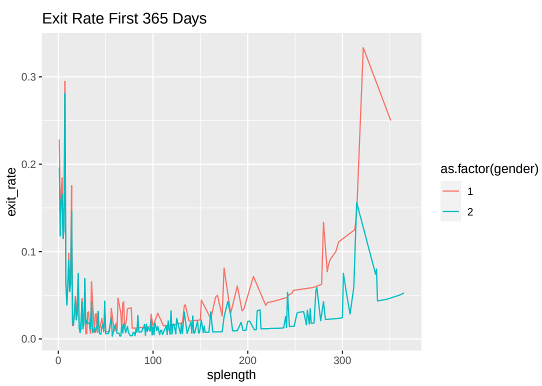
Applied Microeconometrics - Assignment 4

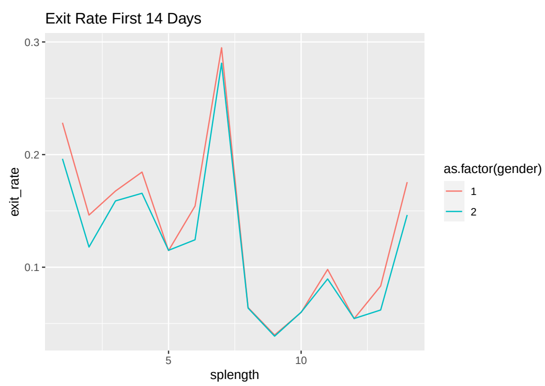
Walter Verwer (589962) & Bas Machielsen (590049)

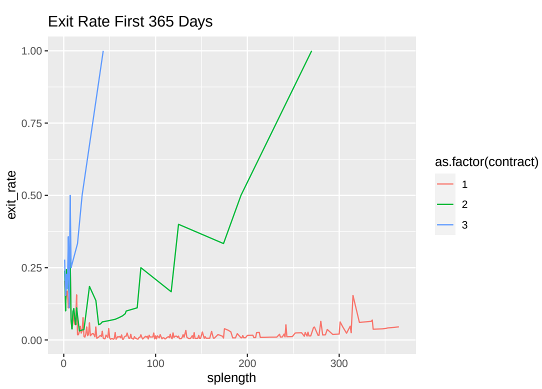
October 5, 2021

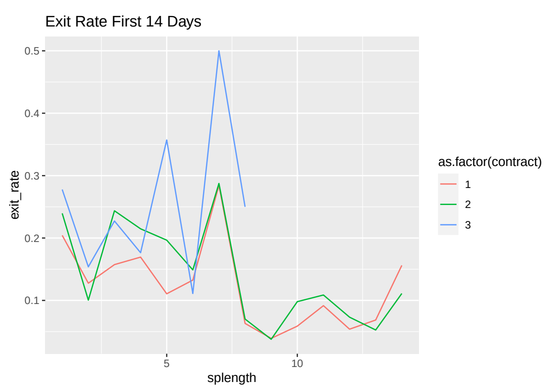
1. Describe the sickness spell data, i.e. do a simple listing of the survivor function and plot the hazard rate and the survivor function. Make separate plots for the first two weeks and for the first year. Also plot the hazard by different subgroups (for instance gender) and test whether the survival curves are the same for the different subgroups.

First, we plot the exit rates:

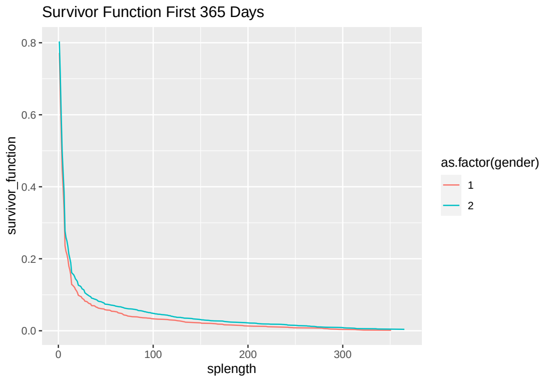


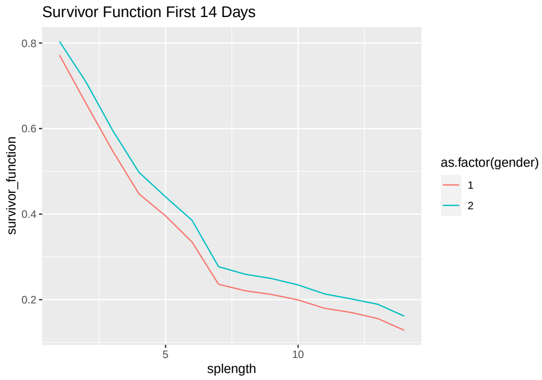


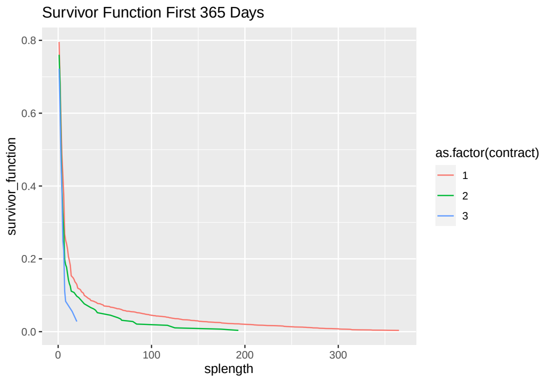


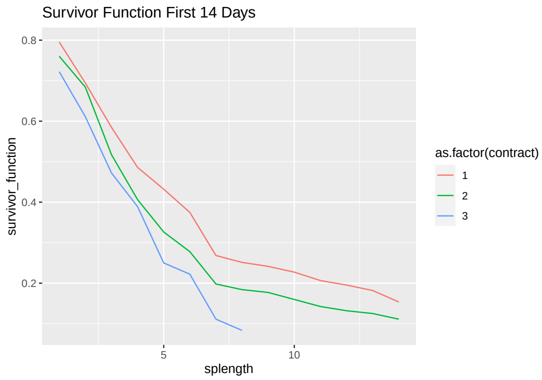


Then, we plot the survivor functions:


Survivor Function First 365 Days
survivor_function
splength
as.factor(gender)
1
2


Survivor Function First 14 Days
survivor_function
splength
as.factor(gender)
1
2


Survivor Function First 365 Days
survivor_function
splength
as.factor(contract)
1
2
3


Survivor Function First 14 Days
survivor_function
splength
as.factor(contract)
1
2
3

2.1 Estimate a Weibull and an Exponential model for sickness spells. Start with a very simple specification and you only include one regressor and subsequently add more regressors. Comment on the change in the Weibull parameters and the regression parameters when you add more variables to the model. Compare the estimates of both models.

Table 1: Weibull Models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	1.938***	1.795***	2.280***	2.279***	2.334***	0.979*
	(0.127)	(0.195)	(0.213)	(0.213)	(0.251)	(0.530)
gender	0.243***	0.254***	0.273***	0.274***	0.278***	0.319***
	(0.077)	(0.078)	(0.078)	(0.087)	(0.087)	(0.092)
Log(scale)	0.490***	0.490***	0.487***	0.487***	0.486***	0.483***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
marstat		0.069	0.055	0.055	0.049	0.031
		(0.066)	(0.066)	(0.066)	(0.066)	(0.068)
contract			−0.467***	−0.467***	−0.467***	−0.423***
			(0.093)	(0.093)	(0.094)	(0.093)
lowgroup				−0.004	−0.005	−0.020
				(0.079)	(0.079)	(0.080)
classize					0.002	0.002
					(0.005)	(0.005)
schsize					0.000	0.000
					(0.000)	(0.000)
public					−0.056	−0.048
					(0.091)	(0.089)
protest					−0.153	−0.078
					(0.106)	(0.108)
merged						0.005
						(0.013)
avgfem						−0.206
						(0.284)
avgage						0.034***
						(0.011)
avglowgr						0.122
						(0.259)
Num.Obs.	6520	6520	6520	6520	6520	6520
AIC	44 443.9	44 442.6	44 415.1	44 417.1	44 416.8	44 383.2
Log.Lik.	−22 218.934	−22 217.295	−22 202.540	−22 202.537	−22 198.402	−22 177.593

