

ECONOMETRÍA

30/07/2020

Junghanss, Juan Cruz

El archivo Word contiene todas las resoluciones correspondientes a cada inciso, junto con desarrollos teóricos, gráficos y complementos visuales a partir de Stata. Está pensado para su lectura sin la necesidad de correr el archivo do de Stata.

EJERCICIO Nº 1: Utilizando los datos del archivo GPA1.dta, estime por MCO un modelo que relacione el promedio de calificaciones en la universidad del estudiante i ($colGPA_i$) con el promedio de calificaciones de ese estudiante en la escuela secundaria ($hsGPA_i$), calificación en el examen de ingreso (ACT_i), la cantidad de clases a las que faltó ($skipped_i$) y una variable dummy que indica si el estudiante posee computadora personal (PC_i).

1. Calcule el caso especial de la prueba de White para heterocedasticidad. En la regresión de \hat{u}_i^2 sobre $colGPA_i$ y $colGPA_i^2$, obtenga los valores ajustados, es decir, \hat{h}_i .

Primero debemos regresar el modelo original

$$colGPA = \beta_0 + \beta_1(hsGPA) + \beta_2(ACT) + \beta_3(skipped) + \beta_4(PC) + u$$

para extraer sus residuos y predicción lineal:

```
reg colGPA hsGPA ACT skipped PC
```

```
predict u1,r
```

```
predict y1,xb
```

Uno de los métodos de detección numérica de la heterocedasticidad es el contraste de White. Para el caso especial de la prueba de White empleamos la variable explicada, es decir, su predicción lineal (y_1), esta elevada al cuadrado ($gen\ y1_2 = y1^2$) y los errores del modelo original elevados al cuadrado también ($gen\ u1_2 = u1^2$). La intuición es ver el comportamiento de las variables independientes y sus términos no lineales o de interacción con los errores del modelo original. Lo que sería u^2 explicado por \hat{y}, \hat{y}^2 . Una vez regresados los errores contra las variables ($reg\ u1_2\ y1\ y1_2$), podemos extraer de allí los valores ajustados ($predict\ h1,xb$), que estarán representados como $h1$. Si mostramos la predicción lineal ($di\ h1$) de esta última regresión tendríamos un valor positivo de 0.054

```
. di h1
```

```
.05444054
```

Procediendo con la computación del test de White armamos el estadístico de prueba LM, que consta de multiplicar el tamaño muestral por el R cuadrado [$scalar\ W = e(r^2) * e(N)$] y lo contrastamos contra una inversa chi cuadrado a dos colas [$invchi2tail(e(df_m),0.05)$], que nos arrojará el siguiente resultado:

```
. scalar W = e(r2)*e(N)
```

```
. di "estadistico W " W
```

```
estadistico W 6.9575701
```

```
. di "valor critico " invchi2tail(e(df_m),0.05)
```

```
valor critico 5.9914645
```

Para un nivel de confianza de 95%, con un valor crítico de 5.991, el estadístico LM de White cae en zona de rechazo, ya que este es mayor que el critical value ($5.991 < 6.957$). Dicho esto,

debemos optar por rechazar la hipótesis nula que plantea que NO hay heterocedasticidad (es decir, que el modelo es homoscedástico) en favor de que si estamos en presencia de heterocedasticidad.

2. Verifique que los valores ajustados del inciso 1) sean todos estrictamente positivos. Después obtenga las estimaciones de MCP usando como ponderadores $1/h_i$. Compare las estimaciones de MCP con las correspondientes estimaciones de MCO para el efecto de número de clases perdidas y el efecto de poseer una computadora. ¿Qué puede decir de su significancia estadística?

En primer lugar, constatamos que los valores ajustados de la consigna (1) son positivos, menos el valor del intercepto. Además, la predicción h_1 posee un valor de 0.054 como demostramos anteriormente.

ul_2	Coef.
y1	.1296036
y1_2	.0029456
_cons	-.3218436

Utilizando la predicción lineal h_1^{-1} como ponderador (*aweight*) en el método de MCP obtenemos las siguientes estimaciones :

(código en Stata: `reg colGPA hsGPA ACT skipped PC[aweight = 1/h1]`)

```
. reg colGPA hsGPA ACT skipped PC[aweight=1/h1]
(sum of wgt is 1.5232e+03)
```

Source	SS	df	MS	Number of obs = 141		
Model	5.70396891	4	1.42599223	F(4, 136) = 15.01		
Residual	12.9238237	136	.095028116	Prob > F = 0.0000		
Total	18.6277926	140	.133055662	R-squared = 0.3062		
				Adj R-squared = 0.2858		
				Root MSE = .30827		

colGPA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hsGPA	.4025059	.0833616	4.83	0.000	.2376532	.5673586
ACT	.013162	.0098273	1.34	0.183	-.0062721	.0325962
skipped	-.0763653	.0221727	-3.44	0.001	-.1202131	-.0325174
PC	.1260048	.0563391	2.24	0.027	.0145909	.2374187
_cons	1.401564	.2984304	4.70	0.000	.8113994	1.991728

Dada la salida de la regresión podemos observar que hay mínimos cambios sobre los coeficientes de cada variable estimada, es decir, no resultan ser diferenciales grandes comparando con los valores de los estimadores MCO. De todas maneras, como la principal consecuencia de la heterocedasticidad es el efecto de sesgidez e invalidación de los errores estándar y estimadores t o F , correspondientemente, implica que lo más relevante para observar en esta nueva regresión es la significatividad estadística de los estimadores:

Específicamente, la variable de cantidad de clases con ausencia (*skipped*) sigue siendo estadísticamente significativa para un p-valor de 0.001 (< 0.05) y un test t de -3.44 . Aumentó su significatividad estadística, teniendo en cuenta que su p-valor disminuyó de 0.008 a 0.001.

De la variable cualitativa de disponer computadora (PC), vemos que su significatividad estadística aumentó teniendo en cuenta que el p-valor también disminuyó y sigue conservando su significancia. Su p-valor anterior era 0.032, por lo que vemos que se cumple:

$$0.05 > 0.032 > 0.027$$

El resto de variables mejoraron también su significatividad, incluyendo la variable ACT. Por el lado del modelo global, podemos decir que también es significativo para un test F de 15.01 y p-valor de cero.

