occur only in outlying regions of the observed covariate space.

3.3 Survival and Hazard Functions

Having described how to estimate the parameters of a entire family of conditional quantile functions of a transformed response $h(T)$, we may now briefly consider how to go about translating these estimates into estimates of conditional survival and hazard functions. The conditional survival function,

$$S_T(t|x) = P(T > t|x) = 1 - F_T(t|x)$$

represents the proportion of those with $X = x$ surviving up to time t. In contrast, quantile regression provides an estimate of the inverse of $F_T(t|x)$, i.e.

$$Q_T(\tau|x) = \inf\{t|F_T(t|x) \geq \tau\},$$

the earliest time by which the proportion τ have died. Even for the transformation model, obtaining the conditional survival function from the conditional quantile function is absolutely transparent: we have estimated the conditional quantile functions, $\hat{Q}_{h(T)}(\tau|x) = x'\hat{\beta}(\tau)$ and we have seen that $Q_{h(T)}(\tau|x) = h(Q_T(\tau|x))$ so we may estimate the conditional quantile functions of the untransformed response by,

$$\hat{Q}_T(\tau|x) = h(x_i\hat{\beta}(\tau)).$$

Then, for any vector of the covariates, x, instead of plotting conditional quantiles as $\tau \mapsto \hat{Q}_T(\tau|x)$, we may plot the conditional survival function as, $\hat{Q}_T(\tau|x) \mapsto 1 - \tau$. As we have noted in Section 2.2, we can interpret the quantile regression parameters as estimates of the horizontal distance between survival curves for subjects differing by one unit in a single covariate and conditional on fixed values for all the other covariates.

For the conditional hazard function

$$\lambda_T(t|x) = f_T(t|x)/S_T(t|x) = -d\log S_T(t|x)/dt.$$

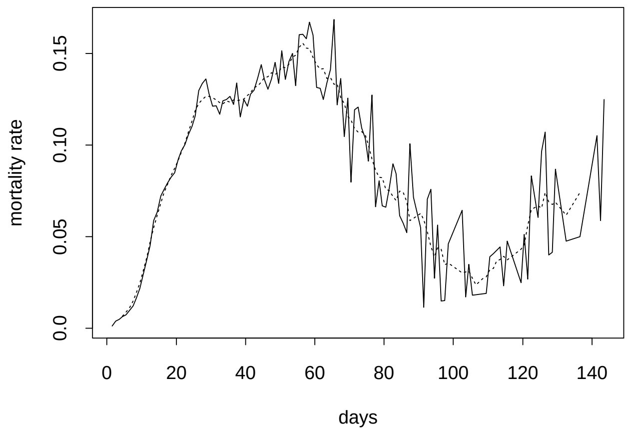

Figure 1. This figure illustrates the raw mortality rate data for the full sample with a smoothed estimate of the hazard as in Carey and Liedo (1995) superimposed as the dotted line. The smooth is a 7-day geometric moving average.

we need to differentiate $\log \hat{S}(t|x)$, and we may use,

$$\hat{\lambda}(\hat{Q}(\tau|x)|x) = \frac{\Delta\tau/\Delta\hat{Q}(\tau|x)}{1-\tau}$$

using a grid of evaluations for τ: $0 < \tau_1 < \ldots < \tau_m < 1$. Note that in the unsmoothed case, since the numerator of the expression for $\hat{\lambda}$ estimates $f_T(t|x)$ between successive time points $t_i = \hat{Q}_T(\tau_i|x)$ and $t_{i+1} = \hat{Q}_T(\tau_{i+1}|x)$ it is advisable to evaluate the denominator at the midpoint $1 - (\tau_i + \tau_{i+1})/2$, and do the plotting versus, $\bar{t}_i = (t_i + t_{i+1})/2$. Thus, we propose plotting $\bar{t}_i$ versus

$$\hat{\lambda}(\bar{t}_i|x) = \frac{\Delta\tau/\Delta\hat{Q}(\tau|x)}{1-(\tau_i+\tau_{i+1})/2}.$$

This approach enables us to estimate the conditional hazard function for a fully general form of the covariate vector, x. Since this derivative is inevitably considerably rougher than the estimates $\hat{Q}$ and $\hat{S}$, it may be reasonable to do some additional smoothing to obtain a final estimate.

To illustrate this consider the proportional hazard model, where,

$$S(t|x) = S_0(t)^{\gamma(x)}$$

with $\gamma(x) = e^{-x'\beta}$, and $S_0(t)$ denoting the baseline survival function. The quantile functions for the survival time T in this model is thus,

$$Q_T(\tau|x) = S_0^{-1}((1-\tau)^{1/\gamma(x)})$$

and therefore,

$$\frac{\partial Q_T(\tau|x)}{\partial x_i} = \frac{(1-\tau)\log(1-\tau)\gamma(x)}{S_0'(Q_T(\tau|x))}\beta_i.$$

So, in the proportional hazard model the marginal effects of the various covariates, viewed as functions of τ, are all identical up to the scalar factors determined by the components of the vector, β. In particular, since the τ-dependent factor multiplying β_i is positive, it is clear that the implicit quantile treatment effects for the Cox model must have the same sign as β_i for all τ, and thus the model inherently prohibits any form of quantile treatment effect that would entail crossings of the survival functions for different settings of the covariates.

