Extending the DSTA Model in Interesting and Useful Ways

(ALDA, Ch. 12)

John Willett & Judy Singer

Harvard University Graduate School of Education

May, 2003

What will we cover?

Alternative Specifications for *TIME*
Thinking Beyond the Dummy Specification for *TIME*
Data Example: *Grade at First Intercourse*

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

The *dummy specification for TIME* is:
1. Completely general.
2. Easily interpretable.
3. *Completely lacking in parsimony*.

Alternative Specifications for TIME
Smooth Polynomial Possibilities for *TIME*?
(cf. ALDA, Table 12.1, p. 411)

It's easy to add polynomial representations of TIME to the person-period dataset, and include them as predictors in the DTSA!

Order of polynomial	Behavior of logit hazard	*n* parameters	Model
0	Constant	1	$\text{logit } h(t_j) = \alpha_0 ONE$
1	Linear	2	$\text{logit } h(t_j) = \alpha_0 ONE + \alpha_1(TIME_j - c)$
2	Quadratic	3	$\text{logit } h(t_j) = \alpha_0 ONE + \alpha_1(TIME_j - c) + \alpha_2(TIME_j - c)^2$
3	Cubic	4	$\text{logit } h(t_j) = \alpha_0 ONE + \alpha_1(TIME_j - c) + \alpha_2(TIME_j - c)^2 + \alpha_3(TIME_j - c)^3$
4	Three stationary points	5	$\text{logit } h(t_j) = \alpha_0 ONE + \alpha_1(TIME_j - c) + \alpha_2(TIME_j - c)^2 + \alpha_3(TIME_j - c)^3 + \alpha_4(TIME_j - c)^4$
5	Four stationary points	6	$\text{logit } h(t_j) = \alpha_0 ONE + \alpha_1(TIME_j - c) + \alpha_2(TIME_j - c)^2 + \alpha_3(TIME_j - c)^3 + \alpha_4(TIME_j - c)^4 + \alpha_5(TIME_j - c)^5$
Completely general		J	$\text{logit } h(t_j) = \alpha_1 D_1 + \cdots + \alpha_J D_J$

Higher order, more complex shape.

More complex shape, better fit!

Re-center *TIME* (pick *c*), to make parameters meaningful.

Alternative Specifications for TIME
Data Example: *Time to Tenure*

- *Research Question*: Whether, and when, recipients of the NAE/Spencer Foundation Post-Doctoral Fellowship received tenure?
- *Citation*: Gamse & Conger, (1997).
- *Sample*: 260 semifinalists and fellowship recipients, who took an academic job after receiving their doctorates.
- *Research Design*:
  - Participants tracked annually for up to 9 years.
    - 166 (64%) received tenure during data collection.
    - 94 (36%) did not (censored).

Alternative Specifications for TIME
How Do You Choose the Correct Specification?
Time to Tenure (ALDA, Table 12.2, p. 413)

Fit a taxonomy of polynomial specifications, "bracketed" by the *constant* and *completely general* models, and compare them.

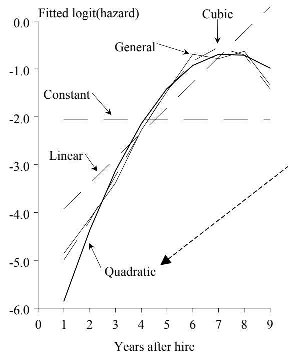
Compare deviance statistics across consecutive models to evaluate the impact of each new polynomial term

Inspect AIC & BIC

Representation for *TIME*	*n* parameters	Deviance	Difference in deviance in comparison to . . . Previous model	General model	AIC	BIC
Constant	1	1037.57	—	**206.36** (8)	1039.57	1042.68
Linear	2	867.46	**170.10** (1)	**36.26** (7)	871.46	877.68
Quadratic	3	836.30	**31.16** (1)	5.10 (6)	841.30	851.63
Cubic	4	833.17	3.13 (1)	1.97 (5)	841.17	853.61
Fourth order	5	832.74	0.43 (1)	1.54 (4)	842.74	858.29
Fifth order	6	832.73	0.01 (1)	1.53 (3)	844.73	863.39
General	9	831.20	—	—	849.20	877.19

