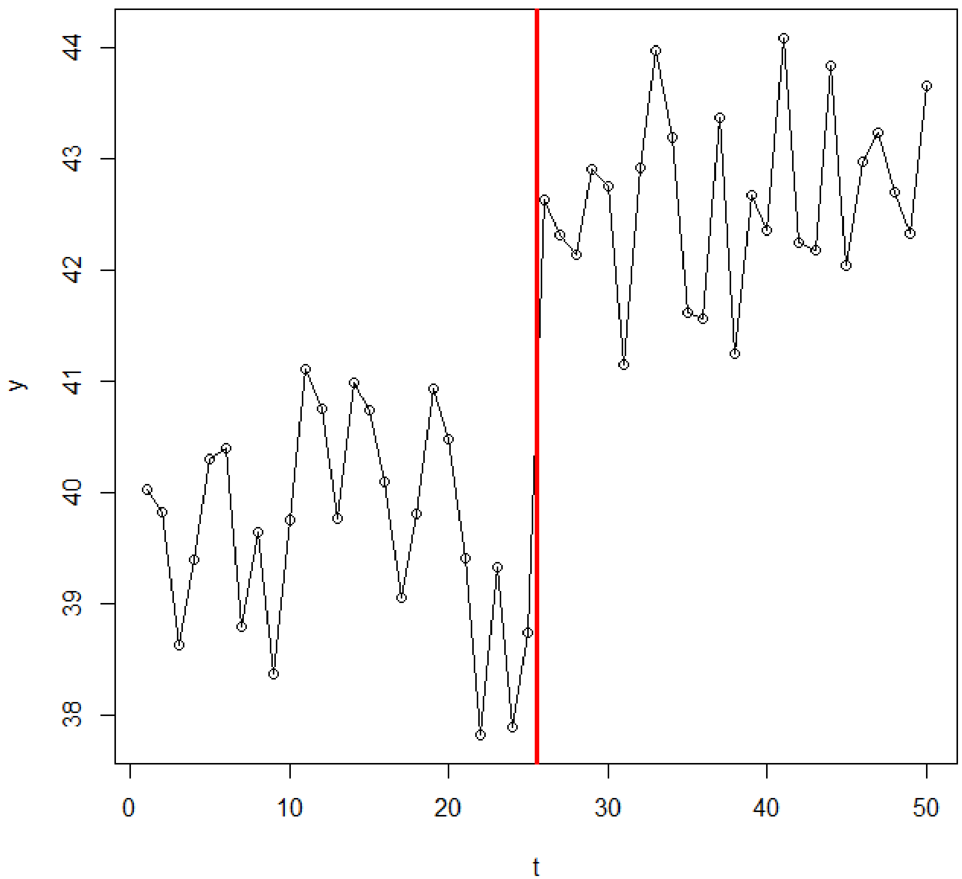
Differences in Differences

Seminar in Criminology, Research and Analysis– Crim 7301
Week 6, 9/29/16
Andrew Wheeler

Class Overview

- Interrupted time series analysis
- Using a control series
- Parallel trend assumption

Interrupted time series



$$\mathbb{E}[y_t] = \beta_0 + \beta_1(D_t)$$

Interrupted time series

$$\mathbb{E}[y_t] = \beta_0 + \beta_1(D_t)$$

D is a *dummy variable* that equals 0 before the intervention, and equals 1 after the intervention.

The pre-mean is then

β_0

the post mean is,

$\beta_0 + \beta_1$

so β_1 tests for the difference between pre and post.

