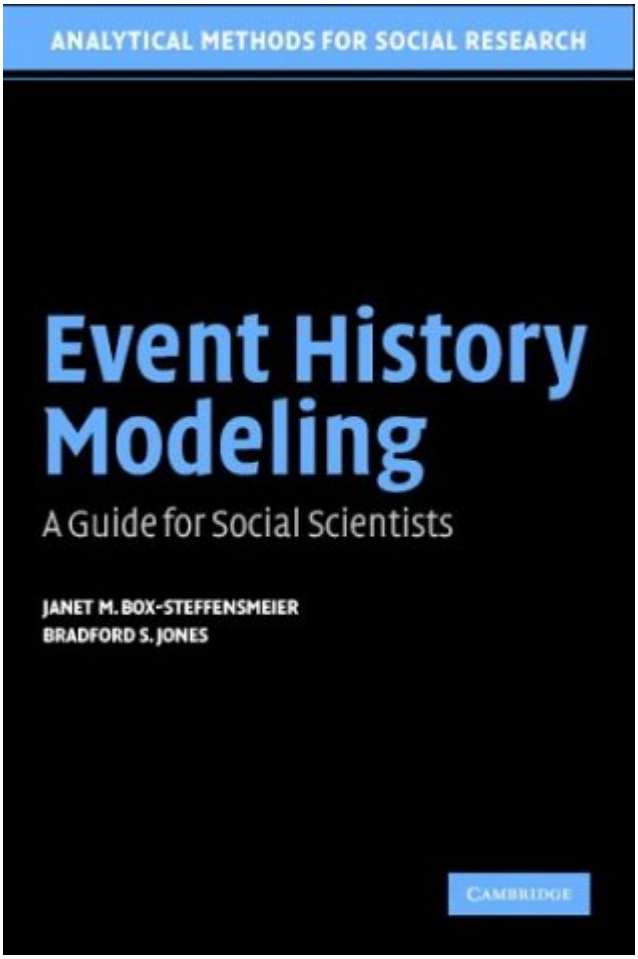
Timing and Social Change: A Stylized Introduction to Duration Analysis

ICPSR
MLE--2
University of Michigan
August 1—12, 2005

Bradford S. Jones
Associate Professor
Department of Political Science
University of Arizona
Tucson, AZ USA

Contact: Brad Jones
315 Social Sciences
Department of Political Science
University of Arizona
Tucson, AZ 85721
USA
bsjones@email.arizona.edu

SHAMELESS PROMOTION



What is the *event* in event history analysis?

- Getting a job
- Losing a job
- Business Failure
- Government Falls
- Politician Loses
- Regime Change
- Policy is Adopted
- Military Conflict Begins
- Political Party Loses Control
- Recidivism
- …and on and on and on

We're naturally interested in events…both as academics and as "everyday" citizens.

Implicitly, events are probabilistic.

(either in a classical sense or in a "Bayesian" sense…)

- ***What are the chances I get this job?***
- ***What are the chances I'll lose this job?***
- ***How likely is it my business will fail?***
- ***Will the government survive the election?***
- ***Will the criminal return to crime?***
- ***Will the Senate confirm Bush's nominee?***

Time is of the Essence

Suppose we incorporate time into the problem? …then "chance" becomes synonymous with "risk."

Event: M.P.'s Legislative Victory
Timing: Number of terms served in the legislature.
The Risk: ***"Given an M.P. has stayed in office over 4 elections, what is the risk she will lose in the subsequent election?"***

Event: Return to Alcoholism
Timing: Sobriety period.
The Risk: ***"Given an alcoholic has stayed sober for 3 months, what are the chances he will return to alcohol in the next month?***

Event: Divorce
Timing: Years Married.
The Risk: ***"Given a couple has remained married 10 years, what is the likelihood they will divorce next year?"***

Timing Implies Risk: Given that something *hasn't happened*, what are the chances *it will happen* subsequently?

RISK IS A KEY INGREDIENT

The "formula" for risk is simple:

$$\frac{\text{"Chance that Something Happens"}}{\text{"Chance that it hasn't Happened Yet"}}$$

= "RISK"

Risk is a "ratio": a relationship between the chances that something *can* happen relative to the chances that it hasn't happened yet.

More succinctly:
RISK=Pr("failure")/Pr("survival")

DEFINITIONS:

Failure: The unconditional probability that an event will occur.

Survival: The probability that "up until now" the event has not yet occurred.

Risk: The conditional failure rate---given that the event has not yet occurred, what are the chances it will occur?

Thus, the chances of the event occurring are conditional on how long the observation has persisted without having experienced the event.

or, as before,

"Given an M.P. has stayed in office over 4 elections, what is the risk she will lose in the subsequent election?"

or "R=F/S"

Conventionally, this "risk ratio" is called a HAZARD RATIO.
Hence, "hazard" implies "risk."

"RISK" "SURVIVAL" "FAILURE"

We See These Concepts Everywhere!

Some Recent News Headlines

"Berlusconi government at risk"

"A poor showing in local elections 11 days ago for Prime Minister Silvio Berlusconi struck home today when a small party in his coalition pulled out of the cabinet, *threatening the government's survival*."---*New York Times*

"Clouds seen ahead for Latin America economic growth"

"Rising global interest rates, competition from China and governments tiring of reform may put the past year's strong economic growth in *Latin America at risk*, analysts said at an investment forum in Miami."---Reuters *Precios y Noticias* (Mexico)

"Risk aversion continues to rule in Australia"

"…Against an inexperienced new Labour party leader, Mark Latham, who ran a spirited campaign, the 65-year-old veteran Liberal-National party coalition leader maintained his tough stance on Iraq and relied on his sound management of the economy to produce a resounding election win. *Australians were not prepared to take the risk* with an untested newcomer and Mr. Howard was able to secure a 2% electoral swing, increase his majority in the House and unexpectedly gain control in the Senate."---*The Banker* (London)

"PM Unable to Elevate Liberals: Poll (Canada)"

Ottawa — Paul Martin's strategy of tackling the sponsorship scandal head-on has burnished his personal popularity but left his party stalled in minority-government territory less than one month before the Prime Minister has been widely expected to call an election, a new poll indicates. Mr. Martin, who took over as Prime Minister late last year after Jean Chrétien stepped down, needs to call an election to win his own mandate. But the recent polling figures — including the current result — *make that a risky prospect.* He could call an election as early as the beginning of April, but some within his party would prefer a vote later in the spring or in the fall. …In Quebec, the Bloc Québécois now leads the Liberals by 18 points. …*That puts the Liberals at risk* of not being able to reclaim the 37 seats they won in 2000 across the province's 75 ridings… . ---*Globe and Mail* (Ottawa)

Getting Some Leverage on Risk is the Goal of

EVENT HISTORY ANALYSIS

Typically, researchers will:

● Have some theory or hypothesis relating timing and other factors (i.e. independent variables or covariates) to some event (i.e. how does time served in Parliament relate to the chances an M.P. will lose (or not lose) his seat? How does the national economy impact the odds the M.P. will lose?).

