

# Survival Analysis of Corporal Punishment Bans

Andy Grogan-Kaylor

16 Jun 2021 08:03:18

## Introduction

Corporal punishment is associated with a number of deleterious outcomes for children, including increases in behavior problems and mental health problems.

A number of countries have instituted country wide bans upon the use of corporal punishment with children.

Below, we employ a variety of empirical strategies to explore the institution of these bans.

## Get The Data

```
. use "../CPBans_w_AllCountries.dta", clear // data with ALL Countries; not just bans
```

NB It is important to have data with the *correct risk set* which includes *all countries*, not just *countries that eventually ban corporal punishment*.

In web versions of this tutorial, click the tabs below to access different sections of the tutorial.

Setup Parametric Survival Models Cox Model Life Table Discrete Time Survival Analysis Cox Model 2  
Compare Estimates

## Setup

### stset The Data

```
. generate current_year = year(today()) // variable with current year

. replace year_of_prohibition = current_year if year_of_prohibition == . // replace missing w/ current year
> nt year
(186 real changes made)

. generate f = type == "CP Ban" // "failure" variable

. stset year_of_prohibition, failure(f = 1) // stset the data with time and failure variables
Survival-time data settings
    Failure event: f==1
Observed time interval: (0, year_of_prohibition]
Exit on or before: failure
```

---

```
248 total observations
0 exclusions
```

---

```

248 observations remaining, representing
62 failures in single-record/single-failure data
500,452 total analysis time at risk and under observation
      At risk from t =      0
    Earliest observed entry t =      0
      Last observed exit t =    2,021

```

## Data Wrangling

```
. encode continent, generate(continent_NUMERIC) // numeric version of continent
```

## Graphs

### Survival Function

```

. sts graph, scheme(michigan) tmin(1970) // Kaplan-Meier Survivor Function
      Failure _d: f==1
    Analysis time _t: year_of_prohibition

. graph export mysurvival.png, width(500) replace
file /Users/agrogan/Box Sync/GitHub/research/cpbans/survival-analysis/mysurvival.png saved as PNG
format

```

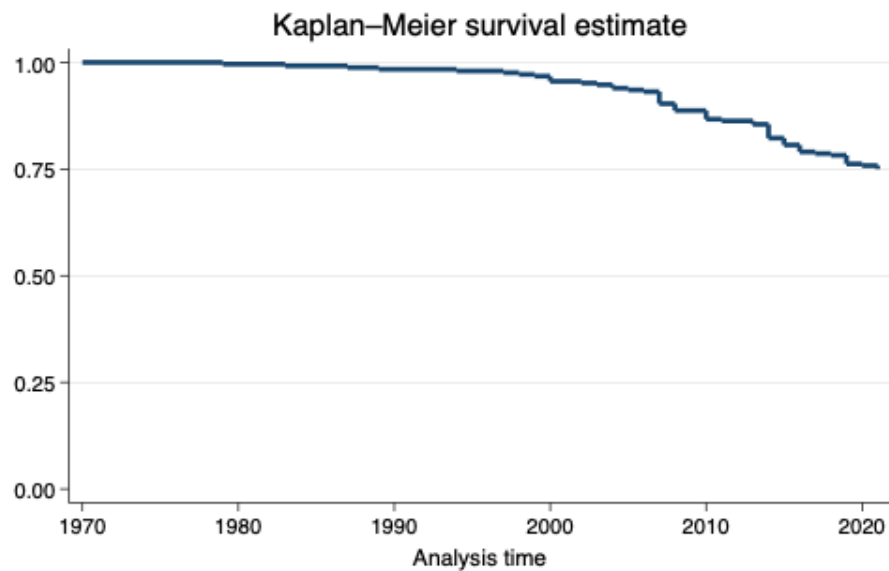


Figure 1: Kaplan-Meier Survivor Function

### Failure Function

```

. sts graph, failure scheme(michigan) tmin(1970) // Kaplan-Meier Failure Function
      Failure _d: f==1
    Analysis time _t: year_of_prohibition

. graph export myfailure.png, width(500) replace
file /Users/agrogan/Box Sync/GitHub/research/cpbans/survival-analysis/myfailure.png saved as PNG
format