Esto nos permite inferir que ahora podemos asegurar con mayor certeza la significancia estadística del modelo habiendo corregido la heterocedasticidad por MCP.

3. En la estimación de MCP del inciso 2), obtenga los errores estándar robustos a heterocedasticidad. En otras palabras, considere la posibilidad de que la función de varianza estimada en el inciso 1) pueda estar mal especificada. ¿Varían mucho los errores estándar con respecto a los del inciso 2)?

Considerando la posibilidad de que la forma de varianza del modelo haya estado mal expresada en su función nos obliga a que la estimemos con los errores estándar robustos a la heterocedasticidad. Código: `reg colGPA hsGPA ACT skipped PC[aweight = 1/h1],r`. Solamente tuvimos que especificar con la letra *r* (de robust) al final del código de *regress*. La salida de la regresión fue la siguiente:

```
. reg colGPA hsGPA ACT skipped PC[aweight=1/h1],r
(sum of wgt is 1.5232e+03)
```

Linear regression

```
Number of obs = 141
F( 4, 136) = 21.23
Prob > F = 0.0000
R-squared = 0.3062
Root MSE = .30827
```

colGPA	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
hsGPA	.4025059	.0863341	4.66	0.000	.231775	.5732368
ACT	.013162	.0104138	1.26	0.208	-.007432	.033756
skipped	-.0763653	.0212092	-3.60	0.000	-.1183078	-.0344227
PC	.1260048	.0590245	2.13	0.035	.0092802	.2427294
_cons	1.401564	.3105961	4.51	0.000	.7873409	2.015786

Podemos observar que en rasgos generales los errores estándar robustos a la heterocedasticidad son levemente mayores que los estimados por MCP. Intuitivamente esto es correcto, ya que empíricamente está comprobado que suelen ser mayores en términos absolutos, aunque no siempre. Como era de esperar, la R-cuadrada sigue siendo la misma y los coeficientes también.

Los errores estándar robustos y los estadísticos *t* robustos se justifican sólo si el tamaño de muestra se hace grande. Con tamaños de muestra pequeños, el estadístico *t* robusto puede tener distribuciones que no estén muy próximas a la distribución *t* y que podrían echar a perder la inferencia, lo que implica que en este caso podemos utilizarlos, ya que $n = 141$.

EJERCICIO Nº 2: Card (1995) (CARD.dta) estimó el impacto de educación sobre el salario de un grupo de hombres en 1976 a partir del siguiente modelo teórico: $\log(\text{salario}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{raza} + \beta_5 \text{capital} + \beta_6 \text{sur} + u$ donde RAZA es una variable binaria para raza negra, CAPITAL una variable binaria por vivir en un área metropolitana y SUR una variable binaria por vivir en el sur. También se incluyen un conjunto completo de variables binarias regionales (reg1-reg9) y una variable binaria de área metropolitana para el lugar donde el hombre vivía en 1966 (reg_smsa).

1. ¿Por qué la variable EDUC podría ser endógena en este caso?

La variable de educación podría llegar a ser determinada endógenamente a partir de las variables exógenas socio-geográficas como capital y raza. Además, cuando trabajamos con diversos grupos sociales corremos el riesgo de incurrir en un problema de Heterogeneidad inobservable. Esto implica que los individuos en una población difieren en características que no son observables y por lo tanto omitidas en la especificación del modelo. Es similar al problema de variables omitidas y sus consecuencias. Por ende, deberemos manipular con atención la variable educación dentro de este modelo.

2. Estime la ecuación por MCO y explique cuáles son las consecuencias de la heterogeneidad inobservable sobre los estimadores MCO.

Como se mencionó anteriormente, un problema de heterogeneidad inobservable es posible causa de endogeneidad en una variable, que posee consecuencias similares al problema de variable omitida. Siguiendo este razonamiento, eso implica que tendremos estimadores sesgados para las variables y un modelo poco fiable. Estimado la ecuación del modelo exponencial (o, en otras palabras, log-nivel) tendríamos:

$$\log(\text{salario}) = \beta_0 + \beta_1(\text{educ}) + \beta_2(\text{exper}) + \beta_3(\text{raza}) + \beta_4(\text{capital}) + \beta_5(\text{sur}) + u$$

El código usado fue “*reg l_salario educ exper raza capital sur*” y la salida de Stata es:

```
reg l_salario educ exper raza capital sur
```

Source	SS	df	MS	Number of obs = 3010		
Model	165.205667	5	33.0411334	F(5, 3004) = 232.21		
Residual	427.435978	3004	.142288941	Prob > F = 0.0000		
Total	592.641645	3009	.196956346	R-squared = 0.2788		
				Adj R-squared = 0.2776		
				Root MSE = .37721		

l_salario	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.073807	.0035336	20.89	0.000	.0668784	.0807356
exper	.0393134	.0021955	17.91	0.000	.0350085	.0436183
raza	-.1882225	.0177678	-10.59	0.000	-.2230607	-.1533843
capital	.1647411	.0156919	10.50	0.000	.1339732	.195509
sur	-.1290528	.0152285	-8.47	0.000	-.1589122	-.0991935
_cons	4.913331	.0631212	77.84	0.000	4.789566	5.037096

3. Card utilizó como variable instrumental para EDUC una variable binaria que era igual a 1 si la persona había crecido cerca de una universidad con carreras de 4 años (CERCANIA4). Verifique el supuesto de relevancia del instrumento.

Código: *corr cercania4 educ, covariance*

```
. corr cercania4 educ, covariance
(obs=3010)
```

	cercan~4	educ
cercania4	.216926	
educ	.179836	7.16586

Recordando la condición de relevancia para el instrumento como $Cov(Z, X) \neq 0$, tenemos entonces que aplicando la matriz de covarianza entre la variable instrumentada y el instrumento podemos verificar el supuesto de relevancia para cercania4 (la variable instrumental). Dado que la covarianza entre estas es de $0.1798 \approx 0.18 \neq 0$ y por ende, se cumple dicha condición de relevancia. Implica que el instrumento tiene un grado de relación con la variable endógena.

Sintetizando, según la comprobación, el instrumento resulta ser válido.

