

Autocorrelación

Gustavo A. García

ggarci24@eafit.edu.co

Econometría II

Programa de Economía

Universidad EAFIT

Link slides en formato **html**

Link slides en formato **PDF**

En este tema

- Naturaleza del problema
- Consecuencias de utilizar MCO en presencia de autocorrelación
- Estructuras de la autocorrelación
- Detección de la autocorrelación
- Solución a la autocorrelación: el estimador MCG
- Ejercicio aplicado en R: relación entre salarios y productividad

Lecturas

- Wooldridge, J. (2013). *Introducción a la econometría*. 5a edición, Cengage Learning. [Caps 10, 11 y 12](#)
- Gujarati, D. y Porter, D. (2010). *Econometría*. 5a edición, Mc Graw Hill. [Cap 12](#)

Naturaleza del problema

- En estudios con datos de corte transversal es más común la presencia de heteroscedasticidad
- Ya que los datos de corte transversal se recopilan con base en una muestra aleatoria de unidades transversales (familias o empresas) no existe razón previa para pensar que el término de error correspondiente a una familia o empresa esté correlacionado con el término de error de otra familia o empresa \implies a menos de que exista **autocorrelación espacial**, es decir una autocorrelación en el espacio más que en el tiempo
- Cuando se trabajan con datos de series de tiempo, estos datos siguen un ordenamiento natural respecto del tiempo, **de modo que es muy posible que las observaciones sucesivas muestren intercorrelaciones, sobre todo si el intervalo entre observaciones sucesivas es muy corto, como un día, una semana o un mes: por ejemplo, los índices bursátiles como el Dow Jones o el S&P500**
- En esta situación de alta interrelación temporal entre los datos es más probable que se **viole el supuesto de no autocorrelación**

Naturaleza del problema

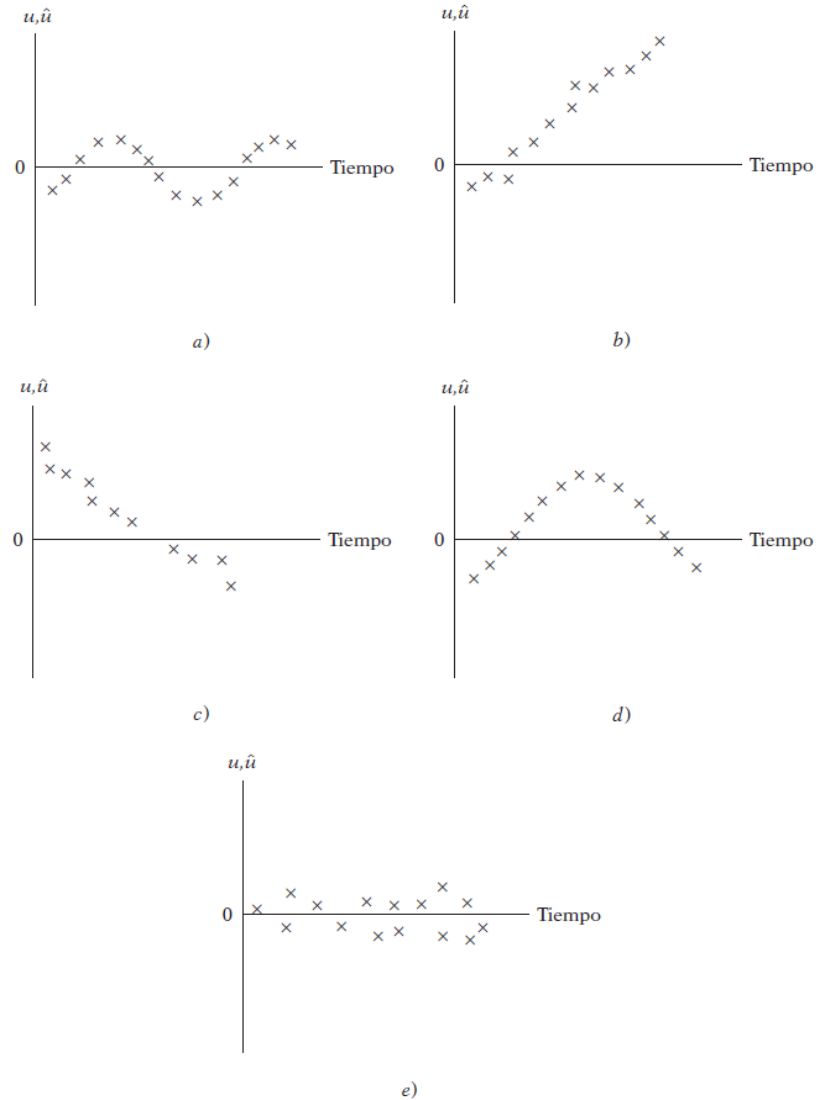
- **Autocorrelación:** correlación entre miembros de series de observaciones ordenadas en el tiempo (datos de series de tiempo) o en el espacio (datos de corte transversal)
- En el contexto de regresión, el modelo clásico de regresión lineal supone que:

$$Cov(u_i, u_j | x_i, x_j) = E(u_i u_j) = 0 \text{ con } i \neq j$$

- Pero en el contexto de series de tiempo principalmente es un supuesto que pueda no se mantenga lo que implica que $E(u_i u_j) \neq 0$

Naturaleza del problema

Visualicemos algunos patrones de autocorrelación y de no autocorrelación



Naturaleza del problema

¿Por qué ocurre la correlación serial?

- **Inercia**

Las series de tiempo económicas se caracterizan por tener inercia o pasividad \implies las observaciones sucesivas sean interdependientes

- **Sesgo de especificación**

- Variables excluidas
- Forma funcional incorrecta

- **"Manipulación" o transformación de datos**

- Pasar de datos que son mensuales a trimestrales
- Interpolación o extrapolación de datos

- **No estacionariedad**

Una serie de tiempo es **estacionaria** si sus características (media, varianza y covarianza) son invariantes respecto al tiempo. Si no es así, tenemos una serie de tiempo **no estacionaria**. Si en un modelo de regresión Y y las X s son no estacionarias, el término de error será no estacionario y por tanto mostrará autocorrelación

Consecuencias de utilizar MCO en presencia de autocorrelación

¿Qué sucede con los estimadores MCO y sus varianzas si introducimos autocorrelación en las perturbaciones?

Partamos del siguiente modelo de RLS

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

Ahora debemos suponer el mecanismo que generan las u_t . Como punto de partida vamos a suponer sigue el siguiente proceso:

$$u_t = \rho u_{t-1} + \epsilon_t$$

donde ρ se conoce como [coeficiente de autocovarianza](#) o [coeficiente de autocorrelación de primer orden](#) y este esquema que sigue u_t se conoce como [esquema autorregresivo de primer orden](#) y suele denotarse como [AR\(1\)](#)

ϵ_t es una perturbación aleatoria que satisface los supuestos estándar:

$$E(\epsilon_t) = 0, \quad Var(\epsilon_t) = \sigma_\epsilon^2, \quad Cov(\epsilon_t, \epsilon_{t+s}) = 0 \quad \text{con } s \neq 0$$

Con estas propiedades, ϵ_t es lo que en la literatura le llaman un [término de error de ruido blanco](#)

Consecuencias de utilizar MCO en presencia de autocorrelación

Es posible demostrar que con el esquema AR(1)

$$Var(u_t) = \frac{\sigma_\epsilon^2}{1 - \rho^2}$$

$$Cov(u_t, u_{t+s}) = E(u_t u_{t+s}) = \rho^s \frac{\sigma_\epsilon^2}{1 - \rho^2}$$

$$Cor(u_t, u_{t+s}) = \rho^s$$

La matriz de varianzas-covarianzas de u_t tiene la forma

$$Cov(\mathbf{u}) = \frac{\sigma_\epsilon^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix} = \sigma_u^2 \mathbf{\Omega} \neq \sigma_u^2 \mathbf{I}_n$$

Note que bajo el esquema AR(1), u_t es:

- homocedástica
- autocorrelacionada

También note que resulta crítico que $|\rho| < 1$. Si por ejemplo $\rho = 1$ (raíz unitaria), las varianzas y covarianzas no están definidas. Si $|\rho| < 1$, se dice que el proceso AR(1) es estacionario, es decir la media, la varianza y la covarianza de u_t no cambian respecto al tiempo

Consecuencias de utilizar MCO en presencia de autocorrelación

Volviendo al modelo de RLS, sabemos que el estimador MCO de la matriz var-cov de $\hat{\mathbf{B}}$ es:

$$Cov(\hat{\mathbf{B}}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Ahora, con el esquema AR(1) la matriz var-cov de $\hat{\mathbf{B}}$ será:

$$Cov(\hat{\mathbf{B}}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\Omega\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

En conclusión, si se aplica MCO a un modelo con autocorrelación:

- Los MCO continúan siendo insesgados y lineales
- Las varianzas y errores estándar de los estimadores son erróneos
- Los MCO no alcanzan la mínima varianza
- Se invalida la inferencia estadística: intervalos de confianza (más amplios) y pruebas de hipótesis (se tiende a no rechazar H_0)

Estructuras de la autocorrelación

1. Proceso Autorregresivo lineal de primer orden AR(1)

El proceso generador de datos AR(1) supone que:

$$u_t = \rho u_{t-1} + \epsilon_t$$

Con ϵ_t siendo ruido blanco

$$E(\epsilon_t) = 0$$

$$Var(\epsilon_t) = \sigma_\epsilon^2$$

$$Cov(\epsilon_t, \epsilon_s) = E(\epsilon_t \epsilon_s) = 0 \quad \forall t \neq s$$

$$Cov(\epsilon_t, u_{t-s}) = E(\epsilon_t u_{t-s}) = 0 \quad \forall t \neq s$$

La matriz de var-cov para u puede representarse como:

$$Cov(\mathbf{u}) = \frac{\sigma_\epsilon^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix} = \sigma_u^2 \mathbf{\Omega} \neq \sigma_u^2 \mathbf{I}_n$$

El análisis puede extenderse a procesos autorregresivos de orden p , AR(p)

Estructuras de la autocorrelación

1. Proceso Autorregresivo lineal de primer orden AR(1)

En este mundo se ha desarrollado una notación eficiente para los rezagos con el operador L . Por ejemplo, el AR(1) se escribe como:

$$\begin{aligned}u_t &= \rho u_{t-1} + \epsilon_t \\&= \rho L u_t + \epsilon_t \\u_t - \rho L u_t &= \epsilon_t \\u_t &= \frac{\epsilon_t}{1 - \rho L}\end{aligned}$$

Así para un proceso AR(p) sería

$$\begin{aligned}u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t \\u_t &= \rho_1 L u_t + \rho_2 L^2 u_t + \dots + \rho_p L^p u_t + \epsilon_t \\u_t(1 - \rho_1 L + \rho_2 L^2 + \dots + \rho_p L^p) &= \epsilon_t\end{aligned}$$

Se define el operador polinomio de rezagos como

$$\rho(L) = 1 - \rho_1 L + \rho_2 L^2 + \dots + \rho_p L^p$$

El proceso AR(p) se expresa como

$$\rho(L)u_t = \epsilon_t$$

Estructuras de la autocorrelación

2. Proceso de Medias Moviles MA(1) de primer orden

Formalmente el proceso se expresa como:

$$u_t = \theta\epsilon_{t-1} + \epsilon_t$$

Con ϵ_t siendo ruido blanco y $|\theta| < 1$ condición de invertibilidad del proceso

La matriz de var-cov para u puede representarse como:

$$Cov(\mathbf{u}) = \sigma_\epsilon^2(1 + \theta^2) \begin{bmatrix} 1 & \theta & \dots & 0 \\ \theta & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma_u^2 \mathbf{\Omega} \neq \sigma_u^2 \mathbf{I}_n$$

El análisis puede extenderse a procesos de medias móviles de orden q, MA(q).

