An introduction to Survival Analysis

Maarten L. Buis
Department of Social Research Methodology
Vrije Universiteit Amsterdam
m.buis@fsw.vu.nl

April 2, 2006

1   Introduction

This paper will give a short introduction to survival analysis. Other names for survival analysis are event history analysis, duration analysis or transition analysis. The techniques of survival analysis are used to analyze things like how long people stay unemployed, how long a cancer patient lives, how long it takes before a lightbulb breaks, etc. What these examples have in common is that they all want to know how long it takes before a certain event (finding a job, dying, breaking of a lightbulb) happens. Table 1 shows the (fictitious) example that will be used in this paper. It contains for ten countries the time it took each of them to ratify a treaty. We may think of this as countries that every year run a 'risk'[1] of ratifying that year. We would expect that the duration is short when this risk is high and long when the risk is low. That risk of ratifying may change from year to year or may be different for countries with different characteristics/explanatory variables. With survival analysis we can estimate the impact of the explanatory variables on the risk of ratifying. Three techniques will be discussed: the non-parametric, the parametric and the semi-parametric.

2   Probability distributions

The risk of ratifying needs to be more specifically defined before the impact of explanatory variables on the risk of ratifying can be discussed. The risk of ratifying is a probability. This probability can be specified in multiple ways, for example: the probability that a country takes longer then 5 years to ratify, or the probability that a country ratifies the treaty after exactly 11 years. These probabilities are of course interrelated, since they are just different representations of the same process. Survival analysis estimates the probability of ratifying and how it changes over time and for different values of the explanatory variables. If we obtain for every possible duration the probability of ratifying than we have obtained the probability distribution. So, the way the probability of ratifying changes over time is captured by the probability distribution. Every probability distribution can be presented in several different ways; for instance the probability that ratification takes longer than some specified amount

---

[1] The word risk is used here because this is the common terminology in survival analysis. Many of the terms are derived from the application of these techniques in medical science where it is used to explain how long patients live after getting a certain illness or receiving a certain treatment. This explains the negative or positive connotations of many of the terms used in survival analysis.

Table 1: example

country	duration
West-Germany	3
Netherlands	7
Canada	12
India	13
USA	15
Thailand	15
USSR	16
Brazil	17
UK	19
Niger	20

Table 2: the survivor function

time	no. countries not ratified	$S(t)$
0	10	10/10=1
3	9	9/10
7	8	8/10
12	7	7/10
13	6	6/10
15	4	4/10
16	3	3/10
17	2	2/10
19	1	1/10
20	0	0/10=0

of time or the probability that ratification occurs at some specified point in time. In this section we will look at four representations of the probability distribution.

The first way to present the probability distribution is obtained by looking at the probability that ratification takes longer than some specified duration. Table 2 presents what happens at different points in time. If we want to know what the probability is that ratification takes longer than 3 years, than all we have to do is look at the number of countries that have "survived" (i.e. not ratified) after 3 years and divide that by the total number of countries. After three years nine countries survived and the total number of countries is ten, so the probability of surviving after three years is 90%. If all these probabilities are graphed against time, as is done in figure 1, than we get the survivor function, $S(t)$.

This Survivor function has three notable features: the first feature is that the graph starts at 1. The reason for this is that at the moment that the country is first able to ratify the treaty ($t = 0$) its probability of not having done so before is by definition equal to 1. The second feature is that the graph does not rise. The reason for this is that those countries who run a risk of not ratifying at $t = 20$ must not have ratified at $t = 10$. Consequently the probability of not ratifying at $t = 20$ can not exceed the probability of not ratifying at $t = 10$. The third feature is that the graph is a step-function. This does not mean that we believe that the 'real' survivor function has this shape, far from it. All that it means is that this is,

Figure 1: the survivor function

given our data, our best estimate of the survivor function. For example, the probability of surviving after five years is the same as the probability of surviving after three years or four years or six years: nine out of ten countries survived, so the probability of surviving is 90% for all these durations.

The survivor function is closely related to another way of representing the same distribution that is commonly used in other statistical techniques, the cumulative probability function ($F(t)$). The cumulative probability function gives for every time t the probability that the duration is less than or equal to t. The survivor function must be the complement to one of the cumulative probability function, since the probability that a country ratifies before, on or after t is necessarily 1 (we assume that all countries will eventually ratify so there are no other possibilities). So the relationship between the survivor function and the cumulative probability function is: $S(t) = 1 - F(t)$. The cumulative probability function is closely related to another way of presenting the probability distribution: the probability density function, $f(t)$. The probability density function is the first derivative of the cumulative probability function. The familiar bell shaped curve of the normal distribution is a probability density function. For a given interval the surface underneath the curve gives the probability that the time it takes to ratify the treaty falls within that interval. Alternatively, it can be thought of as the instantaneous probability of ratifying at time t. The fourth way of representing the probability distribution looks at the probability of ratifying at time t for countries that have not yet ratified, since the countries that have already ratified are clearly no longer at risk of ratifying and the probability density function does not take this into account. So we want to know the probability of ratification conditional on the country surviving to time t. The probability of surviving to time t is the survivor function, as was shown before. We can make the probability conditional on having survived to time t by dividing the probability density function by the survivor function. This is called the hazard function. This measure comes closer to the notion of the instantaneous probability of ratifying than the probability density function. However, the hazard is strictly speaking not a probability. For one thing, the real instantaneous probability of ratifying is necessarily zero.[2] A more correct interpretation of

---

[2]Time is a continuous variable. Being a continuous variable means that time can take an infinite number of specific values. In order to calculate the probability that a continuous variable takes on any one specific value one needs to divided by infinity, and that is zero.

the hazard is the number of times a country would be expected to ratify if the risk of ratifying would remain constant for one unit of time. So if a hazard of 0.1 is found and time is measured in years than a country will on average ratify 0.1 times in the next year if the hazard remains constant during that year. Still, this is close enough to continue with the interpretation of the hazard as the instantaneous probability of ratifying.

To sum up, we are interested in the probability of ratifying at every point in time (and in how these probabilities differ for different kinds of countries, but this will be discussed later in this paper). The set of probabilities of ratifying at every point in time is the probability distribution. Four different ways of presenting the probability distribution have been discussed. The first way is the survival function. This gives at every point in time the probability that ratification takes longer then that point in time. The second way is the cumulative probability function. This gives for every point in time the probability that ratification takes less then that point in time. The third way is the probability density function. This gives for every point in time the probability that ratification occurs on that point in time, only it does not take into account that those countries that have already ratified before are no longer at risk of ratifying. The fourth way, the hazard function, takes care of this omission. It gives for every point in time the probability that ratification occurs on that point in time if the countries has not ratified before.

3   Non-parametric analysis

Several ways of presenting a distribution have been discussed in the previous paragraph. Now it is time to investigate how these can be estimated. Estimating the distribution of the dependent variable without making assumptions about its shape is an important first step in analyzing a dataset. Given the importance of the distribution of the dependent variable it is valuable to "let the data speak for itself" first. Estimating the probabilities without making any assumptions on its shape is called non-parametric analysis. The function used to represent the distribution is the Survivor function. Remember that the Survivor function gives the probability that ratifying takes longer than a certain period of time. The example in table 2 and figure 1 gives the way to estimate the survivor function when there are no censored observations. The survivor function was calculated by dividing the number of survivors by the total number of countries for every time. A censored observation is a country that has not yet ratified at the time the study ended. The reason why the method of the previous paragraph does not work when there are censored observation is best explained by extending our previous example. This extension is presented in table 3. We now assume that India, Thailand, the USSR, the UK and Niger are censored. This is indicated by the value zero for the variable ratified. So we know that India has not ratified before $t = 13$, but we do not know when India will ratify. Suppose that the survivor function of the dataset presented in table 3 has to be calculated. If we wanted to calculate the probability of survival past for example $t = 15$ we would run into trouble. We know for certain that four countries have ratified and five countries have not ratified at $t = 15$. However we do not know whether India has ratified by then. All we know about India is that India has taken longer then 13 years to ratify. If India ratified at $t = 14$ then the probability of survival past $t = 15$ would be 4/10, if India ratified at say $t = 16$ then the probability of survival past $t = 15$ would be 5/10. Since we do not know when India actually ratified we do not know which of these two to choose. So the method used in the previous paragraph of estimating the survivor function

4

Table 3: example with censored observations

country	duration	ratified
West-Germany	3	1
Netherlands	7	1
Canada	12	1
India	13	0
USA	15	1
Thailand	16	1
USSR	16	0
Brazil	17	1
UK	19	0
Niger	20	0

Table 4: the events at each point in time

time	no. at risk	no. ratified	no. censored
3	10	1	0
7	9	1	0
12	8	1	0
13	7	0	1
15	6	1	0
16	5	1	1
17	3	1	0
19	2	0	1
20	1	0	1

by calculating the proportion of countries that have not yet ratified does not work when some of the observations are censored, since there are times when we do not know whether the censored observations have ratified or not.

