Chapter 5

Extended and Stratified Cox

In this chapter we consider two ways of dealing with violations of the PHA. These are:

1. Extending the PHM to include interactions between time and the covariates that do not satisfy the PHA.

2. Stratifying over (categorised) covariates that do not obey the PHA.

The methodology in item 1 also allows us to consider covariates that change with time. This might feasibly be the case for many covariates in a long epidemiological study.

5.1 Extending the PHM

Consider the generalisation of the PHM which has form

$$h(t, x) = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \ldots + \gamma_1 x_1 g_1(t) + \gamma_2 x_2 g_2(t) + \ldots\}.$$
Then the hazards ratio for two individuals \(a\) and \(b\) with \(x_1 = x_a\) or \(x_b\) but all other covariates the same becomes:

\[
\frac{h(t, x_a)}{h(t, x_b)} = \exp\{\beta_1(x_a - x_b) + \gamma_1 g_1(t)(x_a - x_b)\}.
\]

This is no longer constant in time (except for the trivial case when \(g_1(t) = g\)), i.e. it is no longer a proportional hazards model. It is called the extended Cox model.

This generalisation complicates the partial likelihood function. It is derived in the same way as before, but now depends on the time at which each individual fails (rather than just the ordering). Before, the partial likelihood function in the absence of ties was

\[
l_p(\beta, x) = \prod_{i=1}^{m} \left[ \frac{\phi_i}{\sum_{j \in R\{t_i\}} \phi_j} \right]^{\delta_i}
\]

where \(m\) is the number of people in the sample, \(\delta_i = 1\) if \(i\) failed at time \(t_i\) and \(0\) if \(i\) were right censored, \(R\{t_i\}\) is the risk set of time \(t_i\) and \(\phi_i = \exp(\beta^T x_i)\) is proportional to the hazard for \(i\).

For the extended Cox model it becomes

\[
l_p(\beta, x) = \prod_{i=1}^{m} \left[ \frac{\phi_i(t_i)}{\sum_{j \in R\{t_i\}} \phi_j(t_i)} \right]^{\delta_i}
\]

where \(\phi_i(t) = \exp\{\beta^T x_i + \gamma_1 x_1 g_1(t) + \gamma_2 x_2 g_2(t) + \ldots\}\) is proportional to the hazard for individual \(i\) at time \(t\). Note that \(\phi_j(t_i)\) is not a typo (for a change!): this is proportional to the hazard for individual \(j\) at the time \(i\) failed.

The partial likelihood with ties generalises in the same way.
5.1.1 Calculating the partial likelihood

Actually calculating the partial likelihood is not too difficult. It is based on the “counting process” theory developed by Aalen (1978) and others. We won’t go into this in this course, but you may be interested to read Fleming and Harrington (1991) for your personal edification. Basically, we can view an individual’s survival as an ensemble of disjoint segments and obtain equivalent results as when we consider it as a whole. For example, if we had only two individuals $a$ and $b$, with $t_a < t_b$ and $\delta_a = \delta_b = 1$, we could sketch their survival as follows:

We could if we wanted split $b$ into two “individuals” with one being right-censored at time $t_a$ and the other being “born” at $t_a$: 
The partial likelihood for the original Cox PHM in the first scenario is
\[ l_p(β, x_a, x_b) = \frac{ϕ_a}{ϕ_a + ϕ_b} \times \frac{ϕ_b}{ϕ_b}. \]

In the second it is
\[ l_p(β, x_a, x_{b_1} = x_{b_2}) = \frac{ϕ_a}{ϕ_a + ϕ_{b_1}} \times \frac{ϕ_{b_2}}{ϕ_{b_2}} \]
because the risk set at time \( t_a \) is \( R\{t_a\} = \{a, b_1\} \). An instant later, \( b_1 \) leaves the risk set (as does \( a \) as s/he died) and \( b_2 \) enters, i.e. \( R\{t_a + ϵ\} = \{b_2\} \) for \( 0 < ϵ \ll 1 \).

So the approach to calculating the partial likelihood for the extended Cox model is to partition each individual into segments at each earlier failure time. Clearly this will create a lot of data points. This partitioning is needed for models for the interaction \( g_i(t) \) that change at each time, like \( g_i(t) = t \) or \( \log t \).

A simpler approach can be used for the heaviside function \( g_i(t) = 1\{t > T_0\} \).

