

# Survival Analysis Example

Ani Katchova

## Survival Analysis Example

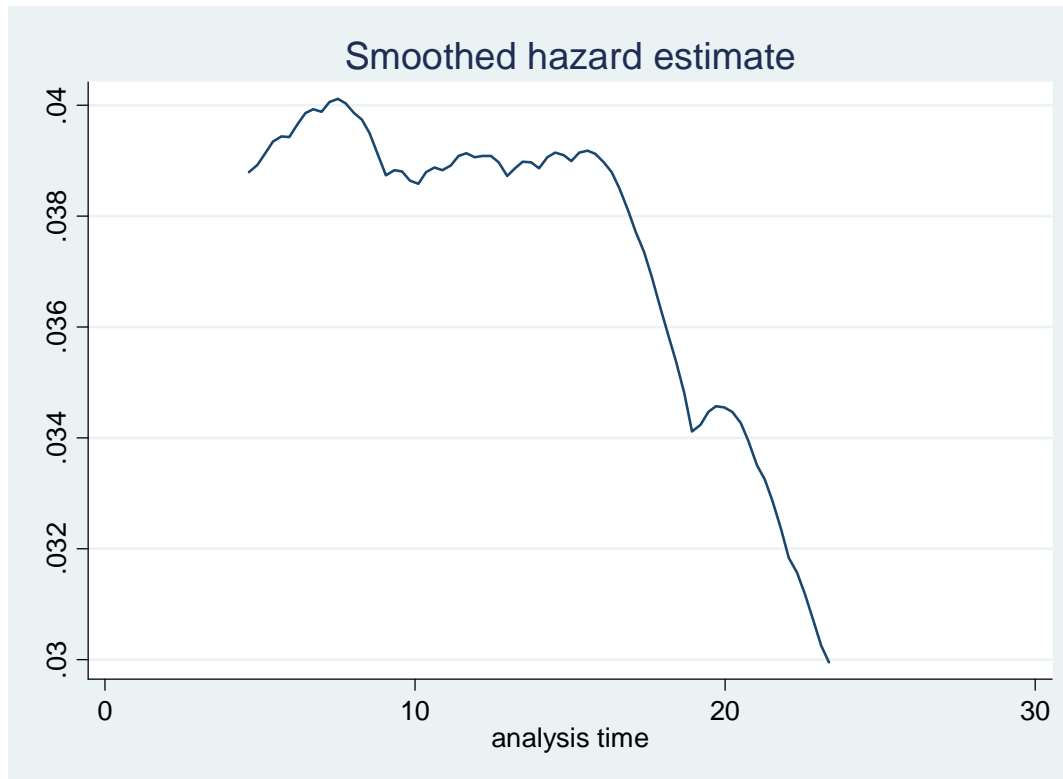
- Survival, hazard, and cumulative hazard functions
  - Nonparametric analysis (Kaplan-Meier survival function)
  - Parametric models (Exponential, Weibull, Gompertz, and Log-logistic)
  - Semi-parametric models (Cox proportional hazard model)
- 
- We want to study the unemployment duration – the length of time it takes someone to find a full-time job.
  - Data from the January Current Population Survey's Displaced Workers Supplements (DWS) for years 1986, 1988, 1990 and 1992.
  - Dependent variable: duration (number of periods being unemployed), event (finding a job)
  - Independent variables: log wage, claim unemployment insurance, and age.
  - Summary statistics: Subjects tracked from 1 to 28 periods. They either find a job (event) or are still looking (censored). Number of subjects is 3,343; time at risk (periods summed over the subjects) is 20,887. Number of failures is 1,073 or 32% of sample has failed. Incidence rate is 5.13% which is the number of failures divided by the time at risk.

Survival function table

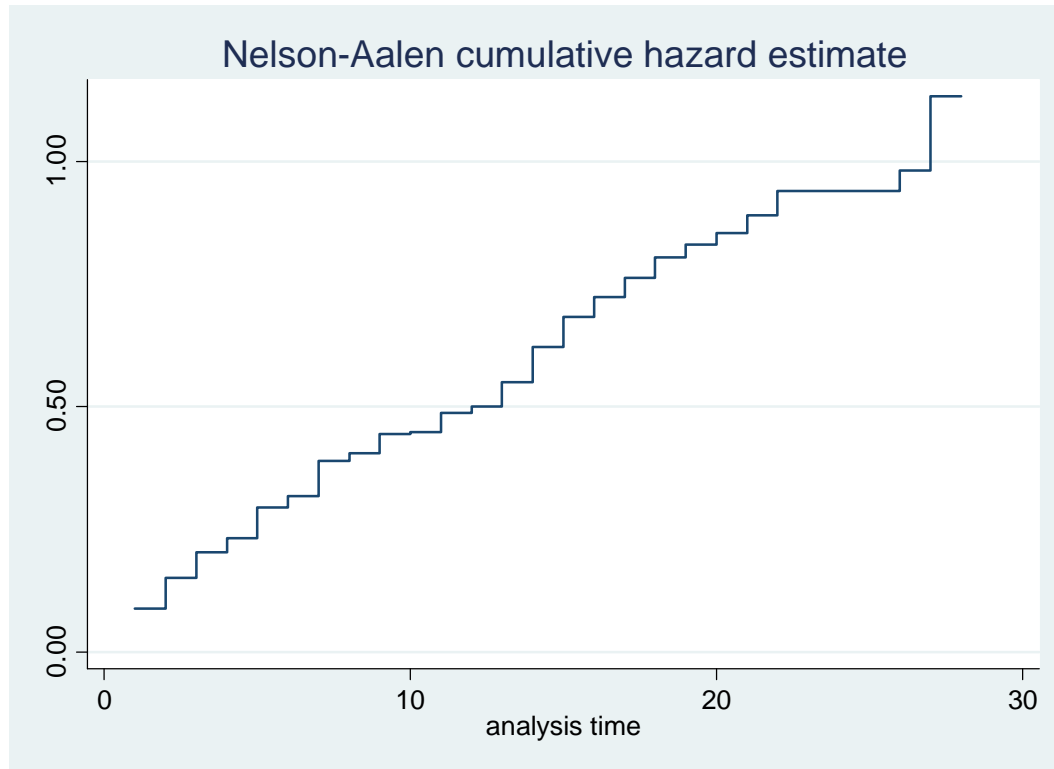
Time	Number of subjects	Failure/event	Net lost/censored	Survival Function
1	3343	294	246	0.91
2	2803	178	304	0.85
3	2321	119	305	0.81
4	1897	56	165	0.79
5	1676	104	233	0.74
6	1339	32	111	0.72
7	1196	85	178	0.67
....				
25	58	0	10	0.38
26	48	2	13	0.37
27	33	5	24	0.31
28	4	0	4	0.31

## Nonparametric methods

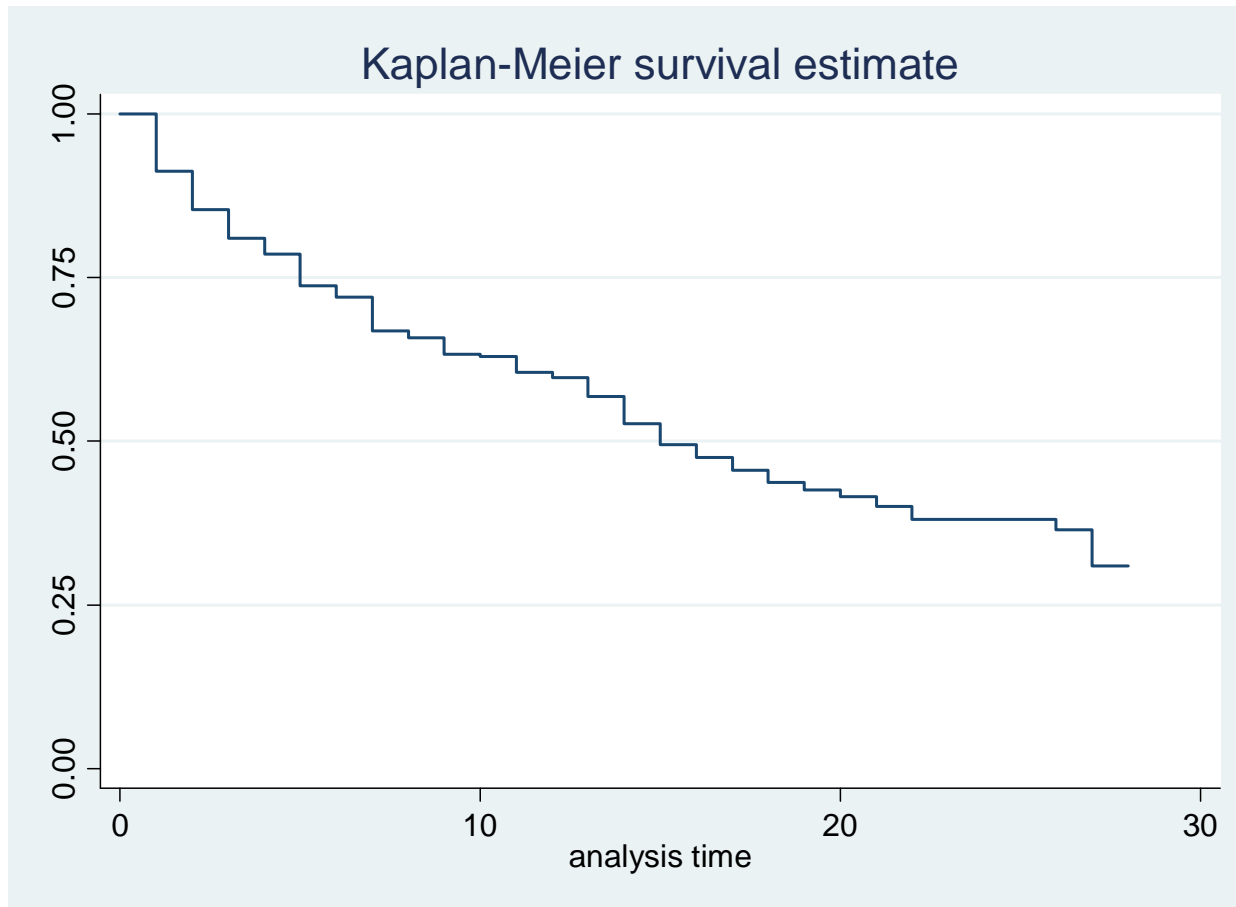
The hazard rate shows the probability of having the event (finding a job) going down from 4% to 3% over 25 time periods.



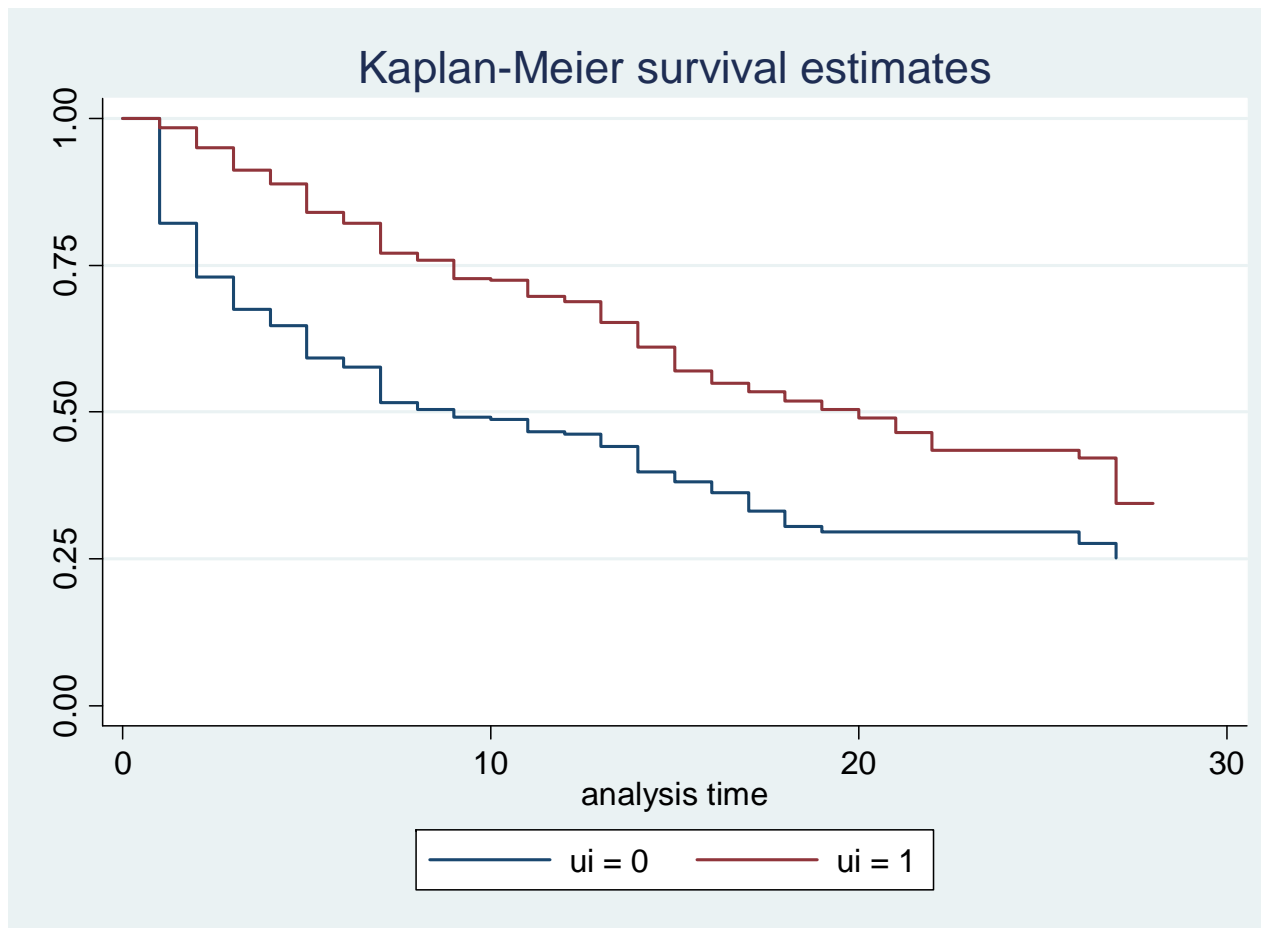
The Nelson-Aalen cumulative hazard estimate is non-decreasing.



The Kaplan-Meier survival function shows that survival probabilities go down to 31% over 28 time periods. This means that 31% of the individuals still have not found a job after 28 time periods.



Survival function for the group that has unemployment insurance ( $ui=1$ ) and the group that does not have unemployment insurance ( $ui=0$ ). The survival functions show at any point in time that claiming unemployment insurance is associated with higher survival rate. This means that if someone receives unemployment benefits he/she is more likely to still be unemployed.



### Parametric regression model coefficients

Duration of unemployment	Exponential regression coefficients	Weibull regression coefficients	Gompertz regression coefficients	Cox proportional hazard model coefficients
Log wage	0.48*	0.49*	0.48*	0.46*
Claim unemployment insurance	-1.08*	-1.11*	-1.06*	-0.98*
Age	-0.01*	-0.01*	-0.01*	-0.01*

Results reported here are Stata results. SAS produces opposite signs for the exponential, Weibull, and Gompertz regression. R produces opposite signs for exponential and Weibull regression. Also, the Weibull regression in SAS and R give different estimates.

- Interpretation of the coefficients: Individuals with higher wages have *lower* unemployment duration, meaning will terminate unemployment faster. Individuals who claim unemployment insurances have *higher* unemployment durations, meaning they terminate unemployment slower.



### Parametric regression model hazard rates

Duration of unemployment	Exponential regression hazard rates	Weibull regression hazard rates	Gompertz regression hazard rates	Cox proportional hazard model hazard rates
Log wage	1.62*	1.63*	1.61*	1.58*
Claim unemployment insurance	0.34*	0.33*	0.35*	0.37*
Age	0.99*	0.99*	0.99*	0.98*

- Interpretation of the hazard rates: A unit increase in the log wage is associated with  $1.62 - 1 = 62\%$  increase in the hazard rates. For individuals who claim unemployment insurance the hazard rates are  $0.34 - 1 = 66\%$  lower. In other words, individuals with higher wages are more likely to find a job and those that claim unemployment insurance are less likely to find a job.