4. Argumente sobre el cumplimiento del supuesto de exogeneidad del instrumento.

Acerca del supuesto de exogeneidad, lo que podemos discutir es que en el contexto de las variables omitidas, la exogeneidad del instrumento significa que el instrumento (llámese Z) no debe tener ningún efecto parcial sobre Y (después de que X y las variables omitidas se han controlado), y que Z no debe estar correlacionada con las variables omitidas. Dado que el supuesto implica la covarianza entre Z y el error inobservable u , generalmente no se espera demostrar este supuesto: en la mayoría de los casos, se debe mantener $Cov(Z, u) = 0$ recurriendo simplemente al comportamiento económico o introspección.

5. Estime la ecuación por MC2E utilizando el instrumento propuesto por Card y compare los coeficientes estimados con los EMC del punto 1). Interprete el coeficiente asociado a la educación y realice una estimación por intervalos para ese parámetro con ambos métodos de estimación.

Para la resolución de la estimación por MC2E debemos utilizar el siguiente código en Stata: *ivregress 2sls l_salario exper raza capital sur (educ = cercania4), first*

Se elige la combinación lineal que presente la mayor correlación la variable endógena, que vendrá dada por la ecuación de forma reducida. Entonces el valor estimado de la endógena en la ecuación reducida será el mejor instrumento, el proceso sería:

- En una primera etapa estimamos la forma reducida de la variable endógena.
- En la segunda etapa estimamos la ecuación estructural.

Instrumental variables (2SLS) regression

Number of obs = 3010
 Wald chi2(5) = 673.47
 Prob > chi2 = 0.0000
 R-squared = 0.2140
 Root MSE = .3934

l_salario	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.1318498	.0495237	2.66	0.008	.0347852	.2289144
exper	.0622698	.0196665	3.17	0.002	.0237241	.1008155
raza	-.1296012	.0532094	-2.44	0.015	-.2338897	-.0253128
capital	.1348259	.0302606	4.46	0.000	.0755162	.1941356
sur	-.1092522	.0231533	-4.72	0.000	-.1546318	-.0638725
_cons	3.939822	.8309333	4.74	0.000	2.311222	5.568421

Instrumented: educ

Instruments: exper raza capital sur cercania4

Aplicando el método de Mínimos Cuadrados en Dos Etapas obtenemos la anterior salida de regresión, de la cual podemos inferir lo siguiente:

- El coeficiente de educación resulta ser diferente al original, gracias a la instrumentación y reducción del sesgo, generando un efecto marginal promedio sobre el logaritmo del salario de 13.18%, para un incremento de 1 año en la educación, siendo esta una variable significativa para un pvalor de $0.008 < 0.05$.

Si quisiéramos calcular el cambio porcentual exacto vía la ecuación $100[\exp(\beta_i) - 1]$ tenemos que $100[e^{0.1318498} - 1]$ es igual a 14.09 %

El intervalo de confianza para el coeficiente de educación es (0.034 ; 0.228).

- El resto de coeficientes sufren cambios de intensificación en su magnitud con la misma dirección, aquellos positivos se incrementan y aquellos negativos disminuyen. Ninguna variable deja de ser significativa estadísticamente.

La primera regresión (de la primera etapa), es decir la forma reducida, queda como:

```
. ivregress 2sls l_salario exper raza capital sur (educ=cercania4),first
```

First-stage regressions

Number of obs = 3010
 F(5, 3004) = 542.34
 Prob > F = 0.0000
 R-squared = 0.4744
 Adj R-squared = 0.4736
 Root MSE = 1.9423

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	-.3955633	.0087186	-45.37	0.000	-.4126584	-.3784683
raza	-1.006665	.0896256	-11.23	0.000	-1.182399	-.8309312
capital	.4028107	.0848419	4.75	0.000	.2364566	.5691648
sur	-.2901097	.0791554	-3.67	0.000	-.4453139	-.1349055
cercania4	.3373683	.0824893	4.09	0.000	.175627	.4991096
_cons	16.60155	.1193692	139.08	0.000	16.36749	16.8356

6. Pruebe si la variable EDUC es exógena.

Para probar la exogeneidad de la variable educación podemos recurrir al contraste de Hausman. Básicamente, este test consta de estimar la forma reducida (endógena en función de exógenas) de la variable que sospechamos por endogeneidad (educación). Salvamos los residuos de la primera etapa, que será la parte de la variable objetivo que no es exógena e introducimos dichos residuos en el modelo original, lo regresamos por MCO y evaluamos la significancia estadística del estimador de los residuos:

- Si el coeficiente es significativo, se rechaza la H_0 de exogeneidad (hay endogeneidad).
- Si el coeficiente no es significativo, no podemos rechazar que sea exógena.

Si bien el proceso descrito es la metodología a aplicar, en Stata podemos hacerlo vía asistente automático con el comando *estat endogenous*, una vez que realizamos primero la regresión por MC2E, como hicimos para el inciso anterior. Haciendo esto, tenemos el siguiente resultado:

```
estat endogenous
```

```
Tests of endogeneity
```

```
Ho: variables are exogenous
```

```
Durbin (score) chi2(1)          =  1.50534   (p = 0.2199)
```

```
Wu-Hausman F(1,3003)           =  1.50259   (p = 0.2204)
```

Dado que para el estadístico de Durbin vía chi-cuadrada, igual a 1.5053 que le corresponde un pvalor de 0.2199, rechazamos la Hipótesis Nula de que la variable sea exógena, en favor de la evidencia de su endogeneidad.

El mismo resultado es extraíble del estadístico Wu-Hausman para un test F, igual a 1.5025 que le corresponde un pvalor de 0.2204, debemos rechazar la Hipótesis Nula de la exogeneidad de educación.

En conclusión, por el contraste de Hausman detectamos que la variable independiente de educación NO es exógena.

EJERCICIO Nº 3: El archivo FISH.dta contiene 97 observaciones de precios y cantidades diarias sobre los precios del pescado en el mercado de Nueva York. Utilice la variable $\log(\text{avgprc})$ el logaritmo del precio promedio del pescado, como la variable dependiente.

1. Realice la regresión de $\log(\text{avgprc})$ sobre cuatro variables binarias (mon, tues, wed, thurs), con el día viernes como base. Incluya una tendencia lineal en el tiempo. ¿Existe evidencia de que los precios varían de manera sistemática a lo largo de la semana?

