Illustration 6

Aim: Explaining unemployment duration, measured in number of two weeks intervals (*spell*), as a function of potential determinants such as:

- ui (UI) = 1 if filed UI claim
- reprate (RR) = eligible replacement rate
- disrate (DR) = eligible disregard rate
- tenure (TENURE) = years tenure in lost job
- logwage (LOGWAGE) = log weekly earnings in lost job (1985$)

The duration is complete when the individual is re-employeed at a full-time job, that is, when CENSOR1 = 1

Details: Cameron and Trivedi (2005), ch. 17.11

Illustration 6 – Question 1

```
. sum spell censor1 censor2 censor3 censor4 ui reprate disrate tenure logwage
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       spell |      3343    6.247981    5.611271          1         28
     censor1 |      3343    .3209692    .4669188          0          1
     censor2 |      3343    .1014059    .3019106          0          1
     censor3 |      3343    .1717021    .3771777          0          1
     censor4 |      3343    .3754113    .4843014          0          1
-------------+--------------------------------------------------------
          ui |      3343    .5527969    .4972791          0          1
     reprate |      3343    .4544717    .1137918       .066      2.059
     disrate |      3343    .1094376    .0735274       .002       1.02
      tenure |      3343    4.114867    5.862322          0         40
     logwage |      3343    5.692994    .5356591    2.70805   7.600402
```

- Non-negative outcome
- 32.01% of the observations present complete durations, which means that they are not censored on the right (censored observations present *censor1=0*)

Illustration 6 – Question 2

```
.stset spell, fail(censor1=1)

     failure event:  censor1 == 1
obs. time interval:  (0, spell]
 exit on or before:  failure

--------------------------------------------------------------------------
     3343  total observations
        0  exclusions
--------------------------------------------------------------------------
     3343  observations remaining, representing
     1073  failures in single-record/single-failure data
    20887  total analysis time at risk and under observation
                                          at risk from t =         0
                                earliest observed entry t =         0
                                     last observed exit t =        28
```

Illustration 6 – Question 2

```
. sts list

         failure _d:  censor1 == 1
   analysis time _t:  spell
```

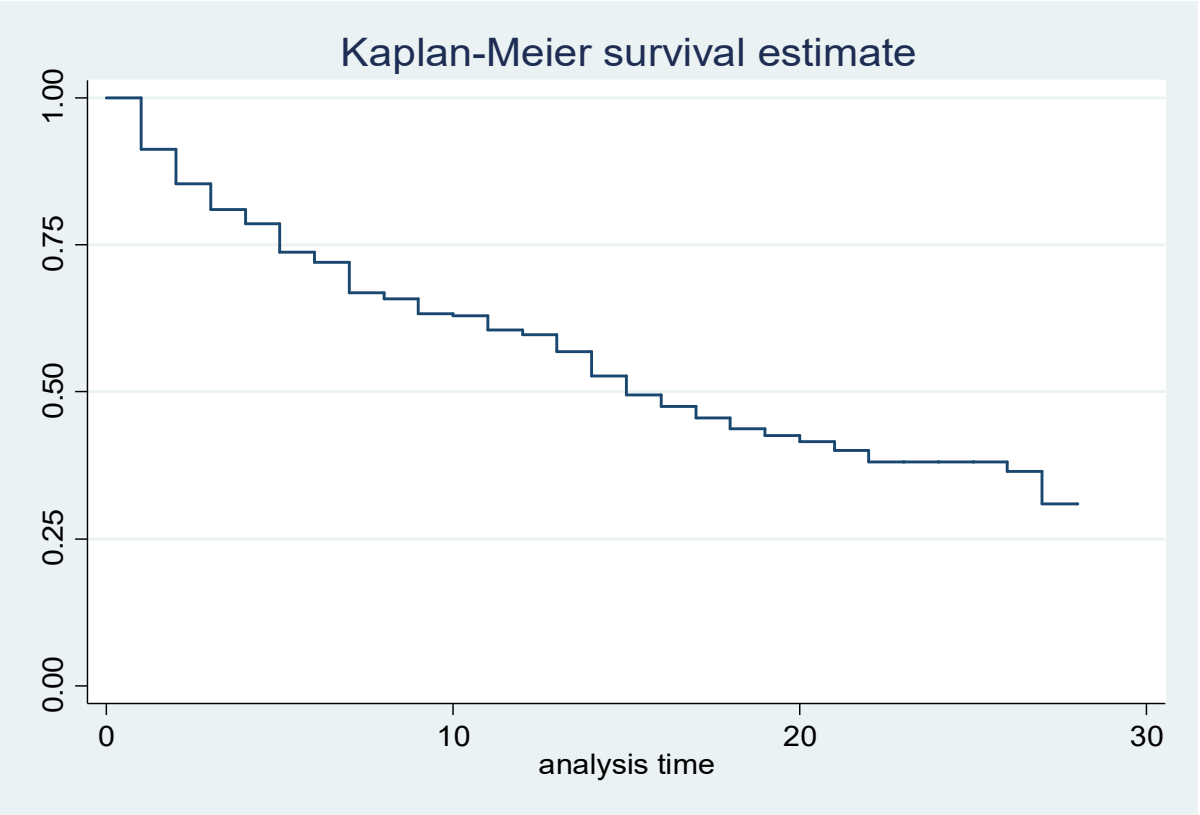
Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
1	3343	294	246	0.9121	0.0049	0.9019	0.9212
2	2803	178	304	0.8541	0.0062	0.8415	0.8659
3	2321	119	305	0.8103	0.0071	0.7960	0.8238
4	1897	56	165	0.7864	0.0076	0.7712	0.8008
5	1676	104	233	0.7376	0.0085	0.7206	0.7538
6	1339	32	111	0.7200	0.0088	0.7023	0.7369
7	1196	85	178	0.6688	0.0098	0.6492	0.6876
8	933	15	70	0.6581	0.0100	0.6380	0.6773
9	848	33	98	0.6325	0.0106	0.6113	0.6528
10	717	3	55	0.6298	0.0106	0.6086	0.6503
11	659	26	77	0.6050	0.0113	0.5825	0.6267

