Reference: An Introduction to Survival Analysis Using Stata (3rd ed.) by M. Cleves, R.G. Gutierrez, W. Gould, Y.V. Marchenko.

Some Basics of Survival Analysis

Continuous Time Duration Models

The conditional probability approach to duration modeling can be exemplified with the exponential distribution:

$$f(t;\lambda) = \lambda \exp(-\lambda t)$$

with $E(t) = \lambda^{-1}$. Therefore, we may parametrize $E(t|x) = \lambda^{-1}$ with $\lambda = \lambda(t|x) = \exp(x'\beta)$ which implies the conditional probability model

$$f(t|x;\beta) = \exp(x'\beta)\exp(-\exp(x'\beta)t)$$

Some observations may wind up being censored either by subjects dropping out or not finishing a spell due to the finite length of time of a duration study.

Some notation:

Let $f(t)$ be the probability density function of t, the time to a single spell event, $t \geq 0$. Furthermore let

$$F(t) = P(T \leq t)$$

(sometimes called the Failure Function) be the cumulative density function of t. This function specifies the probability that a spell will be of t or less duration.

The surviver function shows the probability of surviving past time t which is defined as the complement of the cumulative density function, namely,

$$S(t) = 1 - F(t) = P(T > t).$$

It follows that

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}.$$

For any $h > 0$, the probability of exiting the spell during the interval $(t, t+h)$, given survival up to time t, can be written as $P(t \leq T < t+h \mid T > t)$. The Hazard Function is then defined as

$$\lambda(t) = \lim_{h \to 0} \frac{P(t \leq T < t+h \mid T \geq t)}{h}$$

It can be shown that

$$\lambda(t) = \frac{f(t)}{s(t)} .$$

Hence, the Hazard Function is the limit of the probability that the spell is completed during the interval $(t, t+h)$, given that the spell has not been completed before time t, for the limit $h \to 0$. Thus the Hazard function is the instantaneous exit rate per unit of time

As it turns out, the exponential distribution has a constant hazard function $\lambda(t|X) = \lambda(X'\beta)$ for constant X.

In contrast consider the Weibull conditional probability model for t

$$f(t|X) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^{\alpha})$$

where $\lambda = \exp(X'\beta)$. The Hazard Function for this model is not constant for given X. It is

$$\lambda(t|X) = \lambda\alpha t^{\alpha-1} = \exp(X'\beta)t^{\alpha-1}.$$

Notice that the Weibull distribution collapses to the exponential distribution with $\alpha = 0$. For several Weibull Hazard Functions see Figure 8.5 in the W&S textbook. They are either monotonically increasing or decreasing for given values of α.

For the Hazard Functions of the Gompertz, log-logistic, and log-normal distributions see equations (8.25), (8.26), and (8.27). The Gompertz distribution only offers monotonic (increasing or decreasing) Hazard Functions. In contrast, the log-logistic and log-normal distributions offer non-monotonic Hazard Functions.

For graphs of the various Hazard Functions of these distributions see the online STATA 14 manual on Survival Analysis, p. 246.

http://www.stata.com/manuals14/st.pdf.

