DURATION ANALYSIS OF FLEET DYNAMICS

Garth Holloway, University of Reading, garth.holloway@reading.ac.uk
David Tomberlin, NOAA Fisheries, david.tomberlin@noaa.gov

ABSTRACT

Though long a standard technique in engineering and medical research, duration or survival analysis has become common in economics only in recent decades, and in fisheries economics we are aware of only one previous study. In this paper, we demonstrate the usefulness of duration analysis to understanding fleet dynamics, specifically to predicting fishers' exit decisions. The analysis explores the dependency of these decisions on the length of time during which a fisher has been active and on boat characteristics. We emphasize the development of a Bayesian approach to duration model estimation, and, through an application to the US Pacific salmon fishery, we show the policy relevance of duration analysis to limited-entry fisheries in particular.

Keywords: Duration, Bayesian Estimation, Exit, Commercial Salmon Fishery

INTRODUCTION

Fishers' participation decisions are key to fisheries management, since they reflect conditions in the fishery and also influence both fish stocks and the availability of rents in the fishery. One approach to understanding these participation decisions is duration analysis, also known as survival analysis (in biomedical research) and failure time analysis (in engineering). Duration analysis is a statistical approach to understanding the length of time during which a unit (e.g., a boat) stays in a particular state (e.g., active in a fishery), as well as estimating the probability that a unit will leave that state in a subsequent period. In the context of fisheries, duration analysis is potentially quite useful because it sheds light on the factors that influence participation in the fishery and provides a means for estimating the probability that a given boat will persist in fishing or will exit, information that can help in predicting future effort.

Here, we apply duration analysis to the California commercial salmon fleet to demonstrate the utility of the approach in fisheries analysis. Specifically, we suppose that the length of time during which a boat is active in the California salmon fishery depends on four factors: the boat's average annual revenue from salmon, its share of in-season revenue derived from salmon, its length, and the average number of ports per year at which it lands salmon. We develop and estimate models for two distributions of boats' durations in the fishery, exponential and Weibull.   The exponential model is the simplest commonly used distribution, characterized by a constant hazard rate (the probability of exit at time *t+1*, given participation up to time *t*).   The Weibull model is a generalization in which the hazard rate may be increasing or decreasing with time, as might be the case for example if years of participation in a fishery affects the probability of exit independently of other explanatory variables that are included in the model. We adopt a Bayesian estimation framework, using a Metropolis-Hastings step within a Gibbs sampler to generate parameter estimates based on simulation of the posterior distribution. Because this approach to duration estimation is less common than are frequentist approaches, the paper includes a detailed description of the estimation procedure.

Duration models, as a framework for investigating the factors influencing exit (or entry) in a fishery, are potentially of significant benefit in fisheries management. Understanding the influence of factors such as boat characteristics on participation decisions is helpful in its own right, and may cast light on such issues as changes in fleet composition over time (see [1]) or future fishing effort. In the context of our study fishery, our interest is motivated by the apparent 'hardening' of the fleet into a stable core group after a

long period of large-scale exit. More generally, the approach may prove useful in addressing common fisheries management questions such as the effect of area closures on participation and the likely response of latent effort to fish stock recovery.

The next section provides a brief description of the California commercial salmon fishery and our data set. We then detail the econometric models used, present preliminary results, and conclude with observations on the method and possible applications in fisheries management.

