

Exercise 2

#2

Calculate the correlation between Y and X1.

	Y	u1
Y	1.0000	
u1	0.1160	1.0000

This is quite different from 1.2.

Calculate the coefficients on the regression.

Source	SS	df	MS	Number of obs	=	10,000
Model	315329.133	3	105109.711	F(3, 9996)	>	99999.00
Residual	9947.3343	9,996	.995131483	Prob > F	=	0.0000
				R-squared	=	0.9694
				Adj R-squared	=	0.9694
Total	325276.467	9,999	32.5308998	Root MSE	=	.99756

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
u1	1.175044	.0171802	68.40	0.000	1.141367	1.20872
u2	-.9003673	.0028711	-313.59	0.000	-.9059953	-.8947393
u3	.1005889	.0002157	466.41	0.000	.1001662	.1010117
_cons	.8009973	.6478174	1.24	0.216	-.4688551	2.07085

Calculate the standard errors, using the standard formulas of the OLS.

This can be shown in the chart above.

Using bootstrap with 49 and 499 replications respectively.

Bootstrap replications (49)						
.....						
Linear regression				Number of obs	=	10,000
				Replications	=	49
				Wald chi2(3)	=	358306.09
				Prob > chi2	=	0.0000
				R-squared	=	0.9694
				Adj R-squared	=	0.9694
				Root MSE	=	0.9976

Y	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]	
u1	1.175044	.0168205	69.86	0.000	1.142076	1.208011
u2	-.9003673	.0029568	-304.51	0.000	-.9061625	-.894572
u3	.1005889	.000259	388.30	0.000	.1000812	.1010966
_cons	.8009973	.7839713	1.02	0.307	-.7355582	2.337553

```

-----
Linear regression                                Number of obs   =    10,000
                                                Replications    =      499
                                                Wald chi2(3)    =   283965.16
                                                Prob > chi2     =    0.0000
                                                R-squared       =    0.9694
                                                Adj R-squared   =    0.9694
                                                Root MSE       =    0.9976

```

Y	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]	
u1	1.175044	.017771	66.12	0.000	1.140213	1.209874
u2	-.9003673	.0028295	-318.21	0.000	-.905913	-.8948216
u3	.1005889	.0002219	453.32	0.000	.100154	.1010238
_cons	.8009973	.6634689	1.21	0.227	-.4993779	2.101373

#4

The probit regression:

```

Probit regression                                Number of obs   =    10,000
                                                LR chi2(3)      =   11283.62
                                                Prob > chi2     =    0.0000
Log likelihood = -1287.8187                    Pseudo R2       =    0.8142

```

y dum	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
u1	1.132912	.0560252	20.22	0.000	1.023104	1.242719
u2	-.9120145	.0249007	-36.63	0.000	-.9608189	-.86321
u3	.1018701	.0026759	38.07	0.000	.0966254	.1071148
_cons	-302.3891	7.947723	-38.05	0.000	-317.9664	-286.8119

The logit regression:

```

Logistic regression                            Number of obs   =    10,000
                                                LR chi2(3)      =   11275.15
                                                Prob > chi2     =    0.0000
Log likelihood = -1292.0526                    Pseudo R2       =    0.8135

```

y dum	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
u1	2.057558	.1046452	19.66	0.000	1.852458	2.262659
u2	-1.647092	.0492343	-33.45	0.000	-1.74359	-1.550595
u3	.1839136	.0053234	34.55	0.000	.17348	.1943472
_cons	-545.9487	15.81095	-34.53	0.000	-576.9376	-514.9598

The linear probability model:

Linear regression

Number of obs	=	10,000
F(3, 9996)	=	5524.02
Prob > F	=	0.0000
R-squared	=	0.6088
Root MSE	=	.31275

y dum	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
u1	.0818024	.0053584	15.27	0.000	.0712989	.092306
u2	-.0601577	.0009596	-62.69	0.000	-.0620386	-.0582767
u3	.0071021	.0000648	109.56	0.000	.006975	.0072292
_cons	-20.59574	.1947444	-105.76	0.000	-20.97748	-20.214

Compute the marginal effect of X on Y according to the probit and logit models.

