Establishing the Discrete-Time Survival Analysis Model

(ALDA, Ch. 11)

John Willett & Judy Singer

Harvard University Graduate School of Education

May, 2003

What will we cover?

Specifying the DTSA Model

Data Example: *Grade at First Intercourse*

- *Research Question*: Whether, and when, adolescent males experience heterosexual intercourse for the first time?
- *Citation*: Capaldi, et al. (1996).
- *Sample*: 180 high-school boys.
- *Research Design*:
  - Event of interest is the first experience of heterosexual intercourse.
  - Boys tracked over time, from 7th thru 12th grade.
  - 54 (30% of sample) were virgins (censored) at end of data collection.

Specifying the DTSA Model

Extract from the *Person-Level* & *Person-Period* Datasets

Grade at First Intercourse (ALDA, Fig. 11.5, p. 380)

"Person-Level" data set

ID	*T*	*CENSOR*
193	9	0
126	12	0
407	12	1

Not censored ⇒ *did* experience the event

Censored ⇒ did *not* experience the event

"Person-Period" data set

ID	*PERIOD*	*EVENT*
193	7	0
193	8	0
193	9	1
126	7	0
126	8	0
126	9	0
126	10	0
126	11	0
126	12	1
407	7	0
407	8	0
407	9	0
407	10	0
407	11	0
407	12	0

Boy #193 was tracked until he had sex in 9th grade.

Boy #126 was tracked until he had sex in 12th grade.

Boy #407 was censored, remaining a virgin until he graduated.

Specifying the DTSA Model
Estimating Sample Hazard & Survival Probabilities
Grade at First Intercourse (ALDA, Table 11.1, p. 360)

	Number who . . .				
Grade	Were at risk (virgins) at the beginning of the grade	Had sex during the grade	Were censored at the end of the grade	Hazard probability	Survival probability
7	180	15	0	0.0833	0.9167
8	165	7	0	0.0424	0.8778
9	158	24	0	0.1519	0.7444
10	134	29	0	0.2164	0.5833
11	105	25	0	0.2381	0.4444
12	80	26	54	0.3250	0.3000

Discrete-Time Hazard is the conditional probability that the event will occur in the period, given that it hasn't occurred earlier:

$$h(t_j) = \Pr\{T = j \mid T > j\}$$

Estimated by the corresponding sample probability:

$$\hat{h}(t_j) = \frac{n\ \ events_j}{n\ \ at\ \ risk_j}$$

e.g.

$$\hat{h}(t_9) = \frac{8}{68} = 0.1176$$

Survival probability describes the chance that a person will survive beyond the period in question without experiencing the event:

$$S(t_j) = \Pr\{T > j\}$$

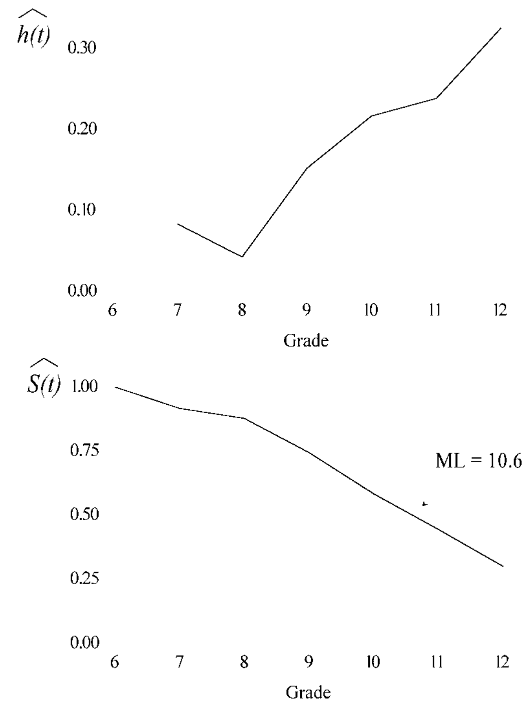
Estimated by cumulating hazard:

$$\hat{S}(t_j) = \hat{S}(t_{j-1})[1 - \hat{h}(t_j)]$$

e.g.,

$$\hat{S}(t_9) = \hat{S}(t_8)[1 - \hat{h}(t_9)]$$
$$= 0.8778[1 - 0.1519]$$
$$= 0.7444$$

Specifying the DTSA Model
Sample Hazard & Survivor Functions
Grade at First Intercourse (ALDA, Fig. 10.2B, p. 340)



Median lifetime, *ML*, is the amount of time it takes for *half* the population (sample) to experience the event of interest.