(…)

- (1-0.9121)*100%=8.8% of the individuals exit unemployement whithin the two first weeks and (1-0.8541)*100%= 14.6% exit unemployement within one month

Illustration 6 – Question 2

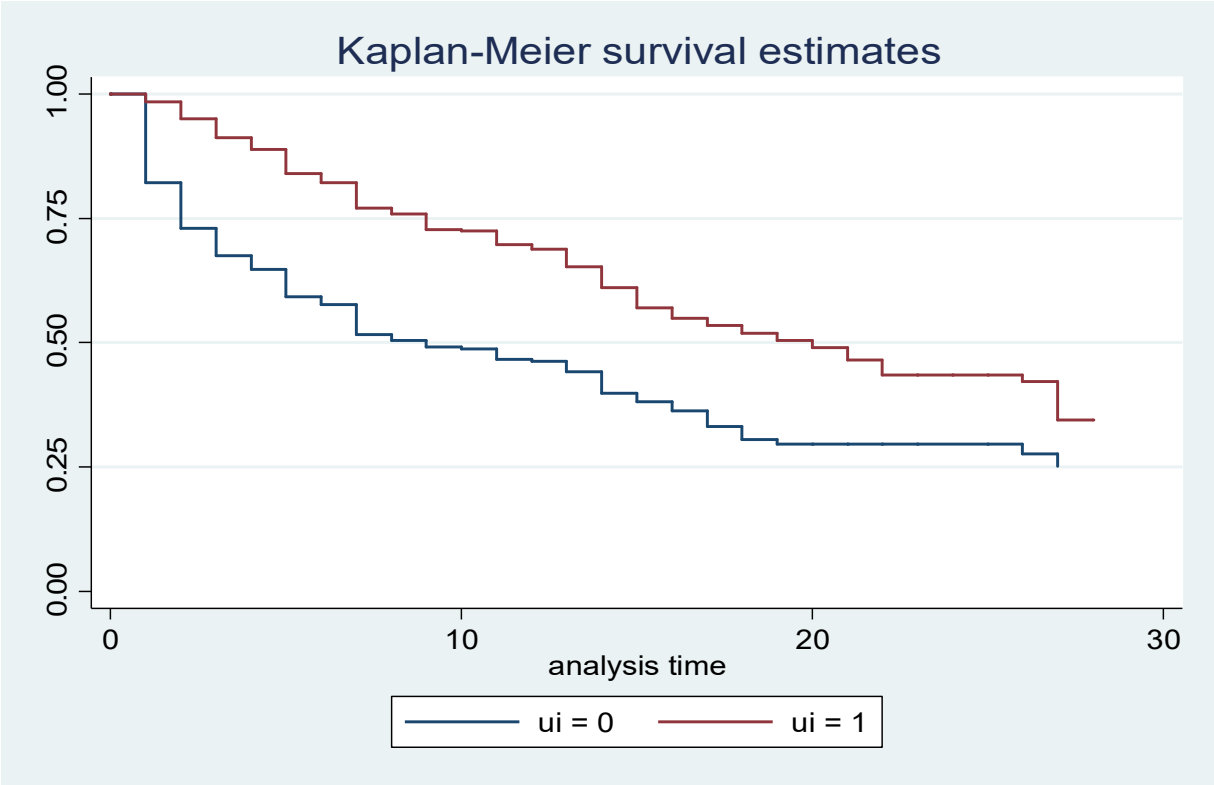
```
. sts graph
```



- Survival decreases more rapidly at first and then slowly

Illustration 6 – Question 2

```
. sts graph, by(ui)
```



- Those claiming unemployement insurance (ui=1) are more likely to remain unemployed

Illustration 6 – Question 3

```
. gen RR = reprate
. gen DR = disrate
. gen UI = ui
. gen RRUI = RR*UI
. gen DRUI = DR*UI
. gen LOGWAGE = logwage
```

```
. global xlist RR DR UI RRUI DRUI LOGWAGE tenure slack abolpos explose stateur
houshead married female child ychild nonwhite age schlt12 schgt12 smsa bluecoll
mining constr transp trade fire services pubadmin year85 year87 year89 midatl
encen wncen southatl escen wscen mountain pacific
```

Illustration 6 – Question 3

```
. streg $xlist, nohr robust dist(exponential)
          failure _d:  censor1 == 1
   analysis time _t:  spell
 (…)
Exponential regression -- log relative-hazard form
 No. of subjects      =         3343                 Number of obs    =      3343
No. of failures       =         1073
Time at risk          =        20887
                                                    Wald chi2(40)    =    565.24
Log pseudolikelihood =   -2700.6903                 Prob > chi2      =    0.0000
 ------------------------------------------------------------------------------
              |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          RR |   .4720235   .6005534     0.79   0.432    -.7050396    1.649087
          DR |  -.5756396   .7624489    -0.75   0.450    -2.070012    .9187327
          UI |  -1.424561   .2493917    -5.71   0.000     -1.91336   -.9357622
        RRUI |   .9655904   .6118408     1.58   0.115    -.2335956    2.164776