Conditional marginal effects

Number of obs	=	10,000
---------------	---	--------

Model VCE : OIM

Expression : Pr(ydum), predict()
dy/dx w.r.t. : u1 u2 u3
at : u1 = 2.00771 (mean)
u2 = 6.012179 (mean)
u3 = 2999.51 (mean)

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
u1	.4516502	.0223285	20.23	0.000	.4078871	.4954132
u2	-.3635866	.0098969	-36.74	0.000	-.3829842	-.344189
u3	.0406119	.0010656	38.11	0.000	.0385234	.0427003

Conditional marginal effects

Number of obs	=	10,000
---------------	---	--------

Model VCE : OIM

Expression : Pr(ydum), predict()
dy/dx w.r.t. : u1 u2 u3
at : u1 = 2.00771 (mean)
u2 = 6.012179 (mean)
u3 = 2999.51 (mean)

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
u1	.5137631	.0261146	19.67	0.000	.4625795	.5649467
u2	-.4112716	.0122423	-33.59	0.000	-.435266	-.3872771
u3	.0459224	.0013272	34.60	0.000	.0433211	.0485237

Exercise 3

#1

The market shares of the ten products:

Variable	Obs	Mean	Std. Dev.	Min	Max
ppk_stk	4,470	.5184362	.1505174	.19	.67
pbb_stk	4,470	.5432103	.1203319	.19	1.01
pfl_stk	4,470	1.01502	.0428952	.95	1.16
phse_stk	4,470	.4371476	.1188312	.19	.64
pgen_stk	4,470	.3452819	.0351661	.25	.55
pimp_stk	4,470	.7807785	.1146461	.33	2.3
pss_tub	4,470	.8250895	.0612116	.5	.98
ppk_tub	4,470	1.077409	.0297261	.98	1.24
pfl_tub	4,470	1.189376	.0140545	.69	1.47
phse_tub	4,470	.5686734	.072455	.33	1.27

2&3

Propose a model for the effects of price on demand. Write the likelihood and optimize the model.

Alternative-specific conditional logit	Number of obs	=	44,700
Case variable: num	Number of cases	=	4470
Alternative variable: price	Alts per case: min	=	10
	avg	=	10.0
	max	=	10
Log likelihood = -7464.9321	Wald chi2(1)	=	1458.85
	Prob > chi2	=	0.0000

dum	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
price						
c	-6.656579	.1742793	-38.19	0.000	-6.99816	-6.314998
1	(base alternative)					
2						
_cons	-.9543068	.0500462	-19.07	0.000	-1.052396	-.856218
3						
_cons	1.296968	.1086515	11.94	0.000	1.084015	1.509921
4						
_cons	-1.717332	.0541582	-31.71	0.000	-1.82348	-1.611184
5						
_cons	-2.904005	.0714605	-40.64	0.000	-3.044065	-2.763945
6						
_cons	-1.515311	.1262303	-12.00	0.000	-1.762718	-1.267904
7						
_cons	.2517684	.079164	3.18	0.001	.0966098	.406927
8						
_cons	1.464868	.1180467	12.41	0.000	1.233501	1.696236
9						
_cons	2.357505	.133774	17.62	0.000	2.095313	2.619697
10						
_cons	-3.896594	.177419	-21.96	0.000	-4.244328	-3.548859

The marginal effect:

. estat mfx							
Pr(choice = 1 1 selected) = .41862592							
variable		dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
c							
	1	-1.62007	.045076	-35.94	0.000	-1.70841 -1.53172	.51844
	2	.38092	.016377	23.26	0.000	.348821 .413019	.54321
	3	.156526	.010709	14.62	0.000	.135537 .177515	1.015
	4	.359811	.016943	21.24	0.000	.326602 .393019	.43715
	5	.202435	.012376	16.36	0.000	.178178 .226691	.34528
	6	.04471	.005301	8.43	0.000	.034319 .0551	.78078
	7	.194866	.011804	16.51	0.000	.171731 .218001	.82509
	8	.12222	.008972	13.62	0.000	.104636 .139804	1.0774
	9	.14162	.009996	14.17	0.000	.122027 .161213	1.1894
	10	.016959	.002973	5.71	0.000	.011133 .022785	.56867
Pr(choice = 2 1 selected) = .13669617							
variable		dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
c							
	1	.38092	.016377	23.26	0.000	.348821 .413019	.51844
	2	-.785545	.030158	-26.05	0.000	-.844654 -.726436	.54321
	3	.051111	.003765	13.57	0.000	.043731 .058492	1.015
	4	.117491	.006448	18.22	0.000	.104853 .130129	.43715
	5	.066102	.004433	14.91	0.000	.057414 .07479	.34528
	6	.014599	.001779	8.20	0.000	.011112 .018087	.78078
	7	.063631	.004253	14.96	0.000	.055295 .071966	.82509
	8	.039909	.003145	12.69	0.000	.033744 .046074	1.0774
	9	.046244	.003507	13.18	0.000	.03937 .053118	1.1894
	10	.005538	.000986	5.62	0.000	.003605 .007471	.56867
Pr(choice = 3 1 selected) = .05617075							
variable		dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
c							
	1	.156526	.010709	14.62	0.000	.135537 .177515	.51844
	2	.051111	.003765	13.57	0.000	.043731 .058492	.54321
	3	-.352903	.02284	-15.45	0.000	-.397668 -.308137	1.015
	4	.048279	.003651	13.22	0.000	.041124 .055434	.43715
	5	.027162	.002319	11.71	0.000	.022618 .031707	.34528
	6	.005999	.000796	7.53	0.000	.004438 .00756	.78078
	7	.026147	.002223	11.76	0.000	.02179 .030504	.82509