The table 4 summarizes what happens at each point in time in the data. At time $t = 3$ all the ten countries were at risk of ratifying, but at that instant only one failed (West Germany). At the next time, $t = 7$, nine countries were at risk of ratifying, and that time one of the nine ratified (the Netherlands). At $t = 13$ one country, India, was censored while no country ratified. After that time India was no longer at risk of being observed to ratify, so the number of countries at risk after $t = 13$ is reduced by one.

The technique used to estimate the survivor function when censoring is present is called the Kaplan-Meier or the product-limit estimator. It uses the principle that although the probability of surviving past $t = 15$ cannot be directly calculated by dividing the number of survivors with the total number of countries, we can calculate the probability of surviving the interval $t = 13$ till $t = 15$. During this interval there were six countries at risk of ratifying of which one ratified, so the probability of surviving that interval is 5/6. We can think of time as a number of intervals between every point in time that at least one country either ratified or censored. The probability of surviving all such interval can be calculated. For instance:

Table 5: probability of surviving interval

interval	no. at risk	no. ratified	no. censored	p
0-3	10	1	0	9/10
3-7	9	1	0	8/9
7-12	8	1	0	7/8
12-13	7	0	1	7/7=1
13-15	6	1	1	5/6
15-16	4	0	1	5/5=1
16-17	3	1	0	2/3
17-19	2	0	1	2/2=1
19-20	1	0	1	1/1=1

- The probability of surviving (not ratifying) the interval $t = 0$ till $t = 3$ is 9/10, since nine out of ten countries survived beyond this interval.

- The probability of surviving the interval $t = 12$ till $t = 13$ is 7/7=1, since seven countries were at risk during this interval, of which none ratified. All that happens is that the number of countries at risk after $t = 13$ is reduced by one.

- The probability of surviving the interval $t = 15$ till $t = 16$ is a bit more difficult. At the end of this interval one country ratified (Thailand) and one country censored (USSR). In order to know the number of countries at risk, we have to make an assumption as to whether the country that censored at time $t = 10$ was still at risk of ratifying at that time. It is common to assume that censoring occurs just a little bit later than ratifying, so that the USSR (the censored country) is still at risk when Thailand ratified. This means that six countries were at risk, and one ratified. As a result the probability is estimated at 5/6.

- And so on.

These probabilities (called p) are added to the table 5.

These probabilities can be used to calculate the survival function. The survival function gives the probability of surviving past every point in time. For instance, the probability of surviving past $t = 7$ equals the probability of surviving the intervals $t = 0$ till $t = 3$ and $t = 3$ till $t = 7$. This is the product of the probabilities of surviving each interval, that is $9/10 \cdot 8/9$ = 4/5. Similarly, the probability of surviving past $t = 12$ equals the probability of surviving the intervals $t = 0$ till $t = 3$, $t = 3$ till $t = 7$ and $t = 7$ till $t = 12$, that is $9/10 \cdot 8/9 \cdot 7/8 =$ 7/10. Thus, the estimate of the Survival function is the running product of the probabilities of surviving the constituent intervals. This can be added to the table, which is done in table 6 where its called $\hat{S}(t)$, or graphed as is done in figure 2. The red ticks in the graphs mark the times when an observation was censored.

This method can easily be extended to compare two groups within the dataset, for instance rich and poor countries. In that case one would calculate separate survivor functions for both groups and graph them. This can be illustrated by extending our example. In table 7 the dataset is split between rich and poor countries. In table 8 the survival functions are calculated using the same method as before, only now different survival functions are calculated for the different groups. These survival functions are shown in figure 3.

Table 6: the Kaplan-Meier survivor function

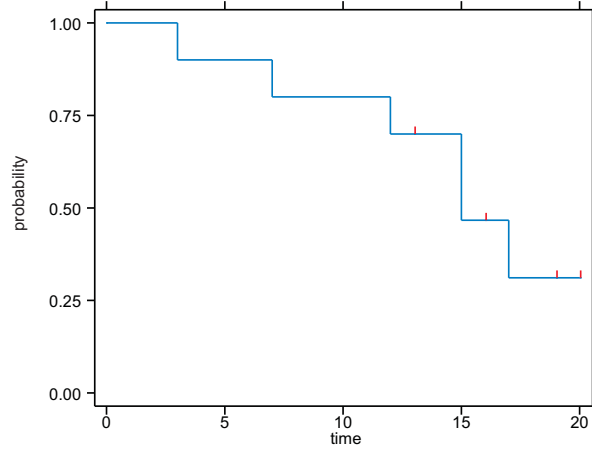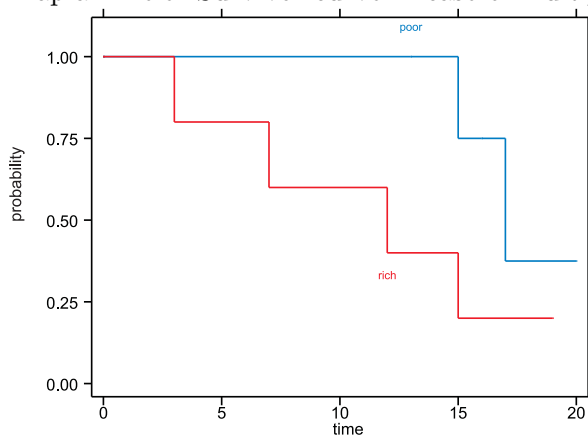| interval	no. at risk	no. ratified	no. censored	p	$\hat{S}(t)$
0-3	10	1	0	9/10	9/10
3-7	9	1	0	8/9	$9/10 \cdot 8/9 = 4/5$
7-12	8	1	0	7/8	$9/10 \cdot 8/9 \cdot 7/8 = 7/10$
12-13	7	0	1	1	$9/10 \cdot 8/9 \cdot 7/8 \cdot 1 = 7/10$
13-15	6	1	1	5/6	$9/10 \cdot 8/9 \cdot 7/8 \cdot 1 \cdot 5/6 = 7/12$
15-16	4	0	1	1	$9/10 \cdot 8/9 \cdot 7/8 \cdot 1 \cdot 5/6 \cdot 1 = 7/12$
16-17	3	1	0	2/3	$9/10 \cdot 8/9 \cdot 7/8 \cdot 1 \cdot 5/6 \cdot 1 \cdot 2/3 = 7/18$
17-19	2	0	1	1	$9/10 \cdot 8/9 \cdot 7/8 \cdot 1 \cdot 5/6 \cdot 1 \cdot 2/3 \cdot 1 = 7/18$
19-20	1	0	1	1	$9/10 \cdot 8/9 \cdot 7/8 \cdot 1 \cdot 5/6 \cdot 1 \cdot 2/3 \cdot 1 \cdot 1 = 7/18$

Figure 2: Kaplan-Meier survivor function

Table 7: multiple groups

country	duration	ratified
rich countries		
West-Germany	3	1
Netherlands	7	1
Canada	12	1
USA	15	1
UK	19	0
poor countries		
India	13	0
Thailand	16	1
USSR	16	0
Brazil	17	1
Niger	20	0