```

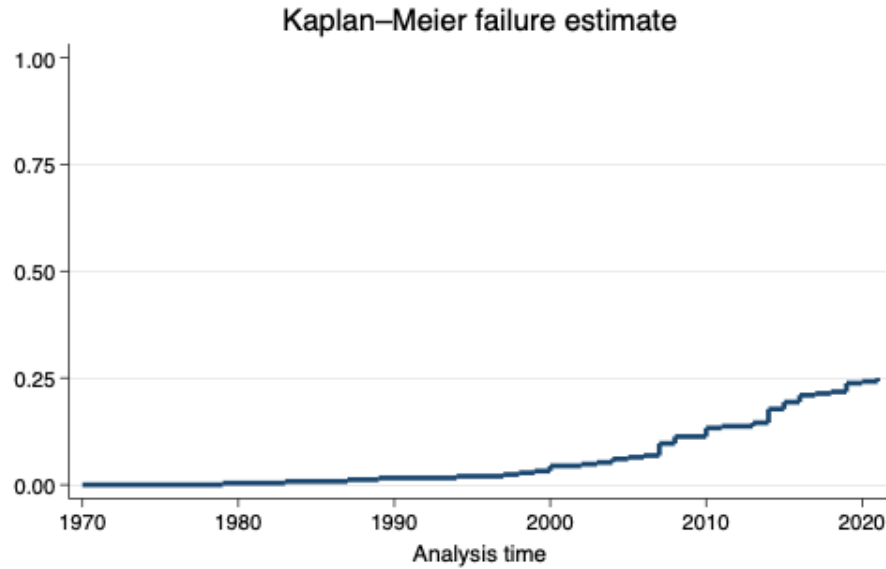


Figure 2: Kaplan-Meier Failure Function

## Hazard Function

```
. sts graph, hazard scheme(michigan) tmin(1970) // Kaplan-Meier Failure Function
    Failure _d: f==1
    Analysis time _t: year_of_prohibition

. graph export myhazard.png, width(500) replace
file /Users/agrogan/Box Sync/GitHub/research/cpbans/survival-analysis/myhazard.png saved as PNG
format
```

## Parametric Survival Models

Unlike other regression commands in Stata, survival analysis commands seem to require covariates. Since Europe is where these bans started, we will use Europe (category 4) as the reference category.

## Weibull

```
. streg ib4.continent_NUMERIC, distribution(weibull) // Weibull distribution
    Failure _d: f==1
    Analysis time _t: year_of_prohibition

Fitting constant-only model:
Iteration 0: log likelihood = -148.2325
Iteration 1: log likelihood = -86.999055
Iteration 2: log likelihood = -27.073844
Iteration 3: log likelihood = 29.365489
Iteration 4: log likelihood = 77.015953
Iteration 5: log likelihood = 106.62899
Iteration 6: log likelihood = 115.32234
Iteration 7: log likelihood = 115.88805
Iteration 8: log likelihood = 115.89021
Iteration 9: log likelihood = 115.89021
```

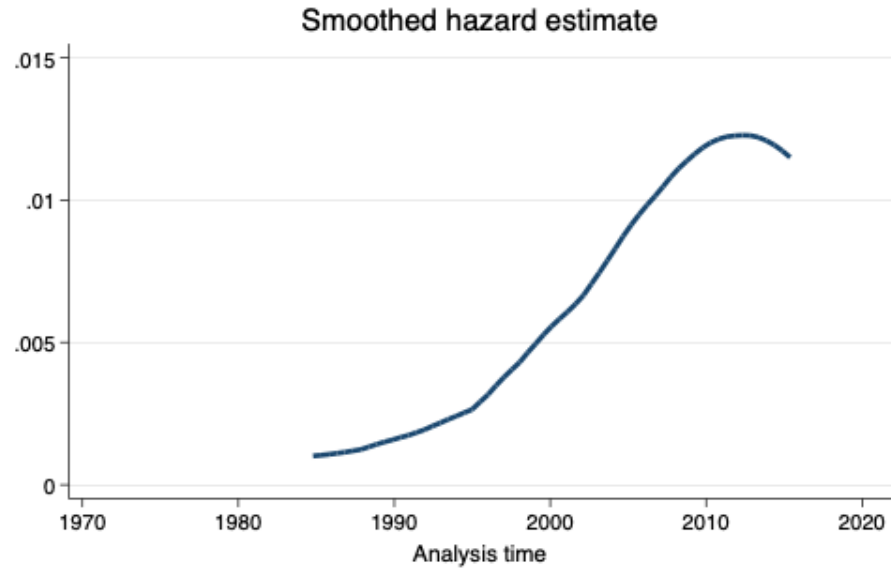


Figure 3: Hazard Function

Fitting full model:

Iteration 0: log likelihood = 115.89021  
 Iteration 1: log likelihood = 139.32561  
 Iteration 2: log likelihood = 142.87372  
 Iteration 3: log likelihood = 143.05492  
 Iteration 4: log likelihood = 143.05732  
 Iteration 5: log likelihood = 143.05732

Weibull PH regression

No. of subjects = 248  
 No. of failures = 62  
 Time at risk = 500,452

Number of obs = 248

LR chi2(5) = 54.33  
 Prob > chi2 = 0.0000

Log likelihood = 143.05732

_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]	
continent_NUMERIC						
Africa	.1684617	.0612563	-4.90	0.000	.0826011	.3435709
Americas	.1938708	.0704541	-4.51	0.000	.0950997	.395226
Asia	.1520997	.0603348	-4.75	0.000	.0698995	.3309653
NA	.0916735	.0931508	-2.35	0.019	.0125119	.6716806
Oceania	.0356574	.0362323	-3.28	0.001	.0048666	.2612621
_cons	0	0	-8.57	0.000	0	0
/ln_p	5.278967	.1166492	45.26	0.000	5.050339	5.507596
p	196.1672	22.88274			156.0754	246.5576
1/p	.0050977	.0005946			.0040558	.0064072

Note: \_cons estimates baseline hazard.

. est store Weibull

## Exponential

. streg ib4.continent\_NUMERIC, distribution(exponential) // Exponential distribution

```

Failure _d: f==1
Analysis time _t: year_of_prohibition
Iteration 0: log likelihood = -148.2325
Iteration 1: log likelihood = -139.40941
Iteration 2: log likelihood = -131.58499
Iteration 3: log likelihood = -131.55897
Iteration 4: log likelihood = -131.55892
Iteration 5: log likelihood = -131.55892

Exponential PH regression
No. of subjects = 248
No. of failures = 62
Time at risk = 500,452

Number of obs = 248

LR chi2(5) = 33.35
Prob > chi2 = 0.0000

Log likelihood = -131.55892

```

_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]
continent_NUMERIC					
Africa	.2736219	.099129	-3.58	0.000	.134516 .5565804
Americas	.3052592	.1105907	-3.28	0.001	.1500692 .6209345
Asia	.2489781	.0984172	-3.52	0.000	.1147345 .5402914
NA	.1586176	.1610769	-1.81	0.070	.0216746 1.160782
Oceania	.061017	.061963	-2.75	0.006	.0083378 .4465293
_cons	.000312	.0000552	-45.67	0.000	.0002206 .0004412

Note: \_cons estimates baseline hazard.

```
. est store Exponential
```

## Cox Proportional Hazards Model

```

. stcox ib4.continent_NUMERIC // Cox Proportional Hazards Model

Failure _d: f==1
Analysis time _t: year_of_prohibition
Iteration 0: log likelihood = -333.92184
Iteration 1: log likelihood = -317.94407
Iteration 2: log likelihood = -308.96171
Iteration 3: log likelihood = -308.00801
Iteration 4: log likelihood = -308.00737
Refining estimates:
Iteration 0: log likelihood = -308.00737
Cox regression with Breslow method for ties
No. of subjects = 248
No. of failures = 62
Time at risk = 500,452

Number of obs = 248

LR chi2(5) = 51.83
Prob > chi2 = 0.0000

Log likelihood = -308.00737