Utilizando el logaritmo del precio medio del pescado como variable explicada y las variables categóricas para los días laborables de lunes a jueves, tomando viernes como base (es decir, agregamos $k - 1$ categorías al modelo), podemos correr una regresión múltiple.

$$\log(\text{average price})_t = \delta_0 + \delta_1(\text{lunes})_t + \delta_2(\text{martes})_t + \delta_3(\text{mierc})_t + \delta_4(\text{jueves})_t + v_t$$

Trabajando en series de tiempo, primero deberemos generar una variable nueva que representa el tiempo o una tendencia, en este caso, lineal (*gen tendencia = _n*) y la configuramos como variable de tiempo para el campo de análisis (*tsset tendencia*), sin especificar ningún formato adicional (daily, monthly, etc). Corriendo la regresión tendríamos

```
reg lavgprc tendencia mon tues wed thurs
```

Source	SS	df	MS	Number of obs = 97		
Model	1.3432706	5	.268654121	F(5, 91) = 1.70		
Residual	14.3698396	91	.157910325	Prob > F = 0.1423		
Total	15.7131102	96	.163678231	R-squared = 0.0855		
				Adj R-squared = 0.0352		
				Root MSE = .39738		

lavgprc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tendencia	-.0039912	.0014444	-2.76	0.007	-.0068603	-.0011221
mon	-.0100699	.1293525	-0.08	0.938	-.2670127	.2468729
tues	-.0088125	.1273075	-0.07	0.945	-.2616932	.2440682
wed	.0376262	.1256956	0.30	0.765	-.2120526	.287305
thurs	.090559	.1256707	0.72	0.473	-.1590703	.3401884
_cons	-.0729573	.1151907	-0.63	0.528	-.3017694	.1558547

Como es un modelo exponencial (logarítmico-lineal), es decir, de semi-elasticidad y observando los valores de los coeficientes obtenidos, podemos inferir a priori de que hay una variación sistemática del precio medio, pero no muy fuerte, dependiendo el día vemos:

- Para el día lunes, tenemos que la variación del logaritmo del precio promedio del pescado respecto el día viernes (grupo base) es en promedio de -1%
- Para el día martes, la variación promedio de la semi-elasticidad del precio del pescado comparado con el día viernes es alrededor de -0.88%
- Los miércoles (variable *wed*) tenemos que el logaritmo del precio promedio del pescado es 3.7% más que los viernes.
- Los días de la semana jueves, podemos ver que respecto los viernes hay una variación aproximada positiva del 9% del logaritmo del precio medio del pescado, siendo el día de mayor variación.

Ninguna de las variables anteriores (de Monday a Thursday, con grupo base Friday) es significativa estadísticamente, presentando cada una p-valores entre 0.473 y 0.938, claramente mucho mayores al valor de referencia $p = 0.05$

2. Añada las variables wave2 y wave3, que son mediciones de la altura de las olas durante varios días pasados. ¿Estas variables son individualmente significativas? Describe un mecanismo mediante el cual el mar tempestuoso aumente el precio del pescado.

Incorporando a la ecuación del modelo original las variables independientes wave2 y wave3, que representan mediciones de altura de las olas durante días anteriores, tendríamos:

$$\log(\text{avgprc})_t = \delta_0 + \delta_1(\text{lun})_t + \delta_2(\text{mar})_t + \delta_3(\text{mier})_t + \delta_4(\text{jue})_t + \delta_5(\text{wave2}) + \delta_6(\text{wave3}) + v_t$$

El código en Stata nos queda como

reg lavgprc tendencia mon tues wed thurs wave2 wave3 y obtenemos la siguiente salida:

```
reg lavgprc tendencia mon tues wed thurs wave2 wave3
```

Source	SS	df	MS	Number of obs = 97		
Model	4.86305046	7	.694721494	F(7, 89) = 5.70		
Residual	10.8500597	89	.121910784	Prob > F = 0.0000		
Total	15.7131102	96	.163678231	R-squared = 0.3095		
				Adj R-squared = 0.2552		
				Root MSE = .34916		

lavgprc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tendencia	-.0011575	.0013908	-0.83	0.408	-.003921	.0016061
mon	-.018158	.1140693	-0.16	0.874	-.2448113	.2084954
tues	-.0085331	.1121208	-0.08	0.940	-.2313147	.2142485
wed	.0500406	.1117406	0.45	0.655	-.1719856	.2720669
thurs	.1225463	.1109771	1.10	0.272	-.0979629	.3430554
wave2	.0908906	.0217536	4.18	0.000	.0476667	.1341145
wave3	.0473748	.0208107	2.28	0.025	.0060243	.0887252
_cons	-.9202542	.1898134	-4.85	0.000	-1.297409	-.5430991

- La variable agregada wave2 muestra significatividad estadística para un p-valor igual a cero, con su estadístico t de 4.18.
- La variable incorporada wave3 nos arroja también significatividad estadística para un p-valor de 0.025, con su estadístico t de 2.28.

Asimismo, tenemos que los coeficientes de las variables categóricas de los días de semana cambian en su mayoría incrementando su efecto marginal en la misma dirección, es decir, aquellos valores negativos (como las estimaciones para lunes y martes), cambian negativamente y aquellos valores positivos (como las estimaciones de miércoles y jueves) se incrementan en alrededor de 1.3~3%. Además, la R-cuadrada de este modelo nuevo se incrementa, respecto el anterior. Todo esto anteriormente mencionado es un indicio de que estábamos omitiendo variables relevantes y, por ende, en el modelo original incurriamos en un sesgo por omisión de variable.

Claramente vemos que la relevancia de la altura de las olas resulta ser un factor importante para nuestra variable dependiente.

