

# ARE213 Problem Set #2A

Peter Alstone & Frank Proulx

November 6, 2013

## 1 Problem #1 and #2

These are hand written and attached separately.

## 2 Problem #3

### 2.1 Part A

Running pooled bivariate OLS, adding a quadratic time trend, and adding the covariates that we expect to belong produces the models shown in Table 1. The pooled OLS is an ill-informed baseline model but nonetheless tells us that there is a statistically significant negative correlation between the states with primary seatbelt laws and those without. In particular, the mean of the log fatalities per capita is reduced by 0.14 for states with primary seatbelt laws. This does not, however, account for the year to year trends that are captured by including a set of quadratic year terms in the regression as shown in the “quadratic time” regression. This basic trend is that deaths have gone down over time. Correcting for these trends may allow our OLS estimates approach the true ATE, and indeed the apparent effect of primary seatbelt laws is reduced. When further covariates are added that are relevant (see Table 1) further reductions in the apparent effect occur and the effect is no longer statistically significant.

### 2.2 Part B

No, the standard errors are most likely not correct (are they ever really, truly correct outside a purely theoretical framework?). In this case the error

Table 1: Pooled Models of Fatalities Per Capita

	bivariate	logfatalpc quadratic time	covariates
	(1)	(2)	(3)
primary	-0.144*** (0.026)	-0.075*** (0.026)	0.015 (0.027)
year		-0.025*** (0.006)	-0.042*** (0.005)
sqr.year		0.0005* (0.0002)	0.001*** (0.0002)
secondary			-0.001 (0.022)
college			-2.902*** (0.176)
beer			0.268*** (0.031)
totalvmt			-0.00000*** (0.00000)
precip			-0.015** (0.006)
snow32			-0.297*** (0.017)
rural_speed			0.016*** (0.002)
urban_speed			0.005*** (0.002)
Constant	-1.703*** (0.011)	-1.506*** (0.031)	-2.153*** (0.143)
<i>N</i>	1,127	1,127	1,127
<i>R</i> <sup>2</sup>	0.027	0.094	0.596
Adjusted <i>R</i> <sup>2</sup>	0.027	0.094	0.590
F Statistic	31.007*** (df = 1; 1125)	39.030*** (df = 3; 1123)	149.510*** (df = 11; 1115)

*Notes:*

2

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

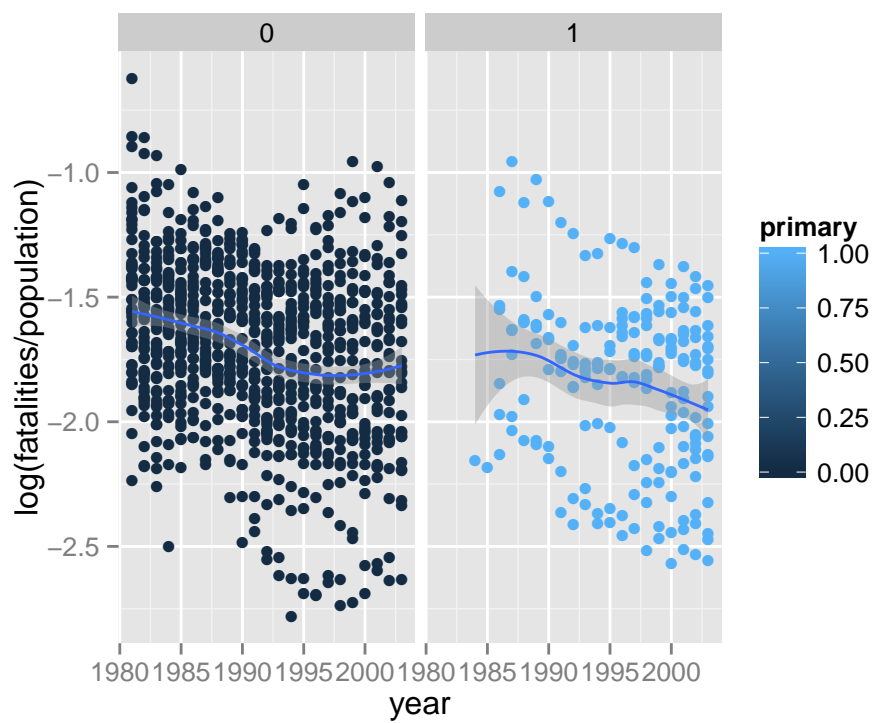


Figure 1: Year to year trends in the log of traffic fatalities per capita, divided by primary seatbelt law presence. A LOESS fit to each dataset is included for reference but is not necessarily indicative of the true underlying function.