STUDY FISHERY AND DATA

The California commercial salmon fishery is a limited-entry fishery dominated by small trollers with a crew of one or two persons, most commonly owner-operators. The fishery takes place during the summer months (generally May-September) and is subject to a variety of regulations including time and area closures and minimum sizes. Prior to the 1970s, the fishery supported a large number of fishermen who operated under essentially open-access conditions, but concern over declining fish stocks led to the imposition of a limited-entry regime in the early 1980s. Since that time, participation in the fishery has required a salmon vessel permit, which must be renewed each year. The listing of some California salmon stocks under federal and state endangered species laws has further restricted the activities of commercial salmon fishermen. While catch and revenues have fluctuated significantly over the last two decades, the overall story is one of declining participation and revenues until the latter 1990s, when the fishery entered a period of relative stability (Figure 1).
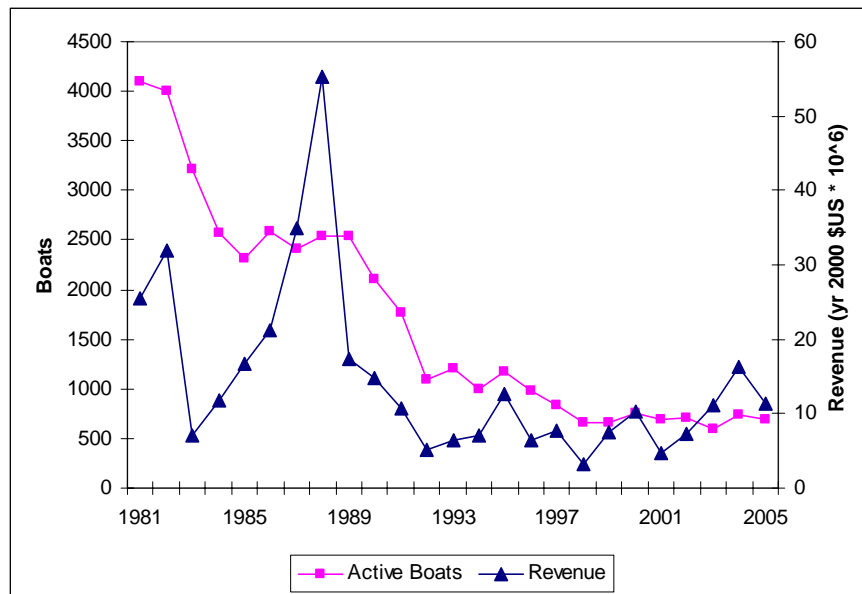

Figure 1.   Number of boats landing salmon in California and fleet real salmon revenue.

While a total of 6541 boats report landing salmon in California at least once during our study period, in this paper we will include in our data set only the 3779 boats with annual revenue time series that are uninterrupted by periods of non-participation. Accounting for boats that switch in and out of the fishery is a significant econometric complication and we can develop our argument here more easily by excluding them from our study. It's worth noting that the 3779 boats in our data set account for only 34% of total salmon revenues in the state during 1981-2005, suggesting that the behavior of boats that switch in and out of the fishery opportunistically is an important topic for later work.

2

Table 1 gives summary statistics on the variables used in our analysis. The dependent variable is duration, or the number of consecutive years in which a boat reports landing salmon in California. The covariates are boat length, average ports per year in which a boat landed salmon, real average salmon revenue per year, and the proportion of each boat's salmon revenue over the years in which is landed salmon to each boat's total revenue during salmon seasons over those years. Boat length appears as a proxy for a boat's catching power and its ability to pursue other fisheries (we have far more data on length than on tonnage). Average ports per year in which a boat lands salmon is included as a measure of mobility, which presumably makes 'survival' in the fishery more likely. Average salmon revenue per year, to the extent that it reflect a more profitable operation, might be expected to have a positive effect on duration. The share of a boat's total revenue due to salmon might reflect a higher degree of dependence on the species, hence a reluctance to exit the fishery. We have not included time-varying covariates, such as ocean conditions or regulations, leaving that for later work.

Table 1. Summary statistics on boats used in the analysis.

	Mean	Std. Dev.	Minimum	Maximum
Years of Salmon Landings	4.2	5.0	1	25
Boat Length (ft)	29	13.4	10	360
Avg. Ports per Year with Salmon Landings	1.3	0.7	1	7
Avg. Salmon Rev. per Year (yr 2000 USD)	4389	8024	3	132,595
Share of Salmon in Total Boat Revenue	0.7	0.4	0	1

Figure 2 shows the distribution of participation duration among our 3779 boats. While by far the most common duration is a single year, every participation period possible during our study horizon is represented, with 75 boats (about 2% of our study fleet) remaining in the fishery during all of 1981-2005.