With this model, we need only partition each survivor at time \( T_0 \). This is easier in practice, which is why I recommend the heaviside function to you.

Note that it is possible to have a categorical covariate defined by multiple time points \( T_0, T_1, T_2, \ldots \). This can be effected in a similar way. This is left as an exercise to the dedicated reader.

### 5.1.2 Doing it in R

An example follows of how this many be done in R. This uses the built in dataset \texttt{aml} on acute myelogenous leukaemia, which has three elements: \texttt{time}, \texttt{status} (i.e. death or censoring) and a binary indicator \( x \) for maintenance chemotherapy being given. Times vary from 5 to 161, with median 23, so we try a split at time 20.

We use the \texttt{survSplit} command to partition survival, then give the \texttt{Surv} command a different form of arguments than we are used to: instead of end time and censoring indicator, we provide start times, end times and censoring indicator.

The commands applied to this example are:
library(survival)
aml2=survSplit(aml,cut=c(20),end='time',event='status',start='start')
aml2$gt=(aml2$start==20)+0

Fitting the PHM to the original and partitioned data yield, respectively,

> coxph(Surv(start,time,status)~x,data=aml2)
Call:
coxph(formula = Surv(start, time, status) ~ x, data = aml2)

    coef exp(coef) se(coef) z  p
xNonmaintained 0.916  2.5  0.512 1.79 0.074

Likelihood ratio test=3.38 on 1 df, p=0.0658 n= 36
> coxph(Surv(time,status)~x,data=aml)
Call:
coxph(formula = Surv(time, status) ~ x, data = aml)

    coef exp(coef) se(coef) z  p
xNonmaintained 0.916  2.5  0.512 1.79 0.074

Likelihood ratio test=3.38 on 1 df, p=0.0658 n= 23

Note that the fit to the newly formed data is the same as the fit to the old format (as we have not actually extended the model to have a time-varying hazard ratio yet), with the singular exception of the sample size, which now counts each partition as a separate individual. We can trivially extend these commands to test the hypothesis that the PHA is satisfied, as follows:

> aml2$x=(aml2$x=='Maintained')+0
> coxph(Surv(start,time,status)~x+gt:x,data=aml2)
Call:
coxph(formula = Surv(start, time, status) ~ x + gt:x, data = aml2)
<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>-0.688</td>
<td>0.503</td>
<td>0.732</td>
<td>-0.940</td>
<td>0.35</td>
</tr>
<tr>
<td>x:gt</td>
<td>-0.431</td>
<td>0.650</td>
<td>1.023</td>
<td>-0.422</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Likelihood ratio test=3.56 on 2 df, p=0.168 n= 36

Note that we do not need an individual term for gt in the model as this would not be identifiable, being swallowed up in the baseline hazard. The p-value for the time-covariate interaction is 0.67, giving us no reason to reject $H_0$. We could thus revert to the PHM for this data set. Note though that the conclusion is dependent on the choice of function $g(t) = 1\{t \geq 20\}$.

### 5.1.3 Example

Recall that in the previous chapter, we had fitted the models

$$h^{(1)}(t, \mathbf{x}) = h^{(1)}_0(t) \times \exp\{-0.36x_{\text{fin}} + 42/x_{\text{age}} + 0.084x_{\text{prio}} - 0.29x_{\text{educ}} - 0.87x_{e2}\}$$

and

$$h^{(2)}(t, \mathbf{x}) = h^{(2)}_0(t) \times \exp\{-0.39x_{\text{fin}} + 42/x_{\text{age}} - 0.24x_{\text{prio}} - 0.59x_{\text{educ}} - 0.90x_{e2} + 0.098x_{\text{prio}}x_{\text{educ}}\}.$$
CHAPTER 5. EXTENDED AND STRATIFIED COX

of the ex-cons were less than 30 and the mean age was just under 25. So let us categorise age into the following binary covariate $x_{\text{age}} = 1\{x_{\text{age}} \geq 25\}$ and use this to define groups for constructing an expected versus observed plot. Here are the commands for the interaction model:

\begin{verbatim}
c.age=age>=25; c.age=c.age+0
mc15=coxph(S~fin+c.age+prio*educ+e2)
d.mc15=coxph.detail(mc15)
times=c(0,d.mc15$t)
h0=c(0,d.mc15$hazard)
S0=exp(-cumsum(h0))
beta = mc15$coef
meanx=c(mean(fin),mean(c.age),mean(prio),mean(educ),mean(e2)
,mean(prio*educ))
z=0; x1=c(mean(fin[c.age==z]),z,mean(prio[c.age==z]),
mean(educ[c.age==z]),mean(e2[c.age==z]),
mean((prio*educ)[c.age==z])) -meanx
z=1; x2=c(mean(fin[c.age==z]),z,mean(prio[c.age==z]),
mean(educ[c.age==z]),mean(e2[c.age==z]),
mean((prio*educ)[c.age==z])) -meanx
Sx1=S0^exp(t(beta)%*%x1)
Sx2=S0^exp(t(beta)%*%x2)
\end{verbatim}

These allow the following graph to be created. Note that we are adjusting the other covariates by those in the two classes: the plot would look very similar in this case if we didn’t bother with this.
(The plot for the no-interaction model is almost identical.) It appears that initially, both age groups have very similar survival curves. After about half a year, the older group starts to become relatively less likely to be rearrested. As a consequence the hazards are not proportional. This plot motivates using a heaviside function with step at 26 weeks.

We thus seek to fit these two models extended to include the following term in the exponent: $\gamma 1\{t \geq 26\} x_{age}$. The output from the fitted models is:

```r
> D2=survSplit(D,cut=26,end='week',event='arrest',start='start')
> D2$gt=(D2$start==26)+0
> attach(D2)
> S=Surv(start,week,arrest)
> em30=coxph(S~fin+I(1/age)+prio+educ+e2+I(gt/age))#no interaction
> em15=coxph(S~fin+I(1/age)+prio*educ+e2+I(gt/age))#interaction
> em30
Call:
coxph(formula = S ~ fin + I(1/age) + prio + educ + e2 + I(gt/age))

 coef exp(coef) se(coef)      z    p
fin   -0.3628   0.696e-01  0.1910 -1.899 0.0580
I(1/age)  3.4955   3.30e+01 16.5504  0.211 0.8300
prio   0.0851   1.09e+00  0.0285  2.990 0.0028
educ  -0.2875   7.50e-01  0.1568 -1.833 0.0670
e2    -0.9243   3.97e-01  0.5580 -1.656 0.0980
I(gt/age) 82.6328 7.71e+35 25.5004  3.240 0.0012

Likelihood ratio test=47.2 on 6 df, p=1.72e-08 n= 810
> em15
Call:
coxph(formula = S ~ fin + I(1/age) + prio * educ + e2 + I(gt/age))

 coef exp(coef) se(coef)   z     p
fin    -0.396   6.73e-01 0.1925 -2.058 0.04000
I(1/age)  2.009   7.45e+00 16.5605  0.121 0.90000
prio   -0.247   7.81e-01 0.1558 -1.583 0.11000
educ  -0.592   5.53e-01 0.2215 -2.671 0.00760
```
Interestingly, the interaction between age and our time-varying variable is highly significant (the most significant of all the parameters remaining, actually) but the main effect is not at all significant. The interpretation is that once s/he passes the first half year, every year of age of an ex-con’s age reduces his or her risk of reoffending, as we saw in the plot above. Because the form of the relationship with age is not linear, we cannot make a nice statement about the amount the risk decreases per annum. However, we can plot this difference, say for the interaction model:

![Plot of relative hazard before and after 26 weeks](image)

There is really a quite astounding difference here. Before our (arbitrary) threshold of 26 weeks, there is hardly any difference in reoffending rates with age: the only difference there is is due to the small coefficient of \(1/\text{age}\) (which we have retained due to the rule that interactions should appear with the main effects present). After this threshold, the older ex-cons are far less likely to reoffend. Note that we’ve plotted the hazard relative to a 15 year old ex-con, but could have put an absolute scale with a bit more work. The commands to create this plot were:
cm=1/2.54;pdf("age_rec.pdf",height=8*cm,width=12*cm)
par(mai=c(1.75,1.75,0.5,0.5)*cm,mgp=c(2,0.75,0))
ages=15:45
pre=exp(em15$coeff[2]/ages)
post=exp(em15$coeff[2]/ages + em15$coeff[6]/ages)
pre=pre/max(pre)
post=post/max(post)
plot(ages,pre,type='l',xlab='age (y)',ylab="relative hazard",
     ylim=range(c(pre,post)))
lines(ages,post,col=2)
legend(30,0.8,c('before 26w','after 26w'),lty=c(1,1),col=c(1,2),
       bty='n')
dev.off()

We can check there are no further non-proportionalities in the same way as we did for the original PHMs.