Intuyendo algún mecanismo por el cual las olas del mar puedan influir sobre un aumento del precio del pescado podemos pensar fácilmente que estas generan mayor esfuerzo para mantener un mismo nivel de producción (pescados capturados, valga la redundancia). Más allá de nuestra ignorancia y falta especialidad en el rubro de la pesca, podemos argumentar una mayor contratación de trabajo y capital para mayores niveles de altura de ola por los siguientes motivos:

- 1 – El capital (botes, redes de pescas y herramientas) no necesariamente pueden estar 100% adaptado a olas tan grandes, lo que disminuye su productividad marginal. Esta baja de la productividad deberá ser compensada con mayor utilización de capital. Lo mismo puede suceder con el factor de mano de obra (pescadores), quienes seguramente no tendrán el mismo rendimiento corporal un día de bajas olas comparado con un día de olas altas.
- 2 – Puede ser también que olas muy altas (y un viento más tempestuoso) genere una redistribución o locación de los peces en el agua. Seguramente estos no estarán cerca de la orilla ni la superficie superior del agua cuando se vean amenazados por un clima más peligroso. Hay que analizar mejor el movimiento del banco de peces.

Si bien los argumentos no son fieles, ya que quedan sujetos a la revisión de un profesional del rubro, pueden ser un primer indicio para justificar el comportamiento estadístico de estas variables dentro de nuestro modelo general.

3. ¿Qué ocurrió con la tendencia en el tiempo cuando wave2 y wave3 se añadieron a la regresión? ¿Qué debe estar pasando?

El comportamiento de la variable de tiempo o tendencia lineal en ambos modelos estaría dada, en cada caso, por:

Primer modelo:

tendencia		-.0039912	.0014444	-2.76	0.007	-.0068603	-.0011221
-----------	--	-----------	----------	-------	-------	-----------	-----------

Segundo modelo:

tendencia		-.0011575	.0013908	-0.83	0.408	-.003921	.0016061
-----------	--	-----------	----------	-------	-------	----------	----------

Podemos observar que en nuestra última regresión la tendencia se vuelve rotundamente insignificativa en términos estadísticos, para un nuevo p-valor de 0.408 ($> 0.05 > 0.007$), siendo esta antes sí significativa. Implica que ante la incorporación de variables relevantes (que estaban omitidas) y la consecuente reducción del sesgo en el resto de variables, que la tendencia pierda significancia puede ser porque realmente no es muy útil para explicarnos el comportamiento de Y en el paso del tiempo, tenía un sesgo positivo.

Por otro lado, que esta variable pierda relevancia para explicar el logaritmo del precio medio del pescado puede deberse a que la variable dependiente e independientes tienen tendencias distintas (por ejemplo, una hacia arriba y otra hacia abajo), pero el movimiento en la variable independiente alrededor de su línea de tendencia provoca un movimiento de la variable dependiente alejándose de su línea de tendencia.

4. Explique por qué se supone que todas las variables explicativas son estrictamente exógenas.

El supuesto de la exogeneidad estricta por parte de las variables independientes deviene de dos requisitos:

(1) del supuesto de Gauss-Markov de la media condicionada nula con el término del error. Para cada t , dadas las variables explicativas para todos los periodos, el valor esperado del error u_t es cero. Matemáticamente: $E(u_t/X) = 0, \quad t = 1, 2, \dots, n$

Cuando se cumple, se dice que las X_{tj} son **contemporáneamente exógenas**. Es decir, es condición necesaria pero no suficiente. Entonces, para la **exogeneidad** estricta se debe cumplir también que:

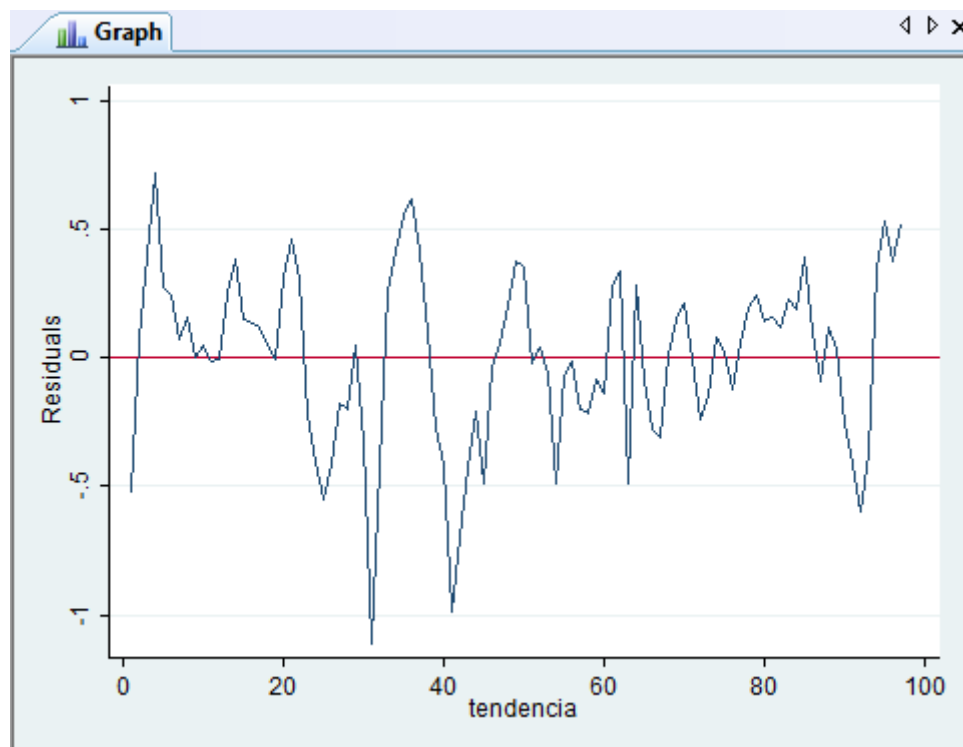
(2) el error esté incorrelacionado con las X_t de **todos** los periodos (pasados/presentes/futuros). Implica que los cambios producidos en el error en el momento actual no provoquen cambios futuros en X , además las variables estrictamente exógenas no pueden reaccionar a lo que le sucedió a Y en el pasado.

Si se cumplen estas dos condiciones, tendremos variables explicativas estrictamente exógenas.