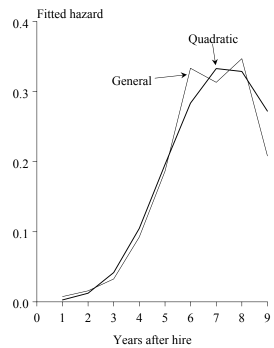
Compare deviance statistic of each polynomial to the deviance statistic for completely general specification to evaluate whether the current model fits "well enough."

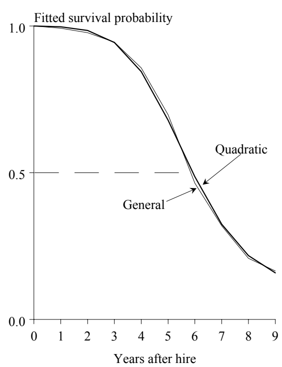
Alternative Specifications for TIME
Comparison of Temporal Specifications
Time to Tenure (ALDA, Figure 12.1, p. 414)



The *quadratic* is the best?

$$logit\ \hat{h}(t_j) = -1.41 + 0.61(t_j - 5) - 0.13(t_j - 5)^2$$





Consider *polynomial TIME* if:
- There are a large number of discrete time periods.
- Hazard may be zero in some time periods.
- Risk set gets really small in some time periods.

Including Time-Varying Predictors in the DTSA Model
Data Example: *Onset of Psychiatric Disorder*

- *Research Question*: Whether, and at what age, adults experience a depressive disorder?
- *Citation*: Wheaton, Rozell & Hall, (1997).
- *Sample*: 1393 adults
- *Research Design*:
  - Retrospective interview to assess age in years at first onset of depression.
  - Huge person-period dataset, with only rare events:
    - 36,997 records, potentially covering 36 years of data on each adult, at ages 4 thru 39.
    - Only 387 people (28%) experienced a first onset.

Including Time-Varying Predictors in the DTSA Model

Extract from the Person-Period Dataset, Case #40

Onset of Psychiatric Disorder

Time-invariant gender (0=male; 1=female)

Time-invariant # of siblings of target adult

id	period	pd	female	nsibs	event
40	4	0	1	4	0
40	5	0	1	4	0
40	6	0	1	4	0
40	7	0	1	4	0
40	8	0	1	4	0
40	9	1	1	4	0
40	10	1	1	4	0
40	11	1	1	4	0
40	12	1	1	4	0
40	13	1	1	4	0
40	14	1	1	4	0
40	15	1	1	4	0
40	16	1	1	4	0
40	17	1	1	4	0
40	18	1	1	4	0
40	19	1	1	4	0
40	20	1	1	4	0
40	21	1	1	4	0
40	22	1	1	4	0
40	23	1	1	4	1