        DRUI |  -.1990635   1.019118    -0.20   0.845    -2.196498    1.798371
     LOGWAGE |   .3508005    .115598     3.03   0.002     .1242327    .5773684
…
       _cons |  -4.079107   .8767097    -4.65   0.000    -5.797426   -2.360788
------------------------------------------------------------------------------

estimates store bexponential
```

Illustration 6 – Question 3

```
. streg $xlist, nohr robust dist(weibull)
Weibull regression -- log relative-hazard form
No. of subjects      =         3343                 Number of obs    =      3343
…
                                                    Wald chi2(40)    =    501.65
Log pseudolikelihood =   -2687.5995                 Prob > chi2      =    0.0000
 ------------------------------------------------------------------------------
              |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          RR |   .4481156   .6381895     0.70   0.483    -.8027127    1.698944
          DR |  -.4269187   .8086983    -0.53   0.598    -2.011938    1.158101
          UI |  -1.496066   .2639679    -5.67   0.000    -2.013434   -.9786984
        RRUI |   1.015226   .6455611     1.57   0.116    -.2500501    2.280503
        DRUI |  -.2988417   1.065384    -0.28   0.779    -2.386956    1.789272
     LOGWAGE |   .3655253     .12212     2.99   0.003     .1261745    .6048761
(…)
       _cons |  -4.357886   .9196792    -4.74   0.000    -6.160424   -2.555347
-------------+----------------------------------------------------------------
       /ln_p |   .1215314   .0194374     6.25   0.000     .0834348    .1596281
-------------+----------------------------------------------------------------
           p |   1.129225   .0219492                      1.087014    1.173075
         1/p |   .8855632   .0172131                      .8524608    .9199511
------------------------------------------------------------------------------
 . estimates store bweibull
```