Specifying the DTSA Model
Introducing a Predictor into the Person-Period Dataset
Grade at First Intercourse (from ALDA, Fig. 11.5, p. 380)

ID	*PERIOD*	*EVENT*	*PT*
193	7	0	1
193	8	0	1
193	9	1	1
126	7	0	1
126	8	0	1
126	9	0	1
126	10	0	1
126	11	0	1
126	12	1	1
407	7	0	0
407	8	0	0
407	9	0	0
407	10	0	0
407	11	0	0
407	12	0	0

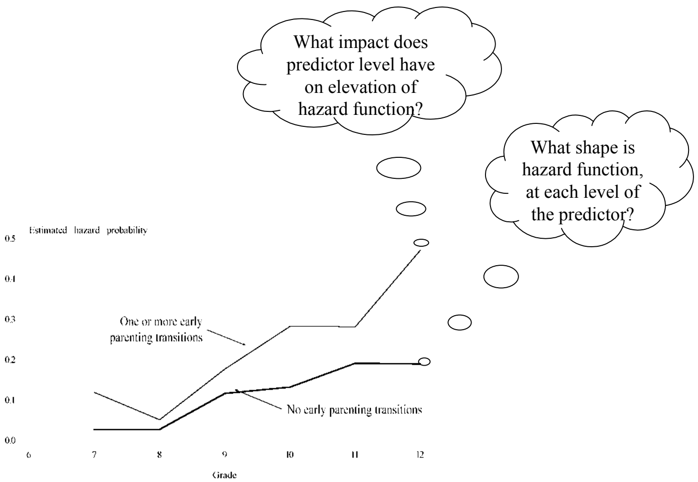
Time-invariant predictor, PT, indicates the presence/absence of a "parenting transition" during the boy's early life:

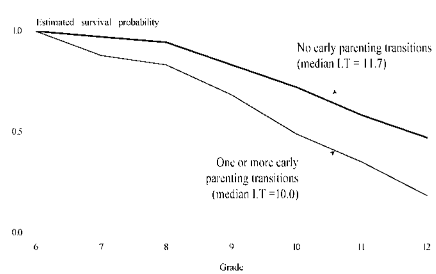
- 0 = boy lived with both biological parents thru 7th grade (n=72, 40% of sample).
- 1 = boy experienced one or more parenting transitions, up thru 7th grade (n=108, 60% of sample).

Specifying the DTSA Model
What Impact Does Predictor PT have?
Computing Sample Hazard & Survivor Probabilities, by *PT*
Grade at First Intercourse (from ALDA, Table 11.1, p. 360)

	Number who . . .				
Grade	Were at risk (virgins) at the beginning of the grade	Had sex during the grade	Were censored at the end of the grade	Hazard probability	Survival probability
No Parenting Transitions (*PT* = 0)					
7	72	2	0	0.0278	0.9722
8	70	2	0	0.0286	0.9444
9	68	8	0	0.1176	0.8333
10	60	8	0	0.1333	0.7222
11	52	10	0	0.1923	0.5833
12	42	8	34	0.1905	0.4722
One or More Parenting Transitions (*PT* = 1)					
7	108	13	0	0.1204	0.8796
8	95	5	0	0.0526	0.8333
9	90	16	0	0.1778	0.6852
10	74	21	0	0.2838	0.4907
11	53	15	0	0.2830	0.3519
12	38	18	20	0.4737	0.1852