Case #40

Onset of depression

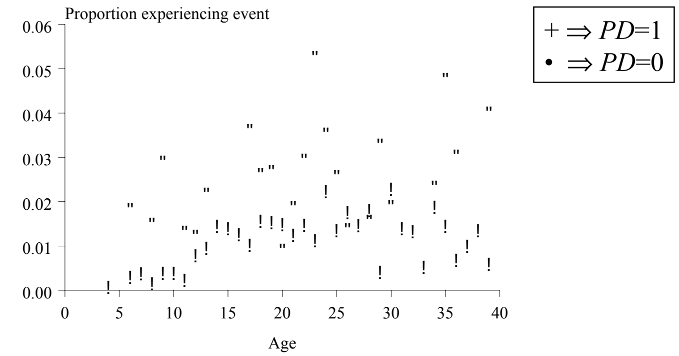
Time Period

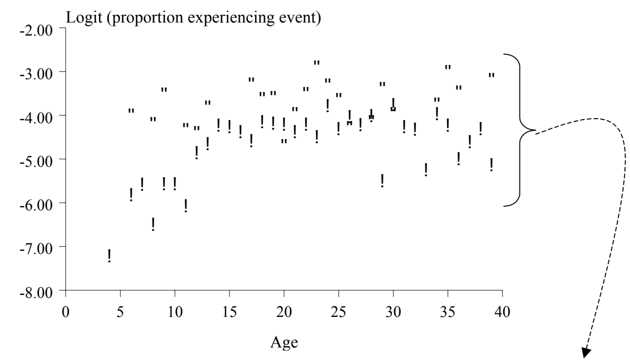
Time-varying indicator of parental divorce:
- *PD* = 0 in periods before parental divorce.
- *PD* = 1 in time periods coincident with, and subsequent to, parental divorce.

Including Time-Varying Predictors

What do the sample data look like?

Onset of Psychiatric Disorder (ALDA, Figure 12.4, p. 432)





Completely general TIME function lacks parsimony.
⇒ **Choose a polynomial specification** (which?)

Including Time-Varying Predictors
Hypothesized DSTA Model for Depression Onset
Onset of Psychiatric Disorder (ALDA, Equ. 12.8, p. 430)

Our exploratory data-analysis suggested that a *cubic* function, with TIME centered at age-18, would do a good job of representing the shape of the logit-hazard function:

$$logit\ h(t_{ij}) = \alpha_0 + \alpha_1(AGE_{ij} - 18) + \alpha_2(AGE_{ij} - 18)^2 + \alpha_3(AGE_{ij} - 18)^3 + \beta_1 PD_{ij}$$

Parameter β_1 :

- Contrasts population logit-hazard for folk who experience, and do not experience, parental divorce.
- But, because PD_{ij} is time-varying, people can switch parental divorce group membership as time passes.

Notice that, although *PD* is a predictor with *time-varying* values, its *effect* (β_1) is hypothesized as constant over time.

Including Time-Varying Predictors
Fitted DSTA Model for Depression Onset
Onset of Psychiatric Disorder (ALDA, Equ. 12.8, p. 435)

The interpretation of the parameter estimates is straightforward:

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
ONE	1	-4.5866	0.1070	1836.5406	<.0001
age_18	1	0.0596	0.0117	26.1547	<.0001
age_18sq	1	-0.00736	0.00122	36.1138	<.0001
age_18cub	1	0.000185	0.000079	5.4655	0.0194
PD	1	0.4151	0.1620	6.5623	0.0104
FEMALE	1	0.5455	0.1094	24.8532	<.0001

Effect of *FEMALE*:

$$e^{\hat{\beta}_2} = e^{0.5455} = 1.73$$

Fitted odds that a female will experience initial onset of depression are 1.73 times the odds for a male.

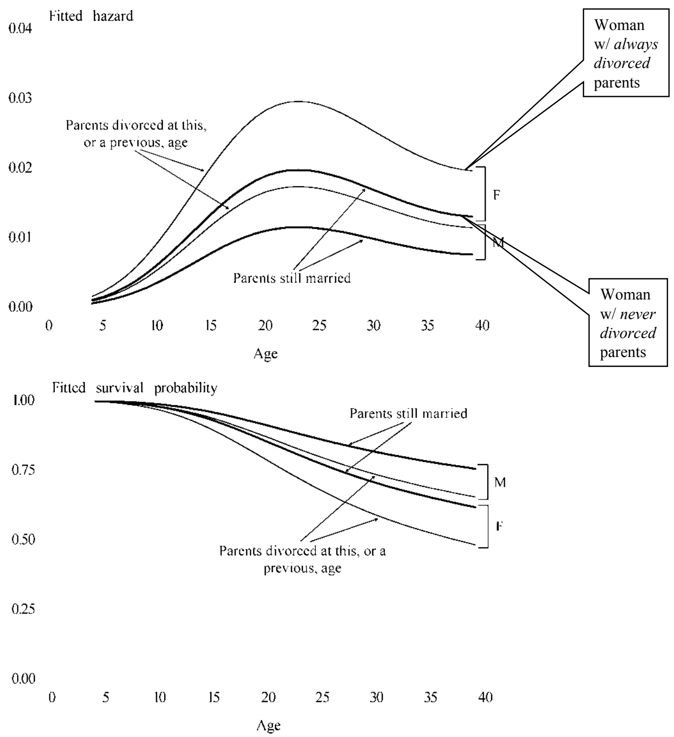
Effect of *PARENTAL DIVORCE*:

$$e^{\hat{\beta}_1} = e^{0.42} = 1.51$$

At every age between 4 and 39, the fitted odds that a person whose parents have concurrently, or previously, divorced will experience initial onset of depression are 1.51 times the odds for a person whose parents have not (yet) divorced.

But, the impact of the time-varying predictor on the hazard profiles of prototypical people is interesting!

Including Time-Varying Predictors
Interpreting Prototypical Fitted H&S Functions
(Parents *Always* Divorced *vs*. Parents *Never* Divorced)
Onset of Psychiatric Disorder (ALDA, Figure 12.5, p. 437)



What about someone whose parents divorced at age-30? Age-40?

Checking the Linear Additivity Assumption
Data Example: *Risk of First Juvenile Arrest*

- *Research Question*: Whether, and at what age, juveniles were first arrested?
- *Citation*: Keiley & Martin, (2002)
- *Sample*: 1553 adolescents.
  - 342 arrested between the ages of 8 and 18.
- *Question Predictors*:
  - $ABUSED_i$, time-invariant record of whether the child had been abused:
    - = 0, no early child abuse.
    - = 1, child had been abused in early life.
  - $BLACK_i$, time-invariant respondent ethnicity,
    - = 0, Caucasian.
    - = 1, African-American.

Checking the Linear Additivity Assumption

Interactions Between Substantive Predictors

Risk of First Juvenile Arrest (ALDA, Figure 12.6, p. 445)

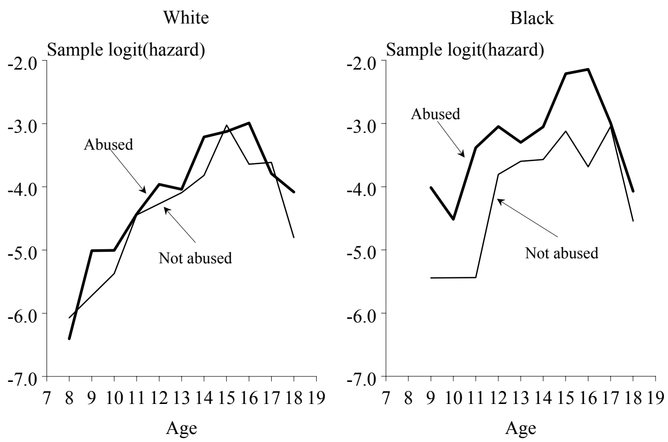
So far, every DSTA model we've fitted has assumed that the effects of the substantive predictors are linearly additive.

Linear additivity ⇒ Unit differences in a predictor, time-invariant or time-varying, correspond to fixed differences in logit-hazard.

Linear additivity can be violated by:

1. *Interactions* among the substantive predictors.
2. *Non-linear effects* of substantive predictors.

Evidence of an *interaction* between race and child abuse?



Checking the Linear Additivity Assumption

Adding Interactions Among Substantive Predictors

Risk of First Juvenile Arrest (ALDA, p. 446)

Hypothesized DSTA model:

$$\text{logit}\, h(t_j) = \alpha_8 D_8 + \ldots + \alpha_{18} D_{18} + \beta_1 ABUSED + \beta_2 BLACK + \beta_3 (ABUSED \times BLACK)$$

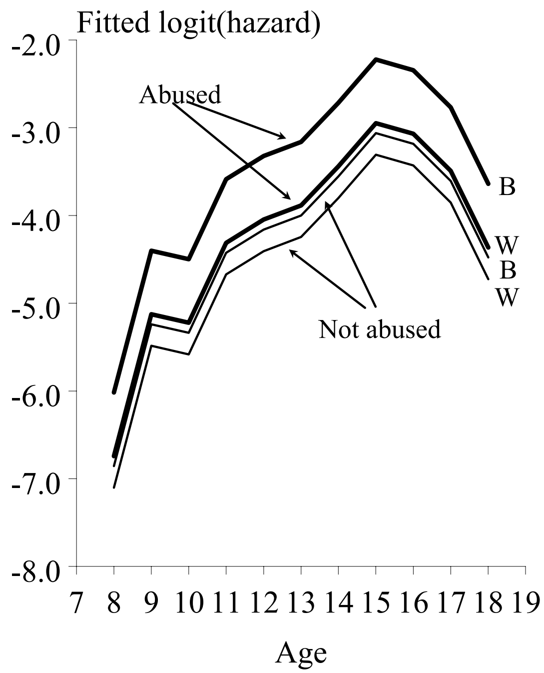
Parameter estimates:

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
D8	1	-7.1013	0.7167	98.1866	<.0001
D9	1	-5.4851	0.3373	264.4353	<.0001
D10	1	-5.5822	0.3533	249.6017	<.0001
D11	1	-4.6712	0.2432	369.0639	<.0001
D12	1	-4.4070	0.2221	393.7051	<.0001
D13	1	-4.2449	0.2115	402.9909	<.0001
D14	1	-3.8010	0.1849	422.4313	<.0001
D15	1	-3.3073	0.1634	409.5355	<.0001
D16	1	-3.4289	0.1712	401.0567	<.0001
D17	1	-3.8521	0.1969	382.8974	<.0001
D18	1	-4.7246	0.2752	294.8176	<.0001
ABUSED	1	0.3600	0.1539	5.4695	0.0194
BLACK	1	0.2455	0.1972	1.5500	0.2131
ABLACK	1	0.4787	0.2391	4.0094	0.0452

Interpretation as *odds-ratios*:

Prototype	*ABUSED*	*BLACK*	Combined Parameter Estimates	Estimated Odds Ratio
White/not abused	0	0	$0 \times 0.3600 + 0 \times 0.2455 + 0 \times 0.4787 = 0.0000$	1.00
White/ abused	1	0	$1 \times 0.3600 + 0 \times 0.2455 + 0 \times 0.4787 = 0.3600$	1.43
Black/not abused	0	1	$0 \times 0.3600 + 1 \times 0.2455 + 0 \times 0.4787 = 0.2455$	1.28
Black/ abused	1	1	$1 \times 0.3600 + 1 \times 0.2455 + 1 \times 0.4787 = 1.0842$	2.96

Checking the Linear Additivity Assumption
Interpretation of an Interaction w/ Prototypical Plots
Risk of First Juvenile Arrest (ALDA, Figure 12.6, p. 445)



Checking the Linear Additivity Assumption
Data Example: *Onset of Psychiatric Disorder*

- *Research Question*: Whether, and at what age, adults experience a depressive disorder?
- *Citation*: Wheaton, Rozell & Hall, (1997).
- *Sample*: 1393 adults
- *Research Design*:
  - Retrospective interview to assess age in years at first onset of depression.
  - Huge person-period dataset, with only rare events:
    - 36,997 records, potentially covering 36 years of data on each adult, at ages 4 thru 39.
    - Only 387 people (28%) experienced a first onset.