in the SE estimates comes from heteroscedasticity in the error terms from the regression. Table 2 shows that introducing robust standard errors (using the HC1 estimator for the OLS covariance matrix). Clustering on the state grouping only increases the SE slightly over the robust pooled version, indicating most of the heteroscedasticity is in the full sample and not within-group. Table 3 shows the results of calculating the standard errors by hand, using the code shown here.

```

1 ## Robust by hand
2 #This calculates the Huber-White Robust standard errors -- code based on
   http://thetarzan.wordpress.com/2011/05/28/heteroskedasticity-robust-and-
   clustered-standard-errors-in-r/
3 s <- summary(pooled.full)
4 X <- model.matrix(pooled.full)
5 u2 <- residuals(pooled.full)^2
6 XDX <- 0
7
8 for(i in 1:nrow(X)) {
9     XDX <- XDX +u2[i]*X[i,]%*%t(X[i,])
10 }
11
12 # inverse(X'X)
13 XX1 <- solve(t(X)%*%X)
14
15 #Compute variance/covariance matrix
16 varcovar <- XX1 %*% XDX %*% XX1
17
18 # Degrees of freedom adjustment
19 dfc <- sqrt(nrow(X))/sqrt(nrow(X)-ncol(X))
20
21 stdh <- dfc*sqrt(diag(varcovar))
22
23 t <- pooled.full$coefficients/stdh
24 p <- 2*pnorm(-abs(t))
25 results.robust <- cbind(pooled.full$coefficients, stdh, t, p)
26 dimnames(results.robust) <- dimnames(s$coefficients)
27 results.robust
28
29 ## cluster by hand -- using many of the same variables as defined in the
   robust section (above), with some modifications:
30 cluster <- "state"
31 clus <- cbind(X, "state"=ps2a.data[,cluster], "resid" = resid(pooled.full))
32                                     #number of clusters
33 m <- dim(table(clus[,cluster]))
34 k <- dim(X)[2]
35
36 uj <- matrix(NA, nrow=m, ncol = k)
37 gs <- names(table(ps2a.data[,cluster]))
38 for (i in 1:m){
39     uj[i,] <- t(matrix(clus[clus[,cluster]==gs[i], 'resid'])) %*% clus[clus
40     [,cluster]==gs[i], 1:k]
41 }
42

```

Table 2: Comparison of Standard Error HC Methods for Full Pooled Model

	Conventional	HC1 Robust	HC1 Robust + Cluster
	(1)	(2)	(3)
primary	0.015 (0.027)	0.015 (0.054)	0.015 (0.054)
year	-0.042*** (0.005)	-0.042*** (0.008)	-0.042*** (0.008)
sqr.year	0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)
secondary	-0.001 (0.022)	-0.001 (0.036)	-0.001 (0.036)
college	-2.902*** (0.176)	-2.902*** (0.446)	-2.902*** (0.448)
beer	0.268*** (0.031)	0.268*** (0.082)	0.268*** (0.082)
totalvmt	-0.00000*** (0.00000)	-0.00000*** (0.00000)	-0.00000*** (0.00000)
precip	-0.015** (0.006)	-0.015 (0.020)	-0.015 (0.020)
snow32	-0.297*** (0.017)	-0.297*** (0.052)	-0.297*** (0.053)
rural_speed	0.016*** (0.002)	0.016*** (0.004)	0.016*** (0.004)
urban_speed	0.005*** (0.002)	0.005 (0.003)	0.005 (0.003)
Constant	-2.153*** (0.143)	-2.153*** (0.277)	-2.153*** (0.278)

*Notes:*

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

Table 3: Comparison of Standard Error HC Methods for Full Pooled Model, as calculated by hand

