

Event history, binary probit, Markov chains: when should we care with applications to comparative politics and international relations

Nathaniel Beck*[†]

Revised for the Spring SCAMP meeting, UCR, May 2,
2003

*Department of Political Science; University of California, San Diego; La Jolla, CA 92093; beck@ucsd.edu

[†]Obvious thanks and debts to David Epstein, Simon Jackman and Adam Przeworski for many instructive conversations, not all of which I took advantage of. Data were kindly provided by Jose Cheibub (the democratic transitions data analyzed in Przeworski, et al.), David Epstein (state failure data, courtesy of the State Failure Task Force) and John Oneal (from data analyzed in Oneal and Russett).

Overview

The data is the same; the analysts differ (Alt, King and Signorino)

- Binary time-series
- Can do for single time-series, but usually won't work
- Here always for time-series–cross-section data (TSCS)
- Most commonly done by (ordinary) probit or logit (OP)
- Event history (how long til something happens) (EH or BKT)
- Transitions (usually Markovian) - Probability of something happening (as a function of what happened before) (TRANS)
- Examples
 - State Failure
 - Transitions to and from democracy
 - Conflict

Ordinary Probit

Ordinary Probit (logit similar, but maths of probit helps some notation and computation, and no reason to prefer logit to probit)

$$y_{i,t}^* = \mathbf{x}_{i,t}\beta + \epsilon_{i,t} \quad (1)$$

$$y_{i,t} = 1 \text{ if } y_{i,t}^* > 0 \quad (2)$$

$$y_{i,t} = 0 \text{ otherwise} \quad (3)$$

$$\epsilon_{i,t} \sim N(0, 1) \quad (4)$$

In epidemiological terms, OP models INCIDENCE, whereas we might want to model ONSET

Transition Model

$$P(y_{i,t} = 1 | y_{i,t-1} = 0) = \text{Probit}(\mathbf{x}_{i,t}\beta) \quad (5)$$

$$P(y_{i,t} = 1 | y_{i,t-1} = 1) = \text{Probit}(\mathbf{x}_{i,t}\alpha) \quad (6)$$

which can be written more compactly as

$$P(y_{i,t} = 1) = \text{Probit}(\mathbf{x}_{i,t}\beta + y_{i,t-1}\mathbf{x}_{i,t}\gamma) \quad (7)$$

where

$$\gamma = \alpha - \beta. \quad (8)$$

- Note that the two parts of the transition model (Eqs. 5 and 6) can be estimated IDENTICALLY by separate probits. Thus we can think of modeling the transition from a 0, then modeling the transition from a 1, and we can do so independently.

- This is a bit different from the continuous DV case, but even there the gains from joint estimation are small
- Of course Equation 7 is convenient for testing hypotheses about whether the two transition processes are similar.
- Note that unlike OP, which assumes that we have the same model for $y_{i,t}$ regardless of $y_{i,t-1}$, here we have two different models depending on the value of $y_{i,t-1}$
- Of course one can test the assumption that the two processes are the same, or they are governed by the same covariates

Lagged Dependent Variable (Restricted Transition)

- People often try to treat dynamics by “throwing in a lagged dv” by analogy to the continuous case

$$y_{i,t}^* = \mathbf{x}_{i,t}\beta + \rho y_{i,t-1} + \epsilon_{i,t} \quad (9)$$

- Note that this is just a very simple restriction on the transition model, since it assumes that all parms relating the iv's to y are the same regardless of whether lagged y is 0 or 1, but the intercept shifts between those two cases (by ρ).
- Thus if one wanted to work with the LDV case, one should start with the transition model and then test the restriction that $\gamma = 0$ in Equation 7 (other than for the constant term).

LDV vs Lagged Latent

- Does the LDV model make sense?
- It says that “failing” makes you more likely to fail, success makes you more likely to succeed.
- Argument in labor economics: true state dependence (Freeman) vs spurious state dependence (Heckman) - does it matter if a teenager accidentally gets a job at McDonald’s
- Does it matter if a state fails by accident (accident is low y^* but $y = 1$)?

