UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN

REVISIÓN DE ECONOMETRÍA I
15 de diciembre de 2011 – 18 horas

EJERCICIO 1 (30 puntos) –

(El ejercicio consta de 3 partes, cada una independiente de las otras)

Contando con datos de corte transversal sobre salarios y otras características individuales, se desea estimar una ecuación de salarios. Para ello se dispone de información sobre 1000 trabajadores norteamericanos en 1997 (Fuente: *Current Population Survey*; tomado de Hill, Griffiths y Lim, 2006).

PARTE A

Considerando las siguientes variables:

WAGE: Salario corriente por hora, en dólares
EDUC: Años de educación formal
EXPER: Años de experiencia laboral del trabajador
EXPER2: Variable EXPER al cuadrado
FEMALE: Binaria que toma el valor 1 si la observación corresponde a mujer

Se estima el Modelo 1:

$$WAGE = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 EXPER2 + \beta_5 FEMALE + \varepsilon$$

Source	SS	df	MS
Model	12180.832	4	3045.208
Residual	26800.6652	995	26.9353419
Total	38981.4972	999	39.0205177

Number of obs	=	1000
F(4, 995)	=	113.06
Prob > F	=	0.0000
R-squared	=	0.3125
Adj R-squared	=	0.3097
Root MSE	=	5.1899

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	1.19825	.0682586	17.55	0.000	1.064303	1.332198
exper	.3427591	.0499702	6.86	0.000	.2447	.4408181
exper2	-.0051238	.0011639	-4.40	0.000	-.0074078	-.0028399
female	-2.546047	.3283586	-7.75	0.000	-3.190402	-1.901692
_cons	-8.422098	1.040664	-8.09	0.000	-10.46425	-6.37995

Se pide 1) El Modelo 1 recoge cierto tipo de discriminación salarial por sexo. Interprete el coeficiente asociado a la variable “female”.

Por otra parte, se sospecha que el comportamiento de los salarios es diferente en áreas metropolitanas y rurales, por lo que se define la siguiente variable:

METRO: Binaria que toma el valor 1 si la observación corresponde al área metropolitana (808 casos) y el valor 0 si corresponde al área rural (192 casos)

Para comprobar la intuición, se estima el mismo modelo separando las observaciones del área rural de las observaciones del área metropolitana, obteniéndose las siguientes salidas:

1a) Área metropolitana:

$$WAGE = \beta_1^M + \beta_2^M EDUC + \beta_3^M EXPER + \beta_4^M EXPER2 + \beta_5^M FEMALE + \varepsilon^M$$

Source	SS	df	MS
Model	10851.934	4	2712.98349
Residual	23682.3416	803	29.4923307
Total	34534.2755	807	42.7934021

Number of obs	= 808
F(4, 803)	= 91.99
Prob > F	= 0.0000
R-squared	= 0.3142
Adj R-squared	= 0.3108
Root MSE	= 5.4307

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	1.212376	.0776569	15.61	0.000	1.059942	1.36481
exper	.3885087	.0583417	6.66	0.000	.2739886	.5030289
exper2	-.0061527	.0013694	-4.49	0.000	-.0088408	-.0034646
female	-2.569077	.3828345	-6.71	0.000	-3.320551	-1.817602
_cons	-8.699911	1.189145	-7.32	0.000	-11.03411	-6.365712

1b) Área rural:

$$WAGE = \beta_1^R + \beta_2^R EDUC + \beta_3^R EXPER + \beta_4^R EXPER2 + \beta_5^R FEMALE + \varepsilon^R$$

Source	SS	df	MS
Model	1211.6449	4	302.911225
Residual	2674.92219	187	14.3043967
Total	3886.56709	191	20.3485188

Number of obs	= 192
F(4, 187)	= 21.18
Prob > F	= 0.0000
R-squared	= 0.3118
Adj R-squared	= 0.2970
Root MSE	= 3.7821

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.9847127	.1295203	7.60	0.000	.729204	1.240221
exper	.1501267	.0826382	1.82	0.071	-.0128963	.3131498
exper2	-.0007924	.0018695	-0.42	0.672	-.0044805	.0028956
female	-2.091492	.5527063	-3.78	0.000	-3.181832	-1.001151
_cons	-5.484085	1.913387	-2.87	0.005	-9.258684	-1.709486

Se pide 2) De acuerdo con ambas salidas, ¿afirmaría Ud. que hay diferencias entre el área rural y el área metropolitana? Justifique su respuesta objetivamente, planteando el contraste que considere adecuado (presente hipótesis nula y alternativa, estadístico de prueba, su distribución bajo H_0, región crítica y concluya con un nivel de significación del 5%)

PARTE B

A la vista de los resultados anteriores, se resuelve incluir la variable "METRO" en el Modelo 1. Por otra parte, como es habitual en la estimación de ecuaciones de salario se sospecha la existencia de heteroscedasticidad, por lo que se decide transformar la variable dependiente en logaritmos, obteniéndose el Modelo 2:

$$Ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 EXPER2 + \beta_5 FEMALE + \beta_6 METRO + \varepsilon$$

Source	SS	df	MS
Model	109.674497	5	21.9348995
Residual	195.61444	994	.196795211
Total	305.288937	999	.305594532

Number of obs	= 1000
F(5, 994)	= 111.46
Prob > F	= 0.0000
R-squared	= 0.3592
Adj R-squared	= 0.3560
Root MSE	= .44362

lnwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.1076051	.0058445	18.41	0.000	.0961361	.119074
exper	.0365293	.0042713	8.55	0.000	.0281475	.044911
exper2	-.0005814	.0000995	-5.84	0.000	-.0007766	-.0003861
female	-.2510979	.0280892	-8.94	0.000	-.306219	-.1959769
metro	.1086895	.0357312	3.04	0.002	.0385724	.1788067
_cons	.3669512	.0924704	3.97	0.000	.1854917	.5484107

Se pide 3) Exponga las consecuencias que generaría la presencia de heteroscedasticidad sobre las propiedades de los estimadores MCO.

Se pide 4) Explique (justificando) con qué objetivo se toma la variable WAGE en logaritmos

Para descartar (o no) la presencia de heteroscedasticidad, se conservan los residuos del Modelo 2, y a partir de sus cuadrados ("RESIDLN2") se estima la siguiente regresión: auxiliar:

Source	SS	df	MS
Model	4.61374244	17	.271396614
Residual	89.2721991	982	.090908553
Total	93.8859415	999	.093979921

Number of obs =	1000
F(17, 982) =	2.99
Prob > F =	0.0000
R-squared =	0.0491
Adj R-squared =	0.0327
Root MSE =	.30151

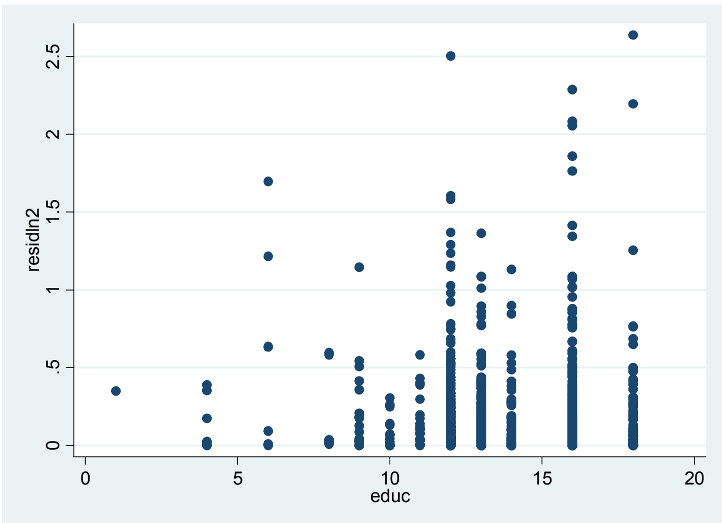
residln2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0096057	.0315374	0.30	0.761	-.0522827	.0714941
exper	-.0118814	.0207183	-0.57	0.566	-.0525386	.0287758
exper2	.0003162	.001124	0.28	0.778	-.0018895	.002522
female	-.0823372	.1258173	-0.65	0.513	-.3292389	.1645644
metro	-.0078853	.1682365	-0.05	0.963	-.3380297	.322259
educ2	.0000753	.000919	0.08	0.935	-.0017281	.0018788
exper4	-1.69e-07	4.39e-07	-0.39	0.700	-1.03e-06	6.92e-07
educ_exper	.0006362	.0013341	0.48	0.634	-.0019818	.0032543
educ_exper2	-.0000384	.0000311	-1.24	0.217	-.0000994	.0000226
educ_female	.0132717	.0080587	1.65	0.100	-.0025425	.0290859
educ_metro	-.0024129	.0113729	-0.21	0.832	-.0247308	.019905
exper3	.0000147	.0000392	0.37	0.708	-.0000622	.0000915
exper_female	.0026011	.0058965	0.44	0.659	-.0089701	.0141724
exper_metro	.015076	.007358	2.05	0.041	.0006367	.0295152
exper2_fem~e	-.0001346	.0001381	-0.98	0.330	-.0004056	.0001363
exper2_metro	-.0002838	.000168	-1.69	0.091	-.0006134	.0000459
female_metro	-.0985429	.0492286	-2.00	0.046	-.1951482	-.0019375
_cons	-.0105208	.300662	-0.03	0.972	-.6005347	.5794931

Se pide 5) En base a la salida de la regresión auxiliar, realice la prueba de detección de heteroscedasticidad. Plantee con qué nombre se conoce la prueba en cuestión, la hipótesis nula y la alternativa. Plantee el estadístico utilizado en la prueba, la distribución del mismo bajo H_0 y la región crítica a un nivel de significación de 5%. Concluya.

PARTE C

Tras analizar distintas alternativas, se resuelve estimar el modelo por MCGF, para lo cual se realiza una inspección gráfica y se opta por asumir que la varianza de la perturbación se incrementa con los años de educación del siguiente modo:

$$E(\varepsilon\varepsilon'|X) = \sigma_\varepsilon^2 diag(EDUC_i^2)$$



Se pide 6) Explique en qué consiste el método de MCGF. Para ello presente la fórmula matricial del estimador $\widehat{\beta}_{MCGF}$, explicitando la forma de las matrices utilizadas. Explique por qué podría considerarse que esta

estimación implica ponderar cada observación por la inversa del desvío estándar esperado de su perturbación.

Se pide 7) ¿Qué otra posibilidad conoce para realizar la estimación del modelo sorteando el problema de la heteroscedasticidad detectado en la Parte B? Explíquela conceptualmente.

EJERCICIO 2 (20 puntos) –

Se desea estimar el modelo:

$$\text{[E]}\ Y_t = \beta X_{1t} + \varepsilon_t,\ \varepsilon_t \sim \text{iid}(0;\sigma^2)\ \ (\beta \neq 1)$$

En forma matricial, el modelo es: $\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1n} \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

Se sabe que, en realidad, X_{1t} se determina simultáneamente con Y_t, de acuerdo con la relación: $X_{1t} = Y_t + X_{2t}$, donde $E(X_{2t}\varepsilon_t) = 0, \forall t$.

Se pide:

a) Demostrar que $E(X_{1t}\varepsilon_t) = (1-\beta)^{-1}\sigma^2, \forall t$

b) ¿Qué ocurre con las propiedades del estimador MCO de β al cumplirse la relación demostrada en **a)**?

c) Escriba explícitamente la expresión de un estimador de β alternativo al de MCO, <u>para este caso concreto</u>. Justifique su elección.

d) Si se dispone de una muestra de 60 observaciones a partir de la cual se han obtenido los siguientes productos mixtos:

Obtenga la estimación de β por el método propuesto en **c)** y por le método de los MCO.

(Ejemplo: $\sum_{t=1}^{t=60} Y_t X_{1t} = 40$)

	Y_t	X_{1t}	X_{2t}
Y_t	100	40	-60
X_{1t}		80	40
X_{2t}			100