Table 8: Kaplan-Meier survivor curve in case of multiple groups

time	no. at risk	no. ratified	no. censored	p	$\hat{S}(t)$
rich countries					
3	5	1	0	4/5	4/5
7	4	1	0	3/4	$4/5 \cdot 3/4 = 3/5$
12	3	1	0	2/3	$4/5 \cdot 3/4 \cdot 2/3 = 2/5$
15	2	1	0	1/2	$4/5 \cdot 3/4 \cdot 2/3 \cdot 1/2 = 1/5$
19	1	0	1	1/1=1	$4/5 \cdot 3/4 \cdot 2/3 \cdot 1/2 \cdot 1 = 1/5$
poor countries					
13	5	0	1	5/5=1	1
16	4	1	1	3/4	$1 \cdot 3/4 = 3/4$
17	2	1	0	1/2	$1 \cdot 3/4 \cdot 1/2 = 3/8$
20	1	0	1	1/1=1	$1 \cdot 3/4 \cdot 1/2 \cdot 1 = 3/8$

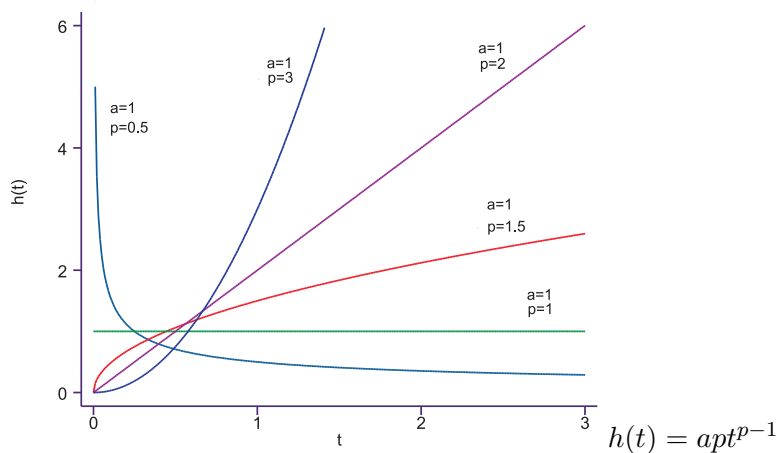Figure 3: Kaplan-Meier Survivor curve in case of multiple groups

If being a rich or a poor country has no influence on the probability of ratifying than the graphs should be more or less equal. If rich countries run a higher risk of ratifying than the risk of surviving (not having ratified) a certain period of time should be lower for rich countries than for pour countries. This means that the survivor function of rich countries should be below the survivor function of the poor countries. This is exactly what we see in figure 3. According to this (fictional) dataset rich countries are likely to ratify faster than poor countries. However each group consists of only five countries, so it could be just coincidence that the slow countries are in the poor group and the fast countries in the rich group. (This is of course still true when the dataset is larger; it will only be less likely). To test whether the observed difference is genuine or just coincidence we want to find the probability of observing the data we have observed if we assume that the two groups are the same. The probability is called the p-value and the assumption the null hypothesis. If the p-value is very small than either the null hypothesis is wrong or we have drawn a very unlikely sample. This is seen as evidence against the null hypothesis, and we reject the null hypothesis that the two groups are the same. A commonly used cut-point to decide whether the observed difference is the result of the difference between the groups or coincidence is 5%. This cut-off point is called level of significance. Tests for the difference between survival functions are the log rank test and the Wilcoxon test. The log rank test is more sensitive to differences at later points in time, while the Wilcoxon test is more sensitive in the beginning. The p-value received from the log rank test for our example data is 11.7%, while the p-value received from the Wilcoxon test is 4.8%. This would suggest that the observed difference between rich and poor countries in the beginning is the effect of a genuine difference between rich and poor countries, while this observed difference at the end could just as well be the result of coincidence.

The advantage of non-parametric analysis is that the results do not depend upon a lot of assumption (since only a small number of assumptions have been made), it just lets the data speak for itself. The disadvantage is that it can only compare a limited number of groups, so it is very difficult to see the impact of one explanatory variable while controlling for other variables. For instance if democratic countries tend to ratify faster than non-democratic countries, that might be the result of the impact of democracy on the time it takes to ratify. However it may also be the result of the fact that democracies are generally richer than non-democracies and richer countries ratify faster than poorer countries. The non-parametric techniques are not particularly good at disentangling these effects, especially when there are many of this type of effects. A second disadvantage of the non-parametric techniques is that it can only deal with qualitative explanatory variables like rich or poor countries. They cannot deal with quantitative variables like GDP per capita (because this would mean that the data has to be split in far too many groups). So, instead of looking at the impact of GDP per capita on the time it takes to ratify a treaty, the non-parametric techniques look at the difference between rich and poor countries.

4   Parametric analysis

We can deal with the disadvantages of non-parametric analysis mentioned at the end of the previous paragraph if we are willing to make assumptions about the functional form of the probability distribution and the way that the explanatory variables influence the risk of ratifying. Techniques that make both assumptions are called parametric techniques. This paragraph will discuss the two assumptions and the way the results of these models can be

Figure 4: Weibull

$$h(t) = apt^{p-1}$$

interpreted. The way in which these models are estimated will be discussed in the next paragraph.

The first assumption deals with the functional form of the probability distribution. Remember that the probability distribution summarizes how the probability of ratifying changes over time. This assumption is, for this reason, also called an assumption on time dependence. One way to represent the probability distribution is the hazard function. The hazard function can be thought of as the instantaneous probability of ratifying, conditional on not having ratified so far. When we choose the functional form of the distribution we are imposing constraints on the shapes the distribution can take, but we are not fixing it completely. For instance, the simplest functional form of the probability distribution is to assume that the hazard is constant over time. This would mean that the risk of ratifying is always the same, regardless of how long a country has been eligible to ratify. The corresponding probability distribution is the exponential model. If the risk of ratifying is constant over time, the distribution of the duration is an exponential distribution. The functional form of the exponential model is $h(t) = a$, whereby a is the constant level of risk. Parametric analysis chooses the level of a that best fits the data. Other distributions are characterized by more then one parameter, one that moves the hazard up or down, like the a in the exponential model, and one or more parameters that determine the shape, or the location of humps, if any. For instance the functional form of the Weibull model is $h(t) = a \times pt^{(p-1)}$, whereby a is the parameter that shifts the hazard up or down, p is a parameter that determines the shape of the hazard function and t is the duration. Parametric analysis now chooses the values of a and p that best fit the data. The different shapes that are possible with a Weibull model are shown in graph 4. Graphs 5 through 7 show the hazard functions of other often used models, the Gompertz, the log-logistic and the log-normal.[3] One often used model, the gamma model, is not shown, because it can have so many shapes that it would not be meaningful to show them here. These graphs show that assuming a functional form is not as restrictive as it may seem, since a wide variety of shapes are possible with only a small number of models.