Lagged Latent

$$y_{i,t}^* = \mathbf{x}_{i,t}\beta + \rho y_{i,t-1}^* + \epsilon_{i,t} \quad (10)$$

- This is the exact analog to continuous LDV case, and makes sense in the way that it makes sense for that case.
- Thus there is some underlying TS process in the latent that behaves like a normal time series, but we can only observe whether that process is below or above some threshold.
- Note that it is very easy to estimate model with lagged y but very hard with lagged latent, which is why folks like lagged y . But modern methods (MCMC) make the lagged latent feasible.
- Note: Can also use MCMC to estimate a model with serially correlated errors (these are not observed, we only know if below or above some threshold, but same argument as for continuous case that these are not useful, so stick to lagged latent).

Is Time Series right?

- If we use time series, typically have same model for low and high y , nothing about different models.
- Democracy: transition model says that transitions from democracy and autocracy different
- Let us ignore issue that continuous measures of democracy, such as Polity Scores, are quite bad
- The TS model says process the same
- This is true if we actually observe $y_{i,t}$ or if we use latent methods and model with $y_{i,t}^*$.
- Thus if we believe the transition model story, we should probably not even use time series models, even if we could
- If we use, should think of switching Markov models, though with latent y^* we now have two latents, asking a lot of the data
- Even with only one latent (y^*), can we really predict the latent knowing only its sign and some economic variables
- Will latent for dems hover near one, latent for autocs hover near zero?
- Can we with Kalman filter type setup, measurement equation uses the Polity scores?

Event History

- How long until something happens?
- How many zeros will we observe until we get a one?
- How many years of peace between militarized conflicts?
- How long do conflicts last (how many years of conflict between peaces)?
- How long til a state fails?
- How long are spells of state failure?
- How long are spells of democracy? autocracy? (Przeworski et al.)
- Often done in continuous time: model the length of spells
- But equally (and probably data is) discrete time (Sueyoshi)
- $P(3 \text{ years peace, war in 4th year}) = P(0001) = P(0)P(0|0)P(0|00)[1 - P(0|000)]$
- Assume that $P(0001)$ is same for every time slice, so $P(0001)$ is same starting in 1952 and 1988 (if not, then length of spell of peace is not what is of interest)
- Right censoring (spells of peace unended at end of data set) easy to deal with

- Left censoring is not trivial - is data set starts in 1950, is 1950 the first year of peace?
- But this is a problem for all types of analyses unless one assumes DURATION INDEPENDENCE

Duration Dependence

- In EH terminology, we have DURATION INDEPENDENCE (with discrete data)
- if $P(\text{fail at } t \mid \text{no fail at } t-1)$ is same for all t .
- Otherwise DURATION DEPENDENCE
- Note the conditional Prob above is the discrete hazard
- Much easier to understand than for continuous data
- IF DURATION INDEPENDENCE, then
- Does NOT matter when spell started (ie can cut in in the middle)
- This is because all we have are probits
($P(y_{i,t} = 1 \mid y_{i,t-1} = 0)$)
- But data setup, which eliminates all data where lagged $y=1$ allows for simple probit estimation
- Thus OP is fine (and same as EH) if Duration Independence
- But OP is bad for duration dependence
- Fix: If data follow proportional hazards (Cox), then
- Just add time counter dummies to probit (Beck, Katz and Tucker or BKT)

Transition vs BKT

- The transition model can be estimated as to separate probits (breaking data into subset with $y_{i,t-1} = 0$ and $y_{i,t-1} = 1$).
- Note in EH, we have a discrete model for the lengths of spells of 0's and a separate model for the lengths of spells of 1's.
- But because there are no time vars in the two models, the transition model assumes these spells are duration independent.
- That is, discrete EH with proportional hazards (BKT) estimates

$$P(y_{i,t} = 1 | y_1 = 0, \dots, y_{i,t-1} = 0) = \mathbf{x}_{i,t}\beta + \gamma I_{i,t} \quad (11)$$

whereas the transition model estimates

$$P(y_{i,t} = 1 | y_{i,t-1} = 0) = \mathbf{x}_{i,t}\beta \quad (12)$$

which has no duration dependence (except, for first period of spell of peace)

- BKT (1998) focused on the transition from 0 to 1 (that is, lengths of spells of peace) because that

was what was interesting in the democratic peace (or other cases with very rare 1's and very short spells of them - this is also the case in state failure, is clearly not always the case).