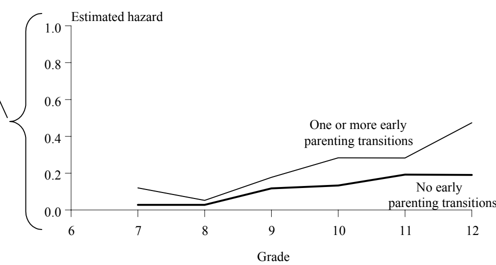
Specifying the DTSA Model
What Impact Does Predictor *PT* have?
Grade at First Intercourse (ALDA. Fig. 11.1, p. 358)





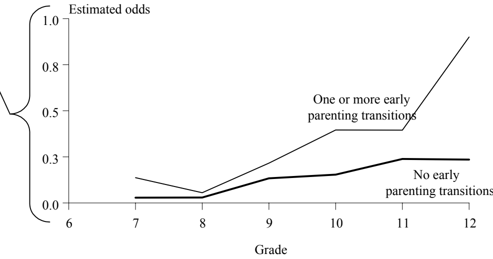
Specifying the DTSA Model
If We Want To Model Hazard What Problems Do We Face?
Grade at First Intercourse (ALDA, Fig. 11.2, 363)

Hazard is a *probability*, bounded by *zero* and *one*

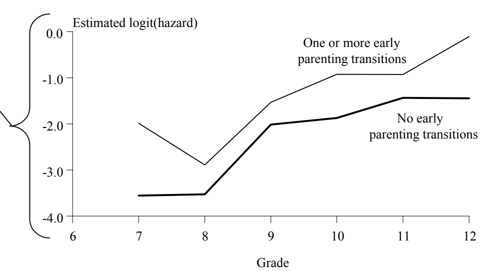


$$Odds = \frac{p}{1-p}$$

are less of a problem, bounded only below, by zero



Log-Odds, or *logits*, are no problem at all, they are without bounds above and below

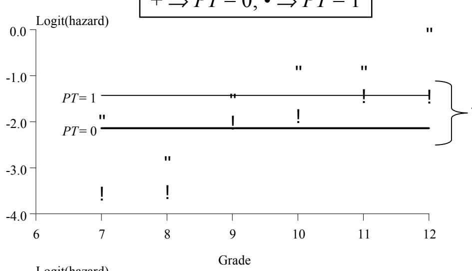


Specifying the DTSA Model

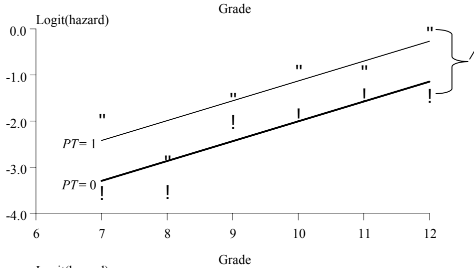
What Population Models Could've Generated These Data?

Grade at First Intercourse Data *(ALDA, Figure 11.3, p. 366)*

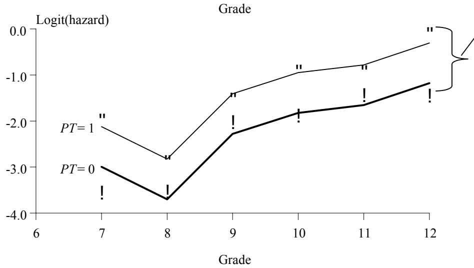
Datapoints:
+ ⇒ *PT* = 0; • ⇒ *PT* = 1



Flat population logit-hazard, elevated when *PT* switches from zero to one.



Population logit-hazard linear with time, elevated when *PT* switches from zero to one.

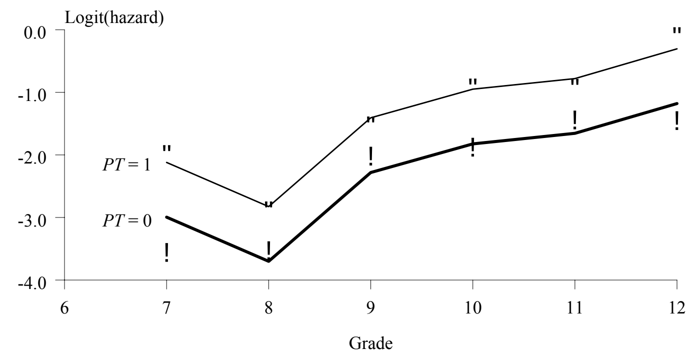


Population logit-hazard has a general shape, elevated when *PT* switches from zero to one.