5. Pruebe si los errores presentan autocorrelación serial AR(1).

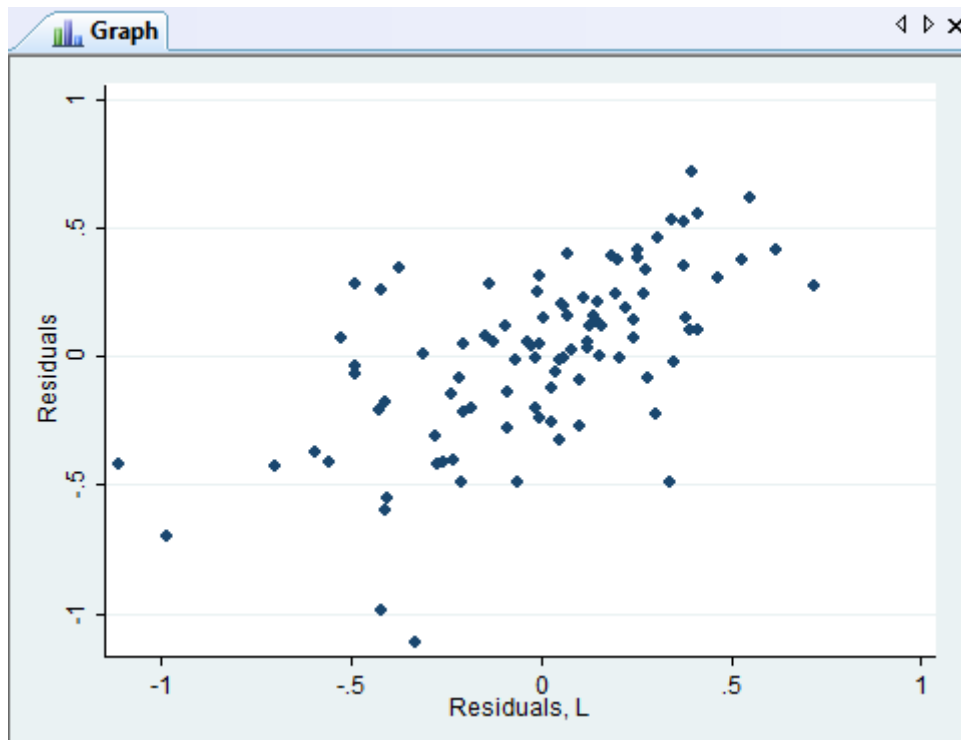
Para detectar correlación serial en los errores de orden uno, por ejemplo, tenemos varias alternativas o herramientas como detección. En primer lugar, podemos optar por graficar los residuos en el tiempo para observar su comportamiento:

Extrayendo los residuos de la regresión (*predict u2,r*) y corriendo el código *tsline u2 if e(sample) == 1, yline(0)* obtenemos el siguiente gráfico:



La línea roja en el cero (traída por el condicional $e(sample) == 1, yline(0)$) es un refuerzo visual para contrastar mejor las variaciones de los residuos cuando cambia su signo.

Si bien no se comportan de manera perfecta en la variación, tampoco podemos detectar a priori algún patrón fuerte específico de distribución en el tiempo que nos permita pensar de autocorrelación. Así que, procedemos a graficar en puntos los residuos contra estos mismos en su período anterior (es decir, con su rezago de un período en el tiempo):
Código en stata: `scatter u2 L1.u2`



Visto de esta manera, ahora sí podríamos pensar con un más de certeza de que sí estamos en presencia de autocorrelación positiva, dada la inclinación y dirección del agrupamiento de puntos en el gráfico. Sintetizando, podemos sospechar de autocorrelación en los residuos, por lo que amerita hacer las pruebas estadísticas formales vía tests de hipótesis:

Un método específico para detectar correlaciones seriales (solo de orden uno) es el de Durbin-Watson. Este test de post-estimación es el cociente de la diferencia al cuadrado de los errores sucesivos sobre la suma de los errores al cuadrado. Corriendo el código correspondiente (`estat dwatson`) tenemos el siguiente resultado:

```
. estat dwatson
```

```
Durbin-Watson d-statistic( 8, 97) = .745231
```

Con 8 parámetros y 97 observaciones, en la tabla estadística (*ver última hoja del parcial para ver la tabla*) podemos ubicarlo por la columna $k' = 8$ y la fila $n = 100$, para un 0.01 nivel de significancia, por lo que para un estadístico DW de 0.7452 resulta ser menor que la cota inferior dL . El resultado es: $0.7452 < 1.378(dL) < 1.717(dU)$.

90	1.496	1.540	1.474	1.563	1.452	1.587	1.429	1.611	1.406	1.636	1.383	1.661	1.360	1.687	1.336	1.714
95	1.510	1.552	1.489	1.573	1.468	1.596	1.446	1.618	1.425	1.642	1.403	1.666	1.381	1.690	1.358	1.715
100	1.522	1.562	1.503	1.583	1.482	1.604	1.462	1.625	1.441	1.647	1.421	1.670	1.400	1.693	1.378	1.717
150	1.611	1.637	1.598	1.651	1.584	1.665	1.571	1.679	1.557	1.693	1.543	1.708	1.530	1.722	1.515	1.737
200	1.664	1.684	1.653	1.693	1.643	1.704	1.633	1.715	1.623	1.725	1.613	1.735	1.603	1.746	1.592	1.757

Determinado esto, como el estadístico cae entre cero y la cota inferior, debemos rechazar la Hipótesis Nula de que NO hay autocorrelación, en favor de la evidencia de correlación serial positiva. Este resultado es nuestra primera verificación estadística de lo que sospechábamos en el segundo gráfico sobre una correlación positiva en los errores.

Por último y como verificación final, tenemos la posibilidad de emplear el contraste de Breusch-Godfrey que sirve para autocorrelaciones en los errores de orden superior (1,2,3,...) De todas maneras, nosotros lo empleamos especificándole que testaremos solo correlación serial de orden uno. Corriendo el correspondiente código en Stata (*estat bgodfrey, lag(1)*) tenemos el siguiente resultado:

```
. estat bgodfrey, lag(1)
```

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	37.348	1	0.0000

H0: no serial correlation

Para un p-valor de literalmente cero, debemos rechazar la hipótesis nula (H_0 : NO hay correlación serial) para un nivel de confianza de 95%, en favor de que el modelo está autocorrelacionado serialmente en los errores.

6. Obtenga la estimación de Prais Winsten para el modelo estimado en el inciso 2). ¿Son wave2 y wave3 conjuntamente significativas en términos estadísticos?