	t	d	y
[1,]	1	0	40.01875
[2,]	2	0	39.81575
[3,]	3	0	38.62867
[4,]	4	0	39.40083
[5,]	5	0	40.29455
[6,]	6	0	40.38979
[7,]	7	0	38.79192
[8,]	8	0	39.63632
[9,]	9	0	38.37333
[10,]	10	0	39.74352
[11,]	11	0	41.10178
[12,]	12	0	40.75578
[13,]	13	0	39.76177
[14,]	14	0	40.98744
[15,]	15	0	40.74139
[16,]	16	0	40.08935
[17,]	17	0	39.04506
[18,]	18	0	39.80485
[19,]	19	0	40.92552
[20,]	20	0	40.48298
[21,]	21	0	39.40369
[22,]	22	0	37.81471
[23,]	23	0	39.32513
[24,]	24	0	37.88094
[25,]	25	0	38.73480
[26,]	26	1	42.62634
[27,]	27	1	42.31244
[28,]	28	1	42.12784
[29,]	29	1	42.89824
[30,]	30	1	42.74622
[31,]	31	1	41.14626
[32,]	32	1	42.92205
[33,]	33	1	43.96857
[34,]	34	1	43.18493
[35,]	35	1	41.62006
[36,]	36	1	41.56449
[37,]	37	1	43.36209
[38,]	38	1	41.24091
[39,]	39	1	42.67546
[40,]	40	1	42.34844
[41,]	41	1	44.08655
[42,]	42	1	42.23746
[43,]	43	1	42.17134
[44,]	44	1	43.83447
[45,]	45	1	42.03235
[46,]	46	1	42.97118
[47,]	47	1	43.23253
[48,]	48	1	42.69879
[49,]	49	1	42.32239

Multiple Time Series

$$\mathbb{E}[y_{it}] = \beta_0 + \beta_1(D_t) + \beta_2(T_i) + \beta_3(D_t \cdot T_i)$$

- Where i indexes each different series, and T is a dummy variable equal to *one* for the treatment series and *zero* for the control.

t	d	y	treat
1	0	0.5993625	0
1	0	4.8375581	1
2	0	1.6654434	0
2	0	7.0848188	1
3	0	4.3679540	0
3	0	8.3280106	1
4	0	6.1377671	0
4	0	11.8493411	1
5	0	5.5058193	0
5	0	9.8726062	1
6	0	6.7863424	0
6	0	12.3495170	1
7	0	6.0977881	0
7	0	11.7587747	1
8	0	8.5328970	0
8	0	11.8748461	1
9	0	8.3541057	0
9	0	14.3822737	1

Multiple Time Series

$$\mathbb{E}[y_{it}] = \beta_0 + \beta_1(D_t) + \beta_2(T_i) + \beta_3(D_t \cdot T_i)$$

- So the pre mean control is:

$$\beta_0$$

- Pre mean treatment:

$$\beta_0 + \beta_2$$

- Post mean control:

$$\beta_0 + \beta_1$$

- Post mean treatment:

$$\beta_0 + \beta_1 + \beta_2 + \beta_3$$

- β_3 is the test for differences in differences

t	d	y	c
1	0	0.5993625	0
1	0	4.8375581	1
2	0	1.6654434	0
2	0	7.0848188	1
3	0	4.3679540	0
3	0	8.3280106	1
4	0	6.1377671	0
4	0	11.8493411	1
5	0	5.5058193	0
5	0	9.8726062	1
6	0	6.7863424	0
6	0	12.3495170	1
7	0	6.0977881	0
7	0	11.7587747	1
8	0	8.5328970	0
8	0	11.8748461	1
9	0	8.3541057	0
9	0	14.3822737	1
10	0	10.2909875	0
10	0	16.4189411	1
11	0	9.7624055	0
11	0	13.4822509	1
12	0	11.5438237	0
12	0	17.6726920	1

Multiple Time Series

$$\mathbb{E}[y_{it}] = \beta_0 + \beta_1(D_t) + \beta_2(T_i) + \beta_3(D_t \cdot T_i)$$

- A trick for interpreting interactions, rewrite the equation.

- If $T_i = 0$:

$$\mathbb{E}[y_{it}] = \beta_0 + \beta_1(D_t)$$

- If $T_i = 1$:

$$\mathbb{E}[y_{it}] = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)(D_t)$$

Parallel Trends Assumption

Example: Ceasefire intervention, treatment are gang shootings, control are non-gang shootings (per week).

Shootings = b0 + b1(Post) + b2(Gang) + b3(Gang*Post)