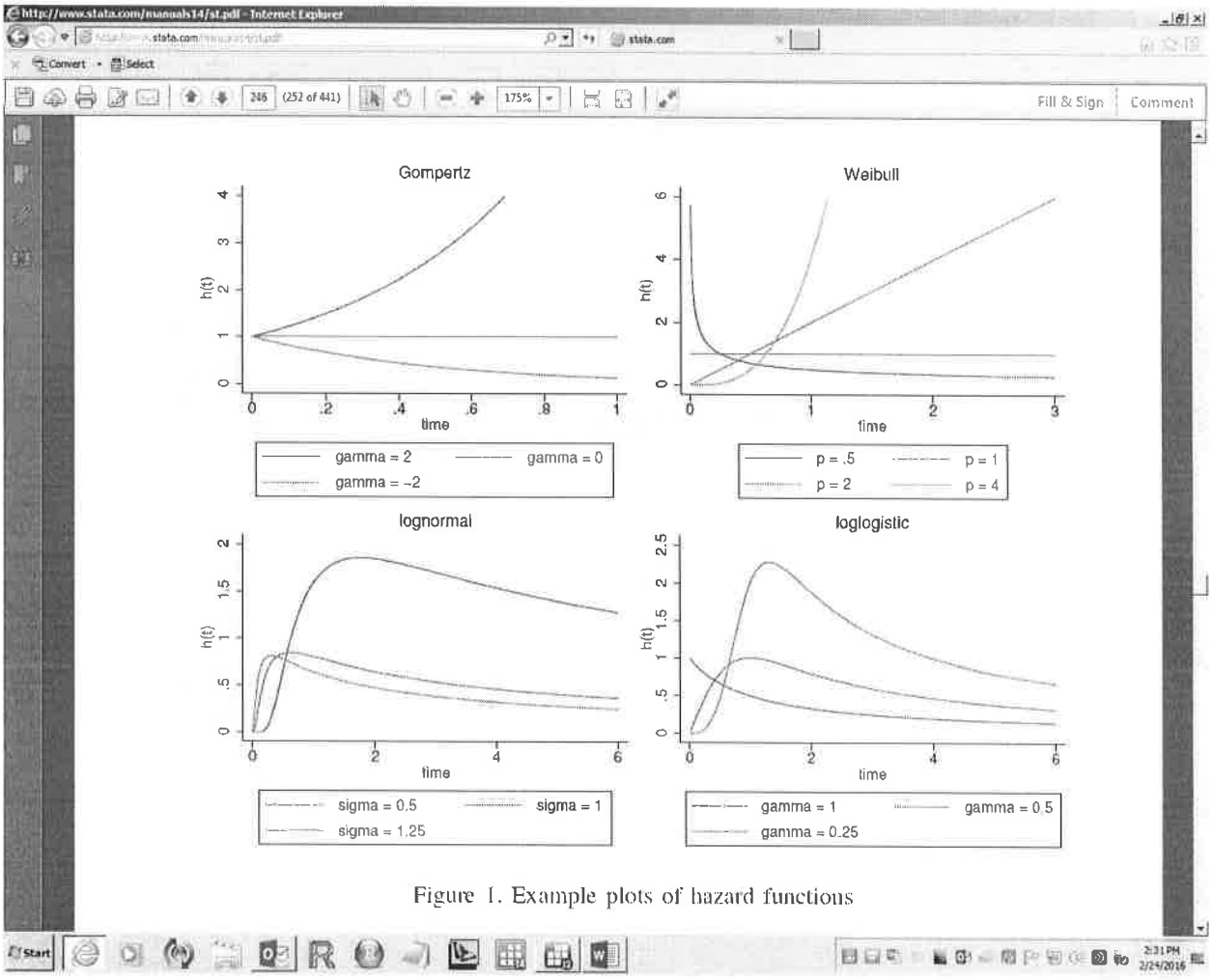

Figure 1. Example plots of hazard functions

Reference: An Introduction to Survival Analysis Using Stata (3rd ed.)
by M. Cleves, R.G. Gutierrez, W. Gould, Y.V. Marchenko

Some Notes on Survival Models

Parametric models for Duration data are written in one of two ways:

The AFT (Accelerated Failure Time) metric (Log-time)

$$\ln(t_j) = X_j'\beta + \varepsilon_j$$

Proportional

and the Hazard metric (= log relative-hazard)

$$h(t|X_j) = h_0(t)\exp(X_j'\beta),$$

where $h_0(t)$ = some function (like exponential, weibull, etc.)

The Stata syntax:

1. Log-time (AFT) parametrization

a. exponential (notice)
   streg, dist(exponential) time

b. weibull (notice)
   streg, dist(weibull) time
   (Note: If $p=1 \Rightarrow$ exponential)

c. lognormal
   streg, dist(lognormal)
d. loglogistic
   streg, dist(loglogistic)
e. gamma
   streg, dist(gamma)

2. Proportional Hazard (= log relative-hazard) parametrization
   a. exponential
      streg, dist(exponential)
   b. weibull
      streg, dist(weibull)
   c. Gompertz
      streg, dist(gompertz)

Note: All of these models must have a preceding stset command that tells stata what the duration variable is and what the "censored" variable is (1 = censored observation, 0 = uncensored observation).

Note: The exponential and Weibull models are available in <u>both</u> metrics. In the case of the AFT parametrization the default presentation of the coefficients

is in the time ratio form whereas for the PH metric the default presentation is in the hazard ratio form. (see pp. 273-274 in W&B)

Note: The AFT parametrization focuses on the mean time to failure while the PH parametrization focuses on the hazard ratio.

Note: If one is concerned about unobserved heterogeneity (frailty) in the duration data one is analyzing, one can use the generalized Gamma distribution as W&B do in Table 8.6 (p. 280). For a discussion of this topic see section 8.2.7 in W&B.

Note: In the case of having a simple duration model with only indicator variables, the empirical hazard functions can be very informative. See for example Figure 8.7 in the W&B text book.

Note: When choosing between the non-nested models within a particular metric (AFT or PH) one can choose the model that provides the minimum AIC or SIC information criterion. For discussion of the information criteria see pp. 90-91 in W & B.

The Cox Proportional Hazards Model (stcox)

Again the Proportional Hazards models are of the form:

$$h(t|x_j) = h_0(t)\exp(x_j'\beta).$$

In the case of the Cox Proportional Hazards model the base hazard rate is left unparametrized. In other words, $h_0(t)$ is modeled as a non-parametric function.

Notice, in the case of the exponential model $h_0(t) = \exp(a)$ and we wind up fitting the parameters (a, β). In fitting a Weibull model we choose

$$h_0(t) = pt^{p-1}\exp(a)$$

with $\rho = 1$ specializing the Weibull model back into the exponential model.

If we assume

$$h_0(t) = \exp(\alpha)\exp(\gamma t)$$

then we obtain the Gompertz model and we wind up estimating the parameter set (α, γ, β). If $\gamma = 0$ then we return to the exponential model.

At any rate all of the models produce results that are directly comparable to those produced by the Cox regression. In all of these models, $X\beta$ is the log relative-hazard and the elements of β have the standard interpretation mean $\exp(\beta_i)$ is the Hazard Ratio for the i-th coefficient (variable). Also parametric models estimates of the ancillary parameters, and from that you can obtain the predicted baseline hazard function $h_0(t)$.

Parametric Models

The direct comparability of the PH exponential, Weibull, and Gompertz

with the Cox model is an attractive feature of the Cox model. That is, when you estimate one or more of the Parametric PH models you can always (and it is prudent to do so) compare the estimated Hazard ratios of the Parametric PH models with the Hazard Ratios obtained from the Cox PH model with the same explanatory variables. If the Hazard Ratios for a chosen PH parametric model are "close" to those of the Cox PH model, then this is evidence that the baseline Hazard function, $h_0(t)$, of the Parametric PH model is probably pretty good. On the other hand, if the Hazard Ratios are quite different, then that would hint that the underlying baseline Hazard function of the Parametric PH model may be misparametrized.