**TPº14: AUTOCORRELACIÓN**

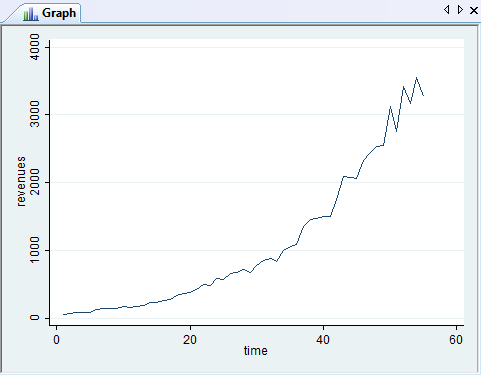
**EJERCICIO Nº 2:**

La base de datos Microsoft.dta contiene 55 observaciones cuatrimestrales sobre ingresos (revenues), gastos de publicidad y marketing (marketing), y gastos de investigación y desarrollo (idexpenditure) de Microsoft para Q1 de 1986 a Q3 de 2000. A partir de ello responda:

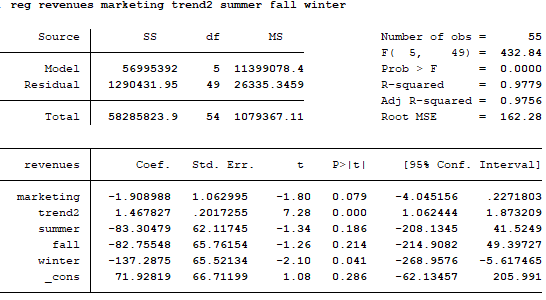
**a) Grafique las series contra el tiempo. ¿Considera oportuno incorporar un componente de tendencia?**

Corriendo una regresión lineal múltiple de la variable ingreso contra publicidad y gastos en R&D, junto con el tiempo, , podemos obtener una R-cuadrada aproximadamente de 0.99, lo que para un grado de bondad de ajusta resulta ser mucho y podría considerarse mandatorio agregar un componente de tendencia.

A partir del gráfico podemos llegar a la misma conclusión, ya que a simple vista se detecta fuerte tendencia positiva creciente:



**b) Estime el efecto del gasto publicitario sobre los ingresos de Microsoft, incorporando una tendencia cuadrática y las variables necesarias para quitar la estacionalidad al modelo.**

Regresando los ingresos de Microsoft (Revenues) contra gasto publicitario, una tendencia cuadrática y los trimestres verano, otoño e invierno (utilizando primavera como grupo base), obtenemos la siguiente salida de regresión:

**c) Pruebe si los errores de la regresión anterior tienen correlación serial AR (1), mediante pruebas gráficas y formales.**

Buscamos un patrón de comportamiento sistemático. Para esto, graficamos los residuos contra el tiempo, y los residuos en un período contra estos en un período anterior. Es la prueba empírica de errores AR(1):

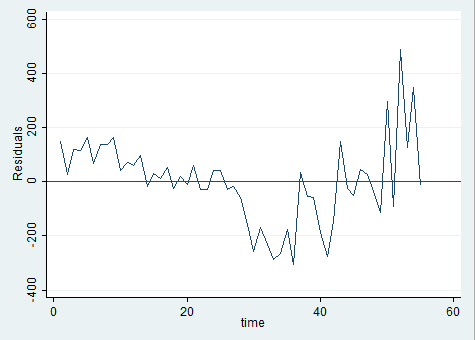
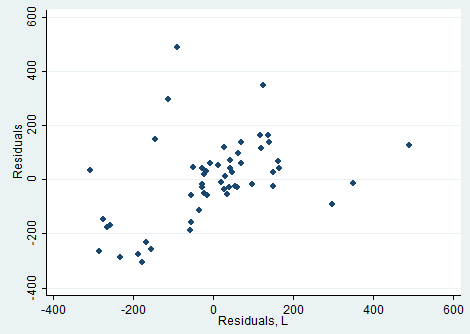
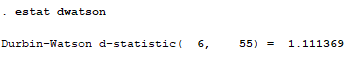
Gráfico errores-tiempo ()  
****