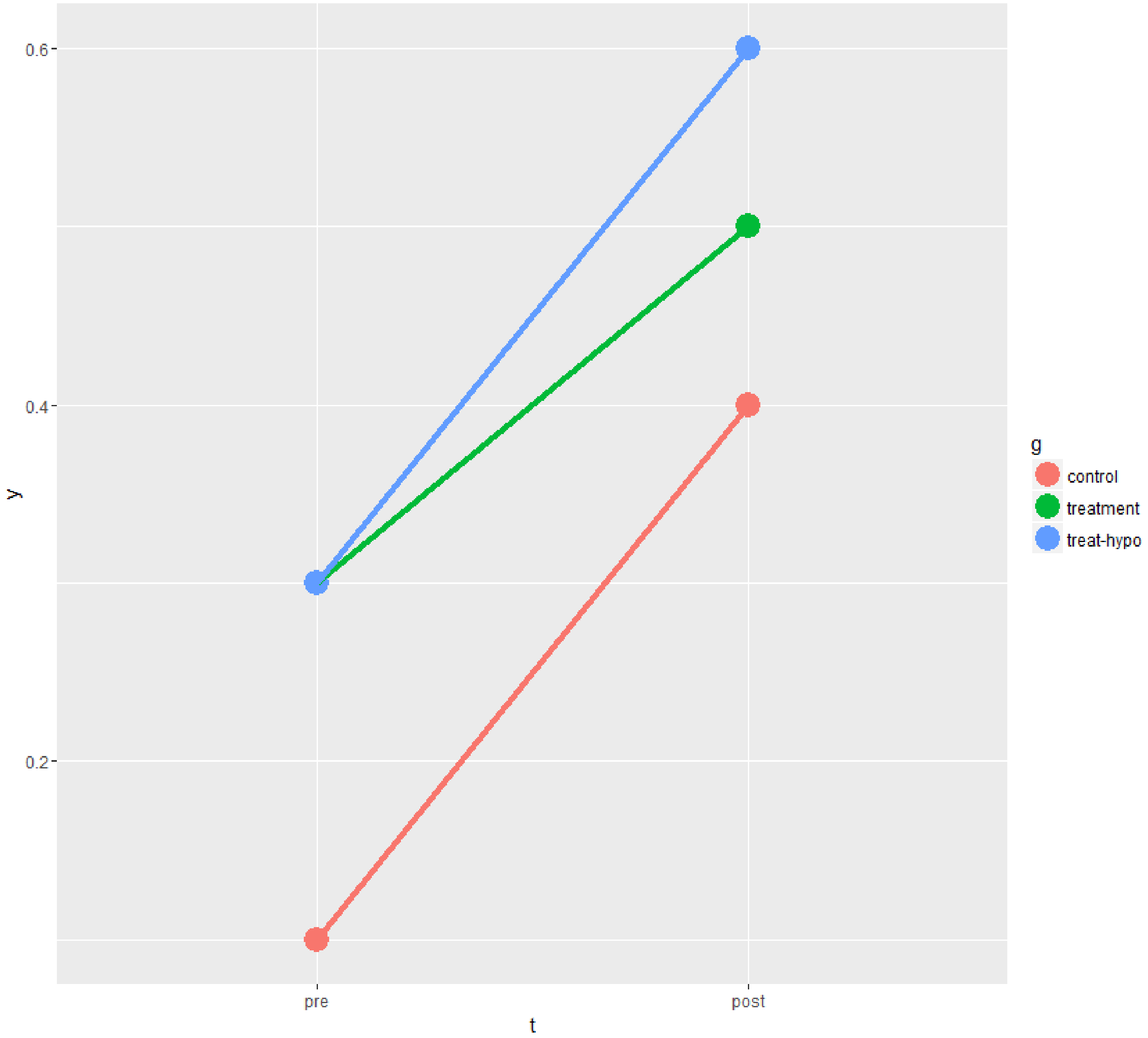
Shootings = 0.1 + 0.3(Post) + 0.2(Gang) – 0.1(Gang*Post)

Means Table

	Non-Gang	Gang
Pre	0.1	0.3
Post	0.4	0.5

The hypothetical mean without the intervention would be [b0 + b1 + b2] = (0.1 + 0.3 + 0.2) = 0.6.

Parallel Trends Assumption



Parallel Trends Assumption

- β_3 is the test for differences in differences
- $\beta_3 = \left(\bar{T}_{post} - \bar{T}_{pre}\right) - (\bar{C}_{post} - \bar{C}_{pre})$
- -0.1 = (0.5 – 0.3) – (0.4 – 0.1)

Parallel Trends Assumption

- If using a non-linear model (such as Poisson), the parallel trends follow that function.

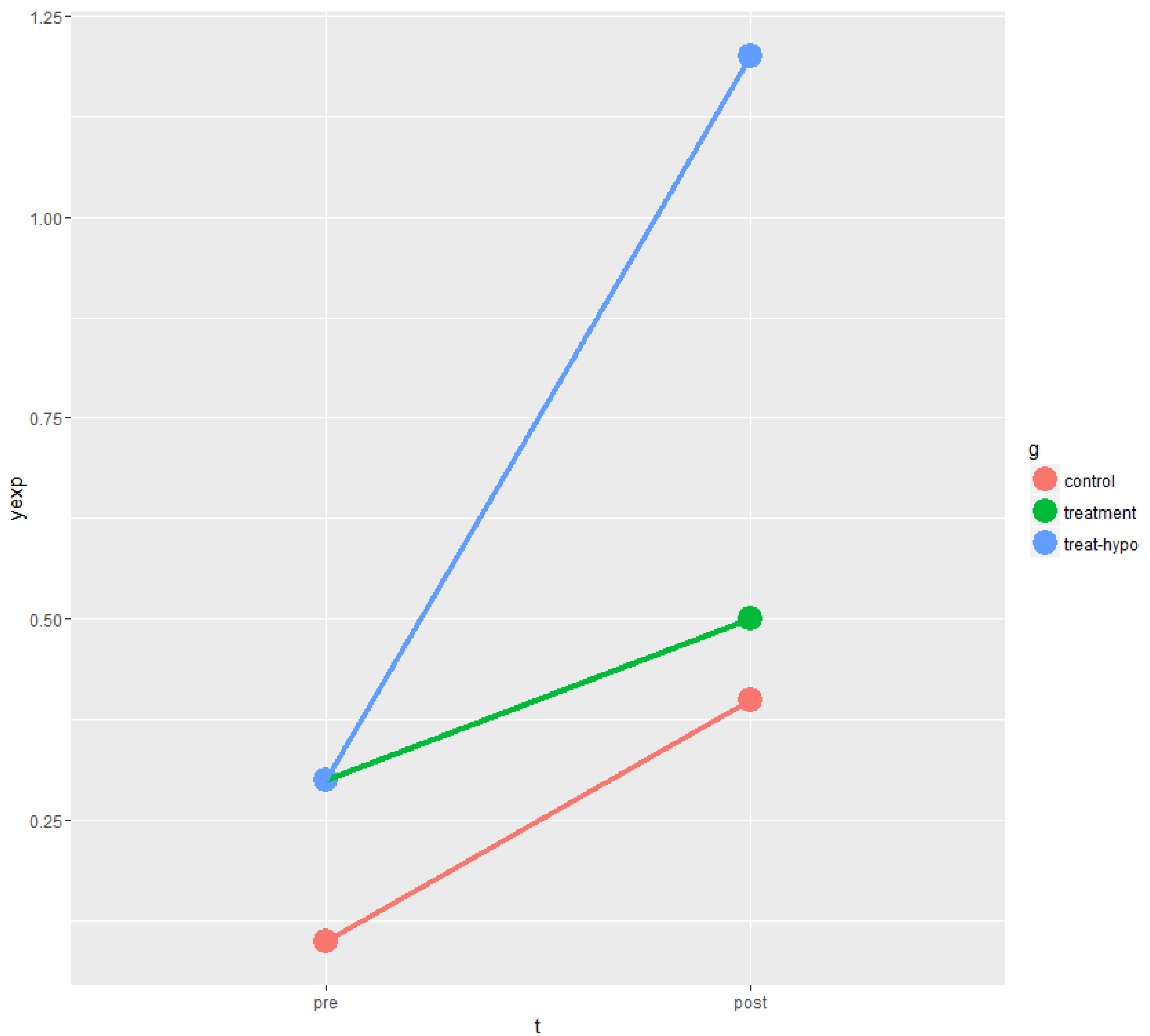
- Shootings = exp[-2.3 + 1.4(Post) + 1.1(Gang) + -0.9(Gang*Post)]