Figure 3 shows the distribution of starting and ending years for our study fleet. Most entry and exit in the fishery took place during the 1980s, after which there has been far less turnover, no doubt at least in part to the earlier drop in participation. Figures 2 and 3 together demonstrate censoring in the data, which is prevalent in duration studies and will strongly influence the econometric development of the next section. In this case we have both left-censoring of the data, since our data set only contains information on landings from 1981 on, and right-censoring of the data, since the total length of time during which still-active boats fish is unobservable.

ECONOMETRIC APPROACH

To answer questions about the duration of fisher enterprises, we first consider the probability of observing an enterprise of length, t. We assume that t, if observed, is a realization from a random process $f(T|\theta)$, where $f(\cdot)$ denotes a probability density function (pdf) for the random variable T and $\theta \equiv (\theta_1, \theta_2, .., \theta_M)'$ denotes an M-vector of parameters conditioning $f(\cdot)$. It follows that the probability of observing the realization t is

$$F(t) = \int f(s)ds = \text{Prob}(T \leq t). \hspace{3cm} \text{(Eq. 1)}$$

As noted by Greene [2 p. 986], we will usually be more interested in the probability that the spell is of length at least as great as t, which is given by the *survival function*

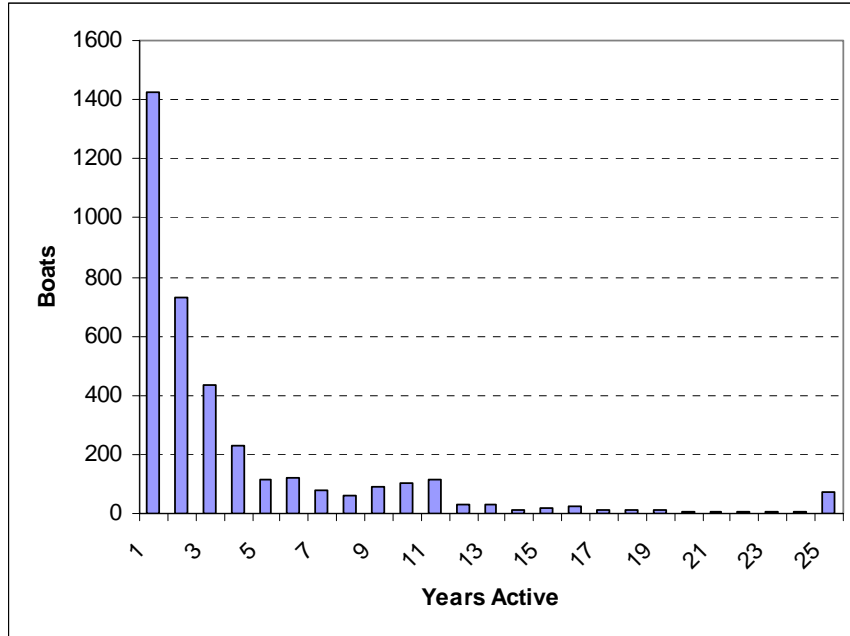$$S(t) = 1 - F(t) = \text{Prob}(t > T). \hspace{3cm} \text{(Eq. 2)}$$

Figure 2.   Duration of participation in California commercial salmon fishery by boats with uninterrupted participation histories (n=3779).
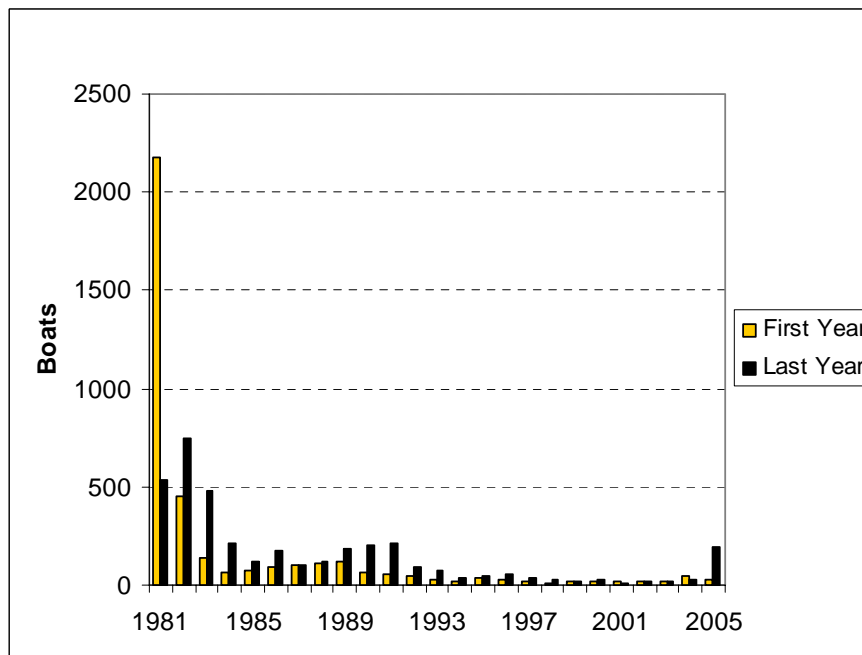

Figure 3. First and last years of reported salmon landings by boats with uninterrupted participation histories (n = 3779).

Correspondingly, we are able to compute the instantaneous probability of failure, given that a vessel has been in operation at least as long as *t*. This probability is referred to as the *hazard rate*, giving rise to the *hazard function*

$$H(t) = \lim_{\Delta \to 0} \text{Prob}(t \leq T \leq t+\Delta \mid T \geq t) \div \Delta \; = \; f(T|\theta) \div S(t). \tag{Eq. 3}$$

The hazard function is important as an aid for interpreting the results of our econometric investigations because the covariate coefficients that we wish to characterize report the response of the hazard rate to changes in the various covariates. In the case of the exponential pdf

$$f^{\text{Exp}}(T|\theta) \equiv \lambda \exp(-\lambda T), \tag{Eq. 4}$$

the survival function assumes the simple form

$$S^{\text{Exp}}(T|\theta) \equiv \exp(-\lambda T), \tag{Eq. 5}$$

giving rise to a hazard function

$$H^{\text{Exp}}(T) \equiv \lambda, \tag{Eq. 6}$$

which the reader will note is time-independent. This feature is the limiting property of the exponential model in duration analysis. As a tool for investigating survival of fishing enterprises, it predicts that the instantaneous rate of failure is independent of the duration of the vessel. As an alternative, the Weibull pdf

$$f^{\text{Wei}}(T|\theta) \equiv \alpha \, T^{\alpha-1} \, \lambda \, \exp(-\lambda T^{\alpha}), \tag{Eq. 7}$$

gives rise to a survival function with form

$$S^{\text{Wei}}(T|\theta) \equiv \exp(-\lambda T^{\alpha}), \tag{Eq. 8}$$

which, in turn, gives rise to the hazard function

$$H^{\text{Wei}}(T) \equiv \alpha \, T^{\alpha-1}. \tag{Eq. 9}$$

In the Weibull context, the parameter α is seen to play a crucial role in the analysis. A value of α exceeding one implies that the hazard function is increasing over time and a value of α less than one implies that the hazard function is decreasing over time. Importantly, when α equals one the Weibull model reduces to the Exponential model, which is nested as a special case among the family of Weibull distributions. Consequently, much interest in empirical applications of the Weibull pdf centers around the scale and location of the posterior distribution for α.

To build a model in which we include covariates – which, for want of a better term we call a *regression model*—employing either the Exponential or Weibull pdfs, we first allow the parameter λ to vary across vessels and focus attention on the vector $\lambda \equiv (\lambda_1, \lambda_2, .., \lambda_N)'$. Second, we parameterize each of the vessel-specific elements of λ in the vector $\lambda \equiv (\exp(x_1'\beta), \exp(x_2'\beta), .., \exp(x_N'\beta))'$, where $x_1 \equiv (x_{11}, x_{12}, \ldots, x_{1K})'$, $x_2 \equiv (x_{21}, x_{22}, \ldots, x_{2K})'$, .., and $x_N \equiv (x_{N1}, x_{N2}, \ldots, x_{NK})'$ are K-vectors of vessel-specific covariates and $\beta \equiv (\beta_1, \beta_2, .., \beta_K)'$ denotes a K-vector of unobserved 'regression' coefficients that is assumed to be the same across the N vessels in the industry.