En términos del operador de rezagos, el proceso MA(q) se expresa como

$$u_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_q\epsilon_{t-q}$$

$$u_t = \epsilon_t + \theta_1 L\epsilon_t + \theta_2 L^2\epsilon_t + \dots + \theta_q L^q\epsilon_t + \epsilon_t$$

$$u_t = (1 - \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)\epsilon_t$$

$$u_t = \theta(L)\epsilon_t$$

Estructuras de la autocorrelación

3. Proceso ARMA(p,q)

Combina las generalizaciones de los dos anteriores procesos

$$u_t = \rho_1 u_{t-1} + \dots + \rho_p u_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

$$\rho(L)u_t = \theta(L)\epsilon_t$$

Detección de la autocorrelación

1. Método gráfico

- Recuerde que el supuesto de no autocorrelación del modelo clásico se relaciona con las perturbaciones poblacionales u_t , las cuales no pueden observarse directamente. En su lugar disponemos de valores sustitutos, los residuos \hat{u}_t , a partir del procedimiento usual MCO
- Un examen visual de las \hat{u} da algunas claves sobre la posible presencia de autocorrelación en las u
- En realidad, un examen visual de \hat{u}_t o (\hat{u}_t^2) proporciona información útil no sólo sobre la autocorrelación, sino también sobre heteroscedasticidad, sobre el grado de adecuación del modelo o sobre el sesgo de especificación
- Hay diversas formas de examinar los residuales:
 - \hat{u}_t contra el tiempo
 - los residuos estandarizados $(\hat{u}_t/\hat{\sigma}_u)$ respecto del tiempo

Detección de la autocorrelación

```
library(gujarati); library(tidyverse) # install.packages("remotes"), remotes::install_github("brunoruas2/gujarati")

data("Table12_4")
data <- Table12_4 |> mutate(year = as.numeric(as.character(Year)),
                             y = as.numeric(as.character(Y)), x = as.numeric(as.character(X))) |>
  select(year, y, x)

modelo <- lm(log(y) ~ log(x), data = data)
data <- data |> mutate(u = resid(modelo), u_e = u/sd(u))

ggplot(data) +
  geom_line(aes(x=year, y=u), color = "blue") +
  geom_line(aes(x=year, y=u_e/46), color = "red") + # 46 = max(data$u_e)/max(data$u)
  geom_hline(yintercept = 0) +
  scale_y_continuous(name = "u", sec.axis = sec_axis(~.*46, name = "u_e"))
```



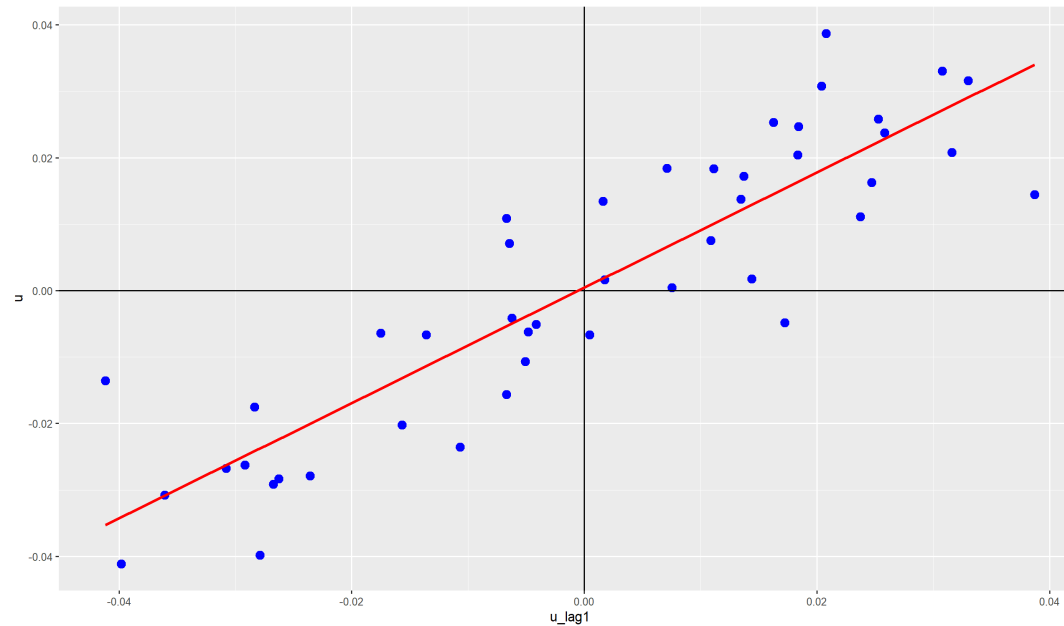
Observamos que tanto \hat{u}_t como \hat{u}_t estandarizados presentan un patrón, lo que indica que tal vez las \hat{u}_t no sean aleatorias