Para intentar solucionar la correlación serial se puede emplear el método de estimación de Mínimos Cuadrados Generalizados Factibles (MCGF) aplicando la técnica de Prais-Winsten (PW). Debemos correr la regresión con las mismas variables pero empleando un comando distinto: *prais lavgprc tendencia mon tues wed thurs wave2 wave3*

Dada la salida de regresión mostrada más abajo, podemos inferir que los coeficientes los estimadores se ven definitivamente alterados, ya que por ejemplo las variables dummies lunes y martes dejan de tener un efecto marginal negativo, comparado con el día viernes, para ahora afectar positivamente al logaritmo del precio promedio del pescado.

Por otro lado, las variables de mediciones de altura de olas pierden un poco de significatividad estadística, incrementando su p-valor e inclusive *wave3* deja de ser significativa

($pvalor = 0.066$). Si probamos un test F de significancia conjunta ($test\ wave2 = wave3 = 0$), tenemos el siguiente resultado:

```
test wave2=wave3=0

( 1) wave2 - wave3 = 0
( 2) wave2 = 0

F( 2, 89) = 5.01
Prob > F = 0.0087
```

Debemos rechazar la hipótesis nula de que son insignificativas ($H_0: wave2 = wave3 = 0$), en favor de su significancia conjunta dentro del modelo.

```
. prais lavgprc tendencia mon tues wed thurs wave2 wave3
```

```
Iteration 0: rho = 0.0000
Iteration 1: rho = 0.6177
Iteration 2: rho = 0.6791
Iteration 3: rho = 0.6864
Iteration 4: rho = 0.6873
Iteration 5: rho = 0.6874
Iteration 6: rho = 0.6874
Iteration 7: rho = 0.6874
Iteration 8: rho = 0.6874
```

```
Prais-Winsten AR(1) regression -- iterated estimates
```

Source	SS	df	MS	Number of obs =	97
Model	.984996757	7	.140713822	F(7, 89) =	1.99
Residual	6.29323192	89	.070710471	Prob > F =	0.0652
Total	7.27822868	96	.075814882	R-squared =	0.1353
				Adj R-squared =	0.0673
				Root MSE =	.26591

lavgprc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tendencia	-.0007329	.0029453	-0.25	0.804	-.0065851	.0051194
mon	.0099479	.0651969	0.15	0.879	-.1195968	.1394927
tues	.0025015	.0743662	0.03	0.973	-.1452626	.1502655
wed	.0624215	.0745511	0.84	0.405	-.0857101	.210553
thurs	.1174177	.0620921	1.89	0.062	-.0059579	.2407934
wave2	.0496817	.0172894	2.87	0.005	.015328	.0840353
wave3	.0323356	.0173988	1.86	0.066	-.0022353	.0669066
_cons	-.6578796	.2387552	-2.76	0.007	-1.132281	-.1834781
rho	.6874011					

```
Durbin-Watson statistic (original) 0.745231
```

```
Durbin-Watson statistic (transformed) 1.648617
```