Gráfico errores-errores rezagados (

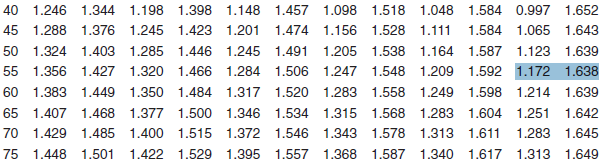


A partir de los gráficos podemos inferir en cierto comportamiento extraño en la distribución de los errores en el tiempo y para si mismos en un gráfico de puntos. Si bien no representan una prueba contundente de presencia de autocorrelación como puede ser algun caso de un patrón muy fuerte, si vemos que tienden a decrecer hasta el período 30 de tiempo y luego crecen, volviendo a decrecer cerca del período 60 de tiempo.   
Sintetizando, podemos sospechar de autocorrelación en los residuos, por lo que amerita hacer las pruebas estadísticas formales:

Primeramente, dado que vamos a testear Autocorrelación de orden 1, podemos utilizar el test de post-estimación basado en residuos, Durbin Watson, que es el cociente de la diferencia al cuadrado de los errores sucesivos sobre la suma de los errores al cuadrado.

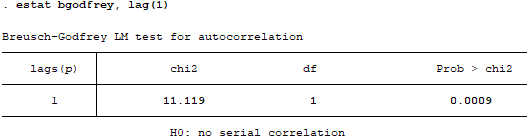


Para un estadístico de 1.111, con 6 parámetros y 55 observaciones, en la tabla estadística podemos ubicarlo por debajo de la cota inferior (dL) ().

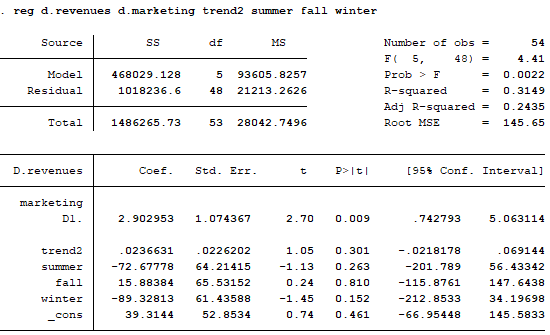


Esto implica que a un 99% de confianza debemos rechazar la Hipótesis Nula (H0) de que NO hay correlación serial, lo que significa que estamos en presencia de autocorrelación positiva en los errores de orden 1, según el test de Durbin-Watson.

Por último, podemos correr un test estadístico de Breusch-Godfrey especificando un rezago (lag) de un período.

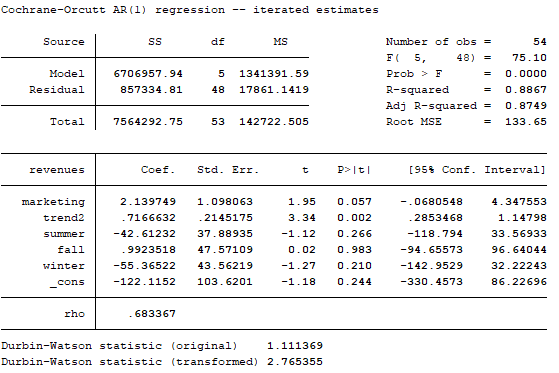


Para un valor chi cuadrado de 11.119, un grado de libertad y un p-valor de 0.0009, debemos rechazar la hipótesis nula de que no hay correlación serial, concluyendo de que estamos en presencia de autocorrelación en los errores de orden uno.