4. DATA

In Carey *et al* (1992) three distinct experiments are analyzed, two of which involved cohorts of 20,000 medflies raised in solitary confinement. Our investigation is restricted to the largest of the three experiments in which roughly 1.2 million medflies were raised in cages each initially containing about 7,200 individuals. The experiments were conducted in a large rearing facility in Metapa, a city located in the Chiapas region of Mexico. Technical details on precise experimental conditions are available in Carey *et al* (1992) and Vargas (1989). The basic conditions, as described by Carey *et al* (1995) were as follows:

"...Pupae were sorted into one of five size classes using a pupal sorter. This enabled size dimorphism to be eliminated as a potential source of sex-specific mortality differences. Approximately, 7,200 medflies (both sexes) of a given size class were maintained in each of 167 mesh covered, 15 cm × 60 cm × 90 cm aluminum cages. Adults were given a diet of sugar and water, *ad libitum*, and each day dead flies were removed, counted and their sex determined ..."

A total of 1,203,646 medflies were studied with survival times recorded in days. Figure 1 displays raw and smoothed mortality (hazard) rates for the entire sample, and illustrates one of the principle findings – the decline in the mortality rate at advanced age. After 50 days only one percent of the original population survives, but mortality declines dramatically from a peak of about 15% to less than 5% at 100 days. Let y_t denote the number alive ("at risk") at day t, then the raw mortality rate, m_t is given by

$$m_t = 1 - y_{t+1}/y_t.$$

Following Carey and Liedo (1995) the smoothed rate is computed as the two sided, seven-day, geometric moving average,

$$\tilde{m}_t = 1 - (\prod_{s=t-3}^{t+3} (1 - m_s))^{1/7}.$$

Figure 2 contrasts the smoothed mortality rates for males and females illustrating the finding of Carey *et al* (1995) that female mortality exceeds males up to about 20 days, after which the male mortality rate exceeds the female rate until about day 60, and rates are indistinguishable thereafter.

The remarkably long right tail of the medfly survival distribution is already apparent in the foregoing plots. By human standards it may appear implausible that individuals could live to age 172 when 99.9% of the population is dead by age 64 (see Carey *et al* (1992), Table 1). So it may be worth remarking at this point that in the smaller experiments in which flies were raised in solitary confinement there were individuals that survived even beyond 200 days.

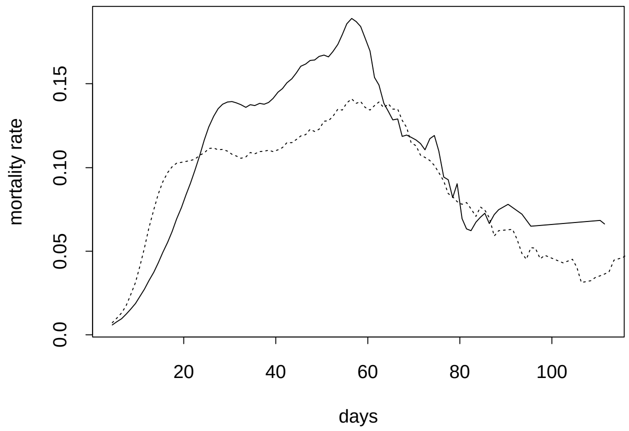

Figure 2. This figure illustrates the smoothed mortality rates for males and females. The female mortality exceeds that of males up to about 20 days, then male mortality is higher than the female rate until about day 60, after which the rates are difficult to distinguish. The smoothing employs the same 7-day geometric moving average approach as the previous figure.

Prior analysis has focused primarily on the age specific pattern of mortality rates, especially at advanced ages, and on gender differences in these rates. We will consider a more extensive catalogue of potential covariates defined below. In doing so, we will be able to explore the effects of these several new influences on survival prospects as well as addressing some criticisms of the prior analyses of these data.

Sex The sex of each fly was determined and recorded at death.

Size Pupae were sorted into five size classes (4-8mm) using a pupal sorter, enabling us to control for size differences as a cause of mortality differences. Pupae size is strongly linked to adult size of the flies. Each cage contained flies of only one size.

Density Initial density of flies varied considerably across cages. The effect of density on longevity has proven to be a somewhat controversial aspect of the interpretation of the experimental results. For a discussion of the effect of density on longevity in the broader context of other medfly experiments, and the rationale for focusing attention on initial density, see Carey, Liedo, and Vaupel (1995).

%Males The initial proportion of males in each cage also varies considerably across cages and merits investigation.