To implement these models, we form a prior pdf over the parameters $\pi(\theta)$, where $\theta \equiv (\beta_1, \beta_2, .., \beta_K)'$ in the case of the Exponential specification and $\theta \equiv (\beta_1, \beta_2, .., \beta_K, \alpha)'$ in the case of the Weibull specification. We form the likelihood $f(\mathbf{y}|\theta)$ for the observed duration data $\mathbf{y} \equiv (y_1, y_2, .., y_N)'$ and study the posterior distribution function for the parameters

$$\pi(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta)\,\pi(\theta), \tag{Eq. 10}$$

where '$\propto$' denotes 'is proportional to.' An important feature of duration studies is that some of the duration observations $\mathbf{y} \equiv (y_1, y_2, .., y_N)'$ will be *censored*. Specifically, if $\mathbf{t} \equiv (t_1, t_2, .., t_N)'$ denote the survival times of the vessels in question and τ denotes the endpoint of the study, then we observe $\mathbf{y} \equiv (\min(t_1,\tau), \min(t_2,\tau), .., \min(t_N,\tau))'$. To formalize the likelihood function in the presence of such censoring, a useful construct is the vector of binary indicators $\mathbf{v} \equiv (v_1, v_2, .., v_N)'$, where, for i = 1, 2, .., N, $v_i = 1$ if $t_i \le \tau$, otherwise $v_i = 0$. Accordingly, the likelihood corresponding to a set of observed durations is

$$f(\mathbf{y}|\theta) \equiv \prod_i f(y_i|\theta)^{v_i}\, S(y_i|\theta)^{(1-v_i)}. \tag{Eq.11}$$

No conjugate priors exist for the either the Exponential or the Weibull regression models. The prior specifications advocated by Ibrahim *et al.* [3] are the multivariate-Normal density $f^{MN}(\beta|\mathbf{E_{\beta o}},\mathbf{C_{\beta o}}) \propto \exp\{-.5(\beta-\mathbf{E_{\beta o}})'\mathbf{C_{\beta o}}^{-1}(\beta-\mathbf{E_{\beta o}})\}$ for the covariate coefficients and a Gamma density $f^G(\alpha|\gamma_o,\kappa_o) \propto \alpha^{(\gamma o-1)} \exp(-\kappa_o\alpha)$ for the scale parameter. Accordingly, by defining $d = \sum_i v_i$, the posterior pdf corresponding to the Exponential specification has the form

$$\pi(\theta|\mathbf{y}) \propto \exp\{\textstyle\sum_i(v_i\mathbf{x_i}'\beta)-y_i\exp(\mathbf{x_i}'\beta)-.5(\beta-\mathbf{E_{\beta o}})'\mathbf{C_{\beta o}}^{-1}(\beta-\mathbf{E_{\beta o}})\} \tag{Eq.12}$$

and the posterior pdf corresponding to the Weibull specification has the form

$$\pi(\theta|\mathbf{y}) \propto \alpha^{(\gamma o+d-1)}\exp\{\textstyle\sum_i\{(v_i\,\mathbf{x_i}'\beta)+v_i(\alpha-1)\log(y_i)-y_i^{\alpha}\exp(\mathbf{x_i}'\beta)\}-\kappa_o\alpha-.5(\beta-\mathbf{E_{\beta o}})'\mathbf{C_{\beta o}}^{-1}(\beta-\mathbf{E_{\beta o}})\} \tag{Eq.13}$$