#3&4

Propose a model for the effects of family income on demand. Write the likelihood and optimize the model.

```
Iteration 3:  log likelihood = -8236.757
```

Alternative-specific conditional logit	Number of obs	=	44,700
Case variable: num	Number of cases	=	4470

```
Alternative variable: price           Alts per case: min =      10
                                       avg =     10.0
                                       max =      10
```

Log likelihood = -8236.757	Wald chi2(9)	=	101.55
	Prob > chi2	=	0.0000

dum	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
1	(base alternative)						
2	income _cons	-.0030887 -.8453241	.003114 .0931354	-0.99 -9.08	0.321 0.000	-.009192 -1.027866	.0030145 -.662782
3	income _cons	.0145862 -2.399858	.0038255 .1335802	3.81 -17.97	0.000 0.000	.0070885 -2.66167	.022084 -2.138045
4	income _cons	.0040504 -1.201326	.0030926 .0971021	1.31 -12.37	0.190 0.000	-.0020109 -1.391643	.0101118 -1.01101
5	income _cons	-.0012536 -1.690582	.0042024 .1269952	-0.30 -13.31	0.765 0.000	-.0094901 -1.939488	.0069829 -1.441676
6	income _cons	.030612 -4.139767	.004674 .210989	6.55 -19.62	0.000 0.000	.0214512 -4.553298	.0397729 -3.726236
7	income _cons	-.0069326 -1.531042	.0044161 .1280434	-1.57 -11.96	0.116 0.000	-.015588 -1.782002	.0017228 -1.280081
8	income _cons	.0228862 -2.848352	.0036217 .1393848	6.32 -20.44	0.000 0.000	.0157878 -3.121541	.0299845 -2.575163
9	income _cons	.017743 -2.575597	.0037623 .13614	4.72 -18.92	0.000 0.000	.010369 -2.842427	.0251169 -2.308768
10	income _cons	.0107909 -4.28227	.01013 .345792	1.07 -12.38	0.287 0.000	-.0090636 -4.96001	.0306454 -3.60453

The marginal effect:

Pr(choice = 1|1 selected) = .39801715

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
casevars								
income	-.001062	.000487	-2.18	0.029	-.002016	-.000108		27.664

Pr(choice = 2|1 selected) = .15691817

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
casevars								
income	-.000904	.000378	-2.39	0.017	-.001645	-.000162		27.664

Pr(choice = 3|1 selected) = .05406295

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
casevars								
income	.000644	.000183	3.53	0.000	.000286	.001002		27.664

Pr(choice = 4|1 selected) = .13391688

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
casevars								
income	.000185	.000329	0.56	0.574	-.00046	.00083		27.664

Pr(choice = 5|1 selected) = .07089742

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
casevars								
income	-.000278	.000264	-1.06	0.291	-.000795	.000238		27.664

Pr(choice = 6|1 selected) = .01478443

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
casevars								
income	.000413	.000066	6.22	0.000	.000283	.000543		27.664

Pr(choice = 7|1 selected) = .07107045

--	--	--	--	--	--	--	--	--

The mixed logit model for the effects of family income on demand.