> cox.zph(em15,transform='rank',global=F)
   rho  chisq     p
fin 0.003510 1.45e-03 0.970
I(1/age) -0.024743 6.85e-02 0.793
prio -0.026407 9.03e-02 0.764
educ -0.102042 1.38e+00 0.240
e2 0.000419 1.86e-05 0.997
I(gt/age) -0.033797 1.37e-01 0.711
prio:educ 0.014147 2.77e-02 0.868
> cox.zph(em30,transform='rank',global=F)
   rho  chisq     p
fin -0.00182 0.00038 0.984
I(1/age) -0.03279 0.12328 0.726
prio -0.11587 1.78055 0.182
educ -0.17512 3.45528 0.063
e2 -0.01313 0.01835 0.892
I(gt/age) -0.03074 0.11531 0.734

The introduction of the interaction between time and age has got rid of all significant non-proportionalities. (The final thing to do is to decide which of the two models to use as the final model! Which would \textit{you} choose?)
5.2 Inherently time-varying covariates

The extension of the model described in the last section was presented as a response to the violation of the PHA by one or more static covariates. It reflects an interaction between the covariates and time, i.e. a change in the effect of the covariates. However, we can also extend the model to reflect dynamic changes in the covariates.

By dynamically changing covariates, we mean covariates that actually change with time (and are either recorded or we can work them out), rather than just their effect changing with time.

As you may have discovered in a tutorial, the recidivism data also contain a large number of columns indicating whether the ex-con worked each week after his or her release for the year the data were collected. These columns represent a covariate $x_{\text{work}}(t)$ for $t = 1, 2, \ldots, 52$. Here are some plots of this covariate for various ex-cons.
To incorporate such time-varying covariates, we need to use the extended Cox model. The hazard function for a model with constant covariate $x_1$ and time-varying covariate $x_2(t)$ can be written thus:

$$h\{t, x_1, x_2(t)\} = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2(t)\}$$

and similarly for other models. In principle, we could have a further interaction with time, replacing $\beta_2 x_2(t)$ by $\beta_3 x_2(t) g(t)$ if the PHA were not satisfied for this model, but the interpretation would be even more complicated.

The approach to dealing with such time-varying data is to split each individual at risk at the time of any change in any individual’s covariates, in much the same way as in the previous section. The extended Cox model may then be fit to these data using the counting process approach.

### 5.2.1 Example 1: recidivism

This example is based upon the excellent description given in the online appendix “Cox Proportional-Hazards Regression for Survival Data” to Fox (2008) (do a google search for Cox Proportion-Hazards Regression for Survival Data—there is also a more recent 2ed by Fox and Weisberg).

The first thing we do is to create a new data set with start and end times at weekly intervals and a single covariate for employment status in each
week. There will be one record per individual per week spent outside prison. Each record will also contain the static covariates, and we shall also have an indicator variable arrest indicating if the individual was arrested that week or “censored”, to be replaced by a Doppelgänger the next week, identical in all respects expect with a new value of employment status.

The code requires a bit of thought:

```r
row=0
D2.start = 0;D2.stop=0;D2.arrest=0
D2.fin=0;D2.age=0;D2.prio=0;D2.educ=0;D2.e2=0;D2.gt=0;D2.empt=0
for(i in 1:nrow(D))
{
  for(j in 11:62)
  {
    if(is.na(D[i,j])==FALSE)
    {
      row=row+1
      D2.start[row]=j-11
      D2.stop[row]=D2.start[row]+1
      D2.arrest[row]=0
      if(D[i,2]==1 && D[i,1]==D2.stop[row])D2.arrest[row]=1
      D2.fin[row]=D[i,3]
      D2.age[row]=D[i,4]
      D2.prio[row]=D[i,9]
      D2.educ[row]=D[i,10]
      D2.e2[row]=D[i,63]
      D2.gt[row]=D[i,64]
      D2.empt[row]=D[i,j]
    }
  }
}
D2=list(start=D2.start,stop=D2.stop,arrest=D2.arrest,fin=D2.fin,
age=D2.age,prio=D2.prio,educ=D2.educ,e2=D2.e2,gt=D2.gt,
empt=D2.empt)
```

