

Economics 312
Daily Problem #18

Spring 2013
March 8

(An extension of HGL Problem 8.4)

The textbook data set `vacation.dta` gives us data on the vacation travel of 200 Chicago households. The variable definition file gives the following definitions and summary statistics:

```
1. miles      miles traveled per year
2. income     annual income in $1000's
3. age        average age of adult members of household
4. kids       number of children in household
```

Variable	Obs	Mean	Std. Dev.	Min	Max
miles	200	1054.23	552.799	0	2609
income	200	63.925	18.33412	19	119
age	200	42.675	9.349139	23	59
kids	200	1.635	1.260842	0	4

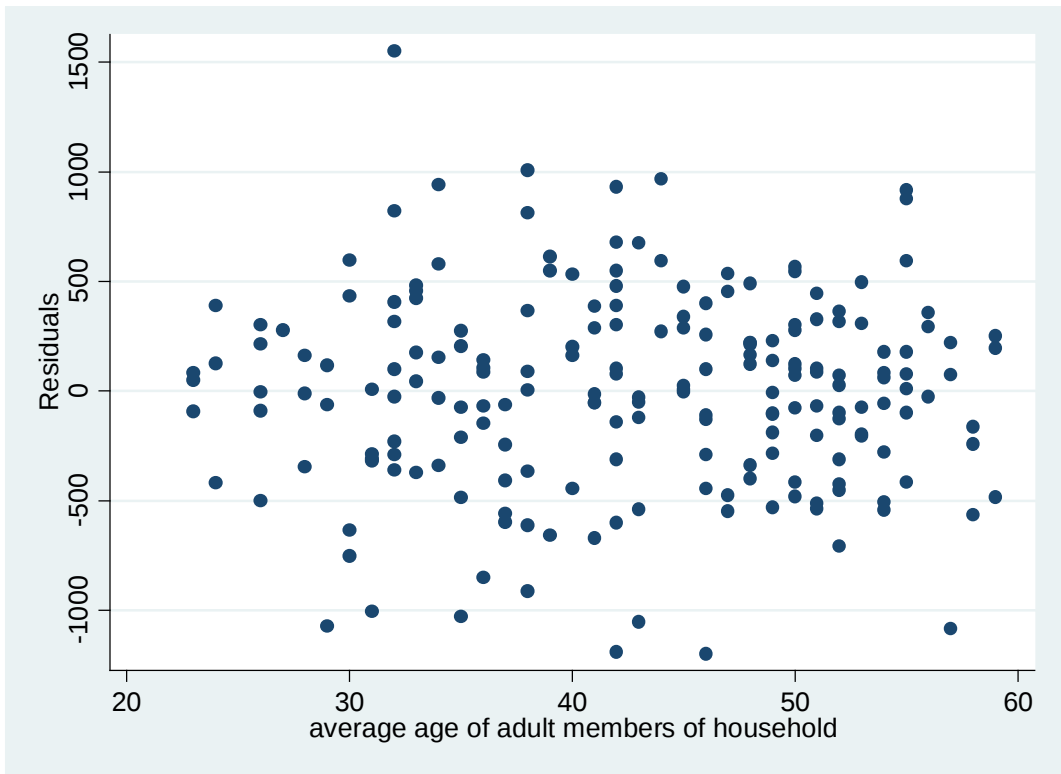
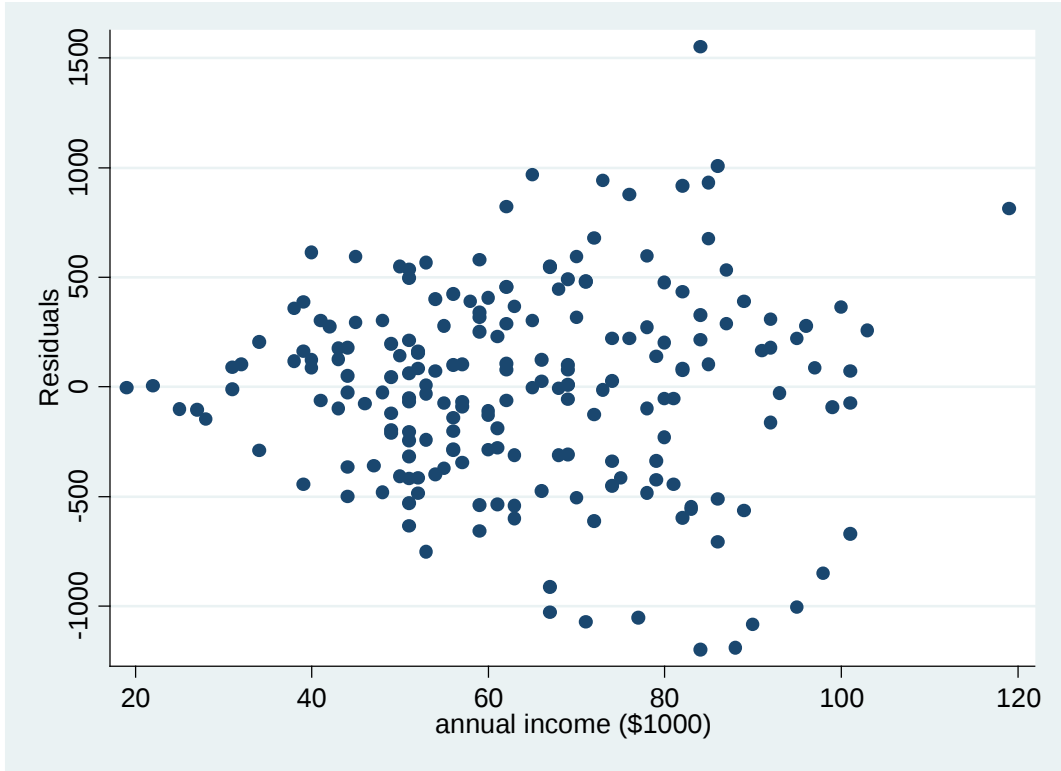
A regression of the model $miles = \beta_1 + \beta_2 income + \beta_3 age + \beta_4 kids$ yields the following:

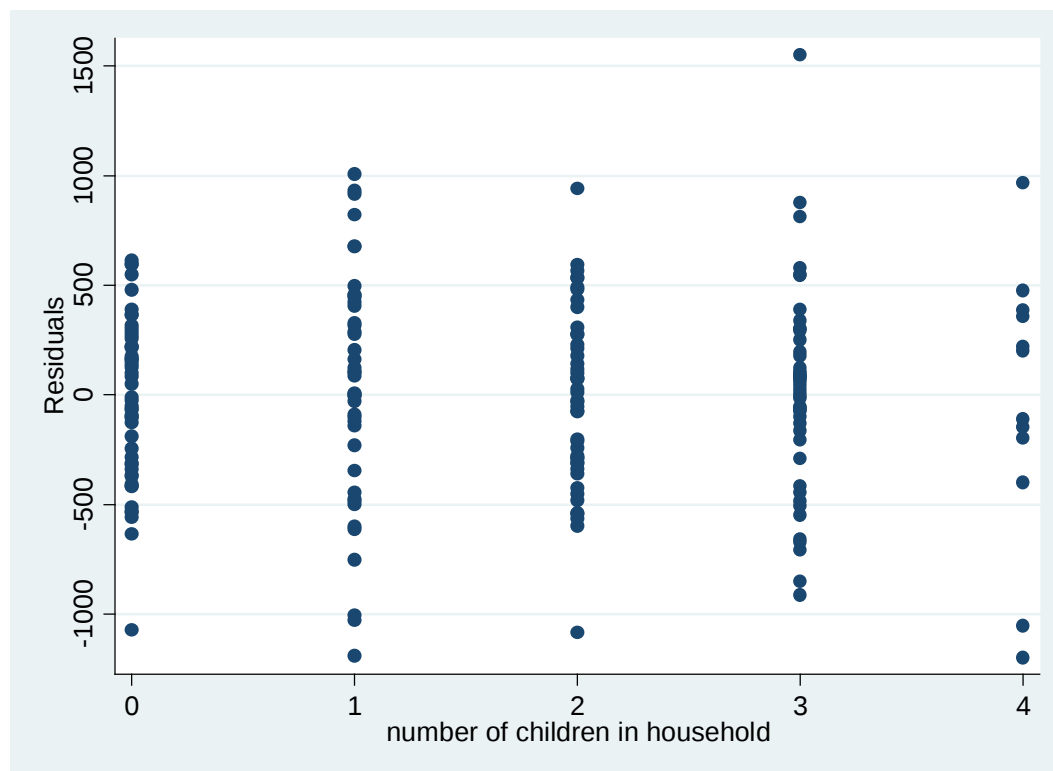
```
. reg miles income age kids
```

Source	SS	df	MS	Number of obs = 200		
Model	20712796.2	3	6904265.41	F(3, 196) = 33.75		
Residual	40098973.2	196	204586.598	Prob > F = 0.0000		
Total	60811769.4	199	305586.781	R-squared = 0.3406		
				Adj R-squared = 0.3305		
				Root MSE = 452.31		

miles	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	14.20133	1.800256	7.89	0.000	10.65097	17.75169
age	15.74092	3.75737	4.19	0.000	8.33086	23.15099
kids	-81.82642	27.1296	-3.02	0.003	-135.3298	-28.32302
_cons	-391.548	169.7752	-2.31	0.022	-726.3687	-56.72731

1. Briefly assess whether these results conform to your expectations.
2. To visually assess heteroskedasticity, we plot the residuals against each of the regressors, which yields





Based on these plots, what would you say about the likelihood of heteroskedasticity? Does this conform to what you would expect?

3. We next want to conduct a Breusch-Pagan test using the three regressors as potential variables to explain error variance. Initially, however, instead of regressing the squared residuals on the regressors we mistakenly use the actual (unsquared) residuals, yielding:

```
. reg ehat income age kids
```

Source	SS	df	MS	Number of obs =	200
Model	0	3	0	F(3, 196) =	0.00
Residual	40098973.4	196	204586.599	Prob > F =	1.0000
Total	40098973.4	199	201502.379	R-squared =	0.0000
				Adj R-squared =	-0.0153
				Root MSE =	452.31

ehat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	5.95e-09	1.800256	0.00	1.000	-3.55036	3.55036
age	-2.64e-08	3.75737	-0.00	1.000	-7.410064	7.410064
kids	-3.77e-07	27.1296	-0.00	1.000	-53.5034	53.50339
_cons	-1.23e-07	169.7752	-0.00	1.000	-334.8207	334.8207

Explain this result.

4. Correcting our procedure, we construct a variable ehat2 for the squared residuals and regress it on the three regressors:

```
. reg ehat2 income age kids
```

Source	SS	df	MS	Number of obs = 200		
Model	2.9205e+12	3	9.7349e+11	F(3, 196) = 11.09		
Residual	1.7210e+13	196	8.7806e+10	Prob > F = 0.0000		
Total	2.0130e+13	199	1.0116e+11	R-squared = 0.1451		
				Adj R-squared = 0.1320		
				Root MSE = 3.0e+05		

ehat2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	6194.121	1179.389	5.25	0.000	3868.199	8520.043
age	-7512.891	2461.539	-3.05	0.003	-12367.39	-2658.388
kids	37255.68	17773.22	2.10	0.037	2204.383	72306.98
_cons	64235.25	111223.6	0.58	0.564	-155113.4	283583.9

Perform the Breusch-Pagan χ^2 test and interpret your conclusion about heteroskedasticity.

5. Based on your results, what are the properties of your OLS estimates? (Are they consistent? Are they efficient?) What (if any) further statistical analysis do you recommend?