These functional forms of the hazard in combination with an assumption on how explana-

---

[3]Beside each graph the functional form of the hazard function is shown to show where the parameters mentioned in the graphs return in the functional form.
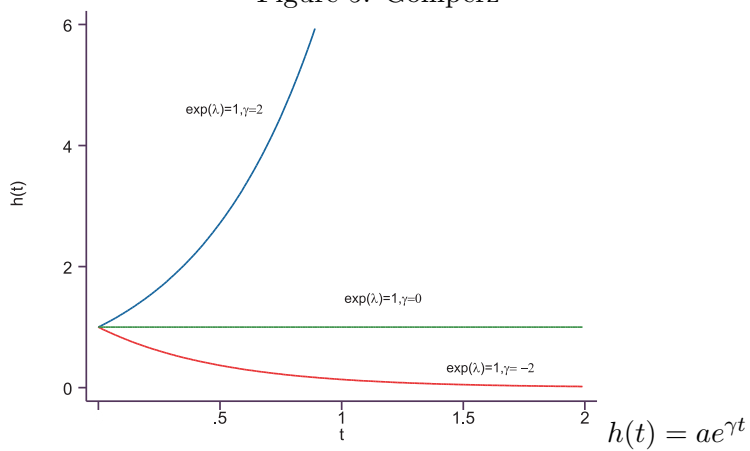
Figure 5: Gomperz

$$h(t) = ae^{\gamma t}$$

Figure 6: log-logistic

$$h(t) = \frac{\lambda^{\frac{1}{\gamma}} t^{1-\gamma}}{\gamma\left(1+(\lambda t)^{\frac{1}{\gamma}}\right)}$$

Figure 7: log-normal

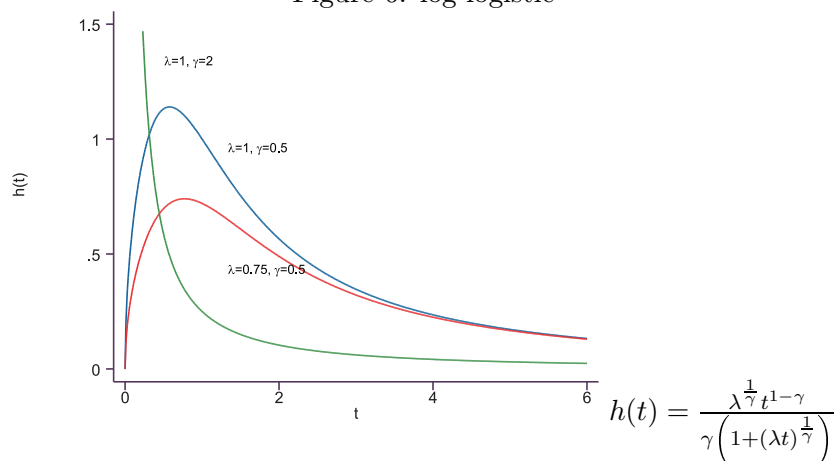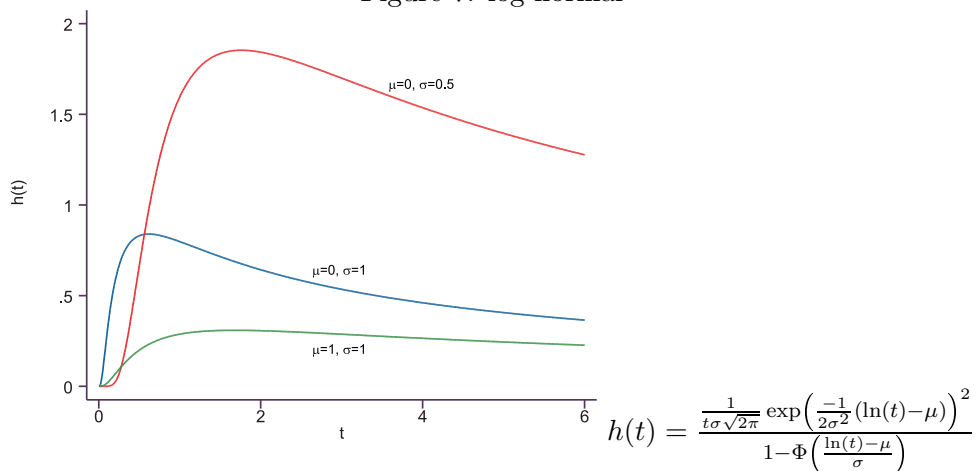$$h(t) = \frac{\frac{1}{t\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2\sigma^2}(\ln(t)-\mu)\right)^2}{1-\Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)}$$

11

tory variables influence the hazard can be used to estimate the impact of the explanatory variables. A simple assumption is the proportional hazard assumption, which can be used in the exponential, the Weibull and the Gompertz models. With the proportional hazard assumption we assume that all countries face a hazard function of the same shape, but that this hazard function is moved up or down with some fixed proportion for different groups of countries. An example of this type of assumption is if we assume that the risk of ratifying for G7 countries is always the same fraction higher or lower than the other countries. The model estimates that fraction. If that fraction is estimated to be 1.2, than the risk of ratifying for G7 countries is always 1.2 times the risk of ratifying for other countries, irrespective of the amount of time that has passed. The hazard function of all of these models have a part that determine its shape and a parameter a that moves the function up or down by some fixed proportion. So when we say that the hazard function for different groups is some fixed proportion higher or lower, we say that these different groups have different values for the parameter a. We can achieve this by replacing the parameter a with a function of the explanatory variables. The following example shows how this assumption is implemented. If we believe that the exponential is the right functional form of the hazard, we can estimate the effect of GDP on the risk of ratifying in the following way. The hazard function of the exponential distribution is $h(t) = a$, whereby a is a constant. We can replace that constant with a function of GDP, however we have to take care that the hazard can not be negative. Replacing a with $e^{\beta_0 + \beta_1 x_1}$, whereby x_1 is the GDP, and the betas the parameters[4], will do the trick. Instead of finding the value of a that fits the data best, parametric analysis now finds the values of the betas that fit the data best. The betas themselves are a bit difficult to interpret, but exponentiated betas are the ratio of the hazards for a unit change in the corresponding covariate. For instance if we find that β_1 is 0.69 then a country will run a risk of ratifying that is $e^{0.69} = 2.0$ times as large if his GDP increases with one unit (say one thousand dollars). The exponentiated parameters are called hazard ratios. Recall that we can extend the proportional hazard assumption to the Weibull and the Gompertz model. Remember that the hazard function of the Weibull is $h(t) = a \times pt^{(p-1)}$. We can again replace the parameter a with $e^{\beta_0 + \beta_1 x_1}$ and the interpretation of the betas is exactly the same. The Gompertz model can also use the proportional hazard assumption. The hazard function of the Gompertz model is $a \times e^{\gamma t}$, and the parameter a is replaced with $e^{\beta_0 + \beta_1 x_1}$.

Another assumption on how explanatory variables influence the risk of ratifying is the accelerated failure time assumption. This assumption is applicable for the exponential, the Weibull, the log-normal, the log-logistic and the gamma. Basically, this assumption assumes that every country faces a hazard curve of the same shape, only time passes by slower or faster for different types of countries. A good example is the conventional wisdom that a year for a dog is equivalent to seven years for a human. So if humans have a 75% chance of surviving past the age of 70, than dogs have a 75% chance of surviving past the age of 10. This example shows that accelerated failure time models are closely related to the survivor function. Basically, the hazard functions are rewritten to a survival function, and the survival function has the following general form: $S(t) = S_0(a \times t)$, whereby a is one of the parameters and S_0 is function which depends on the model. The exponential model can be written in such a way. If the hazard function is h(t) = a, then the survival function of that distribution is $S(t) = e^{-at}$. We can replace the parameter a with a function of the explanatory variables,

---

[4]and e is a number, approximately 2.718. It is the base of the natural logarithm, which like the number π can only be approximated.

just as with the proportional hazard model. We have to take care that $a \times t$ can not be negative, and since t is always positive, a must also be always positive. To achieve this a is again replaced with $e^{\beta_0 + \beta_1 x_1}$. The betas now have a more or less similar interpretation as in a proportional hazard model. The exponentiated beta now is not the ratio of the hazards for a one-unit change in the corresponding covariate, but the ratio of the expected survival time for a one-unit change in the corresponding covariate. So if t is the lifespan of a number of humans and dogs, and x_1 equals 1 if the subject is human and zero if it is a dog, then we expect $e^{\beta 1}$ to be 7. The exponentiated betas are called time ratios.

