Introductory Econometrics

Tutorial 7

Part B

Q1 The purpose of this question is to investigate the influence of various factors on the probability of an individual having had an affair. **The workfile 'affair.wf1' has been placed on Moodle**. The workfile contains 601 observations. The data were collected from a random sample of the Monash University academic staff (Just kidding!!). The variables are:

Variable Code	Description
affair_i	1 if individual i has had at least one affair, 0 otherwise
age_i	age in years of individual i
educ_i	years of schooling of individual i
hapmarr_i	1 if individual i rates their marriage happy, 0 otherwise
kids_i	1 if individual i has children, 0 otherwise
male_i	1 if individual i is male, 0 otherwise
rel_i	1 if individual i is religious, 0 otherwise
yrsmarr_i	number of years married for which individual i has been married

(a) Prove that

$$E(affair_i) = p_i.$$

Ans: Since $affair_i$ is a binary variable

$$\begin{aligned} E(affair_i) &= 1 * p_i + 0 * (1 - p_i) \\ &= p_i. \end{aligned}$$

(b) Assume that

$$\begin{aligned} &E(affair_i|age_i, educ_i, hapmarr_i, kids_i, male_i, rel_i, yrsmarr_i) \\ &= \beta_0 + \beta_1 age_i + \beta_2 educ_i + \beta_3 hapmarr_i + \beta_4 kids_i + \beta_5 male_i + \beta_6 rel_i + \beta_7 yrsmarr_i. \end{aligned} \tag{1}$$

- **i)** What is the probability that an unhappily married, child-free, religious male, aged 30 with 12 years of education who has been married for 10 years, has had an affair?
- **ii)** What is the marginal effect of gender on the probability of having had an affair, controlling for the variables age, educ, hapmarr, kids, rel and yrsmarr?

Ans:

i) From (1)

$$\begin{aligned} &E(affair_i|age_i = 30,, educ_i = 12, hapmarr_i = 0,, kids_i = 0, male_i = 1, rel_i = 1, yrsmarr_i = 10) \\ &= \beta_0 + 30\beta_1 + 12\beta_2 + \beta_5 + \beta_6 + 10\beta_7. \end{aligned}$$

ii) From (1)

$$\begin{aligned} &E(affair_i|age_i, educ_i, hapmarr_i, kids_i, male_i = 1, rel_i, yrsmarr_i) \\ &= \beta_0 + \beta_1 age_i + \beta_2 educ_i + \beta_3 hapmarr_i + \beta_4 kids_i + \beta_5 + \beta_6 rel_i + \beta_7 yrsmarr_i, \end{aligned}$$

and

$$E(affair_i|age_i, educ_i, hapmarr_i, kids_i, male_i = 0, rel_i, yrsmarr_i)$$
$$= \beta_0 + \beta_1 age_i + \beta_2 educ_i + \beta_3 hapmarr_i + \beta_4 kids_i + \beta_6 rel_i + \beta_7 yrsmarr_i.$$

Therefore,

$$\begin{aligned} &E(affair_i|age_i, educ_i, hapmarr_i, kids_i, male_i = 1, rel_i, yrsmarr_i) \\ -&E(affair_i|age_i, educ_i, hapmarr_i, kids_i, male_i = 0, rel_i, yrsmarr_i) \\ &= \beta_5. \end{aligned}$$

(c) Estimate the linear regression equation

$$affair_i = \beta_0 + \beta_1 age_i + \beta_2 educ_i + \beta_3 hapmarr_i + \beta_4 kids_i + \beta_5 male_i + \beta_6 rel_i + \beta_7 yrsmarr_i + u_i \quad (2)$$

by OLS.

i) Report the estimated model.

ii) Are the regressors jointly significant? Briefly explain.

iii) Evaluate the explanatory power of the model.

iv) Is there any evidence that males are more likely than females with the same characteristics to have had an affair?