Illustration 6 – Question 3

```
. streg $xlist, nohr robust dist(gompertz)
(…)
Gompertz regression -- log relative-hazard form
No. of subjects      =         3343                 Number of obs    =      3343
No. of failures      =         1073
Time at risk         =        20887
                                                    Wald chi2(40)    =    529.75
Log pseudolikelihood =    -2700.605                 Prob > chi2      =    0.0000
------------------------------------------------------------------------------
             |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          RR |    .472405   .6033813     0.78   0.434    -.7102005    1.655011
          DR |  -.5627894   .7646131    -0.74   0.462    -2.061404    .9358247
          UI |  -1.428355   .2508349    -5.69   0.000    -1.919982   -.9367272
        RRUI |   .9689413   .6144464     1.58   0.115    -.2353514    2.173234
        DRUI |  -.2112495   1.021112    -0.21   0.836    -2.212593    1.790094
     LOGWAGE |   .3524722   .1162698     3.03   0.002     .1245876    .5803567
(…)
       _cons |   -4.09733   .8802997    -4.65   0.000    -5.822686   -2.371975
-------------+----------------------------------------------------------------
      /gamma |    .002658   .0067759     0.39   0.695    -.0106225    .0159386
------------------------------------------------------------------------------

. estimates store bgompertz
```

Illustration 6 – Question 3

```
 . estimates table bexponential bweibull bgompertz, b star(0.1 0.05 0.01)
-------------------------------------------------------------
    Variable | bexponential      bweibull        bgompertz
-------------+-----------------------------------------------
          RR |   .47202347        .4481156       .47240504
          DR |  -.57563962      -.42691874      -.56278942
          UI |  -1.4245611***    -1.496066***   -1.4283547***
        RRUI |   .96559044       1.0152264       .96894134
        DRUI |  -.19906351      -.29884166      -.21124952
     LOGWAGE |   .35080054***    .36552527***    .35247218***
(…)
       _cons |  -4.0791071***   -4.3578855***   -4.0973303***
-------------+-----------------------------------------------
ln_p   _cons |                   .12153144***
-------------+-----------------------------------------------
gamma  _cons |                                    .00265803
-------------------------------------------------------------
                          legend: * p<.1; ** p<.05; *** p<.01
```

- Weibull is preferred to exponential ($\alpha = \exp(0.122) = 1.129$): the probability of the spell terminating increases for longer spells
- In all the models only UI and LOGWAGE are significant, with little variation across the three models:
  - For those claiming insurance the hazard rate changes $[exp(-1.496) - 1] * 100\% = -77{,}6\%$

Illustration 6 – Question 3

```
. stcox $xlist, nohr robust
(…)
Cox regression -- Breslow method for ties

No. of subjects      =         3343                 Number of obs    =      3343
No. of failures      =         1073
Time at risk         =        20887
                                                    Wald chi2(40)    =    540.98
Log pseudolikelihood =   -7717.2334                 Prob > chi2      =    0.0000
------------------------------------------------------------------------------
             |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          RR |   .5222796   .5711698     0.91   0.361    -.5971926    1.641752
          DR |   -.752507     .72175    -1.04   0.297    -2.167111    .6620971
          UI |  -1.317719   .2372893    -5.55   0.000    -1.782798   -.8526409
        RRUI |   .8822462    .582115     1.52   0.130    -.2586783    2.023171
        DRUI |  -.0951357    .977774    -0.10   0.922    -2.011538    1.821266
     LOGWAGE |   .3352639   .1106483     3.03   0.002     .1183972    .5521306
(…)
------------------------------------------------------------------------------

. estimates store bcox
```