Conventional	Robust	Clustered
0.143	0.126	0.281
0.027	0.024	0.055
0.005	0.006	0.008
0.0002	0.0002	0.0002
0.022	0.021	0.037
0.176	0.161	0.453
0.031	0.028	0.083
0.00000	0.00000	0.00000
0.006	0.006	0.020
0.017	0.018	0.053
0.002	0.002	0.004
0.002	0.002	0.003

```

43 #Compute variance/covariance matrix
44 varcovar <- XX1 %%% crossprod(uj) %%% XX1
45
46 # Degrees of freedom adjustment
47 dfc <- sqrt((m/(m-1)) * (nrow(X)-1)/(nrow(X)-ncol(X)))
48
49 stdh <- dfc*sqrt(diag(varcovar))
50
51 t <- pooled.full$coefficients/stdh
52 p <- 2*pnorm(-abs(t))
53 results.cluster <- cbind(pooled.full$coefficients, stdh, t, p)
54 dimnames(results.cluster) <- dimnames(s$coefficients)
55 results.cluster
56
57 hand.comparison <- cbind(a.typ.full[,2], results.robust[,2], results.cluster
58   [,2])
59 colnames(hand.comparison) <- c("Conventional", "Robust", "Clustered")
60 stargazer(data.frame(hand.comparison),
61   summary = FALSE,
62   title = "Comparison of Standard Error HC Methods for Full Pooled
63     Model, as calculated by hand",
64   style = "qje",
65   out = 'p3b2.tex',
66   font.size = "footnotesize",
67   column.labels = c("Conventional", "HC1 Robust", "HC1 Robust +
68     Cluster"),
69   label="tab:3b2"
70 )

```

## 2.3 Part C

The between estimator will give an unbiased estimate of the effect of primary seat belt laws insofar as variation within states (across time) is uncorrelated with the observables.

Table 4: Between models of effects of primary seatbelt use laws

	<i>Dependent variable:</i>	
	logfatalpc	
	(1)	(2)
primary	−0.071 (0.155)	0.114 (0.165)
secondary		−0.025 (0.159)
college		−2.603*** (0.636)
totalvmt		−0.00000** (0.00000)
snow32		−0.266*** (0.083)
rural_speed		0.063*** (0.012)
Constant	−1.716*** (0.052)	−4.978*** (0.789)
Observations	49	49
R <sup>2</sup>	0.004	0.758
Adjusted R <sup>2</sup>	0.004	0.650
F Statistic	0.212 (df = 1; 47)	21.898*** (df = 6; 42)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

We don't think that this criterion is met here. For example, within a given state, the total vmt per year probably tracks very closely with fatalities, as

the higher VMT within a given year, the more likely there are to be fatal crashes (*ceteris paribus*).

The standard errors are sufficiently large in this model that we cannot rule out the null hypothesis (no effect of primary seatbelt laws) in either the simple case or the case with covariates.

## 2.4 Part D

The RE estimator will give an unbiased estimate so long as the within states variation is uncorrelated with observables. Again, this assumption is probably not met here. In the RE model we find (for the first time) that seat belt laws appear to be correlated with statistical significance with reduced fatalities....but do we trust these? Table ?? summarizes the results for the RE model both with and without covariates.

The Random Effects estimator has the advantage over pooled OLS that it allows for (and assumes) unobserved heterogeneity. OLS has the advantage that it is more efficient than the Random Effects estimator when said heterogeneity does not exist.

## 2.5 Part E

The conventional standard errors for the RE model are likely incorrect because of a built-in assumption that the errors within each individual are correlated equally over time. Because this is a long-term dataset we would prefer to relax this assumption by introducing robust clustered SE estimates. Table 6 shows that this increases the SE of the coefficient on primary seatbelt use, but only by about double. The SE is still small compared to the estimand.