Ans:

i)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.382370	0.138643	2.757941	0.0060
AGE	-0.007686	0.003049	-2.520534	0.0120
EDUC	0.000527	0.007822	0.067329	0.9463
HAPMARR	-0.134809	0.036532	-3.690166	0.0002
KIDS	0.046614	0.046725	0.997616	0.3189
MALE	0.059771	0.038577	1.549396	0.1218
REL	-0.127416	0.037216	-3.423696	0.0007
YRSMARR	0.016990	0.005569	3.050498	0.0024

R-squared	0.076228	Mean dependent var	0.249584
Adjusted R-squared	0.065323	S.D. dependent var	0.433133
S.E. of regression	0.418747	Akaike info criterion	1.110123
Sum squared resid	103.9820	Schwarz criterion	1.168673
Log likelihood	-325.5919	Hannan-Quinn criter.	1.132913
F-statistic	6.990423	Durbin-Watson stat	1.807563
Prob(F-statistic)	0.000000		

Figure 1

ii) The sample value of the F statistic for testing the joint significance of the regressors is 6.990423 and the associated p-value is 0.00000. Consequently, at any significance level, we would reject the null hypothesis that the regressors are jointly insignificant.

iii)

$$R^2 = 0.076228.$$

Therefore, the regressors jointly explain approximately 8% of the **sample variation** in the probability of having an affair. The specified model does not have a great deal of explanatory power.

iv) There is no evidence that males are more likely than females to have had an affair. The sample value of the t statistic for testing the individual significance of the regressor male has a p-value of 0.1218. Since

$$0.1218 > 0.10,$$

we would fail to reject the null hypothesis that the regressor male is insignificant in a two-sided test even if we tested at the 10% significance level.

(**d**) Test the joint significance of the individually insignificant regressors in (2). Specify the unrestricted and restricted models, the null and alternative hypotheses, the form and null distribution of the test statistic, the sample and critical values of the test statistic and your test conclusion.

Ans:

- Based on the output reported in Figure 1 above, the regressors educ, kids and male are individually insignificant in a two-sided test at the 5% significance level, since the t statistic for testing the individual significance of each has a p-value which is greater than 0.05.
- The unrestricted model is given by

$$affair_i = \beta_0 + \beta_1 age_i + \beta_2 educ_i + \beta_3 hapmarr_i + \beta_4 kids_i + \beta_5 male_i + \beta_6 rel_i + \beta_7 yrsmarr_i + u_i.$$

- The restricted model is given by

$$affair_i = \beta_0 + \beta_1 age_i + \beta_3 hapmarr_i + \beta_6 rel_i + \beta_7 yrsmarr_i + u_i.$$

- When we estimate the restricted model by OLS we obtain the output reported in Figure 2 below

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.409330	0.074577	5.488676	0.0000
AGE	-0.006431	0.002930	-2.194705	0.0286
HAPMARR	-0.141906	0.036256	-3.913997	0.0001
REL	-0.124148	0.037104	-3.345941	0.0009
YRSMARR	0.017513	0.004959	3.531125	0.0004

R-squared	0.069739	Mean dependent var	0.249584
Adjusted R-squared	0.063495	S.D. dependent var	0.433133
S.E. of regression	0.419156	Akaike info criterion	1.107139
Sum squared resid	104.7125	Schwarz criterion	1.143733
Log likelihood	-327.6953	Hannan-Quinn criter.	1.121383
F-statistic	11.17003	Durbin-Watson stat	1.816299
Prob(F-statistic)	0.000000		

Figure 2

-

$$
\begin{aligned}
H_0 &: \beta_2 = \beta_4 = \beta_5 = 0 \\
H_1 &: \beta_2 \text{ and/or } \beta_4 \text{ and/or } \beta_5 \neq 0 \\
\text{Significance level} &: \alpha = 0.05 \\
\text{Test stat and null distribution} &: \frac{(SSR_r - SSR_{ur})}{SSR_{ur}} \frac{(n-k-1)}{q} \\
&= \frac{(SSR_r - SSR_{ur})}{SSR_{ur}} \frac{(601-8)}{3} \sim F(3,198) \\
F_{calc} &= \frac{(104.7125 - 103.9820)}{103.9820} \frac{593}{3} \\
&= 1.389 \\
F_{crit} &= 2.13 \\
\text{Decision rule} &: \textit{reject } H_0 \text{ if } F_{calc} > Ft_{crit}
\end{aligned}
$$

- Since

$$1.389 < 2.13,$$

we fail to reject the null hypothesis that the regressors educ, kids and male are jointly insignificant.

