

Compare the adjusted R2 and the coefficient on x1 with those you obtained in part (a).

4. (22 pts) Suppose that you wish to examine the impact of education, gender and experience on the income of individuals by estimating the following specification.

$$\ln \text{income}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{female}_i + \beta_3 \text{exper}_i + \beta_4 \text{exper}_i^2 + u_i \quad (1)$$

where the dependent variable *lnincome* is the natural log of individuals' income; *educ* is the years of education; *female* is a binary variable that indicates whether an individual is a female (i.e. *female*=1 if female individual, zero otherwise); and *exper* is the years of the labor market experience.

The results from estimating this equation are given here.

```
gen exper_sq=exper^2
```

```
reg lnincome female educ exper exper_sq
```

Source	SS	df	MS	Number of obs =	6540
Model	872.165992	4	218.041498	F(4, 6535) =	395.68
Residual	3601.10591	6535	.551049106	Prob > F =	0.0000
				R-squared =	0.1950
				Adj R-squared =	0.1945
Total	4473.2719	6539	.68409113	Root MSE =	.74233

lnincome	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
female	-.3890827	.0185187	-21.01	0.000	-.4253855 -.35278
educ	.1170137	.0037012	31.62	0.000	.1097581 .1242693
exper	.1165762	.0250472	4.65	0.000	.0674755 .1656768
exper_sq	-.0029639	.0011804	-2.51	0.012	-.0052779 -.0006499
_cons	7.316932	.1378585	53.08	0.000	7.046684 7.58718

- Interpret the coefficient on *educ* variable. Is *educ* variable statistically significant at 0.01 significance level? Specify appropriate hypothesis, show essential components of the test, and interpret your result.
- Is the relationship between *lnincome* and *exper* variables **nonlinear**? Explain.
- Draw a graph that shows the relationship between *lnincome* and *exper* variables.
- At what value of *exper* does additional experience actually decrease/increase the predicted *lnincome*?
- Interpret the coefficient on *female* variable.
- Calculate and interpret the coefficient of determination. Be specific by using the numerical value.
- Are the independent variables **jointly** significant at .01 significance level? Specify appropriate hypothesis, show essential components of the test, and interpret your result.
- Now, suppose that you wish to estimate the following specification.

$$\ln \text{income}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{female}_i + \beta_3 \text{exper}_i + \beta_4 \text{exper}_i^2 + \beta_5 (\text{educ}_i * \text{female}_i) + u_i \quad (2)$$

The results from estimating this equation are the following.

gen fmeduc= female*educ

reg lnincome female educ fmeduc exper exper_sq

Source	SS	df	MS	Number of obs = 6540		
Model	881.649777	5	176.329955	F(5, 6534) = 320.79		
Residual	3591.62212	6534	.54968199	Prob > F = 0.0000		
				R-squared = 0.1971		
				Adj R-squared = 0.1965		
Total	4473.2719	6539	.68409113	Root MSE = .74141		

lnincome	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female	-.7952773	.0995248	-7.99	0.000	-.9903783	-.6001762
educ	.1030813	.0049916	20.65	0.000	.0932962	.1128664
fmeduc	.0308856	.0074357	4.15	0.000	.0163092	.045462
exper	.1104828	.0250591	4.41	0.000	.0613588	.1596068
exper_sq	-.002736	.0011802	-2.32	0.020	-.0050497	-.0004224
_cons	7.536249	.1474642	51.11	0.000	7.24717	7.825327

- Interpret the coefficients on variables female, educ, and fmeduc.
- Explain the difference between model (1) and model (2). Explain why one needs to estimate the second model.
- Test the hypothesis that the second model is more/less appropriate than the first model using the “Chow Test” or “the test of exclusion restrictions.”
- Highlight both the economic and statistical significance of the coefficient on variable fmeduc. Carefully discuss the implications of your results.

5. (4 pts) Consider the following output.

. regress rdintens sales profmarg

Source	SS	df	MS	Number of obs = 32		
Model	8.28423732	2	4.14211866	F(2, 29) = 1.19		
Residual	100.549233	29	3.46721493	Prob > F = 0.3173		
				R-squared = 0.0761		
				Adj R-squared = 0.0124		
Total	108.83347	31	3.51075711	Root MSE = 1.862		

rdintens	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sales	.0000534	.0000441	1.21	0.236	-.0000368	.0001435
profmarg	.0446166	.0461805	0.97	0.342	-.0498332	.1390664
_cons	2.625261	.5855328	4.48	0.000	1.427712	3.82281

. estat ovtest

Ramsey RESET test using powers of the fitted values of rdintens
Ho: model has no omitted variables
F(3, 26) = 1.63
Prob > F = 0.2056

Is there a specification error? Explain. Write the null and alternative hypotheses?

6. (4 pts) What are the four moments? Consider the following output.

```
regress rdintens sales profmarg, robust
predict res, r
sum res, d
```

```
.
. sum res, d
```

Residuals				
	Percentiles	Smallest		
1%	-2.222148	-2.222148		
5%	-1.711887	-1.711887		
10%	-1.600882	-1.615433	Obs	32
25%	-1.146509	-1.600882	Sum of Wgt.	32
50%	-.6067708		Mean	2.97e-09
		Largest	Std. Dev.	1.800979
75%	.523954	2.016002		
90%	2.016002	3.039686	Variance	3.243524
95%	3.297106	3.297106	Skewness	1.674078
99%	6.370235	6.370235	Kurtosis	6.170485

Are the residuals from the model skewed and leptokurtic? How do you know?

7. (4 pts) Consider the following output.

```
regress rdintens sales profmarg, robust
```

```
. jbr res
```

```
Jarque-Bera normality test: 28.35 Chi(2) 7.0e-07
```

```
Jarque-Bera test for Ho: normality:
```

Are the errors of the model normally distributed? Explain. Why is it important to have normally distributed errors?

8. (8 pts) Use RDCHEM data and perform the following.

- Draw a scatter plot that shows the relationship between rdintens and sales. Discuss the relationship.
- Run the following OLS regression.

```
regress rdintens sales profmarg
```

and compute the standardized residuals. Are there any outliers?