```

_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]
continent_NUMERIC					
Africa	.1769827	.0643396	-4.76	0.000	.0867938 .3608887
Americas	.2023186	.0735008	-4.40	0.000	.0992661 .4123544
Asia	.1610376	.0638871	-4.60	0.000	.0740009 .3504428
NA	.0969297	.0984941	-2.30	0.022	.0132287 .7102257
Oceania	.0380401	.038653	-3.22	0.001	.0051919 .2787139

```
. est store Cox
```

## Survival Curves

```
. stcurve, survival scheme(michigan) // survival curve

. graph export survival1A.png, width(500) replace
file /Users/agrogan/Box Sync/GitHub/research/cpbans/survival-analysis/survival1A.png saved as PNG
format
```

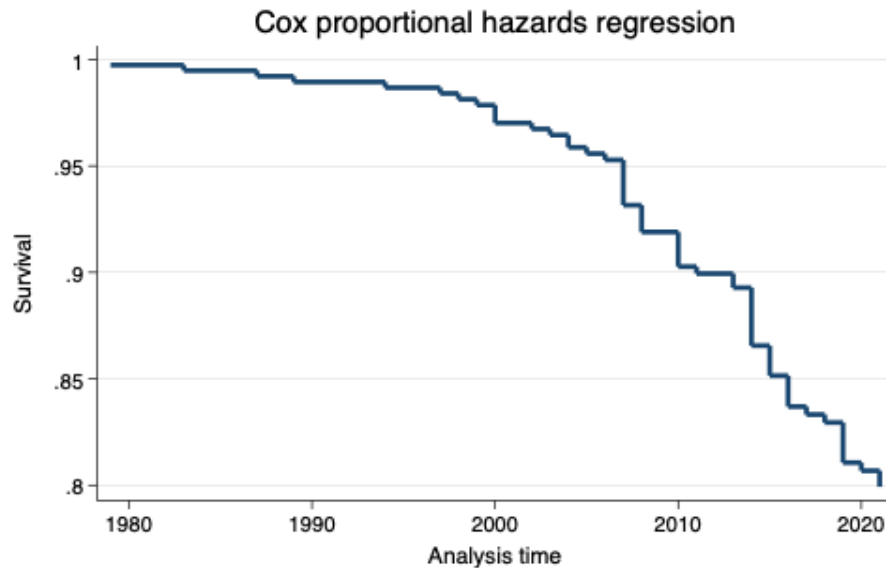


Figure 4: Survival Curve

```
. stcurve, failure scheme(michigan) // failure curve

. graph export survival1B.png, width(500) replace
file /Users/agrogan/Box Sync/GitHub/research/cpbans/survival-analysis/survival1B.png saved as PNG
format

. stcurve, failure at(continent_NUMERIC= (1 2 3 4 5 6)) ///
> legend(order(1 "Africa" 2 "Americas" 3 "Asia" ///
> 4 "Europe" 5 "NA" 6 "Oceania"))) ///
> scheme(michigan) // survival curve by continent

. graph export survival2.png, width(500) replace
file /Users/agrogan/Box Sync/GitHub/research/cpbans/survival-analysis/survival2.png saved as PNG
format
```

## Proportional Hazards Assumption

```
. estat phtest // formal test of PH assumption
Test of proportional-hazards assumption
Time function: Analysis time
```

	chi2	df	Prob>chi2
Global test	6.20	5	0.2870

```
. stphplot, by(continent_NUMERIC) scheme(michigan) // graphical test of PH assumption
```