Illustration 6 – Question 4

```
. estimates table bcox, b star(0.1 0.05 0.01) keep(RR DR UI RRUI DRUI LOGWAGE)

------------------------------
    Variable |     bcox
-------------+----------------
          RR |  .52227964
          DR | -.75250697
          UI | -1.3177194***
        RRUI |  .88224619
        DRUI | -.09513573
     LOGWAGE |  .33526391***
------------------------------
legend: * p<.1; ** p<.05; *** p<.01
```

- The conclusions do not change

Illustration 6 – Question 5

The Figures are produced using the code in CT

```
. streg $xlist, nohr robust dist(exponential)
. predict resid, csnell
. stset resid, fail(censor1)
. sts generate survivor=s
. generate cumhaz = -ln(survivor)
. sort resid

. graph twoway (scatter cumhaz resid, c(J) msymbol(i) msize(small) clstyle(p1))(scatter
resid resid, c(l) msymbol(i) msize(small) clstyle(p2)),scale (1.2) plotregion(style(none))
title("Exponential Model Residuals") xtitle("Generalized (Cox-Snell) Residual",
size(medlarge)) xscale(titlegap(*5)) ytitle("Cumulative Hazard", size(medlarge))
yscale(titlegap(*5)) legend(pos(6) ring(0) col(1)) legend(size(small)) legend( label(1
"Cumulative Hazard") label(2 "45 degree line"))

. graph export exp.wmf, replace

. drop resid survivor cumhaz
```