● Observe some "sample" over time (perhaps a group of MPs).

● Record whether or not some event of interest occurs over time.

● Collect data on important covariates.

● Model the "event" or the "time until the event" as a function of covariates, and perhaps, time itself.

Example of Event History Data (Single-Record)
The Duration of N.Z. Premierships (1856-Present)

Name	Took Office	Left Office	Time	Event	Party
Henry Sewell	7 May 1856	20 May 1856	13	1	None
William Fox	20 May 1856	2 June 1856	13	1	None
Edward Stafford	2 June 1856	12 July 1861	1866	1	None
William Fox	12 July 1861	6 August 1862	390	1	None
Alfred Domett	6 August 1862	30 October 1863	450	1	None
Frederick Whitaker	30 October 1863	24 November 1864	391	1	None
Frederick Weld	24 November 1864	16 October 1865	326	1	None
Edward Stafford	16 October 1865	28 June 1869	1351	1	None
William Fox	28 June 1869	10 September 1872	1170	1	None
Edward Stafford	10 September 1872	11 October 1872	31	1	None
George Waterhouse	11 October 1872	3 March 1873	143	1	None
William Fox	3 March 1873	8 April 1873	36	1	None
Julius Vogel	8 April 1873	6 July 1875	819	1	None
Daniel Pollen	6 July 1875	15 February 1876	224	1	None
Julius Vogel	15 February 1876	1 September 1876	199	1	None
Harry Atkinson	1 September 1876	13 October 1877	407	1	None
George Grey	13 October 1877	8 October 1879	725	1	None
John Hall	8 October 1879	21 April 1882	926	1	None
Frederick Whitaker	21 April 1882	25 September 1883	522	1	None
Harry Atkinson	25 September 1883	16 August 1884	326	1	None
Robert Stout	16 August 1884	28 August 1884	12	1	None
Harry Atkinson	28 August 1884	3 September 1884	6	1	None
Robert Stout	3 September 1884	8 October 1887	1130	1	None
Harry Atkinson	8 October 1887	24 January 1891	1204	1	None
John Ballance	24 January 1891	27 April 1893	824	1	Liberal
Richard Seddon	27 April 1893	10 June 1906	4791	1	Liberal
William Hall-Jones	10 June 1906	6 August 1906	57	1	Liberal
Joseph Ward	6 August 1906	28 March 1912	2061	1	Liberal
Thomas Mackenzie	28 March 1912	10 July 1912	104	1	Liberal
William Massey	10 July 1912	10 May 1925	4687	1	Reform
Francis Bell	10 May 1925	30 May 1925	20	1	Reform
Gordon Coates	30 May 1925	10 December 1928	1290	1	Reform
Joseph Ward	10 December 1928	28 May 1930	534	1	Liberal
George Forbes	28 May 1930	6 December 1935	2018	1	Liberal
Michael Joseph Savage	6 December 1935	27 March 1940	1573	1	Labour
Peter Fraser	27 March 1940	13 December 1949	3548	1	Labour
Sidney Holland	13 December 1949	20 September 1957	2838	1	National
Keith Holyoake	20 September 1957	12 December 1957	83	1	National
Walter Nash	12 December 1957	12 December 1960	1096	1	Labour
Keith Holyoake	12 December 1960	7 February 1972	4074	1	National
Jack Marshall	7 February 1972	8 December 1972	305	1	National
Norman Kirk	8 December 1972	31 August 1974	631	1	Labour
Bill Rowling	6 September 1974	12 December 1975	462	1	Labour
Robert Muldoon	12 December 1975	26 July 1984	3149	1	National
David Lange	26 July 1984	8 August 1989	1839	1	Labour
Geoffrey Palmer	8 August 1989	4 September 1990	392	1	Labour
Mike Moore	4 September 1990	2 November 1990	59	1	Labour
Jim Bolger	2 November 1990	8 December 1997	2593	1	National
Jenny Shipley	8 December 1997	5 December 1999	727	1	National
Helen Clark	5 December 1999	31 May 2005	2004	0	Labour

Modeling Strategy?

*Time-to-Termination=f(Political Party)**
Apply O.L.S.?

```
. reg time Labour Liberal National Reform, robust, if NZ==1

Regression with robust standard errors                Number of obs =      50
                                                      F(  4,    45) =    3.16
                                                      Prob > F      =  0.0226
                                                      R-squared     =  0.2230
                                                      Root MSE      =  1143.2

------------------------------------------------------------------------------
             |               Robust
        time |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Labour |         761   376.1921     2.02   0.049     3.310136     1518.69
     Liberal |   955.8095   627.1627     1.52   0.135    -307.3611     2218.98
    National |   1438.667   589.9617     2.44   0.019     250.4228    2626.911
      Reform |   1470.667   1203.834     1.22   0.228    -953.9801    3895.313
       _cons |   528.3333   107.8366     4.90   0.000     311.1393    745.5274
------------------------------------------------------------------------------
```

*Model is for pedagogical purposes only: don't try this at home!

●Here, the dependent variable is "time" and the covariates are party dummy variables.

●The interpretation is standard: the coefficient tells us how the expected *survival time* of the P.M. increases over the baseline category (in this model, "no political party") conditional on their party affiliation.

● Risk is implied: the expected survival time of Labour Premiers has historically been about 678 days "shorter" than National Premiers (i.e. E[Y]|National PM]=1,967 days, E[Y|Labour PM]=1,289 days; ΔE[Y]=678 days).

● Hence, the model suggests National MPs have had an historically higher "survival" rate implying an historically lower "risk" or "hazard rate." (BUT BE CAREFUL!)

SOME PROBLEMS WITH O.L.S.

● O.L.S. may return negative predicted values---an impossibility: "survival times" must be positive.

● Duration data are often right-skewed, often times, heavily so.

● Hence, modeling the mean function may be less interesting than some other feature of the data, like the median, for example.