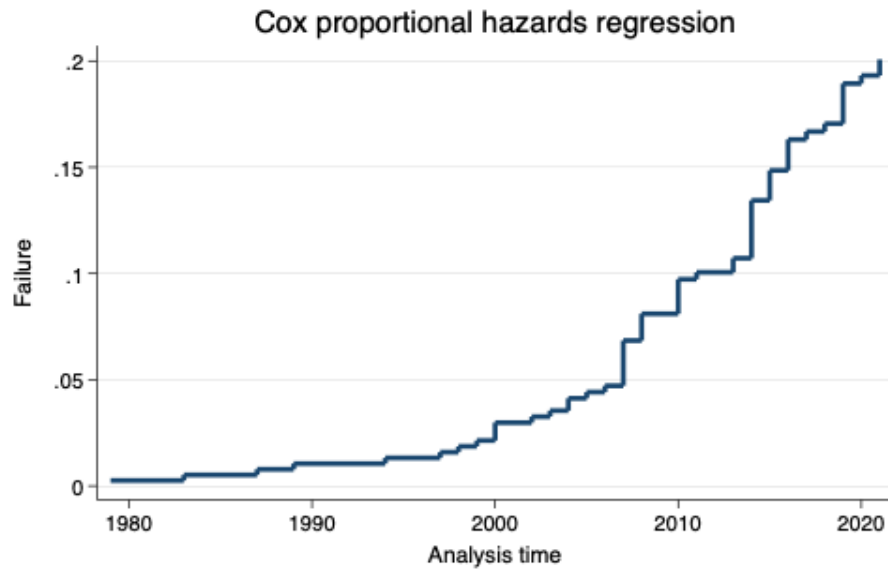


Figure 5: Failure Curve

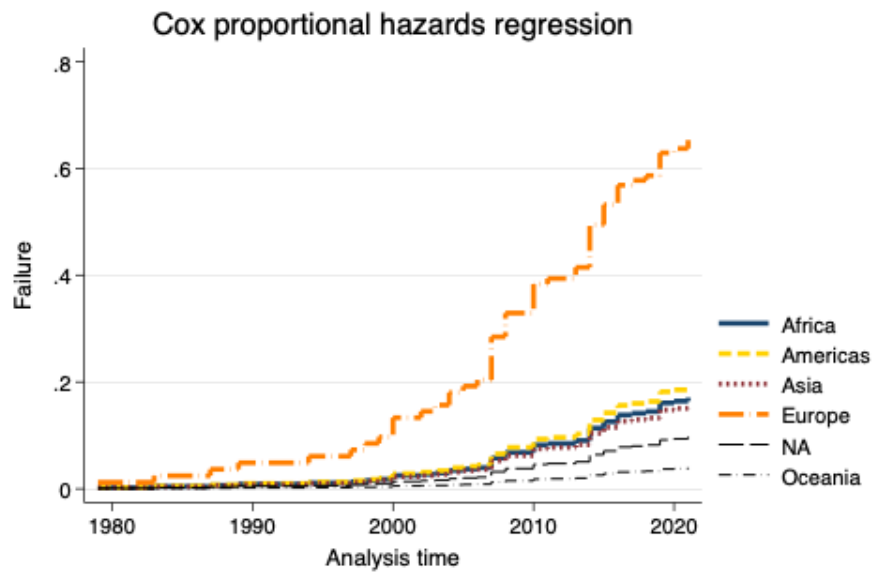


Figure 6: Failure Curve By Continent

```

Failure _d: f==1
Analysis time _t: year_of_prohibition

. graph export ph.png, width(500) replace
file /Users/agrogon/Box Sync/GitHub/research/cpbans/survival-analysis/ph.png saved as PNG format