Illustration 6 – Question 5

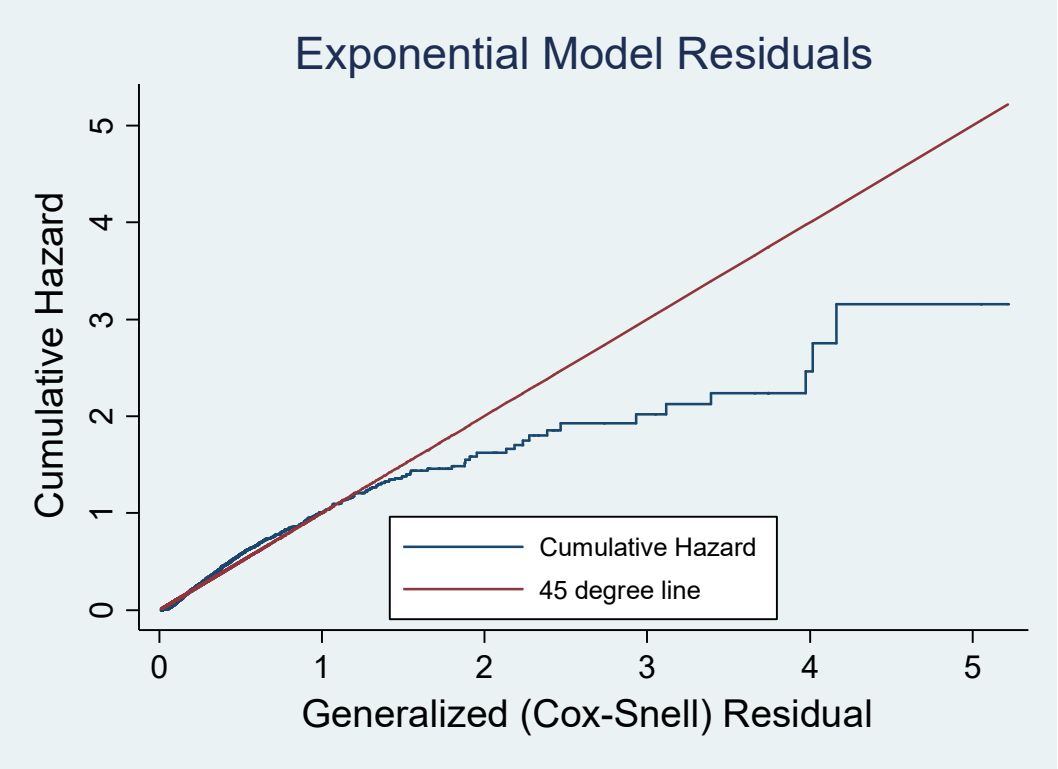
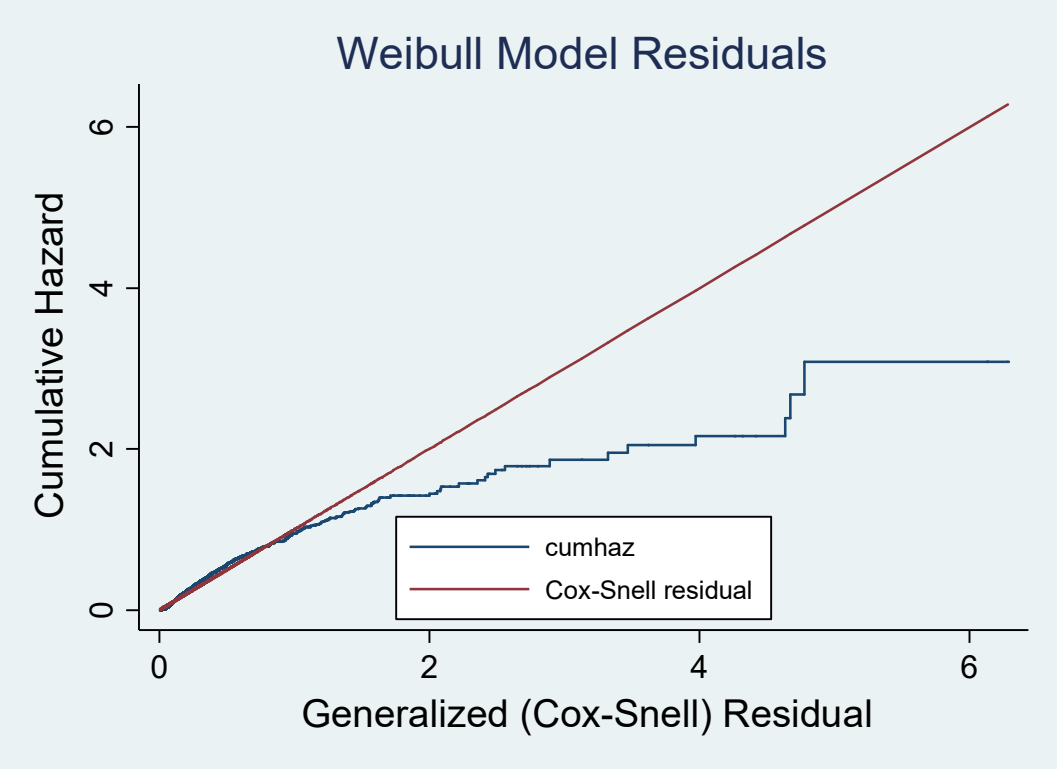


Illustration 6 – Question 5

```
stset spell, fail(censor1=1)
streg $xlist, nolog nohr dist(weibull) robust
estimates store bweib

predict resid, csnell
stset resid, fail(censor1)
sts generate survivor=s
generate cumhaz = -ln(survivor)
sort resid

graph twoway (scatter cumhaz resid, c(J) msymbol(i) msize(small) clstyle(p1))
(scatter resid resid, c(l) msymbol(i) msize(small) clstyle(p2)), scale (1.2)
plotregion(style(none)) title("Weibull Model Residuals") xtitle("Generalized
(Cox-Snell) Residual", size(medlarge)) xscale(titlegap(*5)) ytitle("Cumulative
Hazard", size(medlarge)) yscale(titlegap(*5)) legend(pos(6) ring(0) col(1))
legend(size(small))
legend( label(1 "Cumulative Hazard") label(2 "45 degree line"))

graph export Weibul16.wmf, replace

drop resid survivor cumhaz
```

Illustration 6 – Question 5



- In both the exponential and the Weibull case the residuals are not close to the reference line