We can then fit the model. Commands and output are:
> S=Surv(D2$start,D2$stop,D2$arrest)
> ecox1=coxph(S~fin+I(1/age)+prio*educ+e2+I(gt/age) + empt,data=D2)
> ecox1
Call:
coxph(formula = S ~ fin + I(1/age) + prio * educ + e2 + I(gt/age) +
  empt, data = D2)

          coef exp(coef) se(coef)  z     p
fin     -0.3487  7.06e-01  0.1923 -1.81 7.0e-02
I(1/age) 49.4017 2.85e+21 18.6178 2.65 8.0e-03
prio    -0.2098  8.11e-01  0.1532 -1.37 1.7e-01
educ    -0.5409  5.82e-01  0.2185 -2.48 1.3e-02
e2      -0.8232  4.39e-01  0.5670 -1.45 1.5e-01
I(gt/age) 13.1672 5.23e+05  9.2009 1.43 1.5e-01
empt    -1.3174  2.68e-01  0.2500 -5.27 1.4e-07
prio:educ 0.0866  1.09e+00  0.0449  1.93 5.4e-02

Likelihood ratio test=77.3 on 8 df, p=1.66e-13 n= 19809

We see that employment status is hugely significant. The hazard ratio for an ex-con in work on any given week is around 25% that of one who is out of work. Also, note that including this term means we can get rid of the interaction between time and age that we earlier found to violate the PHA. In other words, the cause of this apparent interaction is due to the relationship between an ex-con’s age and ability to be in employment as a function of time.

However, there is an issue about causality here. Being in prison stops you from working. Therefore it is not clear whether it is work that reduces the risk of rearrest, or being in prison that stops you working and also being arrested (since you’re already in gaol). It would thus be safer to consider a lagged covariate here, that is to replace the term in the model for working at time \( t \) with one for working at time \( t-1 \) (in units of weeks). The starting value we will force to be 0, since the ex-cons start the study already in gaol.

The changes to the code are trivial. I have replaced D2 by D3 and used the following line
to make the lagged covariate. The output of the model is:

> ecox2=coxph(S~fin+I(1/age)+prio*educ+e2+I(gt/age) + empt,data=D3)
> ecox2
Call:
  coxph(formula = S ~ fin + I(1/age) + prio * educ + e2 + I(gt/age) + empt, data = D3)

    coef exp(coef) se(coef)     z     p
fin  -0.3575  0.699e-01  0.192 -1.86 0.06300
I(1/age)  52.9859  1.03e+23  18.551  2.86 0.00430
prio  -0.2172  8.05e-01  0.153 -1.42 0.16000
educ  -0.5559  5.74e-01  0.219 -2.54 0.01100
e2   -0.8782  4.16e-01  0.566 -1.55 0.12000
I(gt/age) 12.8817  3.93e+05  9.155  1.41 0.16000
empt  -0.7775  4.60e-01  0.217 -3.58 0.00035
prio:educ  0.0903  1.09e+00  0.045  2.01 0.04500

Likelihood ratio test=56.4 on 8 df, p=2.32e-09  n= 19809

Note that employment is still highly significant, but its coefficient is not as large as before. After correcting for the causality issue, we find that those in work in one week are around half as likely to be rearrested the next week as those out of work, rather than a quarter as before. Note that the interaction between time and age is still not significant, so we would remove that.

5.2.2 Example 2: treating drug addiction

We revisit an example we considered some time ago, in which 628 American drug addicts were randomly allocated to long or short stays at a residential treatment centre. Consider the simple model without any other terms. The hazard can be written
\[ h(t, x) = h_0(t) \exp(\beta x_1) \]

where \( x_1 \) is 1 if the addict received the long treatment and 0 otherwise.

Fitting the model gives the following output:

Call:
`coxph(formula = S ~ stay)`

```
  coef  exp(coef) se(coef)  z      p
stay -0.233      0.792    0.0891 -2.61 0.009
```

Likelihood ratio test=6.83 on 1 df, p=0.00898 n= 627

indicating fairly strong evidence that the long treatment \((\text{stay}=1)\) reduces
the risk or return to drugs by about 20\%. However, it is not clear which of
these factors this is due to:

- the long treatment has a lasting effect on the addicts’ desire to resume
drug taking; or
- it is harder to access drugs inside the residential centres, and once the
addicts are released, they go back to taking drugs at the same rate.