```

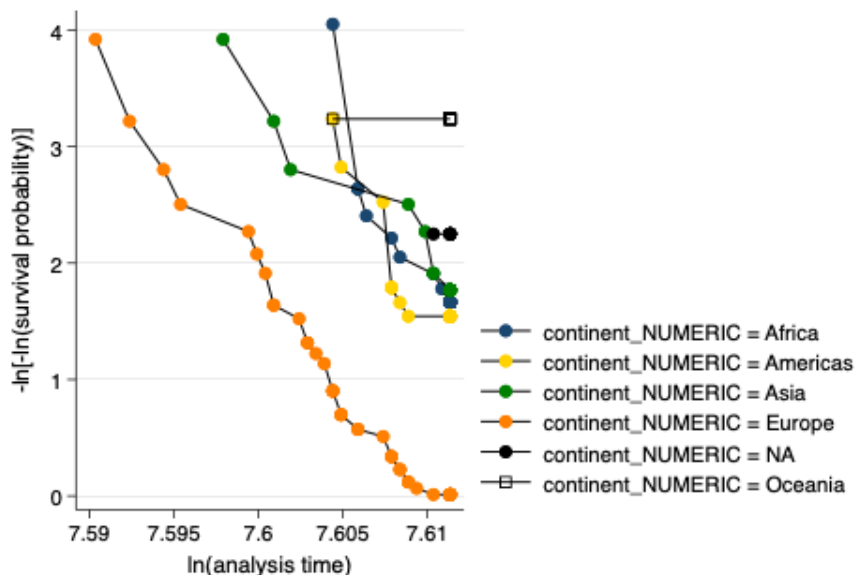


Figure 7: Graphical Test of Proportional Hazards Assumption

## Life Table

```
. ltable year_of_prohibition f, graph failure scheme(michigan) // lifetable
```

Interval		Beg. Total	Deaths	Lost	Cum. Failure	Std. Error	[95% Conf. Int.]	
1979	1980	248	1	0	0.0040	0.0040	0.0006	0.0283
1983	1984	247	1	0	0.0081	0.0057	0.0020	0.0319
1987	1988	246	1	0	0.0121	0.0069	0.0039	0.0370
1989	1990	245	1	0	0.0161	0.0080	0.0061	0.0424
1994	1995	244	1	0	0.0202	0.0089	0.0084	0.0478
1997	1998	243	1	0	0.0242	0.0098	0.0109	0.0531
1998	1999	242	1	0	0.0282	0.0105	0.0136	0.0583
1999	2000	241	1	0	0.0323	0.0112	0.0163	0.0635
2000	2001	240	3	0	0.0444	0.0131	0.0248	0.0787
2002	2003	237	1	0	0.0484	0.0136	0.0278	0.0836
2003	2004	236	1	0	0.0524	0.0142	0.0308	0.0886
2004	2005	235	2	0	0.0605	0.0151	0.0369	0.0983
2005	2006	233	1	0	0.0645	0.0156	0.0400	0.1032
2006	2007	232	1	0	0.0685	0.0160	0.0432	0.1080
2007	2008	231	7	0	0.0968	0.0188	0.0659	0.1409
2008	2009	224	4	0	0.1129	0.0201	0.0794	0.1593
2010	2011	220	5	0	0.1331	0.0216	0.0965	0.1820
2011	2012	215	1	0	0.1371	0.0218	0.1000	0.1865
2013	2014	214	2	0	0.1452	0.0224	0.1069	0.1955
2014	2015	212	8	0	0.1774	0.0243	0.1352	0.2309
2015	2016	204	4	0	0.1935	0.0251	0.1496	0.2484
2016	2017	200	4	0	0.2097	0.0258	0.1641	0.2658
2017	2018	196	1	0	0.2137	0.0260	0.1677	0.2701
2018	2019	195	1	0	0.2177	0.0262	0.1713	0.2745
2019	2020	194	5	0	0.2379	0.0270	0.1897	0.2960



2020	2021	189	1	0	0.2419	0.0272	0.1934	0.3002
2021	2022	188	2	186	0.2579	0.0289	0.2062	0.3196

```
. graph export myltable.png, width(500) replace
file /Users/agrogon/Box Sync/GitHub/research/cpbans/survival-analysis/myltable.png saved as PNG
format
```

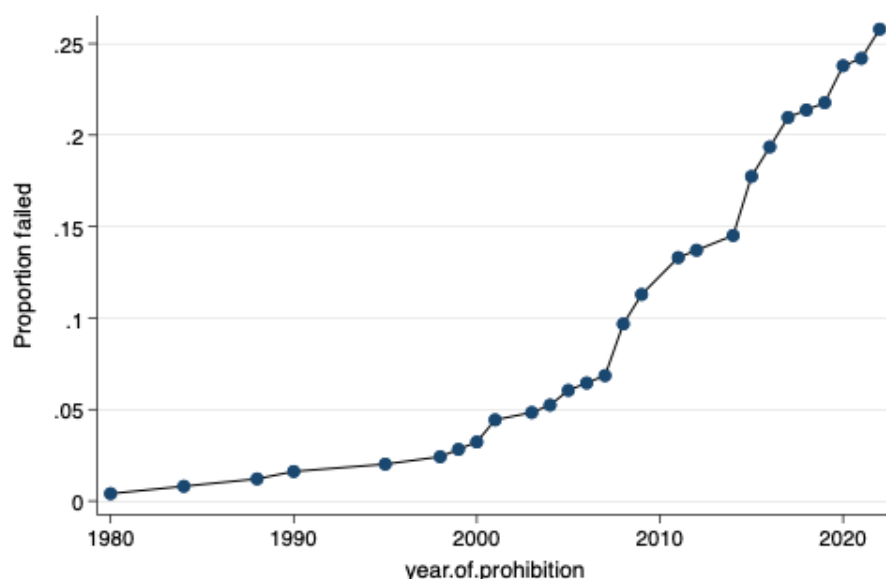


Figure 8: Graph Of Life Table

## Discrete Time Survival Analysis

Discrete time survival analysis is placed at the end because it requires us to restructure the data into a long format, where every row is a *country-year*.