Integration points: 50 Wald chi2(10) = 1211.55
 Log simulated likelihood = -7352.5814 Prob > chi2 = 0.0000

dum		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
price							
	c	-7.547198	.2223547	-33.94	0.000	-7.983005	-7.111391
Normal							
	sd(c)	3.649194	.291882			3.119703	4.268552
1		(base alternative)					
2							
	income	-.0058937	.0036744	-1.60	0.109	-.0130955	.001308
	_cons	-.8886583	.1104599	-8.05	0.000	-1.105156	-.6721609
3							
	income	.0202033	.0046021	4.39	0.000	.0111834	.0292232
	_cons	.3624811	.2026931	1.79	0.074	-.0347902	.7597523
4							
	income	.0025691	.0034659	0.74	0.459	-.004224	.0093622
	_cons	-2.025266	.1126402	-17.98	0.000	-2.246037	-1.804495
5							
	income	-.003862	.0045853	-0.84	0.400	-.012849	.005125
	_cons	-3.383501	.1540533	-21.96	0.000	-3.68544	-3.081562
6							
	income	.0324638	.0049366	6.58	0.000	.0227883	.0421394
	_cons	-2.521205	.2228586	-11.31	0.000	-2.958	-2.08441
7							
	income	-.0061804	.0047636	-1.30	0.194	-.0155168	.003156
	_cons	.4032682	.1520697	2.65	0.008	.105217	.7013193
8							
	income	.0292799	.004687	6.25	0.000	.0200936	.0384662
	_cons	.0961353	.2292829	0.42	0.675	-.353251	.5455217
9							
	income	.0258758	.0051442	5.03	0.000	.0157934	.0359583
	_cons	.7081915	.2768591	2.56	0.011	.1655576	1.250825
10							
	income	.0065357	.010873	0.60	0.548	-.0147751	.0278465
	_cons	-4.412742	.3674048	-12.01	0.000	-5.132842	-3.692641

LR test vs. fixed parameters: **chibar2(01) = 130.70** Prob >= chibar2 = 0.0000

Estimate the model which is extracted choice 10.

Integration sequence: **Hammersley**
 Integration points: **50**
 Log simulated likelihood = **-7171.0302**

avg = **9.0**
 max = **9**
 Wald chi2(9) = **1146.94**
 Prob > chi2 = **0.0000**

dum	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
price						
c	-7.682409	.2324916	-33.04	0.000	-8.138084	-7.226733
Normal						
sd(c)	3.800923	.3038365			3.249721	4.445617
1	(base alternative)					
2						
income	-.0058322	.0036865	-1.58	0.114	-.0130577	.0013933
_cons	-.8990062	.1108781	-8.11	0.000	-1.116323	-.6816892
3						
income	.020762	.004655	4.46	0.000	.0116383	.0298857
_cons	.3657732	.2056176	1.78	0.075	-.0372298	.7687763
4						
income	.002387	.0034768	0.69	0.492	-.0044274	.0092015
_cons	-2.042462	.1133201	-18.02	0.000	-2.264566	-1.820359
5						
income	-.0045297	.0046257	-0.98	0.327	-.0135959	.0045364
_cons	-3.429965	.1567892	-21.88	0.000	-3.737266	-3.122664
6						
income	.0326799	.0049492	6.60	0.000	.0229796	.0423802
_cons	-2.508699	.2233818	-11.23	0.000	-2.94652	-2.070879
7						
income	-.0058333	.0047736	-1.22	0.222	-.0151894	.0035227
_cons	.4157664	.1529337	2.72	0.007	.1160218	.715511
8						
income	.0298916	.0047549	6.29	0.000	.0205722	.039211
_cons	.0940732	.2329927	0.40	0.686	-.3625841	.5507305
9						
income	.0266457	.0052373	5.09	0.000	.0163808	.0369106
_cons	.6893077	.2824188	2.44	0.015	.1357771	1.242838

LR test vs. fixed parameters: **chibar2(01) = 138.35** Prob >= chibar2 = **0.0000**