We can test these hypotheses by setting up a model with four classes of
addicts:

- \( C_{1a} \) if the addict was in the long treatment and is still in the centre
- \( C_{1b} \) if the addict was in the long treatment but has finished and returned
  home
- \( C_{0a} \) if the addict was in the short treatment and is still in the centre
- \( C_{0b} \) if the addict was in the short treatment but has returned home.
The simplest way to do this is to introduce a new variable, \( x_2(t) \), indicating if the addict is still undergoing treatment.

The model will be

\[
h(t, x) = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2(t)\}.
\]

We can set up the following null hypothesis: \( H_0 \) is that the only difference between the treatments is that it is hard to get access to drugs within the centre, i.e. there is no lasting effect. This implies that \( \beta_1 = 0 \), which can be tested in the usual way. The alternative is that the longer centre reducing the chance of returning to drug use after release as well.

You will perform the actual analysis in the next tutorial, so it is not provided here.

### 5.3 Stratified Cox

An alternative to extending the Cox model to deal with non-proportional hazards is to stratify over the covariates that do not satisfy the PHA. In essence, stratification involves fitting a model that has a different baseline hazard in each stratum. The advantage is that you do not have to worry about the (subjective) functional form assumed for the time interaction, \( g(t) \).

There are disadvantages, however:

- the baseline hazards are estimated within strata only: this means that there is more uncertainty in their estimates as information is not pooled over strata as in the extended Cox model;
- by stratifying over a covariate \( x \) we lose the ability to quantify its effect;
- for continuous covariates, we must arbitrarily categorise them.

Nonetheless, stratification is a popular solution to the problem of violation of the PHA.
5.3.1 Formulation

- Partition the covariates into those that satisfy the PHA ($\mathbf{x}$) and those that don’t ($\mathbf{z}$).
- Define a set of strata $\Sigma$ in terms of the non-proportional covariates $\mathbf{z}$.
- Individual $i$ belongs to stratum $\sigma_i \in \Sigma$.
- The “no-interaction” stratified Cox model is
  \[ h(t, \mathbf{x}, \mathbf{z}) = h_0^\sigma(t) \exp\{\mathbf{\beta}^T \mathbf{x}\} \]
  where $h_0^\sigma(t)$ is a stratum-dependent baseline hazard.

5.3.2 Forming the strata

If there are multiple covariates $\mathbf{z}$ not satisfying the PHA, they define a stratum _jointly_. See picture that follows.

If a covariate is continuous, it must be categorised before creating the strata, into two or three (probably not more) categories. The mean, quantiles or a nice round number might be used to partition the real line into categories.
5.3.3 Example

Let us pretend that we haven’t introduced the time-varying covariate employment to the recidivism data, and thus that we still think age does not satisfy the PHA. The mean age was around 25, so we use 25 to partition age into two strata: \( \sigma_i = 1 \) if \( x_{age} < 25 \) and \( \sigma_i = 0 \) if \( x_{age} \geq 25 \). We have already defined such a categorical covariate. The commands and output are:

```r
> scox1=coxph(S~fin+strata(c.age)+prio*educ+e2)
> scox1
Call:
  coxph(formula = S ~ fin + strata(c.age) + prio * educ + e2)

    coef exp(coef) se(coef)      z     p
fin   -0.434   0.648  0.1922 -2.26 0.0240
prio  -0.253   0.776  0.1564 -1.62 0.1100
educ  -0.642   0.526  0.2203 -2.92 0.0035
e2   -1.002   0.367  0.5675 -1.76 0.0780
prio:educ  0.103   1.109  0.0461  2.24 0.0250

Likelihood ratio test=25.5 on 5 df, p=0.000113 n= 432
```

Note that we have identified `c.age` as strata. There is no output for age, as no parameters are estimated for it: instead separate baseline hazards are estimated for the two age groups.

5.3.4 Non-interaction assumption

The effect of \( x \) is assumed to be the same across strata. This is the no-interaction assumption. It is also possible to include interactions between strata and \( x \); indeed, it is possible that these interactions are needed in the model.
CHAPTER 5. EXTENDED AND STRATIFIED COX

Here is a picture of what the interactions mean:

The interaction model can be written mathematically in the following equation:

\[ h(t, x, z) = h_0^\sigma(t) \exp\{\beta_\sigma^T x\} \]

that is, each stratum \( \sigma \) has different parameters \( \beta_\sigma \) (or possibly just a subset of the parameters differ across strata: the notation is a little messier in this case).