A problem with parametric analysis is that we have to choose a model. Ideally, theory should lead to the choice of the model. There are however some options if the theory is silent. The estimated survivor functions can be used to evaluate whether a specific distribution is appropriate for the dataset. Often one uses a manipulation of the survivor function: the cumulative hazard function, $H(t)$, which is $-\ln(S(t))$ For instance, the exponential distribution assumes a constant hazard. If the hazard function is constant than the cumulative hazard function, $H(t)$, is a straight upward sloping line. So, a graph of $-\ln\left(\hat{S}(t)\right)$ against t should yield a straight line if the distribution is indeed exponential. The conclusions that the graph of $\ln\left[-\ln\left(\hat{S}(t)\right)\right]$ against $\ln(t)$ should be straight if the distribution is a Weibull distribution, the graph of $\ln\left[\frac{1-\hat{S}(t)}{\hat{S}(t)}\right]$ against $\ln(t)$ should be straight if the distribution is log-logistic, and the conclusion that the graph of $\Phi^{-1}\left[1 - \hat{S}(t)\right]$ against $\ln(t)$ should be straight if the distribution is log-normal, can be derived in similar ways. (Blossfeld and Rohwer 1995, 199-200)

There are other ways of choosing between models. One way uses the fact that some models are just special cases of other models. For instance, the exponential model is the Weibull model when the shape parameter p equals one. So when choosing between the Weibull and the exponential, all we have to do is estimate a Weibull model and test whether p equals one. The residuals can act as a guide when choosing between models that do not have such a relationship. 'Normal' residuals can not be calculated, but pseudo-residuals can be obtained. Often used pseudo-residuals are Cox-Snell residuals. If the model fits well, than these residuals should follow a standard exponential distribution. The distribution of the actually calculated Cox-Snell residuals can be graphically evaluated. The model that produces Cox-Snell residuals that most closely resemble a standard exponential distribution is the best.

To sum up, in parametric analysis we assume that the probability distribution has a certain functional form. The functional form has one or more parameters that determine its location and/or shape. Parametric analysis finds the values of these parameters that best fit the data. Explanatory variables can be introduced by replacing one of these parameters with a function of the explanatory variables. We make an assumption on how the explanatory variables influence the risk of ratifying by doing so. If we estimate a proportional hazard model we assume that the hazard of one group is always some proportion larger or smaller then the hazard of another group. If we estimate a accelerated failure time model we assume that that every country faces a hazard curve of the same shape, only time passes by slower or faster for different groups of countries.

5   The likelihood function

We stated in the previous paragraph that parametric analysis finds the values of the parameters and betas that best fit the data. This paragraph will show how this is done. The method is called maximum likelihood. Maximum likelihood tries to find the values of the parameters that will maximize the probability of observing the data that were observed. An observation can be thought of as a random draw from a set of possible observations. If we know the probability distribution, we can calculate the probability of "drawing" the observation we have actually drawn. A dataset consisting of two observations can be thought of as two independent draws of the set of possible observations. The probability of observing the dataset is the product of the two probabilities of observing each individual observation. We assume that the probability distribution we have chosen for our parametric model is the real probability distribution. Methods to evaluate which distribution is appropriate were discussed in the previous paragraph. The probability distributions have one or more parameters, and we are interested in finding the values of these parameters that best fit the data. In other words, if we assume that the real distribution has the chosen probability distribution with a set of parameters, than we can calculate the probability of observing the data that we have observed. Maximum likelihood finds those parameters that maximize this probability, but we have to choose the probability distribution we think is applicable. In order to find the best parameters, one should first write down an expression for the probability of the data as a function of the unknown parameters. This function is called the likelihood function. After that one should find the values of the parameters that will maximize the likelihood. The likelihood function will first be discussed for datasets without censoring and explanatory variables. After that censoring and explanatory variables will be added.

The probability of observing a dataset is the product of the probabilities of observing each individual observation, as was discussed before. Because the observation is a duration and a duration is assumed to be measured on a continuum, the probability that it will take on any specific value is 0. Instead, we represent the probability of each observation by the probability density function. This results in likelihood function 1.

$$L\left(\boldsymbol{\theta}\right) = \prod_{i=1}^{n}\left[f\left(t_i|\boldsymbol{\theta}\right)\right] \tag{1}$$

Whereby L is the likelihood, and $\boldsymbol{\theta}$ a vector of parameters, like the a in the exponential or the a and the p in the Weibull. The Π means the product of all values of $f(t_i|\boldsymbol{\theta})$. This functions has to be maximized with respect to $\boldsymbol{\theta}$. This is generally done with an iterative method, which consists of trying repeatedly a number of values of the parameters until they converge to a maximum.

We can use the example from paragraphs 2 and 3 to illustrate this. In this example we want to analyze a dataset of ten countries. We start with the dataset in which all countries ratified. This dataset is repeated in table 9. If we assume that the exponential distribution is the best applicable distribution[5], then we are interested in finding the value of parameter a that best fit the data. The probability density function of an exponential distribution is $f(t) = e^{-at}$, so the likelihood function becomes: $L\left(a\right) = e^{-a3} \times e^{-a7} \times \cdots \times e^{-a20}$. The value

---

[5]A superficial inspection of the survival curve in figure 1 would suggest that the risk of ratifying increases over time. That means that the exponential distribution is probably not the most appropriate, but it is very appropriate for use as an example since it has rather simple hazard, survival and probability density functions.

Table 9: example

country	duration
West-Germany	3
Netherlands	7
Canada	12
India	13
USA	15
Thailand	15
USSR	16
Brazil	17
UK	19
Niger	20

of a that maximizes this function is 0.073.

The likelihood function can be changed to accommodate censored observations. If a case is censored at time t_i, all we know is that this case's duration is greater than t_i. The probability of a duration greater than t_i is given by the survivor function $S(t)$ evaluated at time t_i. Now suppose that we have r uncensored observation and $n-r$ censored observations. If we arrange the data so that all the uncensored cases come first, we can write the likelihood as equation 2

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{r} f(t_i|\boldsymbol{\theta}) \prod_{i=r+1}^{n} S(t_i|\boldsymbol{\theta}) \tag{2}$$

Using the dummy, δ_i, which is one if the case ends in ratification or zero if the case is censored, we can write this likelihood function as function 3.

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [f(t_i|\boldsymbol{\theta})]^{\delta_i} [S(t_i|\boldsymbol{\theta})]^{1-\delta_i} \tag{3}$$

Here the dummy acts as a switch, turning the appropriate functions on or off, depending whether the observation is censored or not. In paragraph 2 we discussed that the hazard rate was the probability density function divided by the survival function. Consequently, the probability density function is the hazard function times the survival function. This means that the likelihood function can be rewritten as function 4

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} [h(t_i|\boldsymbol{\theta})]^{\delta_i} [S(t_i|\boldsymbol{\theta})]^{\delta_i} [S(t_i|\boldsymbol{\theta})]^{1-\delta_i} = \\ &\quad \prod_{i=1}^{n} [h(t_i|\boldsymbol{\theta})]^{\delta_i} S(t_i|\boldsymbol{\theta}) \end{aligned} \tag{4}$$

Again we can use the example from paragraphs 2 and 3 to illustrate this. Table 10 reproduces the dataset in which a number of countries are censored. Again we assume that the exponential distribution is the best applicable distribution. The hazard function of the

Table 10: example with censored observations

country	duration	ratified
West-Germany	3	1
Netherlands	7	1
Canada	12	1
India	13	0
USA	15	1
Thailand	16	1
USSR	16	0
Brazil	17	1
UK	19	0
Niger	20	0

Table 11: example with GDP

country	duration	ratified	GDP
West-Germany	3	1	14341
Netherlands	7	1	13029
Canada	12	1	17173
India	13	0	1264
USA	15	1	18054
Thailand	16	1	3580
USSR	16	0	7741
Brazil	17	1	4042
UK	19	0	13217
Niger	20	0	505

exponential is h(t) = a and the survival function is $S(t) = e^{-at}$.[6] The likelihood function for this example is $L(a) = a^1 e^{-a3} \times a^1 e^{-a7} \times \cdots \times a^0 e^{-a20} = ae^{-a3} \times ae^{-a7} \times \cdots \times 1 \times e^{-a20}$. The value of a that maximizes this function is 0.043.

Explanatory variables can be introduced by replacing one of the parameters with a function of the explanatory variables. This is illustrated with the help of an extension of the examples used before. Table 11 presents the dataset used before, but now the GDP per capita in 1990 has been added as an explanatory variable.