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Exponential Models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	2.435***	2.247***	2.852***	2.848***	2.909***	1.601**
	(0.164)	(0.236)	(0.255)	(0.254)	(0.290)	(0.633)
gender	0.267***	0.280***	0.303***	0.324***	0.332***	0.377***
	(0.097)	(0.098)	(0.097)	(0.104)	(0.103)	(0.108)
marstat		0.091	0.077	0.076	0.077	0.058
		(0.080)	(0.079)	(0.079)	(0.078)	(0.080)
contract			−0.594***	−0.595***	−0.600***	−0.537***
			(0.112)	(0.111)	(0.112)	(0.116)
lowgroup				−0.044	−0.043	−0.063
				(0.096)	(0.097)	(0.098)
classize					0.003	0.003
					(0.005)	(0.005)
schsize					0.000	0.000
					(0.000)	(0.000)
public					−0.116	−0.093
					(0.103)	(0.104)
protest					−0.160	−0.083
					(0.124)	(0.128)
merged						−0.004
						(0.015)
avgfem						−0.237
						(0.335)
avgage						0.033***
						(0.012)
avglowgr						0.091
						(0.297)
Num.Obs.	6520	6520	6520	6520	6520	6520
AIC	49 147.6	49 133.1	49 022.5	49 022.1	48 996.4	48 893.8
Log.Lik.	−24 571.812	−24 563.554	−24 507.235	−24 506.073	−24 489.220	−24 433.900

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.2 Estimate separate Weibull models for males and females. Comment on the results (is it better to estimate separate models for males and females?) Estimate the Weibull duration model for other subgroups that may differ in their behavior and where the baseline hazard may differ.

First, we estimate a model for males:

Table 3: Weibull Models - Males Only

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	2.204***	2.446***	3.162***	3.149***	3.363***	1.558*
	(0.048)	(0.329)	(0.394)	(0.394)	(0.419)	(0.904)
Log(scale)	0.461***	0.460***	0.458***	0.457***	0.456***	0.449***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
marstat		−0.125	−0.138	−0.138	−0.150	−0.149
		(0.168)	(0.165)	(0.166)	(0.161)	(0.157)
contract			−0.675***	−0.685***	−0.631***	−0.559***
			(0.169)	(0.174)	(0.179)	(0.188)
lowgroup				0.083	0.084	0.094
				(0.122)	(0.122)	(0.123)
classize					−0.009	−0.008
					(0.007)	(0.007)
schsize					0.000	0.000
					(0.000)	(0.000)
public					−0.073	−0.036
					(0.128)	(0.132)
protest					0.009	0.032
					(0.180)	(0.174)
merged						0.035
						(0.028)
avgfem						0.457
						(0.400)
avgage						0.041**
						(0.016)
avglowgr						−0.357
						(0.410)
Num.Obs.	2046	2046	2046	2046	2046	2046
AIC	13 422.3	13 422.4	13 416.2	13 417.1	13 421.4	13 405.4
Log.Lik.	−6709.152	−6708.217	−6704.101	−6703.534	−6701.689	−6689.697

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Next, we estimate a model for females:

Table 4: Weibull Models - Females Only

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	2.414***	2.228***	2.715***	2.755***	2.763***	1.729***
	(0.042)	(0.138)	(0.174)	(0.197)	(0.260)	(0.643)
Log(scale)	0.503***	0.502***	0.499***	0.499***	0.498***	0.494***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
marstat		0.106	0.093	0.092	0.076	0.058
		(0.074)	(0.074)	(0.074)	(0.073)	(0.076)
contract			−0.435***	−0.435***	−0.433***	−0.393***
			(0.103)	(0.102)	(0.106)	(0.103)
lowgroup				−0.048	−0.038	−0.081
				(0.099)	(0.099)	(0.103)
classize					0.005	0.006
					(0.006)	(0.006)
schsize					0.000	0.000
					(0.000)	(0.000)
public					−0.039	−0.030
					(0.116)	(0.114)
protest					−0.211*	−0.123
					(0.127)	(0.133)
merged						−0.008
						(0.014)
avgfem						−0.557
						(0.383)
avgage						0.030**
						(0.013)
avglowgr						0.403
						(0.334)
Num.Obs.	4474	4474	4474	4474	4474	4474
AIC	31 018.2	31 013.9	30 994.0	30 995.4	30 991.3	30 969.4
Log.Lik.	−15 507.109	−15 503.936	−15 493.017	−15 492.706	−15 486.663	−15 471.686

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Now, given that under question 1 we observed a difference between the exit rates and the survival functions for the different contract types, we have estimated Weibull models for the contract types. However, our results are difficult to compare. The reason for this is that the models for the temporary and mixed contract have very little observations. Especially, the mixed contract has too little (36 observations). For this reason we can not make an educated comparison for this subgroup.

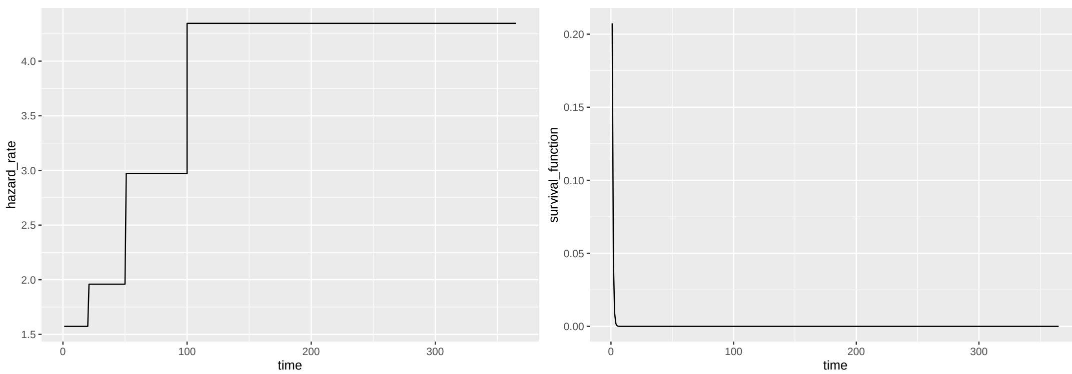
Table 5: Weibull Models - Models for the all contract types

	Fixed contract	Temporary contract	Mixed contract
	Model 1	Model 2	Model 3
(Intercept)	0.526	0.290	2.377
	(0.523)	(1.742)	(2.100)
marstat	0.019	0.439**	−0.323
	(0.070)	(0.220)	(0.225)
gender	0.304***	0.781**	0.616
	(0.095)	(0.309)	(0.629)
lowgroup	−0.010	−0.083	−0.089
	(0.083)	(0.271)	(0.739)
classize	0.003	0.002	0.008
	(0.005)	(0.009)	(0.035)
schsize	0.000	0.000	0.005*
	(0.000)	(0.001)	(0.003)
public	−0.041	−0.087	−0.464**
	(0.090)	(0.264)	(0.189)
protest	−0.085	0.154	0.367
	(0.108)	(0.372)	(0.522)
merged	0.002	0.056	−0.013
	(0.014)	(0.046)	(0.052)
avgfem	−0.172	−0.796	−1.661*
	(0.285)	(0.934)	(1.003)
avgage	0.035***	−0.013	−0.045
	(0.011)	(0.038)	(0.040)
avglowgr	0.080	0.978	0.581
	(0.262)	(0.820)	(1.867)
Log(scale)	0.490***	0.317***	−0.226*
	(0.014)	(0.050)	(0.127)
Num.Obs.	6196	288	36
AIC	42 336.4	1838.7	202.8
Log.Lik.	−21 155.217	−906.333	−88.386