### Restructuring the Data

We first need to generate a variable for the years during which a country was “at risk” of enacting a ban. Countries that have never enacted a ban are at risk up until the current year. Countries that enacted a ban leave the risk set once they have enacted a ban, and are thus at risk for a shorter time period.

```
. generate years_at_risk = year_of_prohibition - 1978 + 1 // years "at risk" of enacting a ban
```

We *expand* the data to generate one row of data for every year that each country is at risk.

```
. expand years_at_risk // "expand" the data; 1 row for every year at risk
(9,908 observations created)
```

We create a year variable.

```
. bysort country_code: generate year = _n + 1977 // generate a year variable for each row
```

Lastly, we generate an indicator of the event, a 0/1 variable, which takes the value 1 for rows in which a ban was enacted, and 0 otherwise.

```
. generate event = type == "CP Ban" & year_of_prohibition == year // generate an event indicator
```

We list out a sample of the data to make sure that the data conform to our expectations. We focus on Norway, a country that *has* enacted a ban, and Great Britain, a country that *has not* enacted a ban.

```
. list country_code year_of_prohibition event continent years_at_risk year ///
> if country_code == "NOR" | country_code == "GBR" , ab(20) // list out a sample of the data
```

	country_code	year_of_prohibition	event	continent	years_at_risk	year
3172.	GBR	2021	0	Europe	44	1978
3173.	GBR	2021	0	Europe	44	1979
3174.	GBR	2021	0	Europe	44	1980
3175.	GBR	2021	0	Europe	44	1981
3176.	GBR	2021	0	Europe	44	1982
3177.	GBR	2021	0	Europe	44	1983
3178.	GBR	2021	0	Europe	44	1984
3179.	GBR	2021	0	Europe	44	1985
3180.	GBR	2021	0	Europe	44	1986
3181.	GBR	2021	0	Europe	44	1987
3182.	GBR	2021	0	Europe	44	1988
3183.	GBR	2021	0	Europe	44	1989
3184.	GBR	2021	0	Europe	44	1990
3185.	GBR	2021	0	Europe	44	1991
3186.	GBR	2021	0	Europe	44	1992
3187.	GBR	2021	0	Europe	44	1993
3188.	GBR	2021	0	Europe	44	1994
3189.	GBR	2021	0	Europe	44	1995
3190.	GBR	2021	0	Europe	44	1996
3191.	GBR	2021	0	Europe	44	1997
3192.	GBR	2021	0	Europe	44	1998
3193.	GBR	2021	0	Europe	44	1999
3194.	GBR	2021	0	Europe	44	2000
3195.	GBR	2021	0	Europe	44	2001
3196.	GBR	2021	0	Europe	44	2002
3197.	GBR	2021	0	Europe	44	2003
3198.	GBR	2021	0	Europe	44	2004
3199.	GBR	2021	0	Europe	44	2005
3200.	GBR	2021	0	Europe	44	2006
3201.	GBR	2021	0	Europe	44	2007
3202.	GBR	2021	0	Europe	44	2008
3203.	GBR	2021	0	Europe	44	2009
3204.	GBR	2021	0	Europe	44	2010
3205.	GBR	2021	0	Europe	44	2011
3206.	GBR	2021	0	Europe	44	2012
3207.	GBR	2021	0	Europe	44	2013
3208.	GBR	2021	0	Europe	44	2014
3209.	GBR	2021	0	Europe	44	2015
3210.	GBR	2021	0	Europe	44	2016
3211.	GBR	2021	0	Europe	44	2017
3212.	GBR	2021	0	Europe	44	2018
3213.	GBR	2021	0	Europe	44	2019
3214.	GBR	2021	0	Europe	44	2020
3215.	GBR	2021	0	Europe	44	2021
6811.	NOR	1987	0	Europe	10	1978
6812.	NOR	1987	0	Europe	10	1979
6813.	NOR	1987	0	Europe	10	1980
6814.	NOR	1987	0	Europe	10	1981
6815.	NOR	1987	0	Europe	10	1982
6816.	NOR	1987	0	Europe	10	1983
6817.	NOR	1987	0	Europe	10	1984