We can rewrite this equation by introducing dummy variables \( y_\sigma \) that equal 1 if an individual belongs to strata \( \sigma \) and 0 otherwise. Assuming that we make \( y_1 \) be the baseline stratum, the model can be written

\[ h(t, x, z) = h_0^\sigma(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \ldots + \gamma_{12} x_1 y_2 + \gamma_{13} x_1 y_3 + \ldots + \gamma_{22} x_2 y_2 + \ldots\}. \]

This formulation is identical to the first and is how we would fit the model in practice.

Clearly, the interaction model is a generalisation of the non-interaction model (the special case is when the \( \gamma \)s are all 0). We can thus test whether there are interactions using the likelihood ratio test. This can be done on \( H_0 : \gamma_{11} = \gamma_{12} = \ldots = 0 \) or on a subset of the \( \gamma \)s.
5.3.5 Example

Consider again the recidivism data. We can attempt to put interactions between our categorised age covariate and the other terms in the model. So doing leads to the following output:

\[
> \text{scox2 = coxph(S ~ fin + strata(c.age) + prio + educ + e2 + prio:educ} \\
> \quad + c.age:fin + c.age:prio + c.age:educ + c.age:e2))
\]

\[
> \text{scox2}
\]

Call:
coxph(formula = S ~ fin + strata(c.age) + prio + educ + e2 + 
       prio:educ + c.age:fin + c.age:prio + c.age:educ + c.age:e2)

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>fin</td>
<td>-0.37983</td>
<td>0.684</td>
<td>0.2208</td>
<td>-1.720</td>
<td>0.08500</td>
</tr>
<tr>
<td>prio</td>
<td>-0.27983</td>
<td>0.756</td>
<td>0.1506</td>
<td>-1.858</td>
<td>0.06300</td>
</tr>
<tr>
<td>educ</td>
<td>-1.08300</td>
<td>0.339</td>
<td>0.2806</td>
<td>-3.859</td>
<td>0.00011</td>
</tr>
<tr>
<td>e2</td>
<td>-1.84324</td>
<td>0.158</td>
<td>1.0485</td>
<td>-1.758</td>
<td>0.07900</td>
</tr>
<tr>
<td>prio:educ</td>
<td>0.11149</td>
<td>1.118</td>
<td>0.0452</td>
<td>2.464</td>
<td>0.01400</td>
</tr>
<tr>
<td>fin:c.age</td>
<td>-0.18599</td>
<td>0.830</td>
<td>0.4497</td>
<td>-0.414</td>
<td>0.68000</td>
</tr>
<tr>
<td>prio:c.age</td>
<td>-0.00906</td>
<td>0.991</td>
<td>0.0691</td>
<td>-0.131</td>
<td>0.90000</td>
</tr>
<tr>
<td>educ:c.age</td>
<td>0.94370</td>
<td>2.569</td>
<td>0.3120</td>
<td>3.025</td>
<td>0.00250</td>
</tr>
<tr>
<td>e2:c.age</td>
<td>1.79451</td>
<td>6.017</td>
<td>1.2752</td>
<td>1.407</td>
<td>0.16000</td>
</tr>
</tbody>
</table>

Likelihood ratio test=35.1 on 9 df, p=5.78e-05 n= 432

There are no significant interactions between the strata and e2, number of prior incarcerations or receipt of financial assistance, but there is an interaction between education and age. Getting rid of the non-significant interactions leads to:

\[
> \text{scox3 = coxph(S ~ fin + strata(c.age) + prio + educ + e2 + prio:educ} \\
> \quad + c.age:educ))
\]

\[
> \text{scox3}
\]
Call:
coxph(formula = S ~ fin + strata(c.age) + prio + educ + e2 + 
prio:educ + c.age:educ)