Again we assume that the exponential distribution is the best fitting distribution. This means that $h(t) = a$ and $S(t) = e^{-at}$, whereby a is a constant. If we think that the risk of ratifying is influenced by a number of explanatory variables than we can substitute a with a function of the explanatory variables. Since the hazard can not be negative, we must take care that the function can not be negative. This is generally achieved by substituting a with $e^{\beta_0 + \beta_1 x_1}$, whereby β_0 is a constant, x_1 the GDP per capita and β_1 the coefficient denoting the influence of GDP. The likelihood function now becomes:

---

[6]A peculiarity of the exponential distribution is that the probability density function is identical to the survival function.

Figure 8: Output parametric analysis

```
Iteration 0:   log likelihood = -10.969684
Iteration 1:   log likelihood = -10.170943
Iteration 2:   log likelihood = -10.102208
Iteration 3:   log likelihood = -10.102071
Iteration 4:   log likelihood = -10.102071


Exponential regression -- log relative-hazard form

No. of subjects =              10             Number of obs   =           10
No. of failures =               6
Time at risk    =             138
                                               LR chi2(1)      =         1.74
Log likelihood  =   -10.102071                 Prob > chi2     =       0.1877


------------------------------------------------------------------------------
         _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
        GDP |   .0000849   .0000667     1.27   0.203    -.0000457     .0002156
      _cons |  -3.984975   .8804094    -4.53   0.000    -5.710546    -2.259404
------------------------------------------------------------------------------
```

$$
\begin{aligned}
L\left(\beta_0\beta_1\right) &= \left[e^{\beta_0+\beta_1 14341}\right]^1 e^{-\left(e^{\beta_0+\beta_1 14341}\right)} \times \left[e^{\beta_0+\beta_1 13029}\right]^1 e^{-\left(e^{\beta_0+\beta_1 13029}\right)} \times \\
&\quad \cdots \times \left[e^{\beta_0+\beta_1 505}\right]^0 e^{-\left(e^{\beta_0+\beta_1 505}\right)}
\end{aligned}
$$

The values of β_0 and β_1 that maximize the likelihood function are -3.98 and 0.0001 respectively. These results were obtained using the statistical analysis program Stata. When estimating this model with Stata you will obtain the output presented in figure 8.

The first five lines illustrate that we are dealing with an iterative method. That is, we tried new values of the betas until the likelihood no longer improved. The first five lines give the natural logarithm of the likelihood for each attempt (iteration). The natural logarithm of the likelihood is used because Stata (and other statistical software packages) actually maximize the logarithm of the likelihood function. This function has the same maximum, but this maximum is easier to find. The left part of the eleventh line tells us that the natural logarithm of the likelihood is –10.102. The tenth and the right part of the eleventh line use this for a test (likelihood ratio test) of the hypothesis that all the betas except the constant are zero. Basically, this test compares the likelihood of a model without explanatory variables with a model with explanatory variables. The level of significance of this test is 0.19, which is above the 0.05 cut-off point, signifying that that a model with only the effect of time and without the explanatory variable GDP works just as well. The beta can be found in the column labeled `Coef.` The beta of GDP, the only explanatory variable, is 0.0000849, which means that the hazard ratio is $e^{0.0000849} = 1.000085$. This can be interpreted as an increase in GDP per capita of one dollar results in a increase in the hazard of ratifying of 0.009%. The column labeled `Std. Err.` gives the standard error of the estimated beta. This is used

17

to test whether the beta is different from zero. The results of this test are presented in the columns labeled `z` and `P>|z|`, whereby the last gives the p-value. Again we find that the beta of GDP is not statistically different from zero. The last two columns give the 95% confidence interval.

The other models are estimated in similar ways. Only these models have multiple parameters that can be replaced by explanatory variables. For instance the Weibull has two parameters: the a and the p. If we replace a with the explanatory variables we get the parameters for the explanatory variables with the interpretation discussed in the previous paragraph. However, there is no fundamental reason why we could not estimate a model in which other parameter(s), in this case the p, are replaced by one or more of the explanatory variables. The only problem would be that the estimated betas for these explanatory variables would be much more difficult to interpret. So, this should only be done when there is strong evidence that this would seriously improve the model and when there is no other model that will produce more or less equally good results but with parameters that are more easily interpretable.
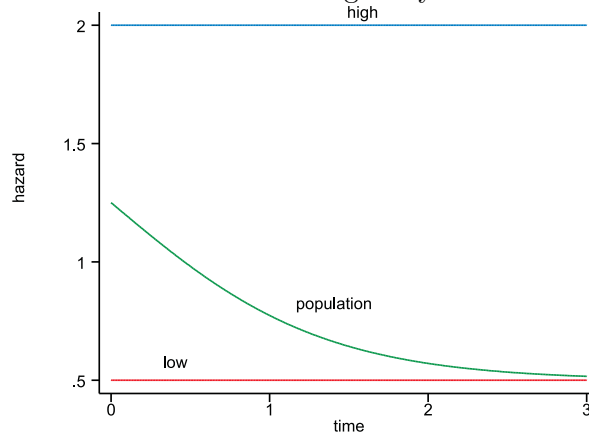
6   Unobserved heterogeneity[7]

An implicit assumption of the models we have considered so far is that if two countries have identical values on the covariates, they also have identical hazard functions. Obviously, this is an unrealistic assumption. Countries differ in so many respects that no set of measured covariates can capture all the variation among them. The problem that countries differ in ways that are not fully captured by the model is called unobserved heterogeneity. One consequence of unobserved heterogeneity is that it tends to produce estimated hazard functions that decline with time, even when the true hazard is not declining for any individual country in the sample. This is most easily explained with the help of an example. Suppose we have a sample of 100 countries, all of whom have hazards that are constant over time. The sample is equally divided between two kinds of countries: those with a high hazard of ratifying ($h = 2.0$) and those with a lower hazard of ratifying ($h = 0.5$). Unfortunately, we do not know which countries have which hazard, so we must estimate a single hazard function for the entire sample. Figure 9 shows what happens. The hazard function for the entire population starts out, as might be expected, midway between .5 and 2. But then it steadily declines until it approaches .5 as an asymptote. What is happening is that the high hazard countries are ratifying more rapidly at all points in time. As a result, as time goes by, the remaining sample is increasingly made up of countries with low hazards. Since we can only estimate the hazard function at time t with those who are still at risk at time t, the estimated hazard will be more and more like the smaller hazard. The basic principle remains the same when the countries can be divided into more than two groups. Those with higher hazards will tend to ratify before those with lower hazards, leaving a risk set that is increasingly made up of low hazard countries. (Allison 1995, 234-35)

The betas of the explanatory variables are also influenced by unobserved heterogeneity. First of all, the coefficients may be severely biased if the unobserved components are correlated with the measured covariates, as is the case with any regression technique. For instance, suppose that democratic countries are generally rich countries and that rich countries ratify faster then poor countries and that democracy has no effect on the speed of ratification and

---

[7]This section relies heavily on (Allison 1995)

Figure 9: effect of unobserved heterogeneity on the hazard function

that we do not know which countries are rich and which are poor. Democracy will in this case be positively correlated with the speed of ratification , because democracies are generally rich and rich countries ratify faster than poor countries, even though democracy has no effect on the speed of ratification. So when we estimate the effect of democracy without controlling for the wealth of countries, the estimated effect of democracy will actually be a combination of the effect of democracy and some of the effect of the wealth of the countries. The estimates are however also biased when the unknown explanatory variables are not correlated with the known explanatory variables. The estimates of the coefficients will in this case be attenuated toward zero. On the other hand, the standard errors and test statistics are not biased. Therefore, a test of the hypothesis that a coefficient is 0 remains valid, even in the presence of unobserved heterogeneity. (Allison 1995, 236)