Batch Pupae were raised in 8 distinct batches with potentially heterogeneous composition. As a consequence we have investigated the possibility of a confounding "batch effect". As with **Size**, cages were allocated pupae from only one batch.

Because survival was recorded in days for males and females separately, and all other covariates were associated with the 167 cages, the data set can be collapsed to 19072 observations and their associated cell counts. This feature leads to some gains in efficiency from a computing standpoint since weighting reduces the effective sample size.

In addition to the statistically almost irresistible attraction of the sheer size of the Carey *et al* (1992) medfly survival experiment, it has the uncommon virtue of being free of censoring. It is thus well suited to classical life table methods of analysis illustrated above. Such methods are, however, not well suited to modeling the effects of continuously measured covariates, so we turn now to a description of the quantile regression formulation.

5. A QUANTILE REGRESSION SURVIVAL ANALYSIS

Our basic model for analysing the medfly survival data takes the traditional accelerated failure time form,

$$Q_{\log(T)}(\tau|x) = x'\beta(\tau). \tag{3}$$

We model the conditional quantile functions of the logarithm of survival times as linear in the observed covariates, x. The choice of the log transformation is primarily dictated by ease of interpretability and the desire to achieve linearity in the parametric specification. Multiplicative covariate effects are widely accepted throughout survival analysis, and certainly seem more plausible in the present application than the assumption of additive linear effects. As we have observed in Section 2.1 above, the conditional quantile functions of other transformations of T can be recovered immediately from the model 3. In particular, we may write the family of conditional quantile functions for the untransformed survival time T as,

$$Q_T(\tau|x) = \exp\{x'\beta(\tau)\}.$$

Mortality (hazard) rates may be recovered by numerical differentiation as described above.

We will consider two distinct models. The first includes the covariates: sex, size, density, and %male as additive linear effects on $\log(T)$. In our second model we introduce the batch effect to control for possible heterogeneity in the pupae allocated to cages in the Carey *et al* experiment. In each case we estimated models for the 42 quantiles: $\{.01, .05, .10, ..., .95\}$, $\{.96, .97, ..., .99\}$, $\{.991, ..., .999\}$, and $\{.9991, ..., .9999\}$. A variety of other models were explored, but none offered convincing evidence of effects not represented in the two models that we now present.

5.1 Model A: No Batch Effects

Figure 3 provides a concise visual summary of the results for the first model that omits the batch effect. There are five estimated coefficients. The lightly shaded region in each panel represents a 90 percent pointwise confidence band for each coefficient. The intercept panel of the figure may be interpreted as the estimated quantiles of log survival time for *male* flies of mean size when evaluated at the experimental mean cage density and the mean initial proportion of males.

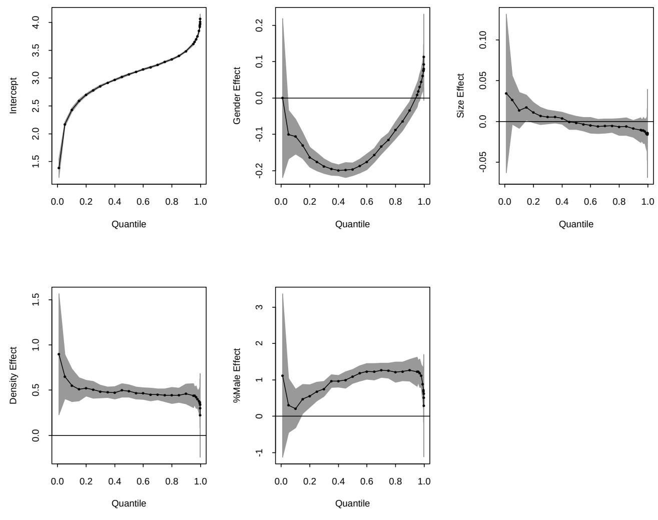

Figure 3. This figure illustrates the quantile regression results for Model A. There are 5 coefficients estimated. The lightly shaded region is a 90 percent pointwise confidence band for the corresponding coefficient. The horizontal line at zero represents the null hypothesis of no effect for each covariate.

The gender effect depicted in the second panel represents the estimated difference in the quantiles of log survival times for female versus male flies holding the other factors constant. It is clear that this effect is considerably more complicated than a simple location shift. Unlike human populations in which females generally outlive males, male medflies have a distinct advantage over females up to about the 95th percentile of longevity. For the longest lived five percent of the population, females appear to have a distinct advantage. More explicitly, we see that the disadvantage of females is maximal near the median, where a coefficient of -0.2 may be translated into a multiplicative effect on the median quantile of longevity of $\exp(-0.2) \approx .82$, implying that the male median lifespan of about 22 days corresponds roughly to a median lifespan for females of 18 days. At the opposite extreme, the coefficient of 0.1 in the extreme right tail of the distribution of survival times implies that females have an $\exp(0.1) \approx 1.1$, or 10 percent longer life span among the oldest old. This cross-over in gender survival distributions is a important finding of the medfly experiments and we will return to it later in this Section.