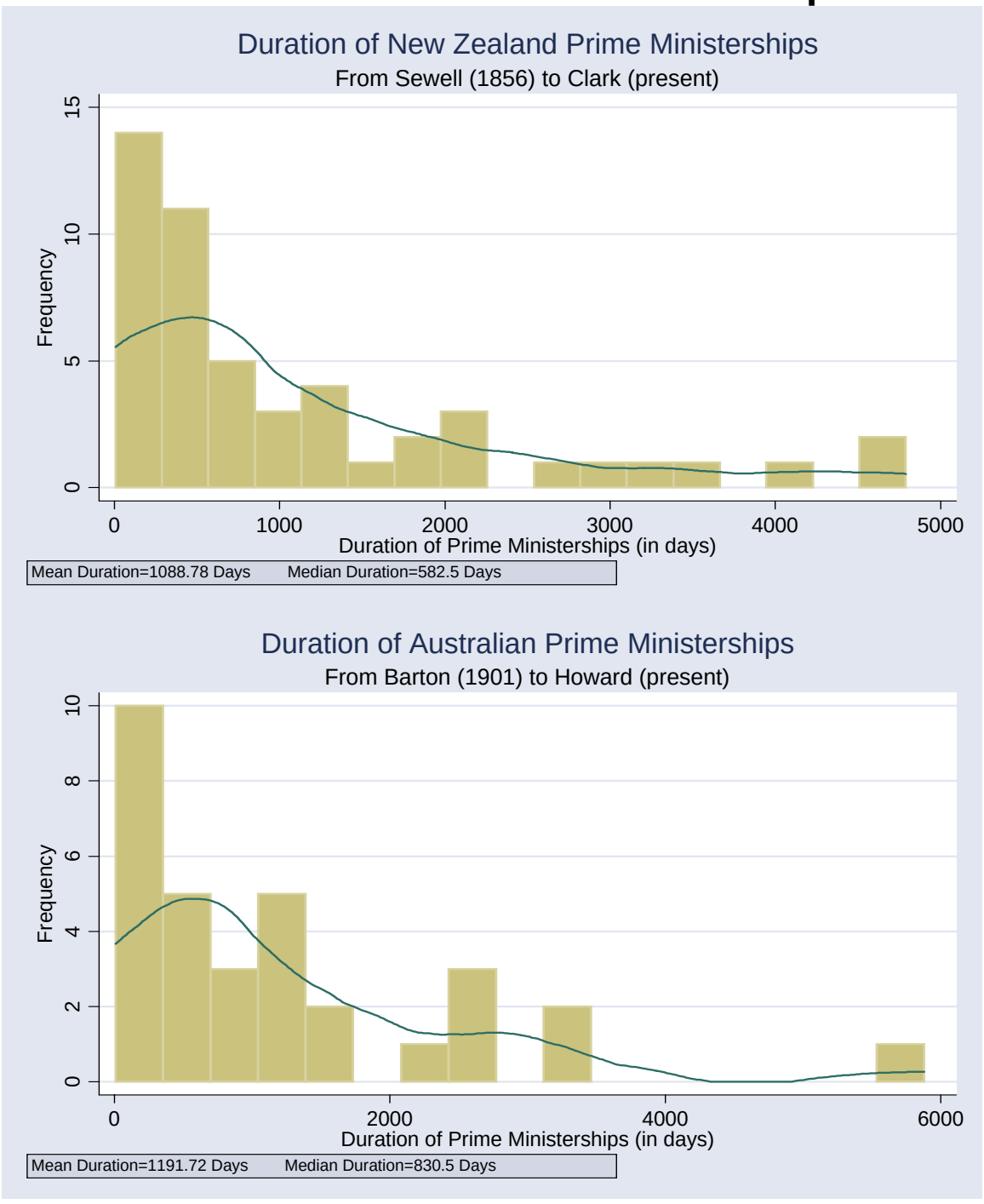
● O.L.S. does not easily distinguish "censored" from "uncensored" cases.

● O.L.S. cannot easily accommodate covariates that change value over time (TVCs).

● Assumed linearity in the survival times may be unrealistic.

THESE ARE REAL ISSUES!

SOME INSIGHTS BY LOOKING AT OUR DATA: N.Z. and Australian Premierships



Data Exhibit Considerable Skew
Median Describes Central Tendency Better than Mean

THE CENSORING PROBLEM

● Right-Censoring is very prevalent in most duration data sets.

● In our pedagogical example, censoring is not prevalent BUT consider these two cases:

New Zealand

P.M.	Time-in-Office	Left Office?
George Forbes	2,018 Days	Yes (6 Dec. 1935)
Helen Clark	2,004 Days	No (still serving)

Australia

P.M.	Time-in-Office	Left Office?
Bob Hawke	3,206 Days	Yes (20 Dec. 1991)
John Howard	3,368 Days	No (still serving)

● In terms of *time*, the cases look similar…

● …unfortunately, they are not: we do not know when the current Prime Ministers will leave (this would require seeing into the future) but we *do* know when the former Primer Ministers exited office. SIMILARITY IS ILLUSORY.

● The problem is, O.L.S. treats them as roughly equivalent, an equivalence that simply does not exist.

● This is a problem…and a really big problem for most event history data sets where censoring is common.

O.L.S. “Fix-Ups”

• Treat log(*t*) as the response variable: mitigates the skewness problem to some degree:

```
. reg logT Labour Liberal National Reform, robust, if NZ==1

Regression with robust standard errors                Number of obs =      50
                                                      F(  4,    45) =    2.11
                                                      Prob > F      =  0.0953
                                                      R-squared     =  0.1423
                                                      Root MSE      =   1.673

------------------------------------------------------------------------------
             |               Robust
        logT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Labour |   1.280525   .5448982     2.35   0.023     .1830437    2.378006
     Liberal |   1.067428   .7036195     1.52   0.136    -.3497348     2.48459
    National |   1.568772   .6528948     2.40   0.020     .2537747     2.88377
      Reform |   .7856566   1.463219     0.54   0.594    -2.161417     3.73273
       _cons |   5.417903   .3642504    14.87   0.000     4.684265    6.151541
------------------------------------------------------------------------------
```

• Interpretation is standard, though E(Survival Time) is now in logged units since coefficients are scaled by log(t).

• What about Censoring? Delete Censored Cases?

```
. reg time Labour Liberal National Reform, robust, if NZ==1  & Event==1

Regression with robust standard errors                Number of obs =      49
                                                      F(  4,    44) =    2.86
                                                      Prob > F      =  0.0342
                                                      R-squared     =  0.2218
                                                      Root MSE      =  1150.5

------------------------------------------------------------------------------
             |               Robust
        time |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Labour |   671.6667    407.952     1.65   0.107    -150.5066     1493.84
     Liberal |   955.8095    627.875     1.52   0.135    -309.5894    2221.208
    National |   1438.667   590.6317     2.44   0.019     248.3266    2629.007
      Reform |   1470.667   1205.202     1.22   0.229    -958.2574    3899.591
       _cons |   528.3333   107.9591     4.89   0.000     310.7561    745.9106
------------------------------------------------------------------------------
```

• “Labour” coefficient is now 672 (*t*=1.65); in previous model, “Labour” coefficient was 761 (*t*=2.02).