Illustration 6 – Question 6

```
stset spell, fail(censor1)
streg $xlist, nolog nohr dist(exponential) frailty(gamma) robust
estimates store bexpgamma

predict resid, csnell
stset resid, fail(censor1)
sts generate survivor=s
generate cumhaz = -ln(survivor)
sort resid

graph twoway (scatter cumhaz resid, c(J) msymbol(i) msize(small) clstyle(p1))
(scatter resid resid, c(l) msymbol(i) msize(small) clstyle(p2)), scale (1.2)
plotregion(style(none)) title("Exponential-Gamma Model Residuals")
xtitle("Generalized (Cox-Snell) Residual", size(medlarge)) xscale(titlegap(*5))
ytitle("Cumulative Hazard", size(medlarge)) yscale(titlegap(*5)) legend(pos(6)
ring(0) col(1)) legend(size(small)) legend( label(1 "Cumulative Hazard")
label(2 "45 degree line"))

graph export exp_gamma.wmf, replace

drop resid survivor cumhaz
```

Illustration 6 – Question 6

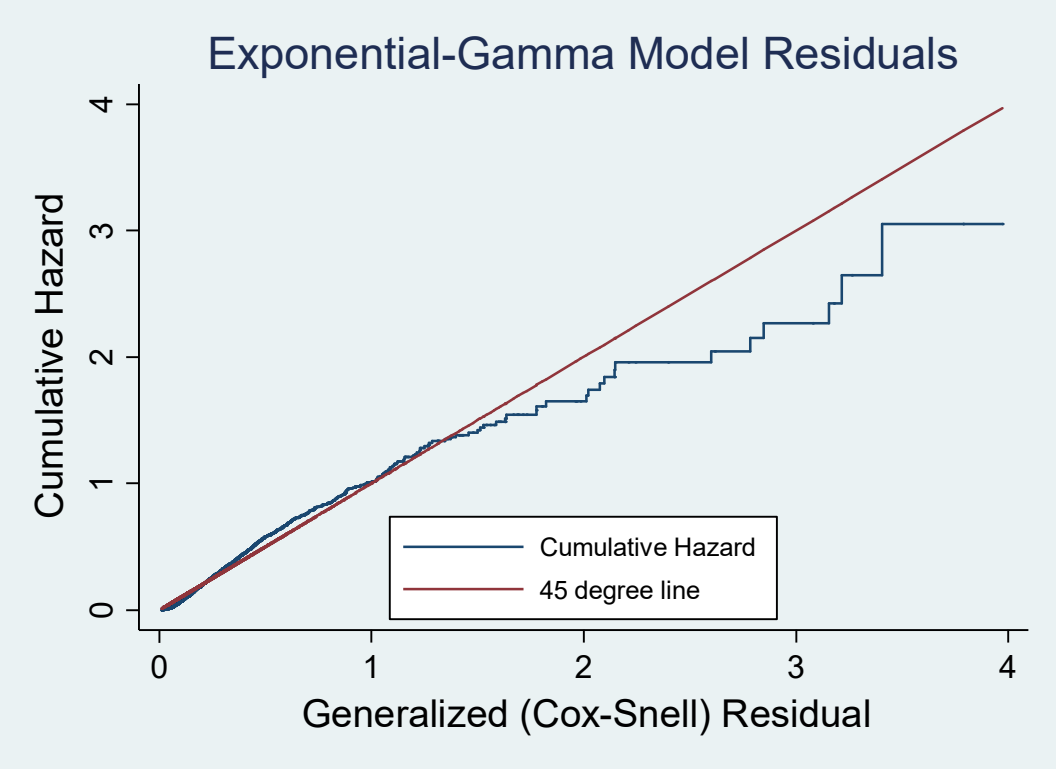
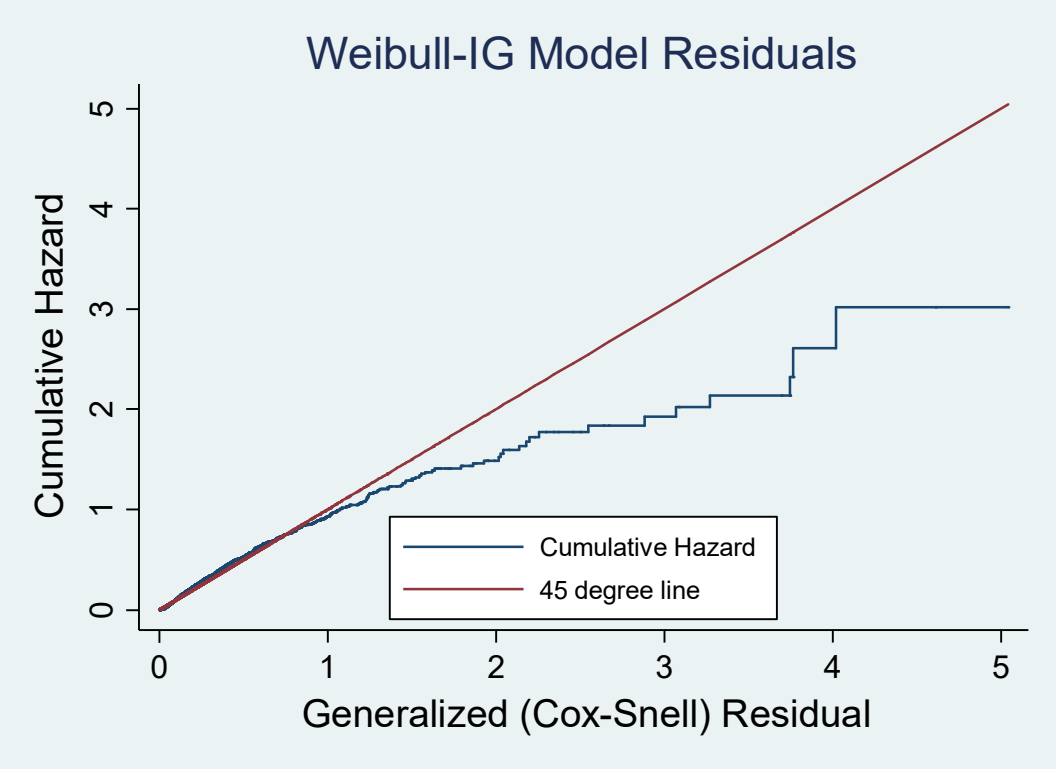