(e) Let

$$affair_i = \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3 rel_i + \alpha_4 yrsmarr_i + u_{i,} i = 1,2,\ldots,n, \tag{3}$$

where

$$E(u_i | age_i, hapmarr_i, rel_i, yrsmarr_i) = 0.$$

i) Extend the model given by (3) to allow for the possibility that the marginal effect of being happily married on the probability of having had an affair may be different for religious and nonreligious people.

Ans:

- To allow for this possibility we extend (3) by introducing the slope (interaction) dummy variable $rel_i * hapmarr_i$. The extended model is given by

$$affair_i = \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3 rel_i + \alpha_4 yrsmarr_i + \alpha_5 rel_i * hapmarr_i + u_{i,} i = 1,2,\ldots,n. \tag{4}$$

- Given the model specified in (4),

$$
\begin{aligned}
p_i &= E(affair_i | age_i, hapmarr_i, rel_i, yrsmarr_i) \\
&= \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3 rel_i + \alpha_4 yrsmarr_i + \alpha_5 rel_i * hapmarr_i, i = 1,2,\ldots,n. \qquad (5)
\end{aligned}
$$

- From (5) the marginal effect of being happily married on the response probability for a **religious** person is

$$
\begin{aligned}
E(affair_i | age_i, hapmarr_i = 1, rel_i = 1, yrsmarr_i) &= -E(affair_i | age_i, hapmarr_i = 0, rel_i = 1, yrsmarr_i) \\
&= \alpha_0 + \alpha_1 age_i + \alpha_2 + \alpha_3 + \alpha_4 yrsmarr_i + \alpha_5 - (\alpha_0 + \alpha_1 age_i + \alpha_3 + \alpha_4 yrsmarr_i) \\
&= \alpha_2 + \alpha_5,
\end{aligned}
$$

- From (5) the marginal effect of being happily married on the response probability for a **nonreligious** person is

$$E(affair_i|age_i, hapmarr_i = 1, rel_i = 0, yrsmarr_i) - E(affair_i|age_i, hapmarr_i = 0, rel_i = 0, yrsmarr_i)$$
$$= \alpha_0 + \alpha_1 age_i + \alpha_2 + \alpha_4 yrsmarr_i - (\alpha_0 + \alpha_1 age_i + \alpha_4 yrsmarr_i)$$
$$= \alpha_2.$$

ii) Is the proposition in i) supported by the data? Briefly explain.

Ans:

- Since the marginal effect of the regressor *hapmarr* on the response probability for a religious person is
$$\alpha_2 + \alpha_5$$
and the marginal effect for a nonreligious person is
$$\alpha_2,$$
the marginal effect will be the same for both types of person if and only if
$$\alpha_5 = 0.$$
- When we estimate (4) by OLS we obtain the output reported in Figure 3 below.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.396993	0.076211	5.209151	0.0000
AGE	-0.006472	0.002932	-2.207579	0.0277
HAPMARR	-0.140232	0.036329	-3.860056	0.0001
REL	-0.077178	0.069985	-1.102783	0.2706
YRSMARR	0.019200	0.005399	3.555853	0.0004
REL*YRSMARR	-0.005373	0.006788	-0.791629	0.4289

Figure 3

- Since the p-value associated with the sample value of the t statistic for testing
$$H_0 : \alpha_5 = 0$$
$$H_i : \alpha_5 \neq 0$$
is
$$0.4289 > 0.05,$$
we fail to reject the null hypothesis that the marginal effect of being happily married on the probability of having had an affair is the same for both religious and nonreligious people.