• In the general case, sample selection problems may (will?) be induced by omitting censored cases.

A Little Side-Trip to Illustrate Censoring

(Data are Pearson-Bauman Data on Interventions)

Duration of Military Intervention=f(Relative Capabilities Index)

```
Regression With All Cases

.  reg durmths pbal

      Source |       SS       df       MS              Number of obs =     586
-------------+------------------------------           F(  1,   584) =    6.65
       Model |  11015.6379     1  11015.6379            Prob > F      =  0.0102
    Residual |  967384.302   584  1656.47997            R-squared     =  0.0113
-------------+------------------------------           Adj R-squared =  0.0096
       Total |   978399.94   585  1672.47853            Root MSE      =    40.7

------------------------------------------------------------------------------
     durmths |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pbal |    13.75079   5.332316     2.58   0.010     3.277937    24.22364
       _cons |   13.19408   3.943963     3.35   0.001     5.448002    20.94016
------------------------------------------------------------------------------

Regression Omitting Censored Cases

.  reg durmths pbal if _d==1

      Source |       SS       df       MS              Number of obs =     559
-------------+------------------------------           F(  1,   557) =    2.04
       Model |  1968.48114     1  1968.48114            Prob > F      =  0.1538
    Residual |  537644.124   557  965.249773            R-squared     =  0.0036
-------------+------------------------------           Adj R-squared =  0.0019
       Total |  539612.605   558  967.047679            Root MSE      =  31.068

------------------------------------------------------------------------------
     durmths |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pbal |    5.925032   4.149014     1.43   0.154    -2.224595    14.07466
       _cons |   14.45036   3.044398     4.75   0.000     8.470458    20.43026
------------------------------------------------------------------------------
```

● The estimated relationship between relative capabilities (i.e. "*pbal*") is about 43 percent lower when we drop the 27 right-censored cases.

● This problem is not atypical; it is common in many event history data sets.

● The Problem: *Both* models are problematic.

TIME-VARYING COVARIATES

● Typical to have covariates that can change values over time (e.g. proportion of seats held in Parliament will vary from election to election).

● O.L.S. cannot easily accommodate these factors.

● Illustration:

Data Without TVCs (N.Z. 1984—2005)

Prime Minister	Time-in-Office	Political Party	Event
David Lange	1839	Labour	1
Geoffrey Palmer	392	Labour	1
Mike Moore	59	Labour	1
Jim Bolger	2,593	National	1
Jenny Shipley	727	National	1
Helen Clark	2,004	Labour	0

The only covariate is party affiliation of the P.M. It is "time-invariant."

Data With TVCs (N.Z. 1984—2005)

Name	Party	Start	Stop	L	N	Time	Event
David Lange	Labour	26 July 1984	15 August 1987	56	37	1115	0
David Lange	Labour	15 August 1987	8 August 1989	59	40	724	1
Geoffrey Palmer	Labour	8 August 1989	4 September 1990	59	40	392	1
Mike Moore	Labour	4 September 1990	2 November 1990	59	40	59	1
Jim Bolger	National	2 November 1990	6 November 1993	29	67	1100	0
Jim Bolger	National	6 November 1993	12 October 1996	45	50	1071	0
Jim Bolger	National	12 October 1996	8 December 1997	37	44	422	1
Jenny Shipley	National	8 December 1997	5 December 1999	37	44	727	1
Helen Clark	Labour	5 December 1999	27 July 2002	49	39	965	0
Helen Clark	Labour	27 July 2002	31 May 2005	52	27	1039	0

Here, two covariates are time-varying: Number of seats held by Labour ("L") and number of seats held by National ("N")

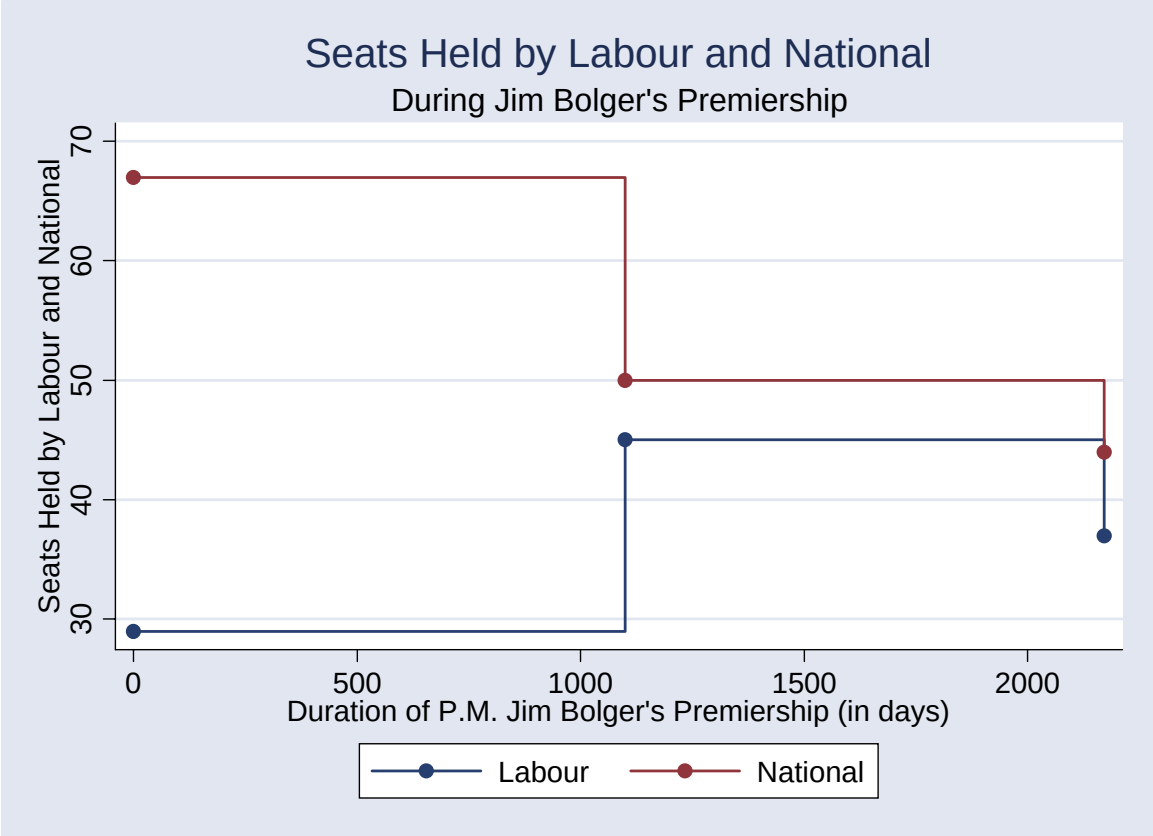
● Implication 1: Second data set requires "spell-splitting"

● Implication 2: "Spell-splitting" will "confuse" O.L.S.: it now looks like there are 10 "stand-alone" cases (but there are not: Lange's total duration as P.M. is 1115 days in "spell 1" and 724 days in "spell 2" which sums to 1,839 days total—there is only one David Lange…not two.).