There are ways to deal with unobserved heterogeneity. Ideally, all relevant variables are included and no unobserved heterogeneity exists, but if that is not possible a way to control for the unobserved variables is a second best option. To do that we can assume that the way countries are different can be captured by an unobserved constant specific for each individual country. This country specific constant is high when the country has a number of unknown characteristics that cause it to ratify relatively fast, and low if the unknown characteristics cause it to ratify relatively slow. This way we assume that the effects of the unobserved variables do not change over time. In our example, in which the unobserved variable splits the sample in two groups, we assume that the high hazard groups will always have a hazard that is a constant proportion larger then the hazard of the low hazard group. This is captured by the following hazard function: $h(t_i|\mathbf{x}_i\alpha_i) = \alpha_i h(t_i|\mathbf{x}_i)$, whereby α_i is the country specific constant. $h(t_i|\mathbf{x}_i)$ is the hazard function for an individual with an average value of the country specific constant, that is the hazard function which is not influenced by the unobserved variables. The α_i is scaled in such a way that, for example, the country specific constant will be 1.2 if the unobserved variables cause a country to ratify 20% faster then average and 0.70 if the unobserved variable cause the country to ratify 30% slower then average. The estimated parameters and betas in $h(t_i|\mathbf{x}_i)$ have been corrected for the unobserved heterogeneity. We do not know the values of these constants but we assume they are random draws from a probability distribution. That is, α_i is a random error term, which captures the effects of the unknown variables. We can of course estimate the correct betas and the correct shape of the hazard function if we know the correct values of the individual error terms. Problem is, we

19

do not know these values. However we can also get the correct estimates of the betas and the shape of the hazard function if we know the probability of observing each value of the error term. The reason for this is that problem with unobserved heterogeneity is that the data consists of two or more unobserved groups and that over time the low hazard group(s) get over-represented. If we know how fast each group ratifies and what proportion of the sample belongs to each group, than we know at each point in time by how much each group is over- or under-represented. The probability distribution of the error tells exactly how fast each group ratifies and what proportion of the sample belongs to each group. Two ways are used to obtain the probability distribution of the error: a parametric and a non-parametric way.

In the parametric method we make an assumption on the functional form of the probability distribution of the error. Much used probability distributions are the gamma and the inverse Gausian. The average value of the error is in both cases assumed to be one. The shape of these distributions is then solely determined by the standard deviation of the error term. We can write a likelihood function, which besides the betas and the parameters also include the standard deviation of the error term[8]. With that likelihood function we can estimate the shape of the probability distribution of the error term together with the corrected betas and parameters. We can also test whether unobserved heterogeneity is a real problem by testing whether the standard deviation of the heterogeneity is zero. The intuition behind this is that all error terms will have the same value if the standard deviation is zero. In other words all observations belong to the same group if the standard deviation is zero.

A major problem with the parametric way of dealing with unobserved heterogeneity is that we make an assumption about the functional form of a distribution of an unobserved variable and this assumption can sometimes have big effects on the results. The non-parametric method ensures that we do not have to make such assumptions. The non-parametric method basically assumes that the error term is not a continuous variable, but that it represents a finite number of different groups of countries. It generally starts with the assumption that there are two unobserved groups of countries, the slow and the fast. That is, the error term can have only two values. This is exactly the case in the example above. The distribution of the heterogeneity is very simple: there are two groups of countries, 50% of the countries is in the fast group, which is 60% faster then average and 50% of the countries is in the slow group, which is 60% slower then average. That means that the α of members of the fast group is 1.6 and the α of members of the slow group is 0.4. Every country has a 50% chance of belonging to the slow group and a 50% chance of belonging to the fast group. If we want to make the likelihood function, we are faced by two likelihood functions: one for the countries that are members of the fast group, and one for the countries that are members of the slow group. We can not make a choice which likelihood function is applicable for which country, since we do not know to which group a country belongs. However, we do know the probability that the likelihood of the fast group is applicable and the probability that the likelihood function of the slow group is applicable: 0.5 each. That means that the likelihood of observing a duration when we do not know which likelihood function is applicable is 0.5 times the likelihood of the slow countries plus 0.5 time the likelihood of the fast countries.

This approach can be generalized and used to estimate the corrected values of the betas and the shape parameters, the proportion of countries belonging to each group and how fast

---

[8]The calculations leading to this likelihood function can get rather complex and do not add to the clarity of this paragraph, so they are not discussed here. Those who are interested in the calculations can find them in (Blossfeld and Rohwer 1995, 247-48)) or in (Cleves et al. 2002, 261-62)

or slow each group is (the alphas). In order to do so we can write the likelihood function, if we assume that the population consists of 2 unobserved groups, but we do not know the probabilities of belonging to each group as $L = pL_{slow} + (1-p)L_{fast}$, whereby p is the probability of belonging to the slow group. This approach can easily be extended to encompass more groups. For instance, the likelihood function can be written as $L_{population} = p_1L_1 + p_2L_2 + (1-p_1-p_2)L_3$, if we assume that the population consists of three groups. Remember that the likelihood function can be written as: $L(\boldsymbol{\theta\beta}) = \prod_{i=1}^{n}[h(t_i|\boldsymbol{\theta\beta})]^{\delta_i}S(t_i|\boldsymbol{\theta\beta})$. We have already determined that the hazard function is $\alpha_i h(t_i|\boldsymbol{\theta\beta})$ in the presence of unobserved heterogeneity. The survivor function can than be written as $[S(t_i|\boldsymbol{\theta\beta}\mathbf{x}_i)]^{\alpha_i}$. The likelihood function of the slow group and the fast group can thus be written as functions 5 and 6.

$$L_{slow}(\boldsymbol{\theta\beta\alpha}) = \prod_{i=1}^{n}[\alpha_{slow}h(t_i|\boldsymbol{\theta\beta})]^{\delta_i}[S(t_i|\boldsymbol{\theta\beta})]^{\alpha_{slow}} \tag{5}$$

$$L_{fast}(\boldsymbol{\theta\beta\alpha}) = \prod_{i=1}^{n}[\alpha_{fast}h(t_i|\boldsymbol{\theta\beta})]^{\delta_i}[S(t_i|\boldsymbol{\theta\beta})]^{\alpha_{fast}} \tag{6}$$

That means that the likelihood function of the entire population can be written as likelihood function 7.

$$\begin{aligned} L_{pop}(\boldsymbol{\theta\beta\alpha}p) &= p\prod_{i=1}^{n}[\alpha_{slow}h(t_i|\boldsymbol{\theta\beta})]^{\delta_i}[S(t_i|\boldsymbol{\theta\beta})]^{\alpha_{slow}} + \\ &\quad (1-p)\prod_{i=1}^{n}[\alpha_{fast}h(t_i|\boldsymbol{\theta\beta})]^{\delta_i}[S(t_i|\boldsymbol{\theta\beta})]^{\alpha_{fast}} \end{aligned} \tag{7}$$

The population likelihood function is a function of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and p. We can maximize this likelihood function with respect to $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and p and find the corrected values of the betas and the shape parameters, the proportion of countries belonging to each group (the p) and how fast or slow each group is (the alphas). The number of groups is subsequently increased until the fit of the model no longer improves. This way any distribution can be approximated.

7 Semi-parametric analysis

The main disadvantage of parametric analysis, as was discussed in section 4, is that the estimates can be influenced by the two assumptions – the assumption on the way the risk of ratifying changes over time and the assumption on the way that the independent variables influence the risk of ratifying. The main disadvantage of non-parametric analysis is that it can only compare the survival functions of a limited number of groups. There is an intermediate technique whereby only an assumption is made about the way that the explanatory variables influence the risk of ratifying and it can still deal with many explanatory variables. This technique is called semi-parametric analysis, or Cox-regression. The advantage is that the results can no longer be influenced by assumptions about time-dependence, since no such assumptions are made. The disadvantages are that hypotheses about time dependence can no longer be tested and that parametric analysis yields more precise estimates than the semi-parametric analysis if the assumptions about the time dependence are correct.