Illustration 6 – Question 6

```
stset spell, fail(censor1=1)
streg $xlist, nolog nohr dist(weibull) frailty(invgauss) robust
estimates store bweibIG

predict resid, csnell
stset resid, fail(censor1)
sts generate survivor=s
generate cumhaz = -ln(survivor)
sort resid
graph twoway (scatter cumhaz resid, c(J) msymbol(i) msize(small) clstyle(p1))
(scatter resid resid, c(l) msymbol(i) msize(small) clstyle(p2)), scale (1.2)
plotregion(style(none)) title("Weibull-IG Model Residuals") xtitle("Generalized
(Cox-Snell) Residual", size(medlarge)) xscale(titlegap(*5)) ytitle("Cumulative
Hazard", size(medlarge)) yscale(titlegap(*5)) legend(pos(6) ring(0) col(1))
legend(size(small)) legend( label(1 "Cumulative Hazard") label(2 "45 degree
line"))

graph export Weibul16_IG.wmf, replace

drop resid survivor cumhaz
```

Illustration 6 – Question 6



- Improvements relative to baseline cases where heterogeneity is overlooked are not relevant. Mispecification is apparent. In fact, the original paper using these data uses a more flexible hazard function

Illustration 6 – Question 6

```
. estimates table bexpgamma bweibIG, b star(0.1 0.05 0.01)
-------------------------------------------
    Variable |   bexpgamma        bweibIG
-------------+-----------------------------
          RR |   .50058284       .73562769
          DR |  -.88244688      -1.0725662
          UI |  -1.5845375***    -2.574752***
        RRUI |   1.0911676*      1.7335706*
        DRUI |   .05740483        -.060621
     LOGWAGE |   .37928053***    .57565599***
...
-------------+-----------------------------
ln_the _cons |  -1.4629948***    1.8526958***
-------------+-----------------------------
ln_p   _cons |                   .56116668***
-------------------------------------------
           legend: * p<.1; ** p<.05; *** p<.01
```

Note that p corresponds to α, which is obtained from ln(p) as exp(0.561)=1.753, σ^2 is theta and thus $\sigma^2 = \exp(1.853) = 6.379$. The duration dependence was underestimated when heterogeneity is ignored.