TABLE D.5B**DURBIN-WATSON d STATISTIC: SIGNIFICANCE POINTS OF d_L AND d_U AT 0.01 LEVEL OF SIGNIFICANCE**

n	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$		$k' = 6$		$k' = 7$		$k' = 8$		$k' = 9$		$k' = 10$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0.390	1.142	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.435	1.036	0.294	1.676	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.497	1.003	0.345	1.489	0.229	2.102	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.554	0.998	0.408	1.389	0.279	1.875	0.183	2.433	—	—	—	—	—	—	—	—	—	—	—	—
10	0.604	1.001	0.466	1.333	0.340	1.733	0.230	2.193	0.150	2.690	—	—	—	—	—	—	—	—	—	—
11	0.653	1.010	0.519	1.297	0.396	1.640	0.286	2.030	0.193	2.453	0.124	2.892	—	—	—	—	—	—	—	—
12	0.697	1.023	0.569	1.274	0.449	1.575	0.339	1.913	0.244	2.280	0.164	2.665	0.105	3.053	—	—	—	—	—	—
13	0.738	1.038	0.616	1.261	0.499	1.526	0.391	1.826	0.294	2.150	0.211	2.490	0.140	2.838	0.090	3.182	—	—	—	—
14	0.776	1.054	0.660	1.254	0.547	1.490	0.441	1.757	0.343	2.049	0.257	2.354	0.183	2.667	0.122	2.981	0.078	3.287	—	—
15	0.811	1.070	0.700	1.252	0.591	1.464	0.488	1.704	0.391	1.967	0.303	2.244	0.226	2.530	0.161	2.817	0.107	3.101	0.068	3.374
16	0.844	1.086	0.737	1.252	0.633	1.446	0.532	1.663	0.437	1.900	0.349	2.153	0.269	2.416	0.200	2.681	0.142	2.944	0.094	3.201
17	0.874	1.102	0.772	1.255	0.672	1.432	0.574	1.630	0.480	1.847	0.393	2.078	0.313	2.319	0.241	2.566	0.179	2.811	0.127	3.053
18	0.902	1.118	0.805	1.259	0.708	1.422	0.613	1.604	0.522	1.803	0.435	2.015	0.355	2.238	0.282	2.467	0.216	2.697	0.160	2.925
19	0.928	1.132	0.835	1.265	0.742	1.415	0.650	1.584	0.561	1.767	0.476	1.963	0.396	2.169	0.322	2.381	0.255	2.597	0.196	2.813
20	0.952	1.147	0.863	1.271	0.773	1.411	0.685	1.567	0.598	1.737	0.515	1.918	0.436	2.110	0.362	2.308	0.294	2.510	0.232	2.714
21	0.975	1.161	0.890	1.277	0.803	1.408	0.718	1.554	0.633	1.712	0.552	1.881	0.474	2.059	0.400	2.244	0.331	2.434	0.268	2.625
22	0.997	1.174	0.914	1.284	0.831	1.407	0.748	1.543	0.667	1.691	0.587	1.849	0.510	2.015	0.437	2.188	0.368	2.367	0.304	2.548
23	1.018	1.187	0.938	1.291	0.858	1.407	0.777	1.534	0.698	1.673	0.620	1.821	0.545	1.977	0.473	2.140	0.404	2.308	0.340	2.479
24	1.037	1.199	0.960	1.298	0.882	1.407	0.805	1.528	0.728	1.658	0.652	1.797	0.578	1.944	0.507	2.097	0.439	2.255	0.375	2.417
25	1.055	1.211	0.981	1.305	0.906	1.409	0.831	1.523	0.756	1.645	0.682	1.776	0.610	1.915	0.540	2.059	0.473	2.209	0.409	2.362
26	1.072	1.222	1.001	1.312	0.928	1.411	0.855	1.518	0.783	1.635	0.711	1.759	0.640	1.889	0.572	2.026	0.505	2.168	0.441	2.313
27	1.089	1.233	1.019	1.319	0.949	1.413	0.878	1.515	0.808	1.626	0.738	1.743	0.669	1.867	0.602	1.997	0.536	2.131	0.473	2.269
28	1.104	1.244	1.037	1.325	0.969	1.415	0.900	1.513	0.832	1.618	0.764	1.729	0.696	1.847	0.630	1.970	0.566	2.098	0.504	2.229
29	1.119	1.254	1.054	1.332	0.988	1.418	0.921	1.512	0.855	1.611	0.788	1.718	0.723	1.830	0.658	1.947	0.595	2.068	0.533	2.193
30	1.133	1.263	1.070	1.339	1.006	1.421	0.941	1.511	0.877	1.606	0.812	1.707	0.748	1.814	0.684	1.925	0.622	2.041	0.562	2.160
31	1.147	1.273	1.085	1.345	1.023	1.425	0.960	1.510	0.897	1.601	0.834	1.698	0.772	1.800	0.710	1.906	0.649	2.017	0.589	2.131
32	1.160	1.282	1.100	1.352	1.040	1.428	0.979	1.510	0.917	1.597	0.856	1.690	0.794	1.788	0.734	1.889	0.674	1.995	0.615	2.104
33	1.172	1.291	1.114	1.358	1.055	1.432	0.996	1.510	0.936	1.594	0.876	1.683	0.816	1.776	0.757	1.874	0.698	1.975	0.641	2.080
34	1.184	1.299	1.128	1.364	1.070	1.435	1.012	1.511	0.954	1.591	0.896	1.677	0.837	1.766	0.779	1.860	0.722	1.957	0.665	2.057
35	1.195	1.307	1.140	1.370	1.085	1.439	1.028	1.512	0.971	1.589	0.914	1.671	0.857	1.757	0.800	1.847	0.744	1.940	0.689	2.037
36	1.206	1.315	1.153	1.376	1.098	1.442	1.043	1.513	0.988	1.588	0.932	1.666	0.877	1.749	0.821	1.836	0.766	1.925	0.711	2.018
37	1.217	1.323	1.165	1.382	1.112	1.446	1.058	1.514	1.004	1.586	0.950	1.662	0.895	1.742	0.841	1.825	0.787	1.911	0.733	2.001
38	1.227	1.330	1.176	1.388	1.124	1.449	1.072	1.515	1.019	1.585	0.966	1.658	0.913	1.735	0.860	1.816	0.807	1.899	0.754	1.985
39	1.237	1.337	1.187	1.393	1.137	1.453	1.085	1.517	1.034	1.584	0.982	1.655	0.930	1.729	0.878	1.807	0.826	1.887	0.774	1.970
40	1.246	1.344	1.198	1.398	1.148	1.457	1.098	1.518	1.048	1.584	0.997	1.652	0.946	1.724	0.895	1.799	0.844	1.876	0.749	1.956
45	1.288	1.376	1.245	1.423	1.201	1.474	1.156	1.528	1.111	1.584	1.065	1.643	1.019	1.704	0.974	1.768	0.927	1.834	0.881	1.902
50	1.324	1.403	1.285	1.446	1.245	1.491	1.205	1.538	1.164	1.587	1.123	1.639	1.081	1.692	1.039	1.748	0.997	1.805	0.955	1.864
55	1.356	1.427	1.320	1.466	1.284	1.506	1.247	1.548	1.209	1.592	1.172	1.638	1.134	1.685	1.095	1.734	1.057	1.785	1.018	1.837
60	1.383	1.449	1.350	1.484	1.317	1.520	1.283	1.558	1.249	1.598	1.214	1.639	1.179	1.682	1.144	1.726	1.108	1.771	1.072	1.817
65	1.407	1.468	1.377	1.500	1.346	1.534	1.315	1.568	1.283	1.604	1.251	1.642	1.218	1.680	1.186	1.720	1.153	1.761	1.120	1.802
70	1.429	1.485	1.400	1.515	1.372	1.546	1.343	1.578	1.313	1.611	1.283	1.645	1.253	1.680	1.223	1.716	1.192	1.754	1.162	1.792
75	1.448	1.501	1.422	1.529	1.395	1.557	1.368	1.587	1.340	1.617	1.313	1.649	1.284	1.682	1.256	1.714	1.227	1.748	1.199	1.783
80	1.466	1.515	1.441	1.541	1.416	1.568	1.390	1.595	1.364	1.624	1.338	1.653	1.312	1.683	1.285	1.714	1.259	1.745	1.232	1.777
85	1.482	1.528	1.458	1.553	1.435	1.578	1.411	1.603	1.386	1.630	1.362	1.657	1.337	1.685	1.312	1.714	1.287	1.743	1.262	1.773
90	1.496	1.540	1.474	1.563	1.452	1.587	1.429	1.611	1.406	1.636	1.383	1.661	1.360	1.687	1.336	1.714	1.312	1.741	1.288	1.769
95	1.510	1.552	1.489	1.573	1.468	1.596	1.446	1.618	1.425	1.642	1.403	1.666	1.381	1.690	1.358	1.715	1.336	1.741	1.313	1.767
100	1.522	1.562	1.503	1.583	1.482	1.604	1.462	1.625	1.441	1.647	1.421	1.670	1.400	1.693	1.378	1.717	1.357	1.741	1.335	1.765
150	1.611	1.637	1.598	1.651	1.584	1.665	1.571	1.679	1.557	1.693	1.543	1.708	1.530	1.722	1.515	1.737	1.501	1.752	1.486	1.767
200	1.664	1.684	1.653	1.693	1.643	1.704	1.633	1.715	1.623	1.725	1.613	1.735	1.603	1.746	1.592	1.757	1.582	1.768	1.571	1.779