Detección de la autocorrelación

Para ver esto en forma diferente, podemos graficar \hat{u}_t respecto de \hat{u}_{t-1} , es decir, el residuo en el tiempo t frente a su valor en el tiempo $(t - 1)$, una clase de prueba empírica del esquema AR(1)

```
data <- data |> mutate(u_lag1 = lag(u, 1))

ggplot(data, aes(x=u_lag1, y=u)) +
  geom_point(color = "blue", size = 3) +
  geom_hline(yintercept = 0) + geom_vline(xintercept = 0) +
  geom_smooth(method = "lm", se = FALSE, formula= y~x, color="red")
```



Como muestra esta figura, la mayoría de los residuos están agrupados en el segundo y el cuarto cuadrantes, lo cual indica una **correlación positiva fuerte en los residuos**

Detección de la autocorrelación

2. Prueba d de Durbin-Watson

- La prueba más conocida para detectar correlación serial es la de los estadísticos Durbin y Watson
- Se le conoce como **estadístico d de Durbin-Watson**, que se define como:

$$d = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2}$$

- Aunque el estadístico d se utiliza ahora en forma rutinaria, **es importante observar los supuestos en los cuales se basa**:
 - El modelo de regresión incluye el término del intercepto
 - Las perturbaciones u_t se generan mediante el esquema autorregresivo de primer orden: $u_t = \rho u_{t-1} + \epsilon_t$. Por tanto, no se pueden utilizar para detectar esquemas autorregresivos de orden superior
 - Se supone que el término de error u_t está normalmente distribuido $\implies d \longrightarrow 2$
 - El modelo de regresión no incluye valor(es) rezagado(s) de la variable dependiente como una variable explicativa
 - No hay observaciones faltantes en los datos

Detección de la autocorrelación

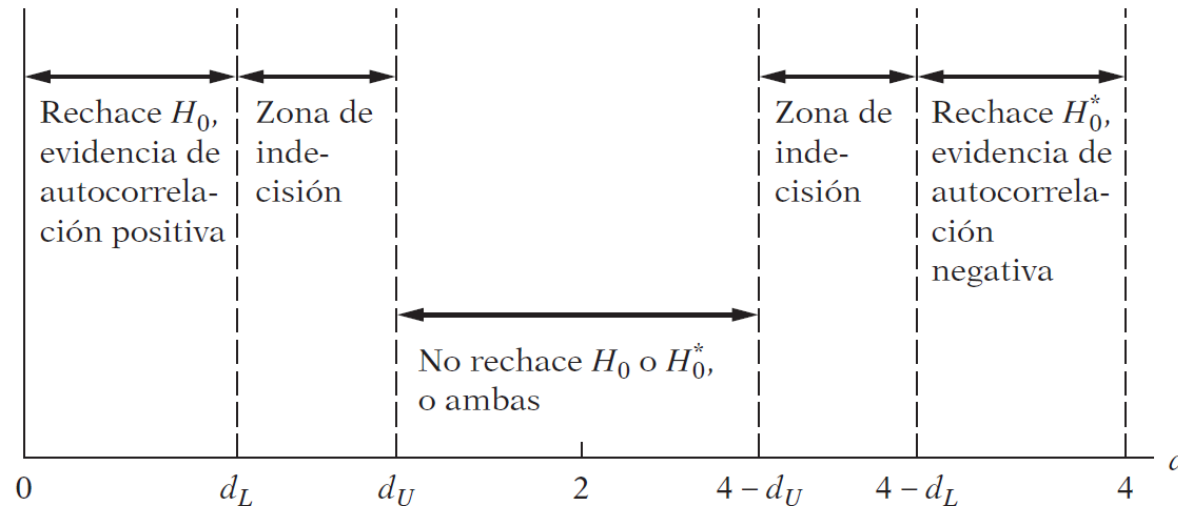
- El muestreo exacto o la distribución de probabilidad del estadístico d es difícil de derivar porque, como demostraron Durbin y Watson, tiene una dependencia compleja de los valores presentes de X en una muestra dada
- Por consiguiente, a diferencia de las pruebas t , F o χ^2 , no hay un valor crítico único que lleve al rechazo o al no rechazo de la hipótesis nula de que no hay correlación serial de primer orden en las perturbaciones u_t
- Sin embargo, Durbin y Watson lograron encontrar un límite inferior d_L y un límite superior d_U tales que si el valor d calculado cae por fuera de estos valores críticos, puede tomarse una decisión respecto de la presencia de correlación serial positiva o negativa
- Además, estos límites sólo dependen del número de observaciones n y del número de variables explicativas, y no de los valores que adquieren estas variables explicativas
- Para determinar la presencia o ausencia de autocorrelación del tipo AR(1) se siguen las siguientes reglas:
 - Ausencia de autocorrelación: $d = 2$
 - Autocorrelación positiva: $0 \leq d \leq 2$
 - Autocorrelación negativa: $2 < d \leq 4$