Cox regression uses the proportional hazard assumption, which was discussed in section 5. Remember that it assumes that all groups of countries face a hazard function of the same shape. The only difference between groups is that the hazard functions of a group can be some constant proportion higher or lower then the hazard function of another group. For instance, if we are interested in the difference between rich and poor countries, then we assume that rich and poor countries both have a hazard function of the same shape, but that the hazard function of rich countries always lie some fixed proportion above or below the hazard function of the poor country. The strength of semi-parametric analysis is that the shape of the hazard function remains unspecified, which means that it can take any shape imaginable.

The proportional hazard assumption is captured by the hazard function of the Cox-regression, which can be written as equation 8.

$$h_i\left(t\right) = h_0\left(t\right) e^{\beta_1 x_{i1} + \cdots + \beta_k x_{ik}} \tag{8}$$

Equation 8 says that the hazard for country i at time t is the product of two factors:

- A hazard function $h_0(t)$ that is equal for all countries and is left unspecified. This hazard function is called the baseline hazard The baseline hazard function captures the shape of the hazard function. It can be thought of as the hazard function for countries whose covariates are all zero.

- A linear function of the set of covariates, which is then exponentiated. The function of the set of covariates is exponentiated to ensure that it can not be negative. The betas have the same interpretation as the betas in the parametric proportional hazard models.

Cox-regression can estimate the values of the betas that best fit the data without having to make an assumption about the baseline hazard. It uses a method called maximum partial likelihood, which is similar to maximum likelihood. Recall that maximum likelihood tries to find the values of the parameters and the betas that will maximize the probability of observing the data that has been observed. Basically maximum likelihood looks at each country individually and calculates the probability that that country ratifies at the time it did. The product of these probabilities is the probability that all the countries ratified at the time they did. This is a measure of the probability of observing the data that actually has been observed. An alternative measure is achieved when we look at each time a country ratifies and calculate the probability that at that time the country that ratified ratifies and not another country at risk of ratifying. The product of these probabilities will also be a measure of the probability of observing the data we have observed. This probability is written as a function of the unknown betas, but the baseline hazard is no longer part of this function, because being common to all countries it can not make a difference. The values of the betas that maximize this partial likelihood function, are the values of the betas that best fit the data.

The example data used in section 5 can also be used to illustrate this method. The data is repeated in table 12. The model we are trying to estimate is $h_i\left(t\right) = h_0\left(t\right) e^{\beta_1 GDP_i}$, because GDP is the only explanatory variable. We want to find the value of β_1 that best fit the data without having to make an assumption about the baseline hazard. So, at $t = 3$ we ask what the probability is that West Germany ratified instead of one of the other countries. The answer is the hazard for West-Germany divided by the sum of the hazards for all the countries at risk (Cleves et al. 2002, 21-24), as is shown in equation 9.

Table 12: example with GDP

country	duration	ratified	GDP
West-Germany	3	1	14341
Netherlands	7	1	13029
Canada	12	1	17173
India	13	0	1264
USA	15	1	18054
Thailand	16	1	3580
USSR	16	0	7741
Brazil	17	1	4042
UK	19	0	13217
Niger	20	0	505

$$L_1 = \frac{h_{West-Germany}(3)}{h_{West-Germany}(3) + h_{Netherlands}(3) + \cdots + h_{Niger}(3)} \tag{9}$$

At $t = 7$ we ask what the probability is that the Netherlands ratified instead of one of the other countries. This probability is given by equation 10. Note that Germany is no longer in the denominator since it is no longer at risk in $t = 7$.

$$L_2 = \frac{h_{Netherlands}(7)}{h_{Netherlands}(7) + h_{Canada}(7) + \cdots + h_{Niger}(7)} \tag{10}$$

Note, that according to our model we can write $h_{westgermany}(3)$ as $h_0(3)e^{\beta_1 GDP_{West-Germany}} = h_0(3)e^{\beta_1 14341}$. If we substitute the hazards with these expressions of the hazard we get equation 11 which represents the probability that West-Germany ratified at moth 3 and not another country.

$$L_1 = \frac{h_0(3)e^{\beta 14341}}{h_0(3)e^{\beta 14341} + h_0(3)e^{\beta 13029} + \cdots + h_0(3)e^{\beta 505}} \tag{11}$$

We can simplify this expression by eliminating the baseline hazard. This can be done because the baseline hazard is common to every term in both the numerator and the denominator, as can be seen in equation 12.

$$L_1 = \frac{h_0(3) \times e^{\beta 14341}}{h_0(3) \times (e^{\beta 14341} + e^{\beta 13029} + \cdots + e^{\beta 505})} = \\ \frac{e^{\beta 14341}}{e^{\beta 14341} + e^{\beta 13029} + \cdots + e^{\beta 505}} \tag{12}$$

We can write down the expression of these probabilities for every time a country ratified. The product of all these probabilities is the partial likelihood function, and the only unknown in this function is the β_1. The value of β_1 that maximizes the partial likelihood function (a measure of the probability of observing the data that have been observed) is the value that best fits the data, and we did not have to specify the baseline hazard.

When we estimate this model using Stata, we get the output shown in table 10. The first five lines illustrate that we are dealing with an iterative method. That is, we tried new values

Figure 10: Output semi-parametric analysis

```
Iteration 0:   log likelihood = -11.079061
Iteration 1:   log likelihood = -9.6611683
Iteration 2:   log likelihood = -9.6603393
Refining estimates:
Iteration 0:   log likelihood = -9.6603393


Cox regression -- no ties

No. of subjects =              10              Number of obs   =          10
No. of failures =               6
Time at risk    =             138
                                               LR chi2(1)      =        2.84
Log likelihood  =    -9.6603393               Prob > chi2     =      0.0921


------------------------------------------------------------------------------
        _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
       GDP |     .00012   .0000759     1.58   0.114    -.0000288    .0002687
------------------------------------------------------------------------------
```

of the β_1 until the likelihood no longer improved. The first five lines give the log likelihood for each attempt (iteration). The sixth line tells that we are doing a Cox-regression and that there are no two or more countries that ratified at the same time. The left part of the eleventh line tells us that the natural logarithm of the likelihood is –9.6603393. The tenth and the right part of the eleventh line use this for a test (likelihood ratio test) of the hypothesis that all the betas are zero. The p-value of this test is 0.09, which is above the 0.05 cut-off point, signifying that none of the betas is different from zero. The β_1 can be found in the column labeled `Coef.` The β_1 of GDP, the only explanatory variable, is 0.00012, which means that the hazard ratio is $e^{0.00012} = 1.00012$. This can be interpreted as an increase in GDP per capita of one dollar results in a increase in the hazard of ratifying of 0.012%. The column labeled `Std. Err.` gives the standard error of the estimated β_1. This is used to test whether the β_1 is different from zero. The results of this test are presented in the columns labeled `z` and `P>|z|`, whereby the last gives the p value. Again we find that the β_1 of GDP is not different from zero. The last two columns give the 95% confidence interval.

8   Conclusion

Three types of techniques where discussed in this chapter: the non-parametric, the parametric and the semi-parametric. All of them have their own advantages and disadvantages. The non-parametric allows us to gain insight with the smallest number of assumptions, but it can only compare a limited number of groups. Consequently it cannot deal with continuous variables or control for other variables. The parametric technique can deal with both discrete and continuous explanatory variables and control for a large number of other explanatory variables. However in order to estimate such a model we have to make assumptions on how the probability of ratifying changes over time (time dependence) and on how the explanatory

24

variables influence the risk of ratifying. The semi-parametric technique requires only the last assumption. However the estimated parameters and betas will be less precise then the ones obtained from parametric analysis (provided that the assumptions made in parametric analysis are correct) and we can no longer test hypotheses about time dependence.

References

Allison, P. D.: 1995, *Survival analysis using the SAS system: A practical guide*, SAS Institute Inc., Cary, NC.

Blossfeld, H.-P. and Rohwer, G.: 1995, *Techniques of event history modeling, new approaches to causal analysis*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Cleves, M. A., Gould, W. W. and Guitierrez, R. G.: 2002, *An introduction to survival analysis using Stata*, Stata Press, College Station, Texas.