● Implication 3: Second data set gives rise to "counting process" framework which is critical for event history analysis (i.e. start-stop data).

● Implication 4: "Jump-Process" interpretation emerges with TVCs.

“Jump-Process” In Action



- **Time path of the covariate “jumps” at elections.**

- **If seats were related to survival, we would find that Bolger’s “risk” or “survival” also *jumps*—or is responsive to—changes in the number of seats held.**

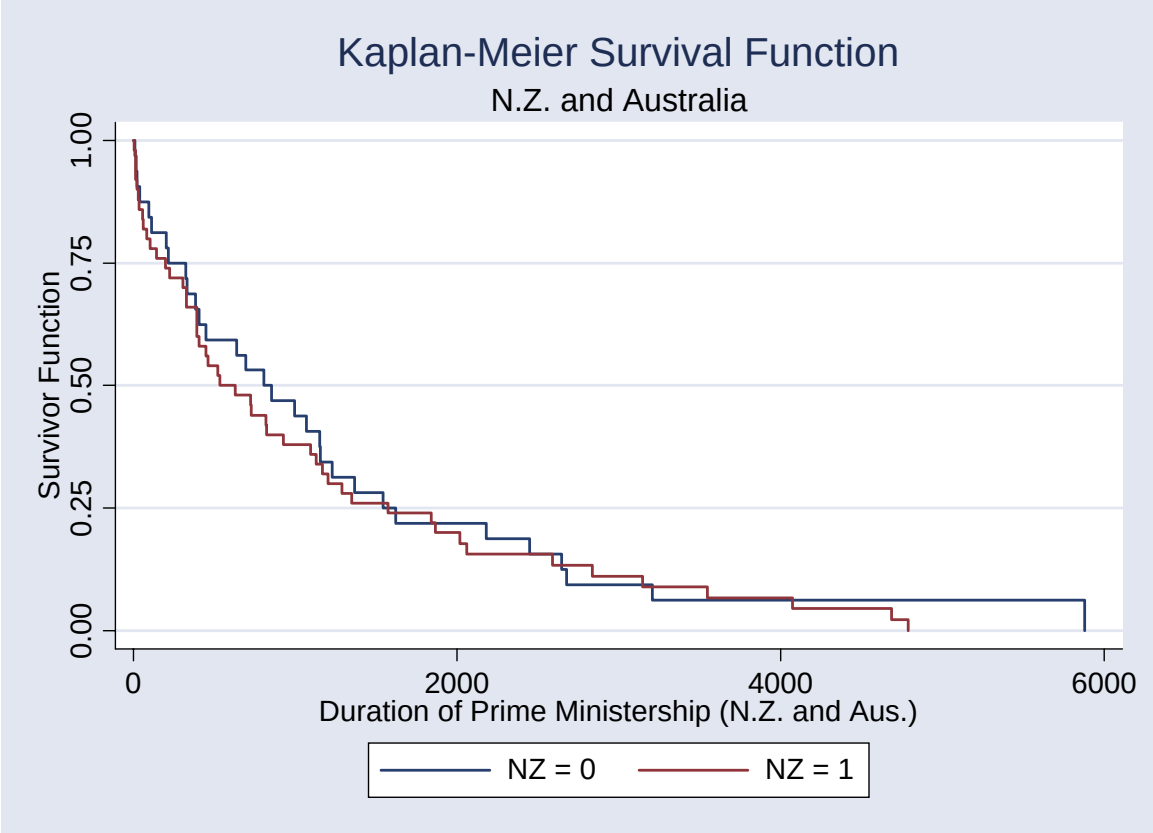
- **TVCs amplify or diminish risk.**

SOME QUICK EVENT HISTORY RESULTS

Kaplan-Meier Results

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
Australia							
7	32	1	0	0.9688	0.0308	0.7982	0.9955
19	31	1	0	0.9375	0.0428	0.7725	0.9840
22	30	1	0	0.9063	0.0515	0.7369	0.9688
40	29	1	0	0.8750	0.0585	0.7004	0.9512
95	28	1	0	0.8438	0.0642	0.6646	0.9318
113	27	1	0	0.8125	0.0690	0.6295	0.9111
201	26	1	0	0.7813	0.0731	0.5952	0.8892
216	25	1	0	0.7500	0.0765	0.5618	0.8663
321	24	1	0	0.7188	0.0795	0.5291	0.8426
331	23	1	0	0.6875	0.0819	0.4971	0.8180
384	22	1	0	0.6563	0.0840	0.4658	0.7927
405	21	1	0	0.6250	0.0856	0.4352	0.7668
450	20	1	0	0.5938	0.0868	0.4052	0.7402
636	19	1	0	0.5625	0.0877	0.3759	0.7130
692	18	1	0	0.5313	0.0882	0.3471	0.6852
806	17	1	0	0.5000	0.0884	0.3190	0.6567
855	16	1	0	0.4688	0.0882	0.2915	0.6277
996	15	1	0	0.4375	0.0877	0.2646	0.5981
1071	14	1	0	0.4063	0.0868	0.2383	0.5679
1152	13	1	0	0.3750	0.0856	0.2128	0.5371
1155	12	1	0	0.3438	0.0840	0.1879	0.5056
1227	11	1	0	0.3125	0.0819	0.1638	0.4734
1367	10	1	0	0.2813	0.0795	0.1404	0.4406
1543	9	1	0	0.2500	0.0765	0.1180	0.4069
1620	8	1	0	0.2188	0.0731	0.0965	0.3724
2183	7	1	0	0.1875	0.0690	0.0761	0.3369
2447	6	1	0	0.1563	0.0642	0.0570	0.3003
2648	5	1	0	0.1250	0.0585	0.0395	0.2623
2677	4	1	0	0.0938	0.0515	0.0240	0.2228
3206	3	1	0	0.0625	0.0428	0.0111	0.1811
3368	2	0	1	0.0625	0.0428	0.0111	0.1811
5882	1	1	0	0.0000	.	.	.
New Zealand							
6	50	1	0	0.9800	0.0198	0.8664	0.9972
12	49	1	0	0.9600	0.0277	0.8494	0.9898
13	48	2	0	0.9200	0.0384	0.8007	0.9692
20	46	1	0	0.9000	0.0424	0.7763	0.9571
31	45	1	0	0.8800	0.0460	0.7522	0.9442
36	44	1	0	0.8600	0.0491	0.7286	0.9307
57	43	1	0	0.8400	0.0518	0.7054	0.9166
59	42	1	0	0.8200	0.0543	0.6826	0.9020
83	41	1	0	0.8000	0.0566	0.6602	0.8870
104	40	1	0	0.7800	0.0586	0.6381	0.8716
143	39	1	0	0.7600	0.0604	0.6163	0.8559
199	38	1	0	0.7400	0.0620	0.5947	0.8399
224	37	1	0	0.7200	0.0635	0.5735	0.8236
305	36	1	0	0.7000	0.0648	0.5525	0.8070
326	35	2	0	0.6600	0.0670	0.5114	0.7730
390	33	1	0	0.6400	0.0679	0.4911	0.7557
391	32	1	0	0.6200	0.0686	0.4711	0.7381
392	31	1	0	0.6000	0.0693	0.4513	0.7204
407	30	1	0	0.5800	0.0698	0.4318	0.7024
450	29	1	0	0.5600	0.0702	0.4124	0.6842
462	28	1	0	0.5400	0.0705	0.3933	0.6658