The density effect illustrated in the third panel of Figure 3 is particularly interesting in view of some of the criticism directed at the conclusions of Carey *et al* (1992) It has been suggested by Kowald and Kirkwood(1993) as well as Nusbaum *et al* (1993) that the initially high density of flies in the Carey experiment may have contributed to higher mortality rates for younger flies and thus distorted the pattern of mortality rates portrayed by Carey *et al.* Indeed Nusbaum *et al* suggest,

> "The results of Carey *et al* may arise from an ecological dependence of fly mortality rate on population density, not because of some undiscovered property of extremely old flies."

The evidence from Figure 3 does not seem to support the Nusbaum *et al* hypothesis. Over the range of densities observed in the cages of the Carey experiment, initial density seems to exert a strongly *positive* effect on longevity. The density effect appears to be quite constant over the entire range of quantiles. This location shift effect in the model for $\log T$ implies, of course, a scale shift in T due to density. The standard deviation of cage density weighted by the cell counts is roughly .127, so the predicted effect of increasing density by one standard deviation is to increase T by a factor of $\exp\{.5 \cdot .127\} = 1.065$, or 6.5 percent. Note that density has been rescaled by 10,000 to facilitate reporting of the results throughout this Section.

Given the disparity in the sex ratios across cages it seemed interesting and worthwhile to explore whether the initial proportion of males in a cage exerted any effect on survival chances. To our surprise, we found that flies in cages with a higher proportion of males tended to live significantly longer. This effect is strongest above the median

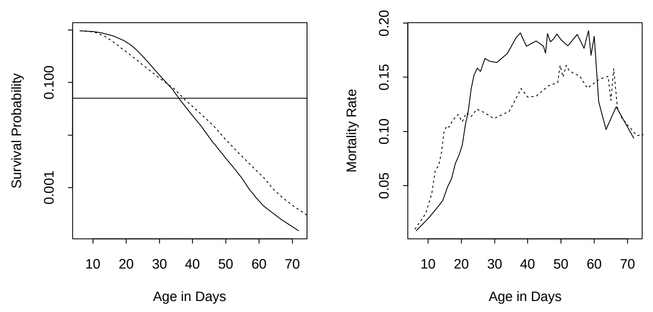

Figure 4. This figure illustrates the estimated survival and hazard functions for Model B. Survival and hazard functions are illustrated males (solid line) and females (dotted line) with the other covariates evaluated at the experimental sample means.

where the coefficient is roughly 1.3. Since the standard error of the %male variable is 0.046, this implies that a one standard deviation increase in the proportion of males increases the third quartile of the survival distribution by roughly 6.2 percent. Clearly, below the median the effect is considerably weaker. It is natural to ask whether this effect is shared equally by males and females. Our attempt to explore this question by adding an interaction effect between the sex and %male variables yielded no significant interaction over the range of quantiles estimated, thus suggesting that both sexes benefited from an excess of males in the initial population. It is also natural to ask whether there is an "optimal" density level or proportion of males. To explore this we attempted to fit quadratic terms in these variables, but we were unable to identify a significant quadratic effect in either case. Of course, it is quite plausible that further experimentation involving a considerably greater range for density, or the proportion of males, might reveal such effects.

Finally, we may consider the effect of size on longevity. Recall that there are 5 initial size categories corresponding to pupal sizes 4mm, 5mm, 6mm, 7mm and 8mm. There does appear to be a slight advantage in being larger in the lower quantiles of the survival distribution, and perhaps a slight disadvantage in being larger in the upper quantiles, but neither of these effects is statistically very compelling. Observe that the horizontal line at $\beta(\tau) = 0$ representing the null effect rarely emerges from the confidence band for this group of coefficients.

5.2 Model B: The Batch Effect

Figure 5 illustrates the results from Model B, which includes the batch effect. The other variables are exactly as in Model A. Results, depicted in Figure 5, exhibit some important differences from those reported for Model A.

First, we should observe that the shape and significance of the gender effect is essentially unchanged by the new specification. Males retain their substantial survival advantage up to the 95th percentile of the survival distribution, but as in Model A female survival prospects exceed males' in the upper 5 percent of the distribution. The effect of initial cage density has, however, changed substantially. There is still a significant positive effect of higher density, but only in the lower third of the distribution. In the upper tail the effect is negligible. The proportion-of-males effect is also substantively altered by the introduction of the batch effect. While in Model A this effect was weak in the left tail and large, $\hat{\beta}(\tau) \approx 1.3$, above the median, in Model B the effect appears roughly constant at about $\hat{\beta}(\tau) \approx 0.5$, over the entire distribution. The confidence band is slightly wider in Model B, rendering the effect marginally significant throughout. The effect of size in Model B is qualitatively similar to the results of Model A. Larger size seems advantageous for survival for younger flies, up to about the 40th percentile, but is disadvantageous thereafter. But note that in this case the introduction of the batch effect seems to have improved the precision of the size effect estimates. Since we have no information on the nature of the experimental batches, it is difficult to interpret the coefficients on the batch effects. It is clear that the batches do constitute a significant factor in assessing survival prospects in the Carey *et al* experiment. Because the batch factor is cage specific, like density, size, and proportion-of-males it is not surprising to encounter some confounding of effects.