Do the hausman test

```
. hausman haust haust2, allegs constant
```

		—— Coefficients ——			
		(b) haust	(B) haust2	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
price	c	-7.547198	-7.682409	.1352107	.
Normal	sd(c)	3.649194	3.800923	-.1517296	.
2	income	-.0058937	-.0058322	-.0000615	.
	_cons	-.8886583	-.8990062	.0103479	.
3	income	.0202033	.020762	-.0005587	.
	_cons	.3624811	.3657732	-.0032921	.
4	income	.0025691	.002387	.0001821	.
	_cons	-2.025266	-2.042462	.0171966	.
5	income	-.003862	-.0045297	.0006678	.
	_cons	-3.383501	-3.429965	.0464637	.
6	income	.0324638	.0326799	-.0002161	.
	_cons	-2.521205	-2.508699	-.0125056	.
7	income	-.0061804	-.0058333	-.000347	.
	_cons	.4032682	.4157664	-.0124982	.
8	income	.0292799	.0298916	-.0006117	.
	_cons	.0961353	.0940732	.0020621	.
9	income	.0258758	.0266457	-.0007698	.
	_cons	.7081915	.6893077	.0188837	.

b = consistent under Ho and Ha; obtained from asmixlogit
 B = inconsistent under Ha, efficient under Ho; obtained from asmixlogit

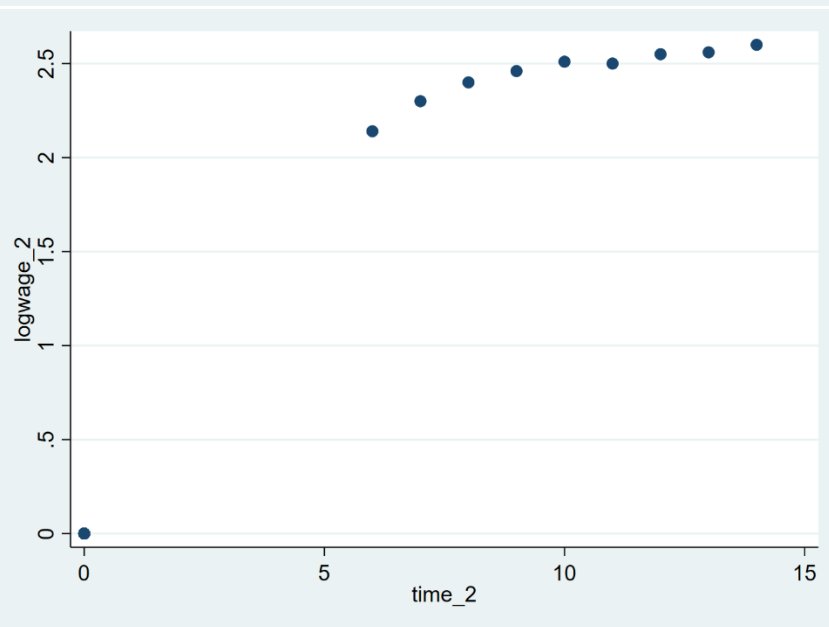
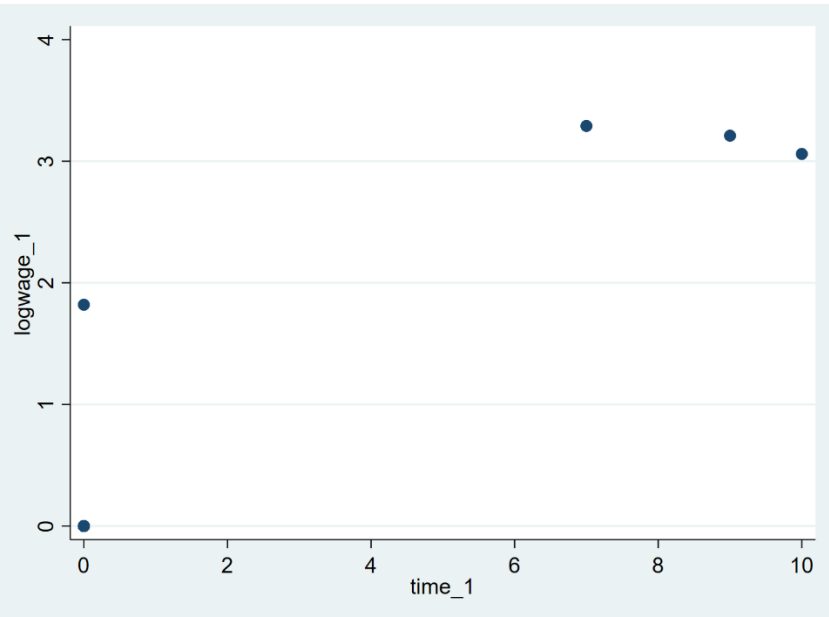
Test: Ho: difference in coefficients not systematic