6818.	NOR	1987	0	Europe	10	1985
6819.	NOR	1987	0	Europe	10	1986
6820.	NOR	1987	1	Europe	10	1987

## Analysis

Lastly, we analyze the data using a straightforward logistic regression model. While there is some discussion on this point, we choose not to **cluster** the standard errors on country, because of the argument from Singer and Willett (2003) that the rows of data are *conditionally* independent.

We ask for *odds ratios* so that our results are roughly comparable to those from the continuous time survival models.

```
. logit event ib4.continent_NUMERIC year, or
Iteration 0:  log likelihood = -377.92887
Iteration 1:  log likelihood = -372.7393
Iteration 2:  log likelihood = -330.01528
Iteration 3:  log likelihood = -328.96762
Iteration 4:  log likelihood = -328.96528
Iteration 5:  log likelihood = -328.96528
Logistic regression                                Number of obs = 10,156
                                                    LR chi2(6)      = 97.93
                                                    Prob > chi2     = 0.0000
                                                    Pseudo R2      = 0.1296
Log likelihood = -328.96528
```

event	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
continent_NUMERIC						
Africa	.1666907	.0611921	-4.88	0.000	.0811775	.3422843
Americas	.1916195	.0703136	-4.50	0.000	.0933462	.3933534
Asia	.1507161	.0603004	-4.73	0.000	.0688019	.3301562
NA	.0906814	.092426	-2.36	0.019	.012301	.6684916
Oceania	.0351902	.0358209	-3.29	0.001	.0047859	.2587488
year	1.088786	.013881	6.67	0.000	1.061917	1.116335
_cons	2.05e-76	5.25e-75	-6.81	0.000	3.38e-98	1.24e-54

Note: \_cons estimates baseline odds.

```
. est store Discrete
```

## Cox Model With Multiple Records Per Observation and Time Varying Covariates

We make use of the fact that the data are structured with multiple records per individual to include the effect of year, which is a *time varying covariate*. We need to newly **stset** the data to account for the multiple records per individual.

**I am not sure how to use the tvc option to program this when there is only one record per *individual*.**

```
// bysort country_y: egen maxyear = max(year) //
// bysort country_y: generate f2 = (f ==1) & (year ==maxyear) // // stset year, failure(f2 =
1) id(country_y) origin(time 1978) // stset the data with time, failure & id variables // // stcox
ib4.continent_NUMERIC year //
// est store CoxTVC2
```

## Compare Estimates

Note the difference in the effect of *year* in the two models where this is included.

```
. estimates table Weibull Exponential Cox Discrete, ///
> b(%9.3f) star stats(N r2_a) equations(1) // nice table of estimates
```

Variable	Weibull	Exponential	Cox	Discrete
#1				
continent__C				
Africa	-1.781***	-1.296***	-1.732***	-1.792***
Americas	-1.641***	-1.187**	-1.598***	-1.652***
Asia	-1.883***	-1.390***	-1.826***	-1.892***
continent__C				
NA	-2.390*	-1.841	-2.334*	-2.400*
Oceania	-3.334**	-2.797**	-3.269**	-3.347**
year				0.085***
_cons	-1492.992***	-8.073***		-174.278***
ln_p				
_cons	5.279***			
Statistics				
N	248	248	248	10156
r2_a				

Legend: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

## References

- Allison, P. D. (1984). *Event History Analysis: Regression for Longitudinal Event Data*. SAGE Publications.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis : modeling change and event occurrence*. Oxford ; New York : Oxford University Press.
- StataCorp. 2021. *Stata 17 Survival Analysis Reference Manual*. College Station, TX: Stata Press