Checking the Linear Additivity Assumption

Extract from the Person-Period Dataset, Case #40

Onset of Psychiatric Disorder

Time-invariant gender (0=male; 1=female)

Time-invariant # of siblings of target adult

id	period	pd	female	nsibs	event
40	4	0	1	4	0
40	5	0	1	4	0
40	6	0	1	4	0
40	7	0	1	4	0
40	8	0	1	4	0
40	9	1	1	4	0
40	10	1	1	4	0
40	11	1	1	4	0
40	12	1	1	4	0
40	13	1	1	4	0
40	14	1	1	4	0
40	15	1	1	4	0
40	16	1	1	4	0
40	17	1	1	4	0
40	18	1	1	4	0
40	19	1	1	4	0
40	20	1	1	4	0
40	21	1	1	4	0
40	22	1	1	4	0
40	23	1	1	4	1

Case #40

Onset of depression

Time Period

Time-varying indicator of parental divorce:
- *PD* = 0 in periods before parental divorce.
- *PD* = 1 in time periods coincident with, and subsequent to, parental divorce.

Checking the Linear Additivity Assumption

Testing for Non-Linear Effects of Substantive Predictors

Onset of Psychiatric Disorder (ALDA, Table 12.4, p. 449)

Use all your usual strategies for checking non-linearity: transform the predictors, use polynomials, *re-bin the predictor*, ….

	Model A	Model B	Model C
Parameter Estimates and Asymptotic Standard Errors			
ONE	−4.3587***	−4.5001*	−4.4828***
	(0.1216)	(0.2067)	(0.1087)
(*AGE*-18)	0.0611***	0.0615***	0.0614***
	(0.0117)	(0.0117)	(0.0117)
$(AGE\text{-}18)^2$	−0.0073***	−0.0073***	−0.0073***
	(0.0012)	(0.0012)	(0.0012)
$(AGE\text{-}18)^3$	0.0002*	0.0002*	0.0002*
	(0.0001)	(0.0001)	(0.0001)
PD_j	0.3726*	0.3727*	0.3710*
	(0.1624)	(0.1625)	(0.1623)
FEMALE	0.5587***	0.5596***	0.5581***
	(0.1095)	(0.1095)	(0.1095)
NSIBS	−0.0814***		
	(0.0223)		
1 OR 2 SIBS		0.0209	
		(0.1976)	
3 OR 4 SIBS		0.0108	
		(0.2100)	
5 OR 6 SIBS		−0.4942~	
		(0.2545)	
7 OR 8 SIBS		−0.7754*	
		(0.3437)	
9 OR MORE SIBS		−0.6685~	
		(0.3441)	
BIGFAMILY			−0.6108***
			(0.1446)
Goodness-of-fit			
Deviance	4124.29	4117.98	4118.78
n parameters	7	11	7
AIC	4138.29	4139.98	4132.78

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

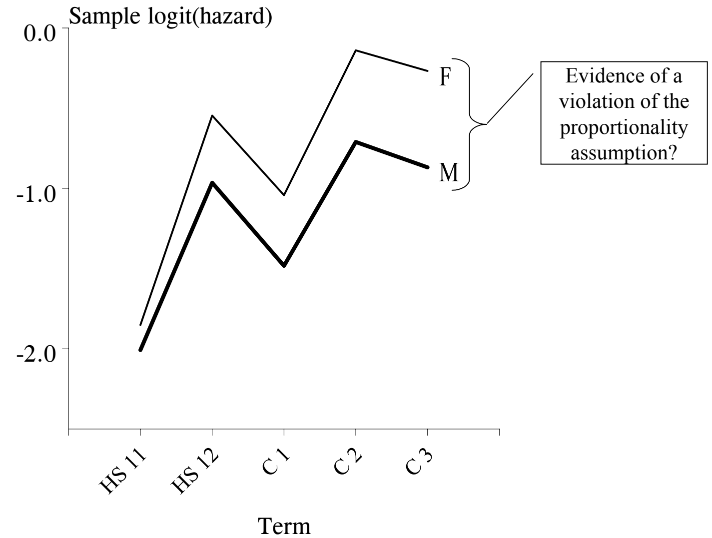
Checking the Proportionality Assumption
Data Example: *Giving Up the Study of Math*

- *Research Question*: Whether, and when, do students terminate their study of math? Does the pattern of termination differ for males and females?
- *Citation*: Graham (1997).
- *Sample*: 3790 high-school students
  - 1875 boys, 1915 girls.
- *Research Design:*
  - Followed for 5 years.
  - Observed annually, in 11th and 12th grade, and during the first three years of college.
- *Question Predictor*:
  - $FEMALE_i$, time-invariant student gender:
    - = 0, male.
    - = 1, female.

Checking the Proportionality Assumption
Sample Evidence of a "Non-Proportional" Relationship?
Giving Up the Study of Math (ALDA, Figure 12.8, p. 458)

In every DTSA model so far, we've assumed that the *proportional odds assumption* holds….

Proportional Odds Assumption	⇒	*A constant difference in the elevation of the logit-hazard profile* among groups defined by constant values of the predictor.



If the *proportionality assumption* is violated for a predictor, then there is an interaction between the predictor and *TIME*.

Checking the Proportionality Assumption

Include an Interaction with *TIME* in the DTSA Model

Giving Up the Study of Math (ALDA, Figure 12.8, p. 458)

Main effect of *FEMALE* only

Completely general interaction between FEMALE & TIME

Linear interaction between FEMALE & TIME

	Model A	Model B	Model C
Parameter Estimates and Asymptotic Standard Errors			
HS11	−2.1308***	−2.0077***	−2.0459***
	(0.0567)	(0.0715)	(0.0646)
HS12	−0.9425***	−0.9643***	−0.9255***
	(0.0479)	(0.0585)	(0.0482)
COLL1	−1.4495***	−1.4824***	−1.4966***
	(0.0634)	(0.0847)	(0.0665)
COLL2	−0.6176***	−0.7100***	−0.7178***
	(0.0757)	(0.1007)	(0.0861)
COLL3	−0.7716***	−0.8690***	−0.9166***
	(0.1428)	(0.1908)	(0.1557)
FEMALE	0.3786***		0.2275**
	(0.0501)		(0.0774)
FEMALE × *HS11*		0.1568	
		(0.0978)	
FEMALE × *HS12*		0.4187***	
		(0.0792)	
FEMALE × *COLL1*		0.4407***	
		(0.1158)	
FEMALE × *COLL2*		0.5707***	
		(0.1445)	
FEMALE × *COLL3*		0.6008*	
		(0.2857)	
FEMALE × *TIME* − *1*			0.1198*
			(0.0470)
Goodness-of-fit			
Deviance	9804.31	9796.27	9797.81
n parameters	6	10	7
AIC	9816.31	9816.27	9811.81

?

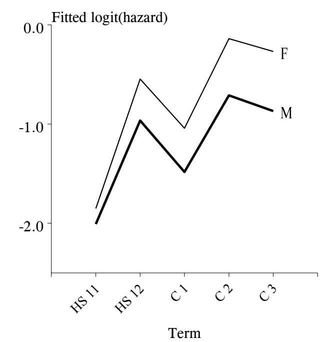
~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Checking the Proportionality Assumption

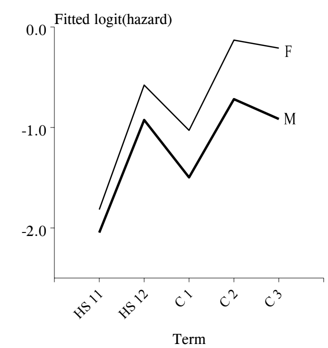
A Fitted "Non-Proportional" Relationship

Giving Up the Study of Math (ALDA, Figure 12.8, p. 458)

Model B: Completely general interaction between FEMALE and time



Model C: Interaction between FEMALE and time



Both interaction (non-proportional) DTSA models do a good job. However, the model with the linear interaction is more parsimonious and almost as good.