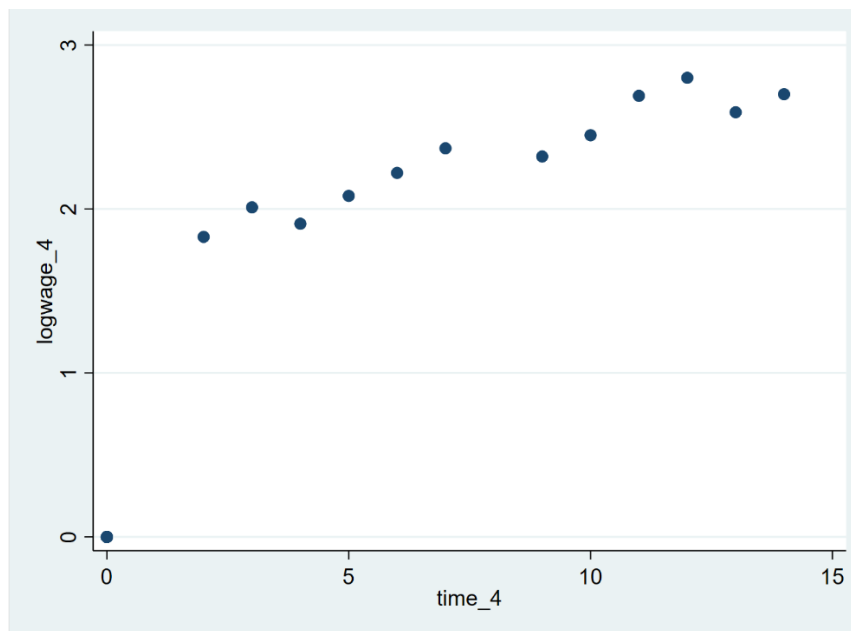
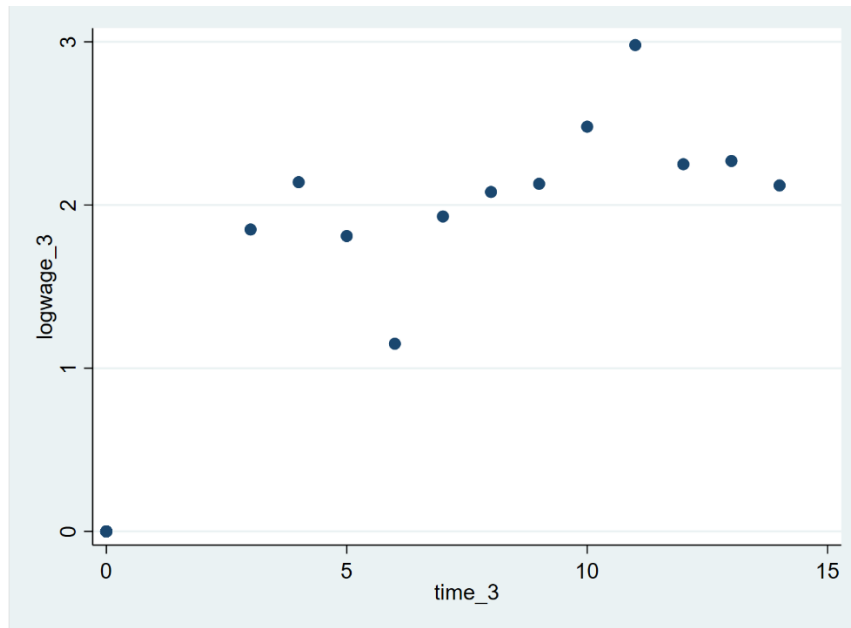
chi2(18) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = -5.59 chi2<0 ==> model fitted on these
 data fails to meet the asymptotic