## 2.6 Part F

The conventional and clustered standard errors differ (as they have for many other approaches) by about a factor of two in the FE method. These are compared in Table 7. The baseline standard errors (or roughly the coefficient of variation) is small for the fixed effects model because all of the unobserved individual state characteristics are wrapped up in the FE model without any between-ness muddying their influence. However, there remain heteroscedastic errors within each state from idiosyncrasies between time periods when



Table 5: Random Effects Models

	<i>Dependent variable:</i>	
	logfatalpc	
	(1)	(2)
primary	−0.230*** (0.016)	−0.138*** (0.015)
secondary		−0.065*** (0.010)
college		−1.420*** (0.169)
unemploy		−0.024*** (0.002)
beer		0.757*** (0.038)
totalvmt		−0.00000*** (0.00000)
precip		−0.024*** (0.006)
snow32		−0.018 (0.014)
rural_speed		−0.006*** (0.001)
urban_speed		0.003*** (0.001)
Constant	−1.688*** (0.044)	−1.902*** (0.092)
Observations	1,127	1,127
R <sup>2</sup>	0.153	0.604
Adjusted R <sup>2</sup>	0.153	0.598
F Statistic	203.173*** (df = 1; 1125)	170.214*** (df = 10; 1116)

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 6: Comparison of Standard Error HC Methods for RE model with covariates

	Conventional SE	HC1 Robust + Cluster
	(1)	(2)
primary	−0.138*** (0.015)	−0.138*** (0.029)
secondary	−0.065*** (0.010)	−0.065*** (0.018)
college	−1.420*** (0.169)	−1.420*** (0.296)
unemploy	−0.024*** (0.002)	−0.024*** (0.003)
beer	0.757*** (0.038)	0.757*** (0.068)
totalvmt	−0.00000*** (0.00000)	−0.00000** (0.00000)
precip	−0.024*** (0.006)	−0.024*** (0.007)
snow32	−0.018 (0.014)	−0.018 (0.020)
rural_speed	−0.006*** (0.001)	−0.006*** (0.002)
urban_speed	0.003*** (0.001)	0.003** (0.001)
Constant	−1.902*** (0.092)	−1.902*** (0.149)

*Notes:*

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

states did and did not have seat belt laws (national-level trends, automobile manufacturing standards, etc.) that are exogenous to the states and link them. Figure 2 shows how the adoption of state-level primary seatbelt laws occurred non-uniformly through the sample.

Table 7: Comparison of Standard Error HC Methods for basic FE model

	Conventional SE	HC1 Robust + Cluster
	(1)	(2)
primary	−0.080*** (0.015)	−0.080** (0.033)
year	−0.025*** (0.002)	−0.025*** (0.004)
sqr.year	0.0005*** (0.0001)	0.0005*** (0.0001)
<i>Notes:</i>		
	***Significant at the 1 percent level.	
	**Significant at the 5 percent level.	
	*Significant at the 10 percent level.	

## 2.7 Part G

The addition of covariate terms to the FE model reduces the estimate of the ATE for enacting primary seatbelt laws to about a 6% reduction in annual fatalities. With robust / clustered standard errors this estimate maintains its significance at a 5% confidence level.