An important aspect of both specifications is that the fully conditional distributions characterizing the joint posteriors do not have known integrating constants. Consequently, in order to estimate the components of θ over which these posteriors are defined requires that we implement a Markov Chain Monte Carlo (MCMC) method. Ibrahim *et al.* [3] suggest exploiting the log-concavity of both of the posterior pdfs and employ the adaptive rejection sampling algorithm of Gilks and Wild [4]. Alternatively, we find that simple Metropolis-Hastings algorithms, employing a multivariate-Normal proposal density for β and a Gamma proposal density for α, work well. In each case these algorithms were run for a total of 100,000 iterations and convergence was assessed by examining the behaviour of a Chi-square statistic formed from the differences in the posterior means of the parameters obtained from the second and fourth quartiles of the sample space. In the case of the Exponential pdf convergence is quite rapid. After about 2,000 iterations posterior mean estimates are indistinguishable to two decimal places. Here we employ a target acceptance rate of the Metropolis steps of 50%. In the case of the Weibull distribution convergence was slightly problematic. Convergence was achieved after about 22,000 iterations. In this case it was found that a target acceptance rate of 75% works well. In addition, the algorithm appeared to work more efficiently when a large start value (in excess of ten) was used for the scale parameter α. In both cases inferences are derived from the second 50,000-observation samples in the 100,000-iteration chains and are obtained by using diffuse priors setting $\kappa_o = 0$ and by setting the K-dimensional square matrix $\mathbf{C_{\beta o}}^{-1}$ equal to the K-dimensional square null matrix.

RESULTS

Results of the two estimations are reported in table 2, with 95% highest posterior density intervals reported in parentheses. In assessing the signs of the effects it is important to re-emphasize that the coefficients report the impacts of changes in the respective covariates on the hazard function, or a monotonic transformation of it. We consider the effects derived from the Exponential specification before turning to those derived from the Weibull model.

Generally speaking, the larger the boat, the larger the average revenue over the period of employment and the larger the salmon share in total revenues of each boat, the lower the instantaneous probability of 'failure.' In contrast the larger the average number of ports visited, the greater the instantaneous rate of departure. These findings are generally what we would expect in a survival study of the establishments comprising the fishery. Larger boats may be more capable of fishing in widely dispersed fisheries, making them more profitable, at the margin. Larger boats also require more capital investment, making owners less inclined to exit in the face of unforeseen shortfalls. Similarly, in the absence of cost information, average revenues are a proxy for the average profitability of the vessels in question. To the extent this proxy is a good one, one expects to observe higher revenues reducing the hazard. Average salmon share of total revenue represents a level of commitment to the fishery in question relative to other potentially gainful alternatives, thus we expect the sign of the coefficient of the salmon-share variable to also be positive. In contrast, we find that higher number of ports visited per year increases the probability of departure, which may reflect high search costs and possibly lack of success in chosen locations. We note that this result is in contrast to Smith's [1] study of the California urchin fishery, which concludes that a higher number of ports visited decreased the instantaneous probability of exit, presumably reflecting greater knowledge or ambition on the part of more mobile divers.

Table 2. Empirical results.

Effect	Exponential			Weibull		
Boat Length (ft)	(-2.01)	-0.80	(0.37)	(-2.67)	-1.05	(0.31)
Avg. Ports per Year with Salmon Landings	(-0.08)	0.42	(0.97)	(0.00)	0.56	(1.39)
Avg. Salmon Rev. per Year (yr 2000 USD)	(-10.06)	-8.95	(-7.90)	(-17.32)	-11.36	(-6.09)
Share of Salmon in Total Boat Revenue	(-0.95)	-0.85	(-.76)	(-1.63)	-1.08	(-0.57)
Constant	(-0.64)	-0.50	(-0.34)	(-1.11)	-0.67	(-0.25)
Scale Parameter α				(0.72)	1.29	(1.75)

Turning to the Weibull results, the reader will note that these results mimic quite closely those obtained from the simpler Exponential specification. Significantly, the Weibull results exhibit the same patterns of significance and signs. The scale parameter, α, which distinguishes the Weibull from the Exponential hazard function, has a posterior mean of 1.29 and a considerable proportion of its mass appears to lie above the value 1.00. A straight-forward calculation from the output of the Weibull estimation reveals that, in fact, 94% of the posterior mass lies above this critical value. This leads to two conclusions. First, it suggests that the hazard is upward sloping in time. That is, the instantaneous rate of failure is positively related to duration; the longer a vessel has been involved in the fishery the more likely it is to exit. Second, short of implementing a formal model comparison along the lines of Chib and Jeliazkov [5], we can safely conclude that the Weibull model, as opposed to the simpler Exponential model, appears to be preferred by the data. In this context an interesting question arising is the magnitude of any difference in the predictions derived from the two models. The natural context within which to investigate this issue is