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>fin</td>
<td>-0.423</td>
<td>0.655</td>
<td>0.1930</td>
<td>-2.19</td>
<td>0.02800</td>
</tr>
<tr>
<td>prio</td>
<td>-0.267</td>
<td>0.766</td>
<td>0.1432</td>
<td>-1.87</td>
<td>0.06200</td>
</tr>
<tr>
<td>educ</td>
<td>-0.979</td>
<td>0.376</td>
<td>0.2608</td>
<td>-3.75</td>
<td>0.00017</td>
</tr>
<tr>
<td>e2</td>
<td>-0.862</td>
<td>0.422</td>
<td>0.5621</td>
<td>-1.53</td>
<td>0.13000</td>
</tr>
<tr>
<td>prio:educ</td>
<td>0.107</td>
<td>1.113</td>
<td>0.0419</td>
<td>2.55</td>
<td>0.01100</td>
</tr>
<tr>
<td>educ:c.age</td>
<td>0.748</td>
<td>2.112</td>
<td>0.2826</td>
<td>2.64</td>
<td>0.00820</td>
</tr>
</tbody>
</table>

Likelihood ratio test = 32.4 on 6 df, p = 1.36e-05 n = 432

A quick check confirms that the PH A is not obviously violated for this new model:

> cox.zph(scox3, global=F, transform='rank')

<table>
<thead>
<tr>
<th></th>
<th>rho</th>
<th>chisq</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>fin</td>
<td>0.00525</td>
<td>0.00329</td>
<td>0.954</td>
</tr>
<tr>
<td>prio</td>
<td>-0.03327</td>
<td>0.11781</td>
<td>0.731</td>
</tr>
<tr>
<td>educ</td>
<td>-0.13122</td>
<td>1.47526</td>
<td>0.225</td>
</tr>
<tr>
<td>e2</td>
<td>0.00317</td>
<td>0.00106</td>
<td>0.974</td>
</tr>
<tr>
<td>prio:educ</td>
<td>0.01445</td>
<td>0.02373</td>
<td>0.878</td>
</tr>
<tr>
<td>educ:c.age</td>
<td>0.07181</td>
<td>0.52284</td>
<td>0.470</td>
</tr>
</tbody>
</table>

### 5.3.6 Plotting strata in R

By stratifying over a covariate, there is no longer a single baseline hazard to be converted into survival curves in a plot, but a set of hazards, one per stratum. Nevertheless, it is easy to make plots from these, as the number of individuals in each stratum define the start and end points of the hazards vector for that stratum in the R output from `coxph.detail()`.

Here are R commands to make a plot for the Australian addicts data, stratifying over clinic.
library(survival)
D=read.table("http://courses.nus.edu.sg/course/stacar/internet/addicts.txt",
    header=TRUE)
attach(D)
S=Surv(t/365.25,depart)
m1=coxph(S~strata(clinic)+dose+prison)
d.phm=coxph.detail(m1)

#strata are Y and Z; number individuals in Y, Z:
nY=d.phm$strata[1]
nZ=d.phm$strata[2]

#h0 for Y
timesY=c(0,d.phm$t[1:nY])
h0Y=c(0,d.phm$hazard[1:nY])
S0Y=exp(-cumsum(h0Y))

#h0 for Z
timesZ=c(0,d.phm$t[(nY+1):(nY+nZ)])
h0Z=c(0,d.phm$hazard[(nY+1):(nY+nZ)])
S0Z=exp(-cumsum(h0Z))

#dose effect:
beta=m1$coef
x1=c(mean(dose[dose<60]),mean(prison))-c(mean(dose),mean(prison))
x2=c(mean(dose[dose>=60]),mean(prison))-c(mean(dose),mean(prison))
Sx1Y=S0Y^exp(beta%*%x1)
Sx1Z=S0Z^exp(beta%*%x1)
Sx2Y=S0Y^exp(beta%*%x2)
Sx2Z=S0Z^exp(beta%*%x2)

xlm=range(timesY,timesZ)
ylm=range(Sx1Y,Sx1Z,Sx2Y,Sx2Z)

plot(0,0,pch='',xlab='t (years)',ylab=expression(hat(S)(t)),
xlim=xlm,ylim=ylm)
lines(timesY,Sx1Y,col=1,type='s')
lines(timesY,Sx2Y,col=2,type='s')
lines(timesZ,Sx1Z,col=1,type='s',lty=2)
This creates the following graph. From this plot we see that a high dose leads to a lower rate of departing the programme (as the red lines are higher than the black: mostly), while those addicts at clinic 2 also have lower drop out rates. After some time, those in clinic 1 with a high dose drop out even more than those at clinic 2 on the low dose. A similar plot can be obtained for the prison covariate, noting that there is no need to categorise it first.