Means Table

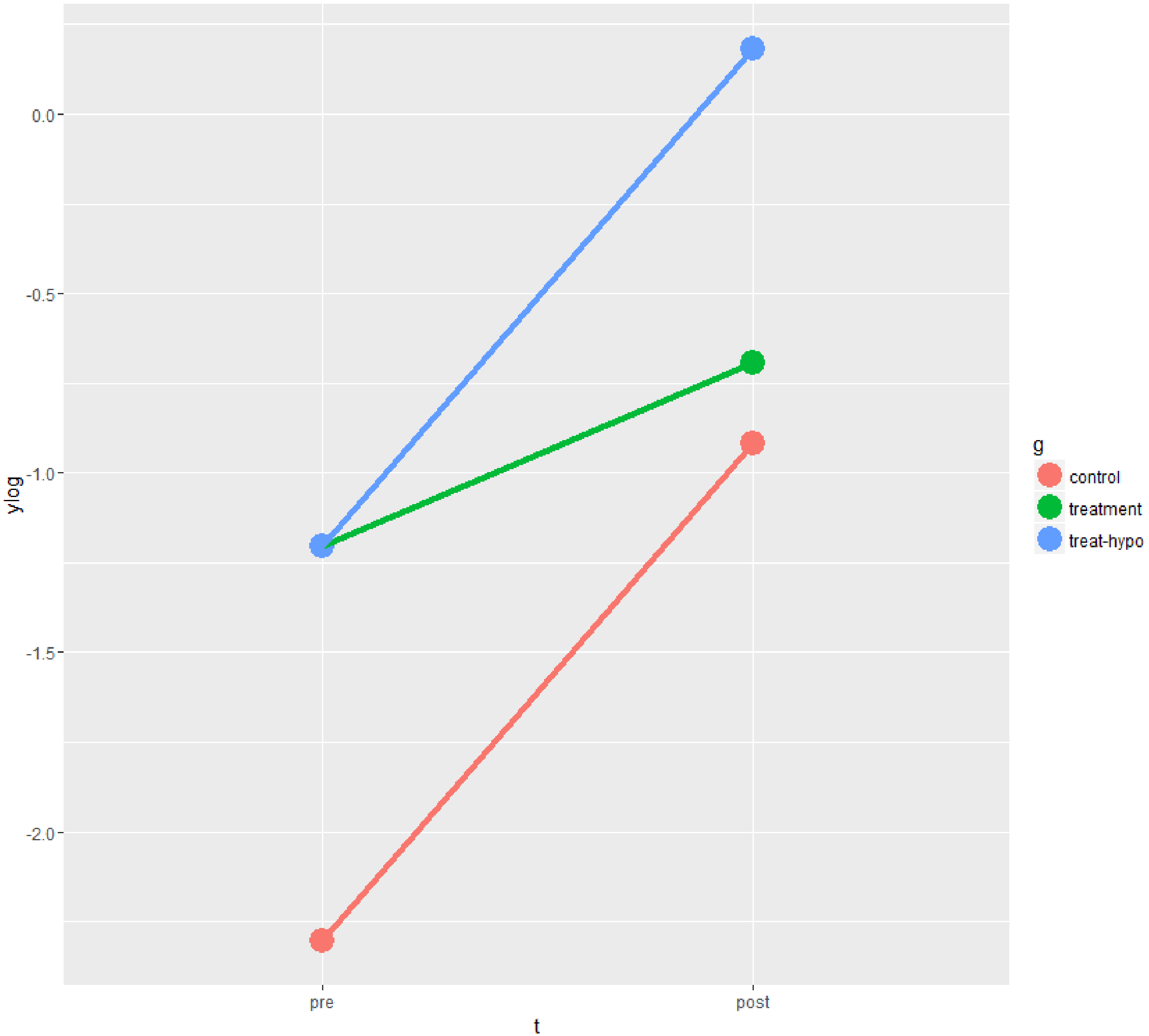
	Non-Gang	Gang
Pre	0.1	0.3
Post	0.4	0.5

- So the hypothetical value would be exp(-2.3 + 1.4 + 1.1) = 1.2
- The hypothetical increase would be 4 times the pre-treatment mean, so would be 0.3*4 = 1.2!

Parallel Trends Assumption



Parallel Trends Assumption



Homework & Next Weeks Class

Lab Assignment

Fit difference in difference models in R or Stata or SPSS. Estimate the hypothetical outcomes for linear and Poisson regression models.

For Next Week

- Mostly Harmless – Chapter 8
- Brame, Bushway, and Paternoster (1999) On the use of panel research designs and random effects models to investigate static and dynamic theories of criminal offending. *Criminology* 37(3):599-642.
- Worrall (2010) A user-friendly introduction to panel data modeling. *Journal of Criminal Justice Education* 21(2): 182-196.