the survival function predictions of the respective formulations. We report these predictions graphically in figure 4. Interestingly, the two survival functions derived from the independent algorithms are almost identical. Hence, at least with respect to the data employed in this analysis, we can safely conclude that the use of the Exponential model to predict survival rates among fishers in the California salmon fishery does not appear to distort model results.
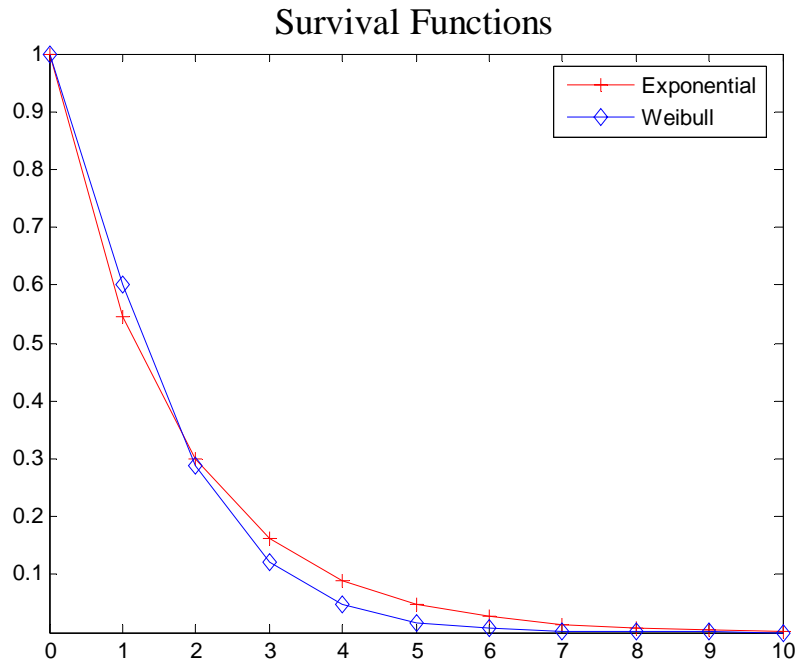

Figure 4. Survival functions inferred from Exponential and Weibull model estimations.

SUMMARY AND CONCLUSIONS

In this paper we introduce two of the fundamental specifications of survival analysis and apply them to data on the California salmon fishery using a Bayesian MCMC simulator. In so doing we extend the excellent work of Smith [1]. The empirical results are intuitive and do not appear to be greatly affected by choice of specification. Whether this conclusion remains robust to other choices of specification, remains to be seen.

Because fishers' participation decisions reflect the financial condition of the fleet and strongly influence the efficacy of measures designed to protect the fish stock, a better understanding of these decisions can contribute directly to management. In the California commercial salmon fishery, the number of boats active in the fishery has stabilized after a long period of steady decline. While we have not yet explored formal predictions of fleet dynamics based on our econometric results, the models presented here appear to have significant explanatory power, providing strong evidence for what intuition would suggest: boats that generate more revenue, have greater capability to pursue other fisheries outside of the salmon season, and that concentrate more on salmon during the salmon season, are less likely to leave the salmon fishery.

Given the importance of effort prediction in the development of regulations for this fishery, this kind of information is potentially quite useful to managers.

REFERENCES

[1] Smith, M. D, 2004, Limited-Entry Licensing: Insights from A Duration Model, *American Journal of Agricultural Economics*, 86:3, pp. 605-618.

[2] Greene, W.H., 1997, *Econometric Analysis*, Upper Saddle River, NJ: Prentice Hall.

[3] Ibrahim, J.G., M.-H. Chen, and D. Sinha, 2001, *Bayesian Survival Analysis*, New York: Springer.

[4] Gilks, W.R., and P. Wild, 1992, Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics* 41:337-348.

[5] Chib, S. and I. Jeliazkov, 2001, Marginal Likelihood From the Metropolis-Hastings Output, *Journal of the American Statistical Association*, 96, pp. 270-81.