5.3 Survival and Hazard Curves

To explore the gender cross-over effect a bit further we plot in Figures 4 the implied survival and hazard functions for Model B. The corresponding plot for Model A is very similar. Male survival is indicated by the solid line, female survival in the dotted line. The cross-over in the survival curves at $\tau = .95$ is indicated by the horizontal line. In the right tail female flies have a rather substantial survival advantage. This is also clear in the hazard plots of the mortality rates for males and females. No attempt was made to smooth the hazard plots based on the quantile After about day 25 female mortality falls below male mortality and remains below until roughly age 60, after which as we have already noted in Section 4, using the life table methods of Carey *et al*, the rates are difficult to distinguish.

In Figure 6 we plot estimated conditional hazard functions, for Model A, for two representative values of initial cage density, *viz.* mean density plus and minus one standard deviation. The plot illustrates the flexibility of the model, since we are able to estimate conditional survival and hazard (mortality) curves for any setting of the covariate vector without imposing stringent conditions on the relationships among these curves.

We should emphasize that in contrast to the smoothed life-table plots presented earlier, the present plots reflect the conditioning of the respective quantile regression models on size, density, percentage males and the batch effect. The crucial finding of Carey *et al* (1992) that mortality rates decline in the right tail is confirmed, thus dispelling suggestions that it may have arisen from inadequate control of confounding factors such as cage density. Of course, it must be recognized that the estimation of the extreme

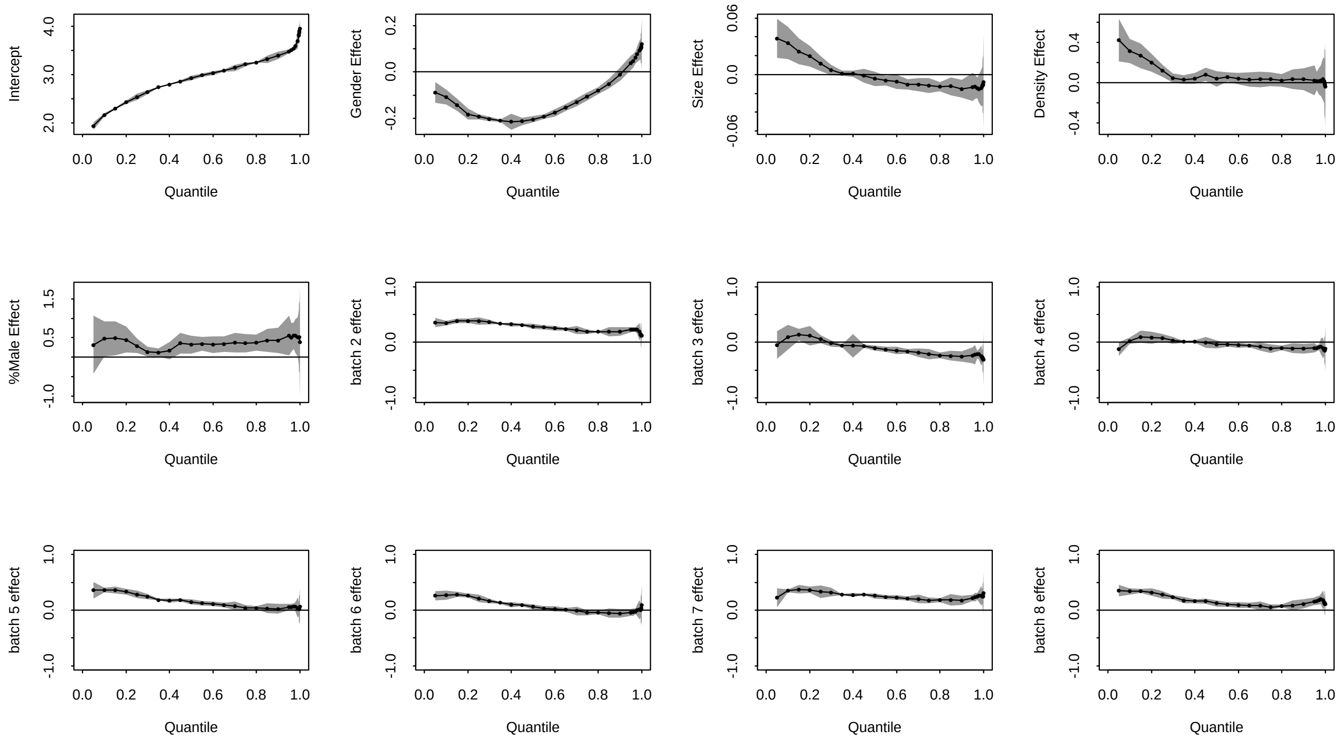

Figure 5. *This figure illustrates the quantile regression results for Model B. There are 12 coefficients estimated. The lightly shaded region is again a 90 percent pointwise confidence band for the corresponding coefficient. Again, the horizontal line at zero represents the null hyypothesis of no effect for each of the covariates.*