522	27	1	0	0.5200	0.0707	0.3743	0.6472
534	26	1	0	0.5000	0.0707	0.3556	0.6283
631	25	1	0	0.4800	0.0707	0.3371	0.6093
725	24	1	0	0.4600	0.0705	0.3188	0.5901
727	23	1	0	0.4400	0.0702	0.3007	0.5707
819	22	1	0	0.4200	0.0698	0.2829	0.5510
824	21	1	0	0.4000	0.0693	0.2652	0.5312
926	20	1	0	0.3800	0.0686	0.2478	0.5112
1096	19	1	0	0.3600	0.0679	0.2306	0.4909
1130	18	1	0	0.3400	0.0670	0.2137	0.4704
1170	17	1	0	0.3200	0.0660	0.1970	0.4497
1204	16	1	0	0.3000	0.0648	0.1806	0.4287
1290	15	1	0	0.2800	0.0635	0.1645	0.4075
1351	14	1	0	0.2600	0.0620	0.1487	0.3860
1573	13	1	0	0.2400	0.0604	0.1331	0.3642
1839	12	1	0	0.2200	0.0586	0.1180	0.3421
1866	11	1	0	0.2000	0.0566	0.1032	0.3197
2004	10	0	1	0.2000	0.0566	0.1032	0.3197
2018	9	1	0	0.1778	0.0545	0.0867	0.2952
2061	8	1	0	0.1556	0.0520	0.0709	0.2702
2593	7	1	0	0.1333	0.0491	0.0560	0.2445
2838	6	1	0	0.1111	0.0457	0.0420	0.2181
3149	5	1	0	0.0889	0.0416	0.0292	0.1907
3548	4	1	0	0.0667	0.0366	0.0177	0.1622
4074	3	1	0	0.0444	0.0304	0.0083	0.1322
4687	2	1	0	0.0222	0.0219	0.0018	0.1009
4791	1	1	0	0.0000	.	.	.



Parametric vs. O.L.S. Estimates

● Let's Contrast O.L.S. with Standard Parametric (using log(*t*) as the dependent variable):

Weibull

```
streg  Labour Liberal National Reform, robust dist(weibull) time nolog, if
NZ==1;

        failure _d:  Event
  analysis time _t:  time

Weibull regression -- accelerated failure-time form

No. of subjects       =           50                 Number of obs   =        50
No. of failures       =           49
Time at risk          =        54439
                                                     Wald chi2(4)    =     18.20
Log pseudo-likelihood =   -88.259242                 Prob > chi2     =    0.0011

------------------------------------------------------------------------------
             |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Labour |   1.051535   .4042052     2.60   0.009     .2593073    1.843763
     Liberal |   1.024581   .4639377     2.21   0.027     .1152803    1.933883
    National |   1.332693    .369094     3.61   0.000     .6092823    2.056104
      Reform |    1.31396   .6623609     1.98   0.047     .0157565    2.612164
       _cons |   6.200884   .2334565    26.56   0.000     5.743318     6.65845
-------------+----------------------------------------------------------------
       /ln_p |  -.1501097   .1254669    -1.20   0.232    -.3960204     .095801
-------------+----------------------------------------------------------------
           p |   .8606136   .1079786                      .672993     1.10054
         1/p |   1.161962   .1457878                     .9086448      1.4859
------------------------------------------------------------------------------
```

O.L.S.

```
reg logT  Labour Liberal National Reform, robust , if NZ==1

Regression with robust standard errors                Number of obs =      50
                                                      F(  4,    45) =    2.11
                                                      Prob > F      =  0.0953
                                                      R-squared     =  0.1423
                                                      Root MSE      =   1.673

------------------------------------------------------------------------------
             |               Robust
        logT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Labour |   1.280525   .5448982     2.35   0.023     .1830437    2.378006
     Liberal |   1.067428   .7036195     1.52   0.136    -.3497348     2.48459
    National |   1.568772   .6528948     2.40   0.020     .2537747     2.88377
      Reform |   .7856566   1.463219     0.54   0.594    -2.161417     3.73273
       _cons |   5.417903   .3642504    14.87   0.000     4.684265    6.151541
------------------------------------------------------------------------------
```

Differences?

●O.L.S. estimates of mean survival (i.e. exp[log(t)]):

Labour: 811 Days
National: 1,082 Days

● Weibull estimates of mean survival time:

Labour: 1,524 Days
National: 2,019 Days

● Weibull estimates of median survival time:

Labour: 922 Days
National: 1,221 Days

● Difference in survival time predictions:

● O.L.S. "under" predicts mean Labour P.M. survival by 713 days.

● O.L.S. "under" predicts mean National P.M. survival by 937 days.

● Obvious Question: which model is preferred?

● Weibull…or some other E.H. variant!

COX MODEL vs. WEIBULL MODEL

● Let's Contrast a Standard Parametric Model with a Cox Model (coefficients are presented in terms of hazard ratios):

First, Weibull:

```
. streg  Labour Liberal National Reform, robust dist(weibull) nolog, if NZ==1

        failure _d:  Event
  analysis time _t:  time

Weibull regression -- log relative-hazard form

No. of subjects       =           50                 Number of obs   =        50
No. of failures       =           49
Time at risk          =        54439
                                                     Wald chi2(4)    =     15.45
Log pseudo-likelihood =   -88.259242                 Prob > chi2     =    0.0038

------------------------------------------------------------------------------
             |               Robust
          _t |  Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Labour |   .4045559   .1372123    -2.67   0.008     .2081031    .7864635
     Liberal |     .41405   .1702806    -2.14   0.032      .184923    .9270745
    National |   .3176091    .108233    -3.37   0.001     .1628641    .6193846
      Reform |   .3227711   .1968266    -1.85   0.064     .0976855    1.066496
-------------+----------------------------------------------------------------
       /ln_p |  -.1501097   .1254669    -1.20   0.232    -.3960204     .095801
-------------+----------------------------------------------------------------
           p |   .8606136   .1079786                      .672993     1.10054
         1/p |   1.161962   .1457878                     .9086448      1.4859
------------------------------------------------------------------------------
```

Second, Cox:

```
    stcox  Labour Liberal National Reform, robust efron nolog, if NZ==1

        failure _d:  Event
  analysis time _t:  time

Cox regression -- Efron method for ties

No. of subjects       =           50                 Number of obs   =        50
No. of failures       =           49
Time at risk          =        54439
                                                     Wald chi2(4)    =     13.68
Log pseudo-likelihood =   -139.89694                 Prob > chi2     =    0.0084

------------------------------------------------------------------------------
             |               Robust
          _t |  Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Labour |   .3906447   .1438303    -2.55   0.011     .1898369    .8038651
     Liberal |   .2920632   .1624966    -2.21   0.027     .0981499    .8690885
    National |   .2766574    .122667    -2.90   0.004     .1160188    .6597147
      Reform |   .2480331   .1884249    -1.84   0.066     .0559593    1.099378
------------------------------------------------------------------------------
```

Differences?

- **Not many**

- **Hazard Ratio for Labour under Weibull is: .40**

- **Hazard Ratio for Labour under Cox is: .39**

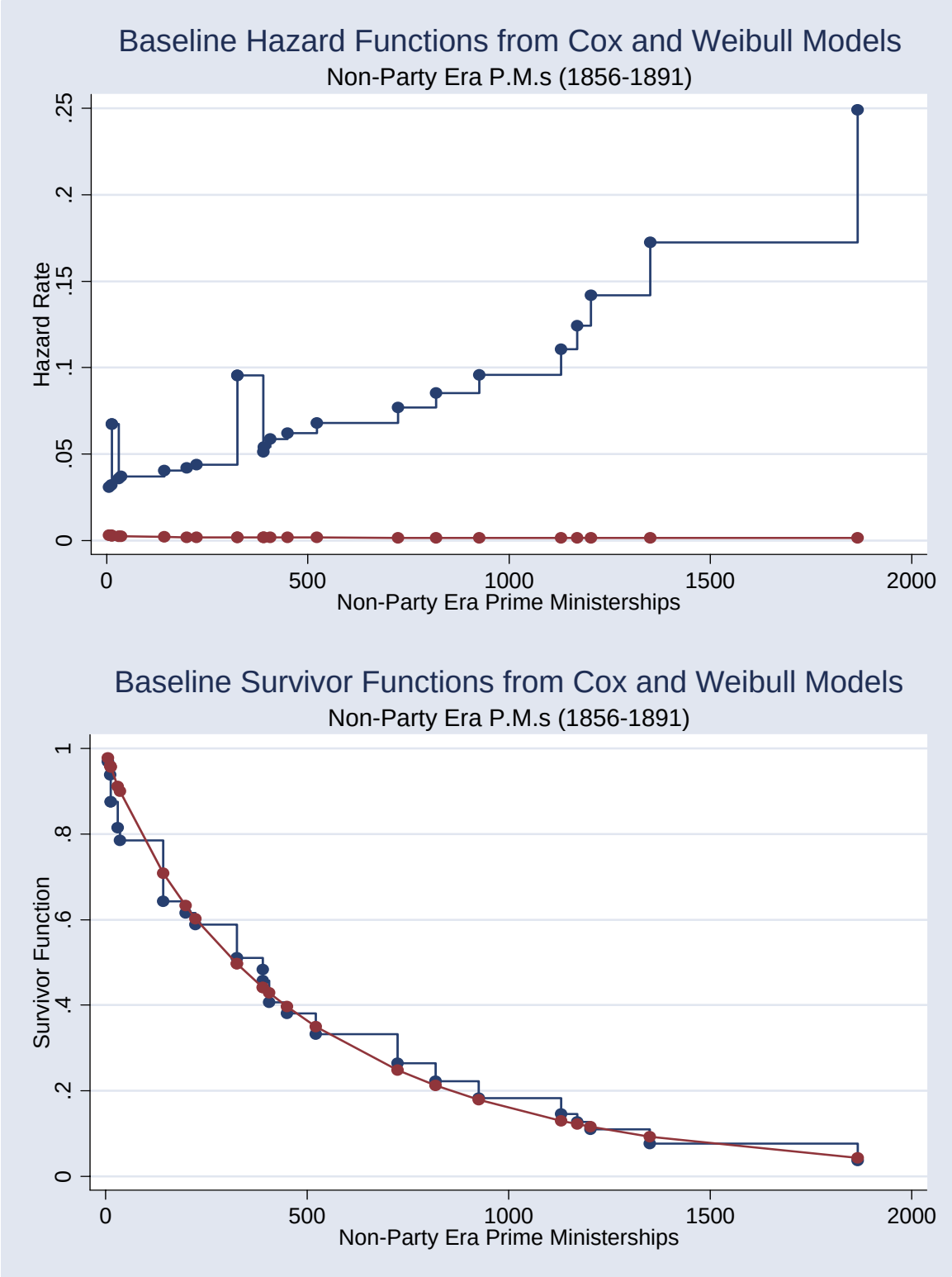
- **Interpretation? The risk of a Labour P.M. exiting office is about 60 percent lower compared to the baseline category ("No Party").**

- **That is: (.40-1)/1=-.60 or "about 60 percent lower."**

- **The hazard ratio for Liberal P.M.s differs between the two models somewhat, but most other estimates are similar.**

- **Major difference in models is how the baseline hazard function is treated: Weibull parameterizes it; Cox does not.**

Illustrating Cox and Weibull Baseline Functions



Stata 8 Code to Replicate Results in these Notes

```
#delimit ;

version 8;


/*Use with PMnzoz.dta*/

/*Creating an observation number for each case*/

gsort -NZ;
gen obsNZ=(_n+1)-1 if NZ==1;

gsort -NZ;
gen obsOZ=(_n+1)-51 if NZ==0;

gen onset_of_partysystem=1 if NZ==1 & obsNZ>=25;
replace onset_of_partysystem=0 if NZ==1 & onset_of_partysystem==.;

/*************************************************************************
Date Functions: Taking a "string" variable and coverting it
 to a numeric variable.  Stata uses the convention, time from Jan. 1, 1960
 to determined the number of days (before or after) .  This is useful because
often, you will have duration data recorded as dates, rather than as time...
but you need time to do the analysis!
*************************************************************************/

gen entry = date(tookoffice, "dmy");
gen exit = date(leftoffice, "dmy");

/*I can turn the variables "entry" and "exit" into something
a bit more readable*/

format entry %d;
format exit %d;

/*Now I can compute my timing variable.  In this case
time is the number of days a PM served*/

gen time=exit-entry;
gen logT=log(time);


/*****************************
Now I've got my timing indicator;
I can stset my data for survival time
analysis.
*******************/

stset time, failure(Event);

/*Creating Party Indicator Variables; the omitted category is "no party" */
```

```
 gen Labour  = 1 if party=="Labour" | party=="Labor";
 replace Labour = 0 if Labour==.;

 gen Liberal  = 1 if party=="Liberal";
 replace Liberal = 0 if Liberal==.;

 gen National  = 1 if party=="National" | party=="Country";
 replace National = 0 if National==.;

gen Reform =1 if party=="Reform";
replace Reform = 0 if Reform==.;

gen NZLabour=NZ*Labour;
gen NZLiberal=NZ*Liberal;
gen NZNational=NZ*National;


/*********************************
The following goes through sequence of models
in stylized introduction.
*********************************/

/*Garden-variety OLS model*/

reg time Labour Liberal National Reform, robust, if NZ==1;


/*****************************************
These graphs are in the lecture notes.
Graphing the time variable to show skewness
*****************************************/

sort NZ;
by NZ: summ time, detail;

histogram time, bin(17) frequency kdensity ytitle(Frequency) xtitle(Duration of Prime
Ministerships (in days))
title(Duration of New Zealand Prime Ministerships) subtitle(From Sewell (1856) to Clark
(present))
caption(Mean Duration=1088.78 Days       Median Duration=582.5 Days, size(small) span
box) saving(nzpmhistogram, replace),
if NZ==1
;
histogram time, bin(17) frequency kdensity ytitle(Frequency) xtitle(Duration of Prime
Ministerships (in days))
title(Duration of Australian Prime Ministerships) subtitle(From Barton (1901) to Howard
(present))
caption(Mean Duration=1191.72 Days       Median Duration=830.5 Days, size(small) span
box) saving(ozpmhistogram, replace),
if NZ==0
;
histogram time, bin(17) frequency kdensity ytitle(Frequency) xtitle(Duration of Prime
Ministerships (in days))
title(Duration of N.Z. and Australian Prime Ministerships)
caption(Mean Duration=1128.95 Days       Median Duration=708.5 Days, size(small) span
box) saving(nzozpmhistogram, replace);
```

```
/***************************************
OLS "fixups" using log(t) and deleting censored
cases.
***************************************/

reg logT Labour Liberal National Reform, robust, if NZ==1 ;

reg time Labour Liberal National Reform, robust, if NZ==1  & Event==1;

/***********************************************
To illustrate TVC "jump process" we sidetrack to
a "data set" featuring the premiership spells for
Jim Bolger
***********************************************/

save, replace;

use bolgerdata.dta;

/*Graphing Jump Process (pedagogical use only!)*/

twoway (scatter Labour Time, connect(J)) (scatter National Time, c(J)), ytitle(Seats Held by
Labour and National)
xtitle(Duration of P.M. Jim Bolger's Premiership (in days)) title(Seats Held by Labour and
National )
subtitle(During Jim Bolger's Premiership) name(jimbolgerjump, replace)
saving(jimbolgerjump.gph, replace);

clear;

use PMnzoz;


/*********************************************************
A very useful descriptive tool for survival analysis
is the Kaplan-Meier estimator.  Below, I use the premiership
data to illustrate the K-M Estimates
*********************************************************/

sts list, by(NZ);

sts graph, by(NZ) ytitle(Survivor Function) xtitle(Duration of Prime Ministerships (N.Z. and
Aus.)) title
(Kaplan-Meier Survival Function) subtitle(Prime Ministerships in N.Z. and Australia)
legend(on) saving(KMnzoz, replace);



/*Fun with Parametric Models*/

/*OLS vs. Weibull*/

/*Weibull*/

 streg  Labour Liberal National Reform, robust dist(weibull) time nolog, if NZ==1;
```

```
predict median, median, if e(sample);
predict mean, mean, if e(sample);
predict hazard, hazard, if e(sample);

table median party if NZ==1;
table mean party if NZ==1;

/*OLS*/
reg logT Labour Liberal National Reform, robust , if NZ==1;

predict xb, xb, if e(sample);
gen expxb=exp(xb);

table expxb party if NZ==1;


/*Cox vs. Weibull (obtain hazard ratio parameterization (i.e. exp(b))*/

/*Cox*/
 stcox  Labour Liberal National Reform, robust efron nolog basehc(hazCOX)
basesurv(survCOX), if NZ==1;

/*Weibull*/

  streg  Labour Liberal National Reform, robust dist(weibull) nolog, if NZ==1;

predict surv, surv;

 twoway (scatter survCOX time if Labour==0 & Liberal==0 & National==0 & Reform==0 &
NZ==1, sort c(J))
(scatter surv time if Labour==0 & Liberal==0 & National==0 & Reform==0 & NZ==1, sort
c(l)), ytitle(Survivor Function)
 xtitle(Non-Party Era Prime Ministerships) title(Baseline Survivor Functions from Cox and
Weibull Models) subtitle(Non-Party Era P.M.s (1856-1891)) legend(off)
saving(survivor,replace);



 twoway (scatter hazCOX time if Labour==0 & Liberal==0 & National==0 & Reform==0 &
NZ==1, sort c(J))
(scatter hazard time if Labour==0 & Liberal==0 & National==0 & Reform==0 & NZ==1, sort
c(l)), ytitle(Hazard Rate)
 xtitle(Non-Party Era Prime Ministerships) title(Baseline Hazard Functions from Cox and
Weibull Models) subtitle(Non-Party Era P.M.s (1856-1891)) legend(off)
saving(hazard,replace);


drop obsNZ obsOZ onset_of_partysystem entry exit time logT _st _d _t _t0 Labour Liberal
National
Reform NZLabour NZLiberal NZNational median mean hazard xb expxb survCOX hazCOX
surv;


save, replace;
```