- But can easily allow for duration dependence in either or both spell lengths. Again, work with the two models separately
- Thus there is no essential difference between the transition approach and BKT, except that BKT, in allowing for duration dependence, is more general.

• First periods

- Note that the transition model (Equation 5 and event history (spells of zeros) is a bit different in the first term
- For event history we have

$$P(y_{i,t} = 1 | y_{i,t-1} = 0) = \text{Probit}(\mathbf{x}_{i,t}\beta) + s(t) \text{ if } t > 1 \quad (13)$$

$$P(y_{i,t} = 1) = \text{Probit}(\mathbf{x}_{i,t}\beta) + s(1) \text{ if } t = 1 \quad (14)$$

$$(15)$$

whereas the Equation 5 only estimates Equation 13.

- This follows from the event history approach of assuming that we observe the beginning of all spells and not having a good treatment of left censoring
- If 1950 is first year of conflict data, can we count a year of peace in 1950 as the first year of a spell of peace
- How square with folks doing analysis starting 1870 or even 1816?
- There are some fixes for this that I have used, and

the $s(1)$ term may sop up any problems with the first observation, and for long spells it is unlikely that adding or dropping the first observation matters a lot, but this is something that needs further investigation.

- Note that for any unit, the transition model drops the first observation, since no data on the lagged y is available. Is this better? Does this matter?
- Also, after a transition (that is $0, \dots, 0, 1, \dots, 1, 0$) the final zero is modeled in the equation based on a lagged value of $y_{i,t-1} = 1$.
- In EH it would both end the spells of 1's but also start the new spell of 0's.
- Problem is caused by annual data; thus the final 1 just tells us that a spell of 1's ended in that year; if we had finer grained data, we could model that.
- The transition model is inherently discrete, so wants such annual data.
- In practice does this matter much
- Not much, if spells are long

EH (BKT) vs TRANS

- Transition model (and BKT) much better than OP
- But it assumes duration independence
- If little duration dependence, then TRANS and BKT similar
- If there is some durdep, then BKT better

Data Analysis

- All results reported are probits
- Missing data (not much except for state failure) dropped
- Can discuss that, but note that King's procedure not right here
- Huber standard errors, clustered on dyad, used for MID analysis
- Other two analyses use ordinary se's

MID's (Militarized Interstate Disputes)

1951–1992 - Oneal and Russett

- Spells of disputes
 - 2048 dyad-years of disputes (spells of dispute with first year of peace)
 - Dyads may have multiple spells
 - 307 dyads
 - 636 different spells of dispute not right censored
 - 52 spells of disputes right censored
 - Typical spell length of disputes is short
 - Mean length of disputes is 3.7, median is 2,
- Spells of peace
 - 32376 dyad-years of peace
 - (new dispute number is new dispute, not second year of previous dispute)
 - 1094 dyads
 - 1570 spells of peace right censored
 - 1061 spells not right censored (ends in dispute)

Transitions from Dem to Aut and Vice-versa 1951-1990 - Przeworski, et al.