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.1 Estimate a Piece Wise Constant (PWC) model for the entire sample. Use the stsplit command to create multiple record data. You can have as many steps as the data allow you to take, but first start with only a few (3 or 4 steps). Next estimate a model with 15-20 steps, or even more. Plot de duration pattern implied by the estimates and comment on these and the regression parameters. How do the regression parameters (β) compare with those of the Weibull model?

First, we estimate a piece wise constant model.



Above, we plot the hazard rate and survival function of our estimated piecewise constant model. The piecewise-constant model coefficient estimates can be found in the table below. As the attrition from the sample is very high in the starting period, our Weibull model predicts a high hazard rate, and thus high attrition in the first period. Afterwards, piecewise, it attempts to catch up the imbalances by fitting a new Weibull distribution and hazard rate (implied by the new constant term), which significantly increases the hazard rate relative to the previous period. This proves the model does catch up with the decay in the data that is not as fastly decaying as implied by a Weibull model under the previous parameters.

One other consequence of including a dummy for the first twenty days is an underestimation of the survival probability in the early period. Even though it mildly realistic, it underestimates survival by only focusing on the first 20 days, under which attrition is very high. We repeat this exercise again with a piecewise constant model with more dummies, but omit the plot for brevity's sake.

Table 6: PWC (Weibull)

	PWC 14 steps	PWC 4 steps
	Model 1	Model 2
(Intercept)	1.714***	1.573***
	(0.019)	(0.018)
timeperiod 10	4.399***	
	(0.198)	
timeperiod 11	4.242***	
	(0.160)	
timeperiod 12	4.665***	
	(0.195)	
timeperiod 13	4.108***	
	(0.020)	
timeperiod 14	5.298***	
	(0.249)	
timeperiod 2	1.954***	1.959***
	(0.030)	(0.030)
timeperiod 3	2.582***	2.972***
	(0.042)	(0.047)
timeperiod 4	3.010***	4.346***
	(0.059)	(0.075)
timeperiod 5	3.261***	
	(0.066)	
timeperiod 6	3.573***	
	(0.097)	
timeperiod 7	3.930***	
	(0.127)	
timeperiod 8	3.630***	
	(0.067)	
timeperiod 9	4.315***	
	(0.208)	
Log(scale)	−0.236***	
	(0.010)	
Num.Obs.	6473	6520
AIC	35 555.5	36 158.3
Log.Lik.	−17 762.772	−18 075.161

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.2 Estimate separate models for males and females.

First we estimate a PWC model for males.

Table 7: PWC (Weibull) - Males only

	PWC 14 steps	PWC 4 steps
	Model 1	Model 2
(Intercept)	-850.507^{***}	-1.016
	(0.322)	(0.958)
marstat	-0.017	0.260^{**}
	(0.060)	(0.130)
contract	853.775^{***}	5.494^{***}
	(0.000)	(0.661)
merged	0.006	0.017
	(0.014)	(0.022)
lowgroup	0.067	-0.082
	(0.045)	(0.097)
classize	0.001	-0.008
	(0.003)	(0.007)
schsize	0.000	0.000
	(0.000)	(0.000)
public	0.075	-0.103
	(0.054)	(0.165)
protest	-0.049	-0.341^{**}
	(0.052)	(0.145)
avgfem	-0.009	-0.445
	(0.190)	(0.457)
avgage	0.003	-0.012
	(0.006)	(0.015)
avglowgr	-0.086	0.172
	(0.180)	(0.338)
timeperiod 10	2.133^{***}	
	(0.102)	
timeperiod 11	2.191^{***}	
	(0.105)	
timeperiod 12	2.437^{***}	
	(0.090)	
timeperiod 14	3.320^{***}	
	(0.151)	
timeperiod 2	0.530^{***}	0.256
	(0.070)	(0.222)
timeperiod 3	1.068^{***}	0.807^{***}
	(0.096)	(0.225)
timeperiod 4	1.217^{***}	1.813^{***}
	(0.087)	(0.240)
timeperiod 5	1.558^{***}	
	(0.103)	
timeperiod 6	1.688^{***}	
	(0.087)	
timeperiod 7	1.832^{***}	
	(0.068)	
timeperiod 8	2.041^{***}	
	(0.131)	
timeperiod 9	1.997^{***}	
	(0.072)	
Log(scale)	-1.998^{***}	-1.072^{***}
	(0.210)	(0.114)
Num.Obs.	2029	2036
AIC	597.1	725.3
Log.Lik.	-273.545	-346.670

