

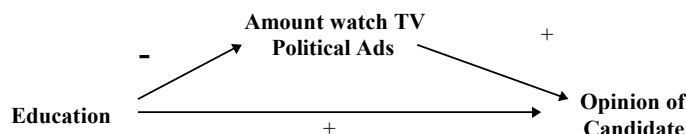
## Soc 63993, Homework #5: Model Mis-Specification/Equality Constraints/Group Comparisons

Richard Williams, University of Notre Dame, <https://academicweb.nd.edu/~rwilliam/>

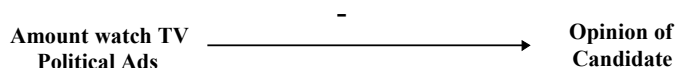
Last revised February 15, 2015

1. **Model mis-specification.** A campaign manager has found that the amount of time spent watching TV political ads is negatively correlated with favorable opinion of her candidate. Two models have been proposed to explain this relationship:

(i)



(ii)



A. Suppose that model (i) is correct. What harm will result from estimating model (ii) and relying on the results? If appropriate, discuss such things as biased coefficients, inflated standard errors, and misguided policy decisions (particularly with regards to the use of TV advertising). Similarly, discuss the harm that will result if Model (ii) is correct and model (i) is mistakenly estimated and relied upon.

B. Model (i) is estimated, yielding the following results. Based on this information, determine what the regression coefficient would be for model (ii). Compute the regression coefficient using both the formula for omitted variable bias and the formula for the slope coefficient in a bivariate regression.

**. sum**

Variable	Obs	Mean	Std. Dev.	Min	Max
opinion	200	79	9.4	57.15274	99.79181
educ	200	14	2.7	6.597328	20.61872
tv	200	15	5.6	-.8872261	34.85936

**. corr**

	opinion	educ	tv
opinion	1.0000		
educ	0.3500	1.0000	
tv	-0.2200	-0.9000	1.0000

```
. corr, cov
```

	opinion	educ	tv
opinion	88.36		
educ	8.883	7.29	
tv	-11.5808	-13.608	31.36

```
. reg opinion educ tv, beta
```

Source	SS	df	MS		Number of obs =	200
Model	2989.21851	2	1494.60926		F( 2, 197) =	20.17
Residual	14594.421	197	74.0833553		Prob > F =	0.0000
					R-squared =	0.1700
					Adj R-squared =	0.1616
					Root MSE =	8.6072
Total	17583.6395	199	88.3599975			

opinion	Coef.	Std. Err.	t	P> t	Beta
educ	2.785185	.5184337	5.37	0.000	.8
tv	.8392856	.2499591	3.36	0.001	.5
_cons	27.41812	10.77459	2.54	0.012	.

C. Based on these results, which model do you think is most plausible? Why?

D. The campaign manager is concerned by the large correlation between educ and tv. Suppose the manager decided to “solve” the problem of multicollinearity by excluding education from the model. What would be the consequence of that decision? Do you think this would be a good idea in this case?

2. Equality constraints. From the course web page, download gender.dta. This is yet another modified version of our income/education/job experience example. The sample now consists of 225 men and 275 women. Regress income on education and job experience. Test the following hypotheses:

$$H_0: \beta_{\text{Educ}} = \beta_{\text{Jobexp}}$$

$$H_A: \beta_{\text{Educ}} \neq \beta_{\text{Jobexp}}$$

Perform a Wald test, an incremental F test, and a likelihood ratio chi-square test. The results should all be identical or nearly identical.

3. Group comparisons. Using the same data as in problem 2, do the following:

(a) Do T-tests of whether the means of men and women significantly differ on education, job experience, and income. If using Stata, use commands such as

```
. ttest educ, by(female)
```

(b) Test the following. Use a likelihood ratio chi square test. Performing an incremental F test and/or a Wald test using `suest` is optional.

$H_0$ : Model parameters are exactly the same for both men and women

$H_A$ : Model parameters are not exactly the same for both men and women.

(c) Based on your results, explain whether men make more than women and if so why. [Note: these are hypothetical data, and the results are a little peculiar in some respects!]

(d) Suppose there were no gender-related compositional differences, i.e. women had the same levels of education and job experience as men did. If education and job experience continued to have the same effects on women that they do now, how much would the gap in income between men and women be affected?