quantiles of the survival distribution is inherently difficult. By age 70 less than .01 percent of the sample remains: 127 males and 374 females. We are reluctant to push the analysis too much further out into the tail, but given these sample sizes we can be reasonably confident that the finding of declining mortality rate at advanced ages is not an artifact of density, size, or batch effects. We would like to underscore that, in contrast to the life table methods employed in Carey *et al* (1992), quantile regression methods permit us to compute estimated survival and hazard functions for any settings of our covariate vectors including contrasts in the continuous covariates like density and the proportion of males.

5.4 A Proportional Hazard Model Comparison

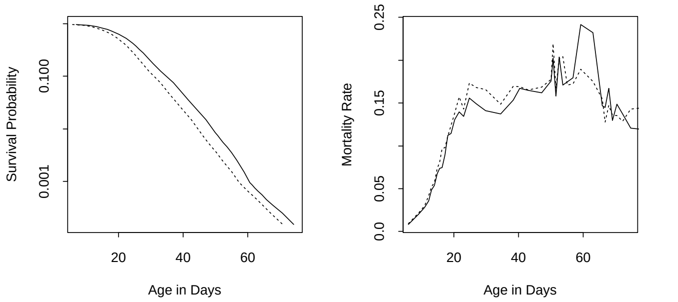

Figure 6. *This figure illustrates estimated survival and hazard functions for male medfies based on Model A for two different settings of the initial cage density. Low initial density defined as mean density minus one standard deviation is indicated by the dotted line, and high initial density defined as mean density plus one standard deviation by the sold line. Other covariates are set at sample means.*

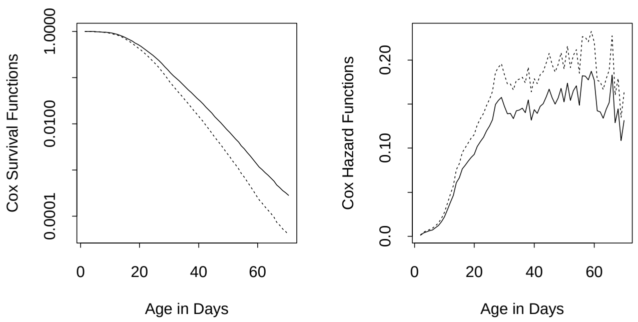

Figure 7. *This figure illustrates estimated survival and hazard functions for males and females based on the reported Cox proportional hazard model. Note that in contrast to the prior plots, the PH model prohibits the survival and hazard functions from crossing.*

The proportional hazard model of Cox(1972) also offers natural extension of life table methods to regression-type models for survival data. To contrast our quantile regression results with the Cox model we estimated the following proportional hazard (PH) model corresponding to model A.

$$\log \lambda(t|x) = \log \lambda_0(t) + \underset{(0.0018)}{0.2165}\mathsf{SEX} + \underset{(0.00084)}{0.0124}\,\mathsf{SIZE}$$
$$-\underset{(0.00072)}{1.021}\,\mathsf{DENSITY} - \underset{(0.0203)}{2.625}\,\mathsf{\%MALE}.$$

All four effects are obviously highly significant. Females, and flies from larger pupae, have higher hazard, while higher cage density and a larger percentage of males reduce the hazard. These results are roughly as anticipated from the prior analysis. However, the PH specification imposes a severe restriction on the nature of these effects. By requiring that the covariates act as a scale shift of the baseline hazard, $\lambda_0(t)$, several of the most interesting observations we have made about the quantile regression results are effectively assumed away. Since, $\lambda_0(t) \geq 0$, the sign of the coefficient in the PH model determines the sign of the effect over the entire distribution. In the quantile regression results both size and gender had ambiguous effects. The coefficient on the indicator variable for gender was significant and negative throughout most of the distribution, but turned significant and positive in the upper tail of the distribution. Similarly, pupal size appears, particularly in Model B, to exert a significantly beneficial effect on survival in the lower tail, but was significantly disadvantageous in the upper tail. Cross-overs of this type are potentially very important. But they are rendered invisible in the conventional PH analysis.

In Figure 7 we plot the estimated survival hazard functions for males and females based on the estimated Cox model, evaluating the other covariates at their means. The baseline hazard function estimation is complicated slightly by the weighted nature of the observations and we have used the methods described by Kalbfleisch and Prentice (1980, p. 85) for this purpose. The hazard curves have the same characteristic shape as those presented earlier with mortality rates declining after age 60. As we have emphasized above, the effect of each covariate on the quantiles of the survival distribution is exactly the same up to a rescaling. In the present case this means that all the covariates must conform to the monotone decreasing pattern over the range of the distribution, with the effects strictly positive and decreasing for covariates for which $\hat{\beta}_i > 0$ and strictly negative and increasing for those with $\hat{\beta}_i < 0$. The consequence for the survival and hazard plots in Figure 4 is that contrary to our earlier quantile regression findings it appears that females have poorer survival prospects *throughout* the distribution.

6. DISCUSSION

Large scale experiments on lower animals will continue to yield important insights into population biology and the nature of the aging process. To fully exploit the evidence offered by such experiments, we have argued that quantile regression methods provide a useful complement to the existing toolkit of survival analysis. By permitting the researcher to focus attention on covariate effects at particular quantiles of the survival distribution a more complete picture of the varied effects on survival may emerge.

We believe that these methods will also eventually prove useful in analyzing clinical trials where researchers may wish to focus, for example, on long-term survival prospects, without imposing stringent assumptions concerning the nature of short run treatment effects. This point is graphically illustrated in the Doksum's (1974) discussion of guinea pig experiments where injection of tubercle bacilli had an apparently beneficial effect on short-run survival prospects, but had disastrous long-term survival consequences. Such findings are difficult to reconcile with many conventional survival models that implicitly assume that covariate effects exert a pure location shift effect on some monotone transformation of survival times.

6.1 Quantile Treatment Effects

The Lehmann-Doksum quantile treatment effect is a useful way to view covariate effects in a wide range of quantile regression applications. By highlighting the possibility that covariates can alter not only location and scale, but may change the entire *shape* of the conditional distribution as well, this view broadens the scope of empirical analysis. In the two-sample treatment control model, Doksum suggests that we may interpret control subjects in terms of a latent characteristic: a control subject may be called frail if he is prone to die at an early age, and robust if he is prone to die at an advanced age. This latent characteristic is thus implicitly indexed by τ, the quantile of the survival distribution at which the subject would appear if untreated, i.e., $(Y_i|D_i = 0) = \alpha(\tau)$. And the treatment, under the Lehmann model, is assumed to alter the subjects control response, $\alpha(\tau)$, making it $\alpha(\tau) + \delta(\tau)$ under the treatment. If the latent characteristic, say, propensity for longevity, were observable *ex ante*, then we might view the treatment effect $\delta(\tau)$ as an explicit interaction with this observable variable. However, in the absence of such an observable variable, the quantile treatment effect may be regarded as a natural measure of the treatment response. Of course, there is no way of knowing whether the treatment actually operates in the manner proscribed by the Lehmann model. In fact, the treatment may miraculously make weak subject especially robust, and turn the strong into jello. All we can observe from experimental evidence, however, is the difference in the two marginal survival distributions, and so it is natural to associate the treatment effect with the difference in the corresponding quantiles of the two distributions. This is what the quantile treatment effect does.

6.2 Censoring

It would be highly irregular to conclude a paper on methods of analysis of survival data without any mention of the word "censoring". In the Carey *et al* medfly experiment we have the rare luxury of complete data, but this is obviously atypical. Fortunately, there is already a rather extensive literature on quantile regression with censoring. Powell (1986) treats the case of fixed censoring common in many econometric applications. Buchinsky and Hahn (1998) treat certain forms of random censoring, as does Lipsitz *et al* (1998). Fitzenberger (1998) and Buchinsky (1998) provide recent surveys of this literature.

6.3 Empirical Findings

Finally, we should make some concluding comments about what we have learned about medflies. Even a brief exposure to fruitflies seems to be sufficient to begin to see them as "little people with wings," so there is an irresistible temptation to offer some pithy life lessons: Males are tough, but only until 40. Crowds are salutary, especially of guys. And most paradoxically, life gets safer, but only after 60.

More seriously, the central findings of Carey *et al* (1992) that mortality rates decline at advanced ages, and that female survival prospects are better than males in the upper tail of the distribution are strongly sustained in estimated models that incorporate the effects of pupal size, initial cage density, the initial proportion of males, and pupal batches. We find that higher initial cage density, contrary to some early criticism of the Carey *et al* findings, exerts a positive effect on survival at least over the range of densities observed in the experiment. Rather surprisingly, the initial proportion of males also has a positive effect on survival prospects. The effect of size is rather weak, but there is some evidence, particularly from Model B, that larger flies have better early survival prospects, but poorer survival prospects at advanced ages. The most intriguing finding – that mortality actually declines at advanced age, survives after conditioning on a larger set of covariates, and thus clearly warrants the extensive further investigation it is receiving.

References

Bassett, G. and Koenker R. (1982) An empirical quantile function for the linear model with iid errors, *Journal of the American Statistical Society*, **77**, 407-415.

Bennett, S., (1983), Analysis of Survival Data By the Proportional Odds Model, *Stat. Medicine*, **2**, 273–277.

Buchinsky, M. (1998) Recent advances in quantile regression models – A practical guideline for empirical research, *J. Human Resources*, **33**, 88-126.

Buchinsky, M. and J.Y. Hahn, (1998) An alternative estimator for censored quantile regression, *Econometrica*, **66**, 653-71.

Carey, J.R. (1993) *Applied Demography for Biologists*, Oxford: New York.

Carey, J.R. and Liedo, P. (1995) Sex mortality differentials and selective survival in large Medfly cohorts, *The Gerontologist,* **35**, 588-96.

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**, 457-61.

Carey, J.R., Liedo, P., Orozco, D. and Vaupel, J.W. (1995) A male-female longevity paradox in medfly cohorts, *J. of Animal Ecology*, **64**, 107-116.

Carey, J.R., Liedo, P. and Vaupel, J.W. (1995) Mortality dynamics of density in the Mediterranean fruit fly, *Experimental Gerontology,* **30**, 605-29.

Chaudhuri, P., K. Doksum, and A. Samarov (1997) On average derivative quantile regression, *Annals of Statisitics*, **25**, 715-44.

Cox, D.R. (1972) Regression Models and Life Tables (with discussion), *J. Royal Stat. Soc (B)*, **34**, 187-220.

Doksum, K. (1974) Empirical probability plots and statistical inference for nonlinear models in the two sample case, *Annals of Statistics*, **2**, 267-77.

Doksum, K. and M. Gasko (1990) On a correspondance between models in binary regression and survival analysis, *Intl. Stat. Rev.*, **58**, 243-52.

Fitzenberger, B. (1998) The moving block bootstrap and robust inference for linear least squares and quantile regression, *J. of Econometrics*, **82**, 235-87.

Hahn, J. (1995) Bootstrapping quantile regression estimators, *Econometric Theory*, **11**, 105-121.

Hall, P., and S. Sheather (1988): On the distribution of a studentized quantile, *J. of Royal Stat Society. (B)*, 50, 381–391.

Horowitz, J.L. (1998) Bootstrap methods for median regression models, *Econometrica*, **66**, 1327-42.

Kalbfleisch, J.D. and Prentice, R.L. (1980) *The Statistical Analysis of Failure Time Data,* Wiley, New York.

Koenker, R. (1997): Rank Tests for Linear Models, in *Handbook of Statistics*, ed. by G. Maddala, and C. Rao, vol. 15. North Holland.

Koenker, R. and Bassett, G. (1978) Regression quantiles, *Econometrica,* **46**, 33-50.

Koenker, R. and D'Orey, V. (1987) Computing regression quantiles, *Journal of the Royal Statistical Society, Applied Statistics*, **36**, 383-393.

Koenker, R. and D'Orey, V. (1994) Remark on Alg. AS 229: Computing dual regression quantiles and regression rank scores, *Applied Statistics,* **43**, 410-414.

Koenker, R. and J. Machado (1999) Goodness of Fit and Related Inference Processes for Quantile Regression, *J. Am. Stat. Assoc.*, **94**, 1296-1311.

Kowald, A. and Kirkwood, T.B.L. (1993) Explaining fruit fly longevity. Technical comments, *Science,* **260**, 1664-1665.

Lehmann, E. (1974) *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day: San Francisco.

Lipsitz, S.R., G.M Fitzmaurice, G. Molenberghs, and L.P. Zhao (1997), Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with HIV, *Applied Statistics*, **46**, 463-76.

Nusbaum, T.J., Graves, J.L., Mueller, L.D., and Rose, M.R. (1993) Fruit fly aging and mortality. Letters to the Editor, *Science,* **260**, 1567.

Parzen, M. I. Wei, L.J., and Ying, Z. (1994) A resampling method based on pivotal estimating functions, *Biometrika*, **81**, 341-350.

Portnoy, S. and R. Koenker (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-error vs. Absolute-error Estimators, with discussion, *Statistical Science*, 12, 279-300.

Powell, J.L. (1986) Censored regression quantiles, *J. of Econometrics*, **32**, 143-55.

Thatcher, A.R. (1999) The long term pattern of adult mortality and the highest attained age, with discussion. *J. of the Royal Stat. Soc. (A)*, **162**, 5-43.

Vargas, R.I. (1989) Mass production of tephritid fruit flies, in *World Crop Pests: Fruit Flies: Their Biology, Natural Enemies and Control*, **2B**, A.S. Robinson and G. Hooper, (eds.), Elsevier: Amsterdam.

Vaupel, J.W., Carey, J.R., Christensen, K., Johnson, T.E., Yashin, A.I. et al. (1998) Biodemographic trajectories of longevity, *Science,* **280(5365)**, 855-860.