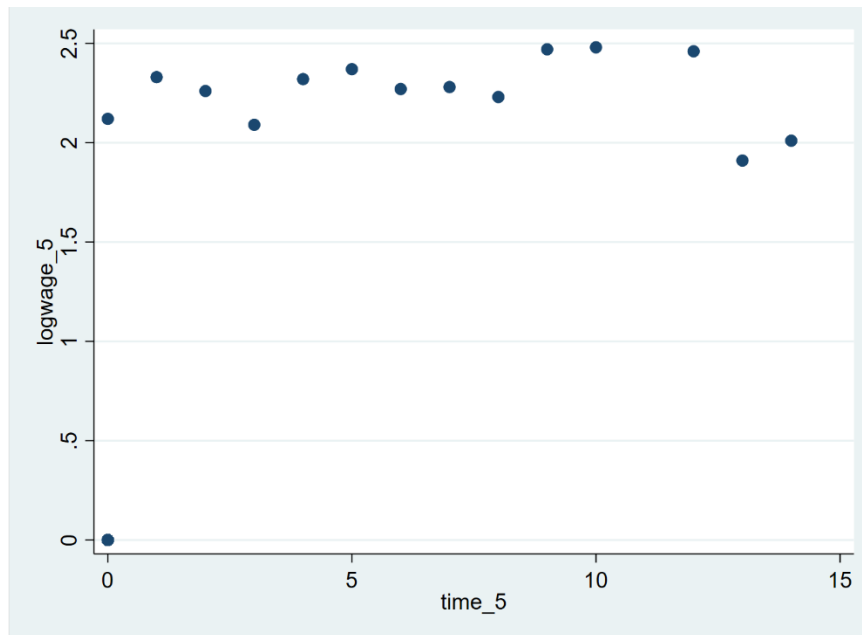
Exercise 4

#1

Represent the panel dimension of wages for 5 randomly selected individuals.







2#

Estimate the random effect model under the normality assumption of the disturbance terms.

```

Random-effects GLS regression              Number of obs   =    17,919
Group variable: personid                 Number of groups  =     2,178

R-sq:                                     Obs per group:
    within = 0.1961                        min =             1
    between = 0.1533                       avg  =            8.2
    overall  = 0.1578                       max =            15

corr(u_i, X)   = 0 (assumed)                Wald chi2(2)      =    4209.96
                                           Prob > chi2       =     0.0000

```

logwage	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.107938	.0033832	31.90	0.000	.1013071	.114569
potexper	.0387645	.0007178	54.00	0.000	.0373576	.0401714
_cons	.5635206	.0438846	12.84	0.000	.4775083	.6495328
sigma_u	.37207276					
sigma_e	.33545728					
rho	.5516129	(fraction of variance due to u_i)				

3#

Between Estimator:

Between regression (regression on group means) Number of obs = **17,919**
Group variable: **personid** Number of groups = **2,178**

R-sq: Obs per group:

within = 0.1962	min = 1
between = 0.1553	avg = 8.2
overall = 0.1518	max = 15

sd(u_i + avg(e_i.)) = **.3991313** F(2,2175) = **200.01**
Prob > F = **0.0000**

logwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.0930999	.0046685	19.94	0.000	.0839447	.1022551
potexper	.0259987	.0036049	7.21	0.000	.0189294	.0330681
_cons	.8455688	.0770179	10.98	0.000	.6945324	.9966052

Within Estimator:

Fixed-effects (within) regression Number of obs = **17,919**
Group variable: **personid** Number of groups = **2,178**

R-sq: Obs per group:

within = 0.1964	min = 1
between = 0.1550	avg = 8.2
overall = 0.1551	max = 15

corr(u_i, Xb) = **-0.1273** F(2,15739) = **1923.47**
Prob > F = **0.0000**

logwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.123662	.0057619	21.46	0.000	.1123681	.1349559
potexper	.0385611	.0007585	50.84	0.000	.0370744	.0400478
_cons	.4068016	.0717348	5.67	0.000	.2661931	.54741
sigma_u	.40290853					
sigma_e	.33545728					
rho	.59059603	(fraction of variance due to u_i)				

F test that all u_i=0: F(2177, 15739) = **9.95** Prob > F = **0.0000**

First time difference estimator:

Fixed-effects (within) regression
Group variable: **personid**

Number of obs = **15,741**
Number of groups = **2,095**

R-sq:
within = **0.0008**
between = **0.0010**
overall = **0.0010**

Obs per group:
min = **1**
avg = **7.5**
max = **14**

corr(u_i, Xb) = **-0.0017**

F(**2,13644**) = **5.32**
Prob > F = **0.0049**

logwage_D	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ_D	.0218726	.0097279	2.25	0.025	.0028047	.0409406
potexper_D	.0114422	.0047082	2.43	0.015	.0022133	.020671
_cons	.0419012	.0064347	6.51	0.000	.0292883	.0545141
sigma_u	.1752942					
sigma_e	.39181966					
rho	.16677308	(fraction of variance due to u_i)				

F test that all u_i=0: F(**2094, 13644**) = **0.49**

Prob > F = **1.0000**