(f) Let
$$affair_i = \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3 rel_i + \alpha_4 yrsmarr_i + u_i, i = 1, 2, \ldots, n, \quad (3)$$
where
$$E(u_i|age_i, hapmarr_i, rel_i, yrsmarr_i) = 0.$$

i) Define the population of category A individuals to be the population of unhappily married, nonreligious people ,aged 30, who have been married for 2 years. What is the probability that a category A person has had an affair?

ii) Define the population of category B individuals to be the population of unhappily married, religious people, aged 30, who have been married for 1 year. What is the probability that a category B person has had an affair?

iii) What are the restrictions on the parameters of (3) implied by the null hypothesis that the probability that a category A person has had an affair is the same as the probability that a category B person has had an affair.

Ans:

i)

$$p^A = \alpha_0 + 30\alpha_1 + 2\alpha_4.$$

ii)

$$p^B = \alpha_0 + 30\alpha_1 + \alpha_3 + \alpha_4.$$

iii) Setting

$$p^A = p^B$$

we obtain

$$\alpha_0 + 30\alpha_1 + 2\alpha_4 = \alpha_0 + 30\alpha_1 + \alpha_3 + \alpha_4,$$

which implies that

$$\alpha_4 = \alpha_3$$

(**g**) Use the model specified in (3) to test the null hypothesis that the probability that a category A person has had an affair is the same as the probability that a category B person has had an affair, against the alternative hypothesis that the probability of a category A person having had an affair is greater.

Ans:

- Under the alternative hypothesis

$$\begin{aligned} p^A &> p^B \\ &\Rightarrow \alpha_0 + 30\alpha_1 + 2\alpha_4 > p^B = \alpha_0 + 30\alpha_1 + \alpha_3 + \alpha_4 \\ &\Rightarrow \alpha_4 > \alpha_3. \end{aligned}$$

Therefore,

$$\begin{aligned} H_0 &: \alpha_4 = \alpha_3 \\ H_i &: \alpha_4 > \alpha_3. \end{aligned}$$

- Define

$$\begin{aligned} \delta &= \alpha_4 - \alpha_3 \\ &\Rightarrow \alpha_4 = \delta + \alpha_3. \end{aligned} \tag{6}$$

- Substituting (6) into

$$affair_i = \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3 rel_i + \alpha_4 yrsmarr_i + u_i, i = 1,2,\ldots,n \tag{3}$$

and rearranging we obtain

$$\begin{aligned} affair_i &= \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3 rel_i + (\delta + \alpha_3) yrsmarr_i + u_i, i = 1,2,\ldots,n \\ &= \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3 (rel_i + yrsmarr_i) + \delta yrsmarr_i + u_i, i = 1,2,\ldots,n \end{aligned} \tag{7}$$

- Since

$$\delta = \alpha_4 - \alpha_3,$$

testing

$$\begin{aligned} H_0 &: \alpha_4 = \alpha_3 \\ H_i &: \alpha_4 > \alpha_3 \end{aligned}$$

in (3) is equivalent to testing

$$H_0 : \delta = 0$$
$$H_i : \delta > 0$$

in (7).

- When we estimate

$$affair_i = \alpha_0 + \alpha_1 age_i + \alpha_2 hapmarr_i + \alpha_3(rel_i + yrsmarr_i) + \delta yrsmarr_i + u_i, i = 1, 2, \ldots, n, \tag{7}$$

by OLS we obtain the output reported in Figure 4 below.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.409330	0.074577	5.488676	0.0000
AGE	-0.006431	0.002930	-2.194705	0.0286
HAPMARR	-0.141906	0.036256	-3.913997	0.0001
REL+YRSMARR	-0.124148	0.037104	-3.345941	0.0009
YRSMARR	0.141660	0.037904	3.737327	0.0002

Figure 4

- Since the sample value of the t statistic for testing

$$H_0 : \delta = 0$$

is

$$3.737327 > 0$$

and

$$p - value = 0.0002 < 0.10,$$

we reject

$$H_0 : \delta = 0$$
$$H_i : \delta > 0$$

when we test at the 5% significance level. (See the discussion of p-values in one-sided tests in the lecture notes for Topic 6).

- That is, we reject the null hypothesis that the probability that a category A person has had an affair is the same as the probability that a category B person has had an affair, in favor of the alternative hypothesis that the probability of a category A person having had an affair is greater.