- 135 countries
- Spells of democracy
 - 1683 country years
 - 72 spells of democ
 - 38 spells of democ end in autoc
 - 34 spells of democ right censored
- Spells of autocracy
 - 2530 country yeas item 101 spells of autoc
 - 49 spells end in democ
 - 52 spells right censored

State Failure (State Failure Task Force) - 1955-1999

- 151 countries
- Spells of nofailure
 - 4143 country years
 - 93 spells end in failure
 - 122 censored spells
- Spells of failure
 - 959 country years
 - 92 spells end in non-failure
 - 29 censored spells

MIDs

Table 1: MIDs (Huber SE's)

Variable	ALL		PEACE		MID		BKT	
	b	SE	b	SE	b	SE	b	SE
DEM	-.06	.01	-.06	.01	-.02	.02	-.05	.01
TRADE	-93.83	31.82	-25.68	14.3	-115.62	42.9	-24.72	14.00
MAJPOW	.74	.25	.62	.22	.00	.27	.59	.17
LCAPRAT	-.29	.06	-.14	.05	-.23	.07	-.27	.05
LDIST	-.38	.09	-.31	.07	-.12	.10	-.21	.07
CONTIG	1.59	.24	1.61	.22	-.08	.30	1.19	.19
ALLIES	-.85	0.19	-.51	.15	-.52	.21	-.44	.14
C	-.86	.69	-2.53	.58	1.55	.81		PY
N	32727		29745		1360		31174	

Note: SE's are Huber, clustered on dyad

MIDs - What does GEE do?

Table 2: MIDs - Logit, Robust. GEE

Variable	b	Logit		GEE		
		SE	Robust SE	b	SE	Robust SE
DEM	-.06	.01	.01	-.08	.01	.02
TRADE	-93.83	11.63	31.82	-89.70	18.03	51.77
MAJPOW	.74	.09	.25	1.11	.12	.29
LCAPRAT	-.29	.02	.06	-.37	.03	.07
LDIST	-.38	.03	.09	-.32	.05	.12
CONTIG	1.59	.09	.24	1.82	.12	.28
ALLIES	-.85	.07	.19	-1.10	.12	.27
C	-.86	.24	.69	-1.50	.36	.99
ρ				.11		

N=32727

Democracy/Autocracy

Table 3: Transitions from Autocracy and Democracy TO DEMOCRACY

Variable	ALL		DEMLAG		From AUT		From DEM	
	b	SE	b	SE	b	SE	b	SE
GDPLAG	.33	.01	.16	.02	.12	.03	.22	.05
GDPLAG%	-.57	.35	-.18	.69	-1.97	.85	3.96	1.38
DEMLAG			3.75	.10				
C	-1.32	.04	-2.41	.08	-2.30	.10	1.12	.14
N	4126		3991		2407		1584	

NOTE: While significant durdep in both transition equations, coefficients and se's change by under 10%

NOTE: GDPLAG in thousands of dollars, GDPLAG% as decimal

Democracy/Autocracy/TS

Table 4: Time Series Using Polity 21 point Scale

Variable	ALL		Pol < -3		Pol > -4	
	b	SE	b	SE	b	SE
GDPLAG	.07	.01	.10	.02	.06	.01
GDPLAG%	-.47	.42	-1.41	.45	1.28	.79
Polity Lag	.96	.004	1.04	.02	.99	.02
C	-.25	.04	.30	.19	-.46	.09
N	3717		1836		1881	

NOTE: Years with Polity Scores of -66, -77 or -88 dropped (not right!) - some countries have no Polity Scores (not in “world system”)

Note: GDPLAG and GDPLAG% scaled as in previous slide

State Failure

Table 5: Prob of State Failure

Variable	ALL		FAILLAG		From FAIL		From NonFAIL	
	b	SE	b	SE	b	SE	b	SE
OPEN	-.94	.08	-.47	.13	-.59	.20	.43	.15
LINFMORT	.19	.03	.21	.06	.14	.10	.22	.07
AUTO	-.22	.06	-.18	.09	-.30	.16	-.09	.12
DEMOC	-.55	.07	-.18	.12	.26	.23	-.34	.16
FAILLAG			3.09	.07				
C	-.94	.17	-2.39	.27	1.07	.48	-2.46	.34
N	4908		4757		838		3919	