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Now we estimate the same model for females.

Table 8: PWC (Weibull) - Females only

	PWC 14 steps	PWC 4 steps
	Model 1	Model 2
(Intercept)	3.084***	3.780***
	(0.220)	(0.615)
marstat	0.019	0.097*
	(0.019)	(0.059)
contract	0.150	0.309
	(0.099)	(0.251)
merged	0.002	0.001
	(0.007)	(0.019)
lowgroup	−0.073*	−0.172
	(0.038)	(0.125)
classize	−0.001	−0.002
	(0.001)	(0.004)
schsize	0.000	0.000
	(0.000)	(0.000)
public	0.045	0.014
	(0.032)	(0.103)
protest	0.104**	0.137
	(0.045)	(0.143)
avgfem	−0.151	−0.160
	(0.103)	(0.347)
avgage	0.004	0.008
	(0.004)	(0.010)
avglowgr	0.198**	0.226
	(0.086)	(0.326)
timeperiod 10	2.116***	
	(0.057)	
timeperiod 11	2.302***	
	(0.064)	
timeperiod 12	2.437***	
	(0.058)	
timeperiod 13	42.067***	
	(0.000)	
timeperiod 14	3.169***	
	(0.119)	
timeperiod 2	0.591***	0.045
	(0.051)	(0.168)
timeperiod 3	1.002***	0.522***
	(0.060)	(0.174)
timeperiod 4	1.297***	1.500***
	(0.054)	(0.195)
timeperiod 5	1.509***	
	(0.053)	
timeperiod 6	1.627***	
	(0.058)	
timeperiod 7	1.749***	
	(0.053)	
timeperiod 8	2.137***	
	(0.099)	
timeperiod 9	2.069***	
	(0.070)	
Log(scale)	−1.882***	−0.874***
	(0.141)	(0.060)
Num.Obs.	4444	4456
AIC	1619.9	1922.2
Log.Lik.	−783.936	−945.106

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4. Estimate a Cox model and compare the most elaborate specification with the results of the PWC model
5. Repeat the procedure of question 2 for a Weibull model with (e.g., gamma) unobserved heterogeneity. Compare the estimates of the regression coefficients across the models with and without unobserved heterogeneity.

As above, but now with a Piecewise Constant (PWC) specification, where you have an elaborate specification of the baseline hazard (say, 20 dummies).

Compare the estimates of the Cox model (question 4) with the results of the PWC model with unobserved heterogeneity.

6. Multiple Spells

Estimate a standard Cox model (PL) and estimate Stratified Cox models (SPL). Concerning the latter, estimate SPL models, where the school is the stratum and estimate one where the teacher is the stratum. Comment on the teacher SPL approach. Compare the PL and the school SPL estimates. Can you think of a test to test for the relevance of using the school SPL (rather than doing the PL)?

Estimate a model with school specific dummies and compare these estimates with those obtained from the school SPL.

Observed sickness patterns vary between schools. This may be due to sorting effects (bad teachers are the reason why the school scores bad in absenteeism) and/or the school effects (it is elements of the school that make some schools worse than others. Can you think of a test/procedure to shed some more light on this issue?