Specifying the DTSA Model

What Statistical Model Could Have Generated The Data?

Grade at First Intercourse (ALDA, Figure 11.3, p. 366)



What are the necessary features of a reasonable statistical model for discrete-time logit-hazard?

They include:

- For each predictor value, there is a *population logit-hazard function*.
- Each population logit-hazard function has an *identical shape*, regardless of predictor value.
- Differences in predictor value "*shift*" the logit-hazard function "*vertically*"
  - So, the vertical "distance" between pairs of hypothesized logit-hazard functions is the same in every time period.

Specifying the DTSA Model

How Do We Specify A Model That has These Features?

Grade at First Intercourse

Add a set of dummy predictors to the person-period dataset, to represent time *generically*

ID	PERIOD	D7	D8	D9	D10	D11	D12	EVENT	PT
193	7	1	0	0	0	0	0	0	1
193	8	0	1	0	0	0	0	0	1
193	9	0	0	1	0	0	0	1	1
126	7	1	0	0	0	0	0	0	1
126	8	0	1	0	0	0	0	0	1
126	9	0	0	1	0	0	0	0	1
126	10	0	0	0	1	0	0	0	1
126	11	0	0	0	0	1	0	0	1
126	12	0	0	0	0	0	1	1	1
407	7	1	0	0	0	0	0	0	0
407	8	0	1	0	0	0	0	0	0
407	9	0	0	1	0	0	0	0	0
407	10	0	0	0	1	0	0	0	0
407	11	0	0	0	0	1	0	0	0
407	12	0	0	0	0	0	1	0	0

And specify a *statistical model for discrete-time logit-hazard* that looks like this:

$$logit\ h(t_{ij}) = [\alpha_7 D_7 + \ldots + \alpha_j D_j + \ldots + \alpha_{12} D_{12}] + \beta_1 PT_i$$

How does the model work? How do we fit it to data?

Specifying the DTSA Model

How Does It Work?

Grade at First Intercourse Data *(ALDA, Fig. 11.4, p. 374)*

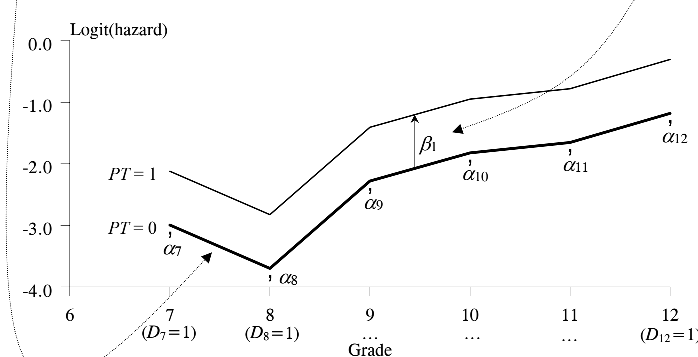
$$logit\ h(t_{ij}) = [\alpha_7 D_7 + \ldots + \alpha_j D_j + \ldots + \alpha_{12} D_{12}] + \beta_1 PT_i$$

When *PT*=0, you get the *Baseline Logit-Hazard Function*:

$$logit\ h(t_{ij}) = [\alpha_7 D_7 + \ldots + \alpha_{12} D_{12}]$$

When *PT*=1, the *Baseline Function* shifts "vertically":

$$logit\ \ h(t_{ij}) = [Baseline\ \ Function\ ] + \beta_1$$



Specifying the DTSA Model

Other Ways of Writing the DTSA Model

Grade at First Intercourse Data *(ALDA, Table 11.2, p. 376)*

Original scale	Desired scale	Use the transformation
Logit	Odds	$Odds = e^{\text{logit}}$
Odds	Probability	$Probability = \dfrac{odds}{1+odds} = \dfrac{e^{\text{logit}}}{1+e^{\text{logit}}}$
Logit	Probability	$Probability = \dfrac{1}{1+e^{-\text{logit}}}$

So, you can "de-transform" the entire logit-hazard model, and it starts to look *recognizable* (?):

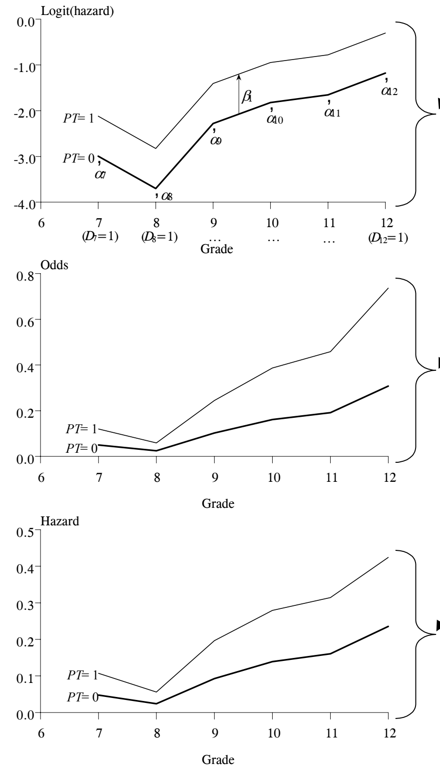
$$h(t_{ij}) = \frac{1}{1+e^{-\{[\alpha_1 D_{1ij}+\ldots+\alpha_J D_{Jij}]+[\beta_1 X_{1ij}+\beta_2 X_{2ij}+\ldots]\}}}$$

Notice how we've begun to generalize the model?

Specifying the DTSA Model

Other Ways of Displaying the DTSA Model

Grade at First Intercourse Data *(ALDA, Fig. 11.4, p. 374)*



When hazard is expressed on a *logit scale*, the vertical distance between functions is *identical* in every time period.

When hazard is expressed on an *odds scale*, the functions are *magnifications* or *diminutions* of each other -- they are *proportional*

?

The "standard" DTSA model is a *proportional odds* model!

Fitting the DTSA Model to Data

First, Add a Continuous Predictor to the *pp* Dataset

Grade at First Intercourse Data *(ALDA, Fig. 11.5, p. 380)*

"Person-Period" data set

ID	*PERIOD*	*D7*	*D8*	*D9*	*D10*	*D11*	*D12*	*EVENT*	*PT*	PAS
193	7	1	0	0	0	0	0	0	1	1.16
193	8	0	1	0	0	0	0	0	1	1.16
193	9	0	0	1	0	0	0	1	1	1.16
126	7	1	0	0	0	0	0	0	1	0.12
126	8	0	1	0	0	0	0	0	1	0.12
126	9	0	0	1	0	0	0	0	1	0.12
126	10	0	0	0	1	0	0	0	1	0.12
126	11	0	0	0	0	1	0	0	1	0.12
126	12	0	0	0	0	0	1	1	1	0.12
407	7	1	0	0	0	0	0	0	0	-0.96
407	8	0	1	0	0	0	0	0	0	-0.96
407	9	0	0	1	0	0	0	0	0	-0.96
407	10	0	0	0	1	0	0	0	0	-0.96
407	11	0	0	0	0	1	0	0	0	-0.96
407	12	0	0	0	0	0	1	0	0	-0.96

PAS is a *continuous* time-invariant measure of parents' antisocial behavior during the child's formative years. Scores on the measure have been standardized to mean 0, standard deviation 1.

Fitting the DTSA Model to Data

Use Logistic Regression Analysis in the PP Dataset

Grade at First Intercourse

Use *logistic regression* analysis to fit the hypothesized DTSA model in the *person-period dataset*.

Treat *EVENT* as the outcome, and regress it on the predictors:

- Time indicators, D_1 thru D_J,
- Substantive predictors, *PT* and *PAS*.

Using sensible data-analytic practices.

All parameter estimates, standard errors, t- and z-statistics, goodness-of-fit statistics, and tests will be correct for the discrete-time hazard model

Fitting the DTSA Model to Data

Using Sensible Data-Analytic Skills to Produce a Taxonomy of Fitted DTSA Models

Grade at First Intercourse (ALDA, Table 11.3., p. 386)

	Model A	Model B	Model C	Model D
Parameter Estimates and Asymptotic Standard Errors				
D_7	−2.3979***	−2.9943***	−2.4646***	−2.8932***
	(0.2697)	(0.3175)	(0.2741)	(0.3206)
D_8	−3.1167***	−3.7001***	−3.1591***	−3.5847***
	(0.3862)	(0.4206)	(0.3890)	(0.4231)
D_9	−1.7198***	−2.2811***	−1.7297***	−2.1502***
	(0.2217)	(0.2724)	(0.2245)	(0.2775)
D_{10}	−1.2867***	−1.8226***	−1.2851***	−1.6932***
	(0.2098)	(0.2585)	(0.2127)	(0.2647)
D_{11}	−1.1632***	−1.6542***	−1.1360***	−1.5177***
	(0.2291)	(0.2691)	(0.2324)	(0.2757)
D_{12}	−0.7309**	−1.1791***	−0.6421**	−1.0099***
	(0.2387)	(0.2716)	(0.2428)	(0.2811)
PT		0.8736***		0.6605**
		(0.2174)		(0.2367)
PAS			0.4428***	0.2964*
			(0.1140)	(0.1254)
Goodness-of-fit				
LL	−325.98	−317.33	−318.59	−314.57
Deviance	651.96	634.66	637.17	629.15
n parameters	6	7	7	8
AIC	663.96	648.66	651.17	645.15
BIC	681.00	668.54	671.05	667.87
Deviance-based Hypothesis Tests				
$H_0: \beta_{PT} = 0$		17.30*** (1)		8.02** (1)
$H_0: \beta_{PAS} = 0$			14.79*** (1)	5.51* (1)
Wald Hypothesis Tests				
$H_0: \beta_{PT} = 0$		16.15*** (1)		7.79** (1)
$H_0: \beta_{PAS} = 0$			15.10*** (1)	5.59* (1)

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Interpreting Parameter Estimates

Interpreting Parameters Associated w/ the Time Dummies

Grade at First Intercourse (ALDA, Figs. 11.3 & 11.4, 386-8)

	Model A
Parameter Estimates and Asymptotic Standard Errors	
D_7	−2.3979***
	(0.2697)
D_8	−3.1167***
	(0.3862)
D_9	−1.7198***
	(0.2217)
D_{10}	−1.2867***
	(0.2098)
D_{11}	−1.1632***
	(0.2291)
D_{12}	−0.7309**
	(0.2387)
PT	
PAS	

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Estimates of the parameters associated with the *time dummies* provide the *fitted hazard function* for the *baseline group*.

Time period	Predictor	Parameter estimate ($\hat{\alpha}_j$)	Fitted odds $e^{(\hat{\alpha}_j)}$	Fitted hazard $\frac{1}{1+e^{(-\hat{\alpha}_j)}}$
7	D_7	−2.3979	0.0909	0.0833
8	D_8	−3.1167	0.0443	0.0424
9	D_9	−1.7198	0.1791	0.1519
10	D_{10}	−1.2867	0.2762	0.2164
11	D_{11}	−1.1632	0.3125	0.2381
12	D_{12}	−0.7309	0.4815	0.3250

In Model A, everyone is in the "baseline group"

Interpreting Parameter Estimates
Substantive Dichotomous Predictors
Grade at First Intercourse

	Model B
Parameter Estimates and Asymptotic Standard Errors	
D_7	−2.9943***
	(0.3175)
D_8	−3.7001***
	(0.4206)
D_9	−2.2811***
	(0.2724)
D_{10}	−1.8226***
	(0.2585)
D_{11}	−1.6542***
	(0.2691)
D_{12}	−1.1791***
	(0.2716)
PT	0.8736***
	(0.2174)
PAS	

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

As in regular logistic regression analysis, *anti-logging a parameter estimate* gives the *fitted odds-ratio associated with a unit difference in the predictor*:

$$e^{\hat{\beta}_{PT}} = e^{0.8736} = 2.4$$

The fitted odds of first intercourse for boys who have experienced a parenting transition are 2.4 times the odds for boys who did not experience such a transition.

Interpreting Parameter Estimates
Substantive Continuous Predictors
Grade at First Intercourse

	Model C
Parameter Estimates and Asymptotic Standard Errors	
D_7	−2.4646***
	(0.2741)
D_8	−3.1591***
	(0.3890)
D_9	−1.7297***
	(0.2245)
D_{10}	−1.2851***
	(0.2127)
D_{11}	−1.1360***
	(0.2324)
D_{12}	−0.6421**
	(0.2428)
PT	
PAS	0.4428***
	(0.1140)

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Anti-logging again provides a *fitted odds-ratio*, now associated with a unit difference in the continuous predictor, *PAS*:

$$e^{\hat{\beta}_{PAS}} = e^{0.4428} = 1.56$$

The fitted odds of first intercourse for boys whose parents exhibited a particular level of antisocial behavior are 1.56 times the odds for boys whose parental antisocial behavior was one standard deviation lower.

Displaying Fitted Functions

First, Compute the Relevant Fitted Hazard Probabilities

Grade at First Intercourse (ALDA, Table 11.5, p. 392)

	Model B
Parameter Estimates and Asymptotic Standard Errors	
D_7	−2.9943*** (0.3175)
D_8	−3.7001*** (0.4206)
D_9	−2.2811*** (0.2724)
D_{10}	−1.8226*** (0.2585)
D_{11}	−1.6542*** (0.2691)
D_{12}	−1.1791*** (0.2716)
PT	0.8736*** (0.2174)
PAS	

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

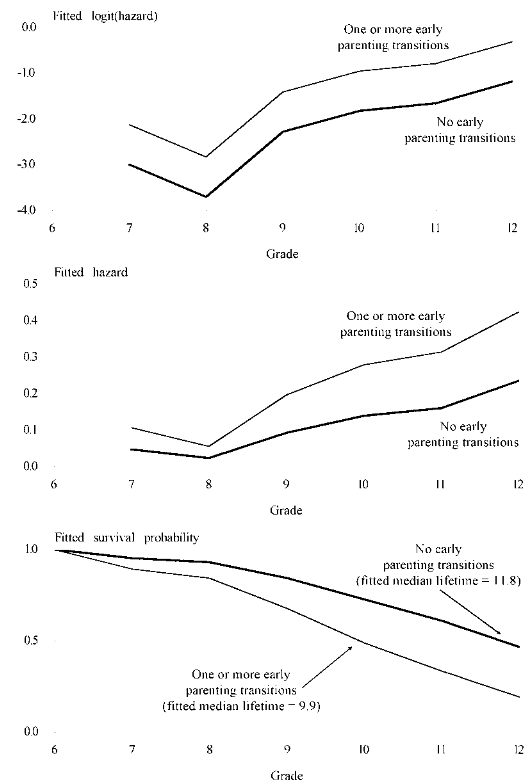
			Fitted value of					
			Logit hazard		Hazard		Survival	
Time period	$\hat{\alpha}_j$	$\hat{\beta}_1$	*PT* = 0	*PT* = 1	*PT* = 0	*PT* = 1	*PT* = 0	*PT* = 1
7	−2.9943	0.8736	−2.9943	−2.1207	0.0477	0.1071	0.9523	0.8929
8	−3.7001	0.8736	−3.7001	−2.8265	0.0241	0.0559	0.9293	0.8430
9	−2.2811	0.8736	−2.2811	−1.4075	0.0927	0.1966	0.8432	0.6772
10	−1.8226	0.8736	−1.8226	−0.9490	0.1391	0.2791	0.7259	0.4882
11	−1.6542	0.8736	−1.6542	−0.7806	0.1605	0.3142	0.6094	0.3348
12	−1.1791	0.8736	−1.1791	−0.3055	0.2352	0.4242	0.4660	0.1928

Fitted values for two prototypical boys

Displaying Fitted Functions

Plotting Prototypical Fitted Functions, by *PT*

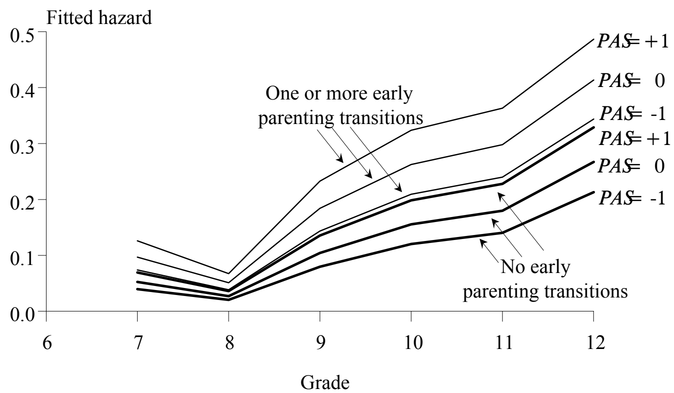
Grade at First Intercourse (ALDA, Fig. 11.6, p. 393)

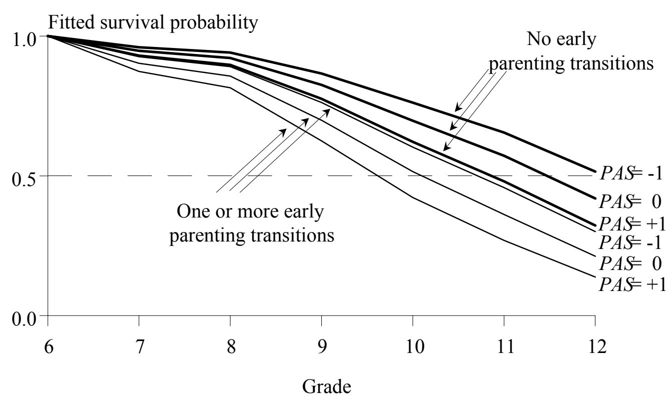


Displaying Fitted Functions

Plotting Prototypical Fitted Functions, by *PAS*

Grade at First Intercourse (ALDA, Fig. 11.7, p. 395)





Comparing DTSA Models

Grade at First Intercourse (ALDA, Table 11.3., p. 386)

	Model A	Model B	Model C	Model D
Parameter Estimates and Asymptotic Standard Errors				
D_7	−2.3979***	−2.9943***	−2.4646***	−2.8932***
	(0.2697)	(0.3175)	(0.2741)	(0.3206)
D_8	−3.1167***	−3.7001***	−3.1591***	−3.5847***
	(0.3862)	(0.4206)	(0.3890)	(0.4231)
D_9	−1.7198***	−2.2811***	−1.7297***	−2.1502***
	(0.2217)	(0.2724)	(0.2245)	(0.2775)
D_{10}	−1.2867***	−1.8226***	−1.2851***	−1.6932***
	(0.2098)	(0.2585)	(0.2127)	(0.2647)
D_{11}	−1.1632***	−1.6542***	−1.1360***	−1.5177***
	(0.2291)	(0.2691)	(0.2324)	(0.2757)
D_{12}	−0.7309**	−1.1791***	−0.6421**	−1.0099***
	(0.2387)	(0.2716)	(0.2428)	(0.2811)
PT		0.8736***		0.6605**
		(0.2174)		(0.2367)
PAS			0.4428***	0.2964*
			(0.1140)	(0.1254)
Goodness-of-fit				
LL	−325.98	−317.33	−318.59	−314.57
Deviance	651.96	634.66	637.17	629.15
n parameters	6	7	7	8
AIC	663.96	648.66	651.17	645.15
BIC	681.00	668.54	671.05	667.87
Deviance-based Hypothesis Tests				
H_0: $\beta_{PT} = 0$		17.30*** (1)		8.02** (1)
H_0: $\beta_{PAS} = 0$			14.79*** (1)	5.51* (1)
Wald Hypothesis Tests				
H_0: $\beta_{PT} = 0$		16.15*** (1)		7.79** (1)
H_0: $\beta_{PAS} = 0$			15.10*** (1)	5.59* (1)

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

AIC, BIC -- *smaller* value, *better* fit, compare non- nested models

Deviance -- *smaller* value, *better* fit, χ^2 dist., compare nested models