**d) Estime el modelo mediante Mínimos Cuadrados Generalizados Factibles (MCGF). En particular, utilice los métodos de: i) cuasi-diferencias; ii) Cochrane-Orcutt y iii) Prais-Winsten. Compare los resultados obtenidos con la estimación por MCO, ¿qué sucede con los estimadores? ¿y con los errores estándares?**

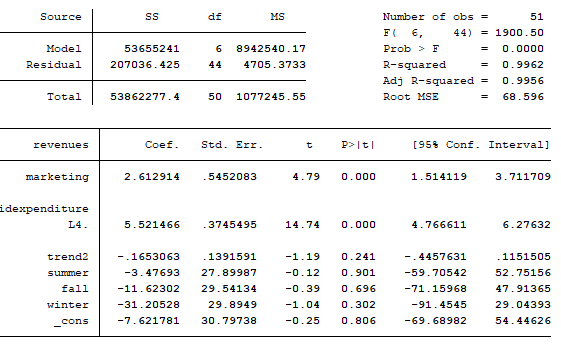
*Método en diferencias*: a partir del modelo regresado en diferencias, con el método de MCG, vemos que cambian los estimadores respecto el modelo original, cobrando mucho más sentido según la teoría económica. Para empezar, el coeficiente de gasto en marketing deja de ser negativo y toma un valor positivo, siendo ahora una variable significativa. Los errores estándar no se alteran mucho, por lo que tampoco hubo mucho costo por utilizar este método, desviándonos del original.

Cochrane-Orcutt: aplicando este método vía el siguiente código , obtenemos la siguiente salida de regresión:

****

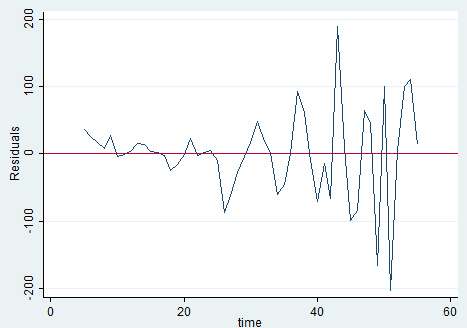
Podemos observar que los coeficientes cambian, de manera similar al método aplicado de diferencias, considerando que en esta ocasión sigue siendo insignificativa la variable de marketing para un p-valor de 0.057 (<0.05), aunque su coeficiente no está tan alejado del modelo anterior, como del original (2.13<2.90)

**e) Incorpore al modelo el gasto en I+D rezagada 4 períodos. ¿Qué pasa con los resultados del modelo, en términos de los parámetros estimados y los residuos? ¿A qué se pueden deber estos resultados?**

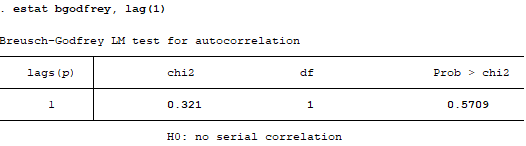


Una vez que incorporamos el gasto en investigación y desarrollo, rezagado en 4 periodos, vemos que nuestro modelo ajusta sus coeficientes y significatividad rotundamente. Para empezar, el coeficiente de marketing toma un valor positivo de 2.612, lo que resulta ser un intermedio entre los valores que tomaba según los modelos que intentaban corregir la correlación serial. La tendencia toma un valor negativo y todas las variables binarias quedan con coeficientes de valor negativo también. Más allá de los coeficientes de los parámetros estimados, vemos que aumenta la significatividad global del modelo, para todos los p-valores, incluyendo el test F de significancia global.

Mirando los nuevos residuos, a partir de un gráfico podemos ver lo siguiente:



Claramente toman un comportamiento en el tiempo mucho más aleatorio que los residuos del modelo anterior. Esto resulta ser otra señal positiva. Si corriéramos los tests estadísticos de detección de correlación serial tenemos que:



Para un p-valor de 0.57 (>0.05) fallamos al rechazar la hipotesis nula de ausencia de correlación serial, lo que nos aproxima a poder afirmar de que no hay autocorrelación de orden uno en los errores del modelo.

Todo esto puede seguramente deberse a las consecuencias del sesgo por variable omitida, ya que vemos un gran impacto positivo para cuando incorporamos esta variable que anteriormente estábamos omitiendo.