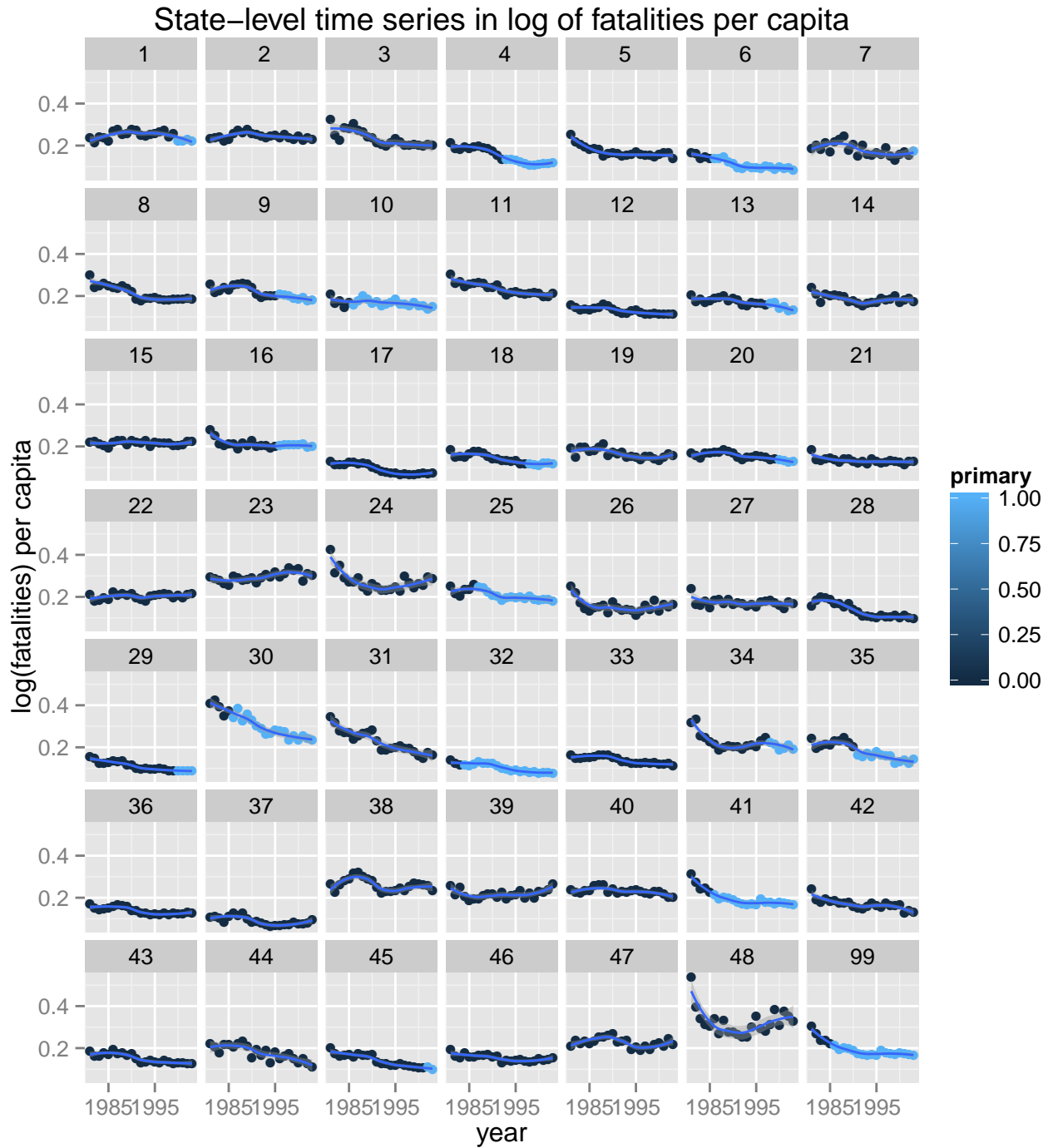


Figure 2: Year to year trends in the log of traffic fatalities per capita, with facets by state ID. A LOESS fit to each dataset is included for reference but is not necessarily indicative of the true underlying function.

Table 8: Fixed Effects Models

	<i>Dependent variable:</i>		
	logfatalpc		
	Simple FE	<i>panel</i> <i>linear</i>	<i>coefficient</i> <i>test</i>
	(1)	FE with Cov.	Robust/Clustered SE
	(1)	(2)	(3)
primary	−0.080*** (0.015)	−0.063*** (0.016)	−0.063** (0.032)
year	−0.025*** (0.002)	−0.017*** (0.003)	−0.017*** (0.004)
sqr.year	0.0005*** (0.0001)	0.0002* (0.0001)	0.0002 (0.0002)
secondary		0.006 (0.011)	0.006 (0.021)
college		0.377 (0.236)	0.377 (0.462)
beer		0.788*** (0.041)	0.788*** (0.079)
totalvmt		−0.00000 (0.00000)	−0.00000 (0.00000)
precip		−0.032*** (0.006)	−0.032*** (0.005)
snow32		−0.007 (0.014)	−0.007 (0.017)
rural_speed		0.004*** (0.001)	0.004*** (0.001)
urban_speed		0.001 (0.001)	0.001 (0.001)
Observations	1,127	1,127	
R <sup>2</sup>	0.459	0.625	
Adjusted R <sup>2</sup>	0.438	0.591	
F Statistic	304.360*** (df = 3; 1075)	161.455*** (df = 11; 1067)	

Note: