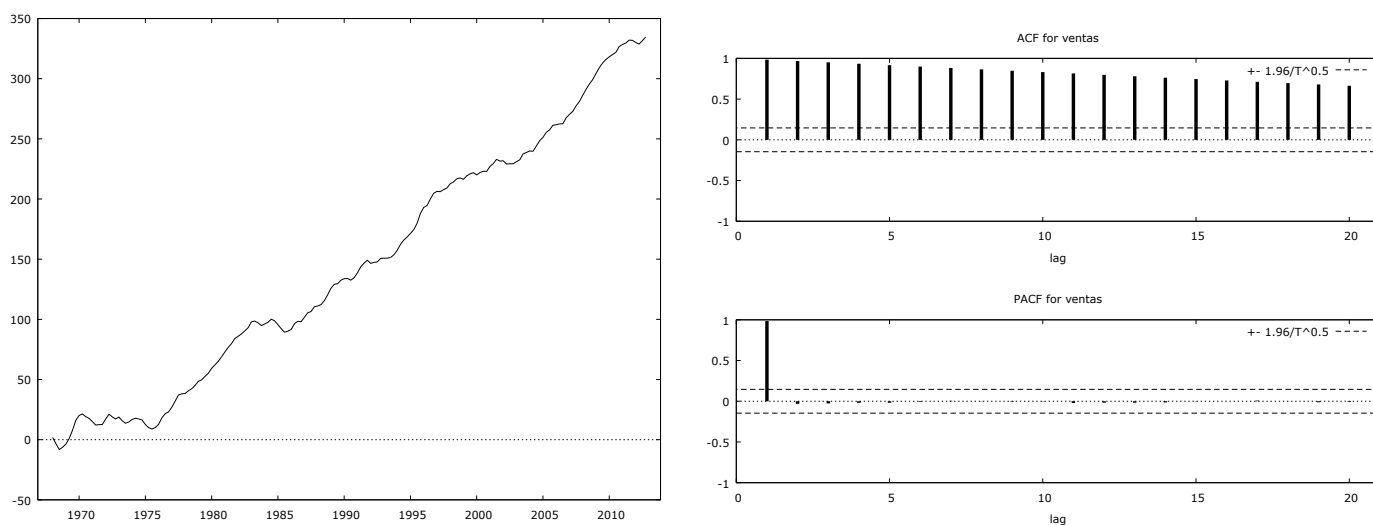


EXAMEN DE ECONOMETRÍA II
25 de julio de 2014

Ejercicio 1 (35 puntos)

Se cuenta con la siguiente serie de datos sobre las ventas trimestrales entre 1968.1 y 2012.4 (180 observaciones) de una empresa que se quiere modelizar utilizando la metodología Box-Jenkins vista en clase.



Parte I

- 1.1 Defina que entiende por estacionariedad en sentido débil. Comente la gráfica en niveles de las ventas y el correlograma correspondiente. ¿Qué puede mencionar respecto a la estacionariedad de la serie? ¿El gráfico le sugiere alguna estructura posible para el proceso generador de datos?
- 1.2 Se realiza el contraste de Dickey-Fuller aumentado utilizando modelos alternativos, a continuación se presenta la información correspondiente a dichas salidas.
- a) Plantee la regresión utilizada en el contraste para la especificación correspondiente al Modelo A, especifique hipótesis nula y alternativa y estadístico de prueba.
 - b) Concluya respecto al contraste de Dickey-Fuller al 5% de significación, indicando qué modelo A, B o C considera apropiado para describir el proceso generador de datos (en cada etapa especifique las hipótesis contraste, estadístico de prueba, región crítica, conclusión).

Modelo A

Augmented Dickey-Fuller regression
 OLS, using observations 1968:3-2012:4 (T = 178)
 Dependent variable: d_ventas

	coefficient	std. error	t-ratio	p-value
ventas_1	0.00401408	0.00106054	3.785	1.0000
d_ventas_1	0.590770	0.0606297	9.744	3.26e-018 ***

AIC: 759.862 BIC: 766.225 HQC: 762.442

Modelo B

Augmented Dickey-Fuller regression
 OLS, using observations 1968:3-2012:4 (T = 178)
 Dependent variable: d_ventas

	coefficient	std. error	t-ratio	p-value
const	0.740203	0.271779	2.724	0.0071 ***
ventas_1	0.00103335	0.00151094	0.6839	0.9919
d_ventas_1	0.544801	0.0618992	8.801	1.27e-015 ***

AIC: 754.472 BIC: 764.018 HQC: 758.343

Modelo C

Augmented Dickey-Fuller regression
 OLS, using observations 1969:3-2012:4 (T = 174)
 Dependent variable: d_ventas

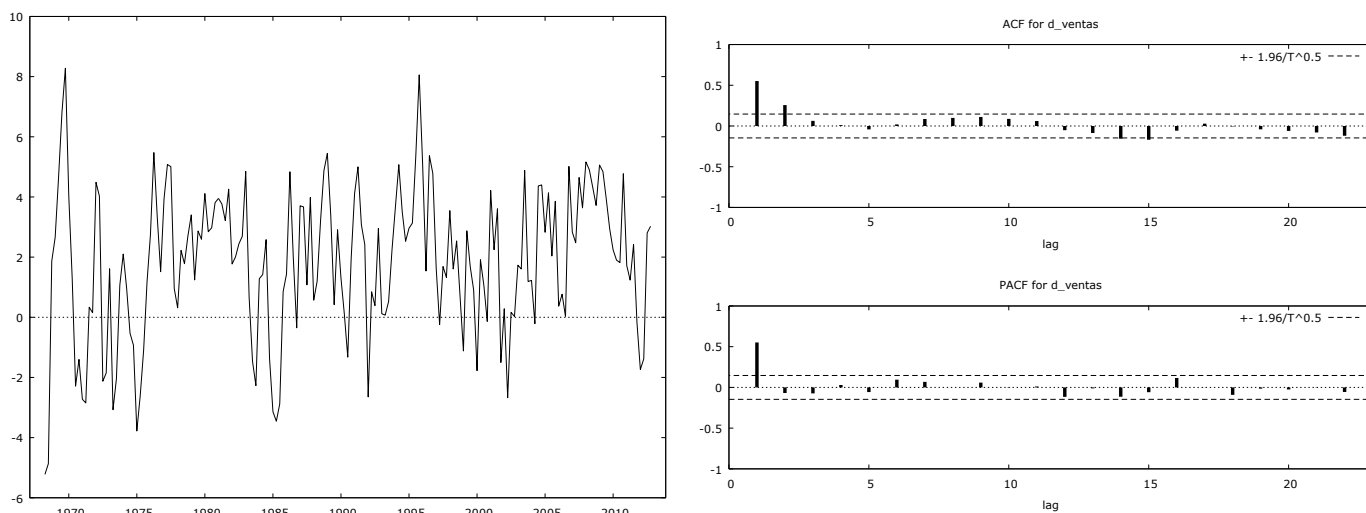
	coefficient	std. error	t-ratio	p-value
const	-7.77316	3.41834	-2.274	0.0242 **
ventas_1	-0.0288045	0.0123667	-2.329	0.4174
d_ventas_1	0.563960	0.0758618	7.434	5.32e-012 ***
d_ventas_2	0.0172816	0.0876317	0.1972	0.8439
d_ventas_3	-0.0853823	0.0871592	-0.9796	0.3287
d_ventas_4	0.0826365	0.0864334	0.9561	0.3404
d_ventas_5	-0.0619098	0.0749614	-0.8259	0.4101
time	0.0606703	0.0242267	2.504	0.0132 **

AIC: 734.671 BIC: 759.943 HQC: 744.923

Modelo	Hipótesis nula	Est.	Valores críticos T=50		Valores críticos T=100		Valores críticos asintóticos	
			95%	99%	95%	99%	95%	99%
$\Delta y_t = \gamma_a y_{t-1} + \varepsilon_t$	$\gamma_a = 0$	τ	-1,95	-2,62	-1,95	-2,6	-1,95	-2,58
$\Delta y_t = \alpha_b + \gamma_b y_{t-1} + \varepsilon_t$	$\gamma_b = 0$	τ_μ	-2,93	-3,58	-2,89	-3,51	-2,86	-3,43
	$\alpha_b = \gamma_b = 0$	ϕ_1	4,86	7,06	4,71	6,7	4,59	6,43
	$\alpha_b = 0$ dado $\gamma_b = 0$	$\tau_{\alpha\mu}$	2,56	3,28	2,54	3,22	2,52	3,18
$\Delta y_t = \alpha_c + \beta_c t + \gamma_c y_{t-1} + \varepsilon_t$	$\gamma_c = 0$	τ_τ	-3,5	-4,15	-3,45	-4,04	-3,41	-3,96
	$\alpha_c = \beta_c = \gamma_c = 0$	ϕ_2	5,13	7,02	4,88	6,5	4,68	6,09
	$\beta_c = \gamma_c = 0$	ϕ_3	6,73	9,31	6,49	8,73	6,25	8,27
	$\beta_c = 0$ dado $\gamma_c = 0$	$\tau_{\beta\tau}$	2,81	3,60	2,79	3,53	2,78	3,46
	$\alpha_c = 0$ dado $\gamma_c = 0$	$\tau_{\alpha\tau}$	3,14	3,87	3,11	3,78	3,08	3,71

Parte II

Se realiza la primer diferencia de la serie ventas y se presentan a continuación el gráfico correspondiente, su correlograma y un test ADF.



```
Augmented Dickey-Fuller test for d_ventas
including 5 lags of (1-L)d_ventas
(max was 3, criterion modified AIC)
sample size 175
unit-root null hypothesis: a = 1
```

```
test without constant
model: (1-L)y = (a-1)*y(-1) + ... + e
1st-order autocorrelation coeff. for e: -0.009
lagged differences: F(3, 171) = 1.246 [0.2946]
estimated value of (a - 1): -0.214586
test statistic: tau_nc(1) = -3.5042
asymptotic p-value 0.000451
```

2.1 ¿Qué puede concluir respecto a la primer diferencia de la serie ventas? ¿Y sobre la serie en nivel?

2.2 En función de lo observado en el correlograma de la primer diferencia de ventas se proponen tres modelos. A continuación se reporta las respectivas estimaciones e información seleccionada sobre el estadístico Q del contraste de Ljung-Box de sus residuos. De considerar la información sobre los residuos relevante, especifique hipótesis nula y alternativa, estadístico de prueba concluya al respecto. Con toda esta información disponible, ¿qué modelo ARIMA(p,d,q) sugeriría para la serie ventas?

Model 1: ARMA(1,2), using observations 1968:3-2012:4 (T = 178)
Dependent variable: d_ventas

	<i>Coefficient</i>	<i>Std. Error</i>	<i>z</i>	<i>p-value</i>	
Const	1.02925	0.275281	3.7389	0.00018	***
phi_1	0.46875	0.119583	3.9199	0.00009	***
theta_1	0.0988948	0.14132	0.6998	0.48406	
theta_2	0.0810543	0.0982401	0.8251	0.40934	
Mean dependent var	1.898430	S.D. dependent var		2.412073	
Mean of innovations	0.002345	S.D. of innovations		1.979705	
Log-likelihood	-374.1358	Akaike criterion		758.2716	
Schwarz criterion	774.1805	Hannan-Quinn		764.7231	

Model 2: ARMA(1,1), using observations 1968:3-2012:4 (T = 178)
Dependent variable: d_ventas

	<i>Coefficient</i>	<i>Std. Error</i>	<i>z</i>	<i>p-value</i>	
Const	0.899745	0.236105	3.8108	0.00014	***
phi_1	0.539041	0.10095	5.3397	<0.00001	***
theta_1	0.0223417	0.137255	0.1628	0.87070	
Mean dependent var	1.898430	S.D. dependent var		2.412073	
Mean of innovations	0.000391	S.D. of innovations		1.983211	
Log-likelihood	-374.4508	Akaike criterion		756.9015	
Schwarz criterion	769.6286	Hannan-Quinn		762.0627	

Model 3: ARMA(1,0), using observations 1968:3-2012:4 (T = 178)
Dependent variable: d_ventas

	<i>Coefficient</i>	<i>Std. Error</i>	<i>z</i>	<i>p-value</i>	
Const	0.874838	0.187094	4.6759	<0.00001	***
phi_1	0.552667	0.0607293	9.1005	<0.00001	***
Mean dependent var	1.898430	S.D. dependent var		2.412073	
Mean of innovations	-1.05e-16	S.D. of innovations		1.994706	
Log-likelihood	-374.4738	Akaike criterion		752.9476	
Schwarz criterion	759.3112	Hannan-Quinn		755.5282	

	Q-stat. [p-value]					
	Modelo 1		Modelo 2		Modelo 3	
1	0.0046	[0.946]	0.0002	[0.989]	0.0240	[0.877]
6	2.3658	[0.883]	3.3844	[0.759]	3.3398	[0.765]
12	6.3359	[0.898]	7.2734	[0.839]	7.1315	[0.849]
18	12.2627	[0.833]	13.0568	[0.788]	13.0059	[0.791]
24	13.2625	[0.926]	14.0355	[0.900]	13.9638	[0.903]

Ejercicio 2 (35 puntos)

Tomando datos de una encuesta realizada por la revista *Psychology Today* a 610 personas, se busca encontrar cómo afectan diferentes factores a la infidelidad dentro de un matrimonio. Para ello se cuenta con las siguientes variables:

- **nrelac:** cantidad de relaciones extramatrimoniales en el año anterior
- **hombre:** vale 1 si la persona encuestada es hombre
- **edad:** edad en años
- **anioscdo:** años de matrimonio
- **hijos:** vale 1 si la persona encuestada tiene hijos
- **rankmat:** valoración subjetiva de la calidad del matrimonio que hace la persona encuestada, varía entre 1 y 5, siendo 5 la mejor
- **relig:** refleja la religiosidad de la persona encuestada, varía entre 1 y 5 siendo 5 el mayor grado de práctica religiosa
- **educ:** años de educación formal

Estadísticas Descriptivas

Número de relaciones en el año	Freq.	Percent	Cum.
0	451	75.04	75.04
1	34	5.66	80.70
2	17	2.83	83.53
3	19	3.16	86.69
7	42	6.99	93.68
12	38	6.32	100.00
Total	601	100.00	

. sum hombre edad anioscdo hijos rankmat relig educ

Variable	Obs	Mean	Std. Dev.	Min	Max
hombre	601	.4758735	.4998336	0	1
edad	601	32.48752	9.288762	17.5	57
anioscdo	601	8.177696	5.571303	.125	15
hijos	601	.7154742	.4515641	0	1
rankmat	601	3.93178	1.103179	1	5
relig	601	3.116473	1.167509	1	5
educ	601	16.16639	2.402555	9	20

Parte I

1. Analice brevemente las estadísticas descriptivas.
2. ¿Entiende que la variable “nrelac” puede presentar algún tipo de censura o truncamiento? Explique
3. Si entiende que se presenta un problema de censura o truncamiento, define qué tipo de censura o truncamiento se observa.

Parte II

Se estima un modelo Tobit para la variable dependiente *nrelac*

```
. *Modelos 3  
. tobit nrelac hombre edad anioscdo hijos rankmat relig educ, r ll(0)
```

```
Tobit regression               Number of obs   =       601  
                               F(      7,      594) =      10.68  
                               Prob > F       =      0.0000  
Log pseudolikelihood = -704.9511   Pseudo R2      =      0.0534
```

nrelac	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
hombre	1.1831	.9802659	1.21	0.228	-.7421091	3.108308
edad	-.1904162	.0887664	-2.15	0.032	-.3647503	-.0160822
anioscdo	.537995	.1469876	3.66	0.000	.2493164	.8266736
hijos	.8984287	1.330698	0.68	0.500	-1.715017	3.511875
rankmat	-2.289961	.3923566	-5.84	0.000	-3.060536	-1.519386
relig	-1.709846	.4069273	-4.20	0.000	-2.509038	-.9106551
educ	.092386	.2025922	0.46	0.649	-.3054982	.4902701
_cons	7.365336	4.26185	1.73	0.084	-1.004792	15.73546
/sigma	8.270711	.4558289			7.375479	9.165943

```
obs. summary:      451 left-censored observations at nrelac<=0  
                   150 uncensored observations  
                   0 right-censored observations
```

1. Derive la función de verosimilitud del modelo.
2. Comente los resultados del modelo
3. En la página siguiente se presentan los efectos marginales asociados al modelo Tobit
 - a. Explique cuáles son los diferentes tipos de efectos marginales que se pueden obtener en modelos y relaciónelos con las salidas presentadas debajo
 - b. Señale las diferencias en las interpretaciones de cada efecto marginal (ilustre con fórmulas donde le parezca más relevante).
 - c. Explique cuál efecto marginal le parece más interesante en este problema en particular.
 - d. Comente los resultados obtenidos en los efectos marginales que a usted le resulten más relevantes para el problema bajo análisis.

Parte III

Luego de estimados los modelos anteriores, se plantea la posibilidad de que exista un problema de sesgo de selección que esté afectando los resultados.

1. Explique qué se entiende por sesgo de selección en el marco de modelos con variable dependiente limitada y cuál sería la interpretación de un sesgo de este tipo en este ejercicio.
2. Proponga una estrategia de estimación alternativa que permita corregir este sesgo, detallando los pasos a realizar.
3. Explique qué se entiende por restricción de exclusión y proponga una variable que podría utilizarse en este caso para cubrir esa restricción. ¿Necesita información extra para asegurar que la variable propuesta es correcta?

Pregunta 1 (10 puntos)

Considere el siguiente modelo de regresión:

$$y_i = \alpha + \beta x_i^* + u_i \text{ donde se cumple que } \text{cov}(x_i^*, u_i) = 0$$

Se tiene una muestra $\{y_i, x_i, w_i\}_{i=1, \dots, N}$ siendo

$$x_i = x_i^* + \varepsilon_i \quad w_i = x_i^* + v_i$$

Se supone $\text{cov}(\varepsilon_i, u_i) = 0$ $\text{cov}(v_i, u_i) = 0$ y $\text{cov}(\varepsilon_i, v_i) = 0$ $\text{cov}(\varepsilon_i, x_i^*) = 0$ $\text{cov}(v_i, x_i^*) = 0$

- 1) Encuentre el sesgo que existe al estimar Beta en una regresión de y utilizando como regresor x. ¿En general, será dicho sesgo de igual magnitud al que se producirá cuando se utiliza w como regresor en lugar de x? Justifique
- 2) Explique rigurosamente si es posible obtener a partir de los datos disponibles una estimación consistente del parámetro Beta. En caso afirmativo explique brevemente como procedería para obtener dicha estimación.

Pregunta 2 (10 puntos)

Considere el siguiente modelo con variable dependiente binaria

$$y_i^* = \beta_0 + \beta_1 x_i + \beta_2 D_i + \beta_3 D_i x_i + u_i$$

$$y_i = 1(y_i^* > 0)$$

1) Siendo x una variable continua y D una variable binaria. Obtenga la formula de los efectos parciales asociados a la variable x:

a) para los individuos para los cuales D=1

b) el individuo "promedio" de la muestra

Suponga alternativamente que el modelo es: a) el modelo de probabilidad lineal; b) el modelo logit.

2) Explique las diferencias y señale en qué condiciones los dos efectos son similares.

Pregunta 3 (10 puntos)

Sean:

$$y_t = y_{t-1} + e_t \text{ con } e_t \sim iid(0, \sigma_e^2) \text{ y } \text{cov}(e_t, e_s) = 0 \forall t \neq s$$

$$x_t = x_{t-1} + u_t \text{ con } u_t \sim iid(0, \sigma_u^2) \text{ y } \text{cov}(u_t, u_s) = 0 \forall t \neq s$$

- 1) ¿Qué tipo de procesos son y_t y x_t . ¿Son estacionarios? Justifique su respuesta.
- 2) Se estima la regresión $y_t = \alpha + \beta x_t + \varepsilon_t$, $t=1, 2, \dots, T$ La estimación resulta en que el valor del estadístico t de la prueba de significación individual del parámetro Beta es 2.8.
 - a) ¿Qué es lo que cabe esperar de la hipótesis $\beta=0$?
 - b) En base a la información disponible ¿podemos concluir que estamos frente a una regresión espúrea? ¿y a una relación de cointegración? Justifique rigurosa y detalladamente.