Detección de la autocorrelación

Como se mencionó, d no tiene una distribución exacta y, por tanto, se trabaja con umbrales críticos. La distribución se tabula por d_L y d_U :

- $0 \leq d < d_L$: autocorrelación lineal positiva
- $d_L \leq d < d_U$: resultado incierto - test no concluyente
- $d_U \leq d < 4 - d_U$: no hay autocorrelación
- $4 - d_U \leq d < 4 - d_L$: resultado incierto - test no concluyente
- $4 - d_L \leq d < 4$: autocorrelación lineal negativa

Gráficamente:



Leyendas

H_0 : No hay autocorrelación positiva

H_0^* : No hay autocorrelación negativa

Detección de la autocorrelación

3. Test basados en el correlograma

El correlograma de una serie temporal se basa en dos funciones:

- la función de autocorrelación simple (*Autocorrelation* - AC)
- la función de autocorrelación parcial (*Partial autocorrelation* - PAC)

Autocorrelation - AC

Se construye con base en las correlaciones simples de la serie con sus rezagos

$\hat{\rho}_1$: primer coeficiente de autocorrelación, es la correlación simple de la serie en t y en $t - 1$

$$\rho_1 = Cov(u_t, u_{t-1}) = \frac{Cov(u_t, u_{t-1})}{\sigma_{u_t} \sigma_{u_{t-1}}} = \frac{Cov(u_t, u_{t-1})}{Var(u_t)}$$

Empíricamente

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^T \hat{u}_t^2}$$

Y así en general el $\rho_s = \frac{Cov(u_t, u_{t-s})}{Var(u_t)}$ y en términos empíricos:

$$\hat{\rho}_s = \frac{\sum_{t=2}^T \hat{u}_t \hat{u}_{t-s}}{\sum_{t=2}^T \hat{u}_t^2}$$

Detección de la autocorrelación

Partial autocorrelation - PAC

Se construye con base en los coeficientes de correlación parcial entre las series y sus rezagos. La forma más sencilla de calcularlos son a partir de las ecuaciones de Yule y Walker

Se estiman las siguientes regresiones:

$$\hat{u}_t = \hat{\phi}_{11}\hat{u}_{t-1} + \textit{residuo}$$

$\hat{\phi}_{11}$: primer coeficiente de autocorrelación parcial (coincide con $\hat{\rho}_1$)

$$\hat{u}_t = \hat{\phi}_{12}\hat{u}_{t-1} + \hat{\phi}_{22}\hat{u}_{t-2} + \textit{residuo}$$

$\hat{\phi}_{22}$: segundo coeficiente de autocorrelación parcial

En general, $\hat{\phi}_{ss}$ en la regresión de $\hat{u}_t = f(\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-s})$ es el s -ésimo coeficiente de correlación parcial

Para construir el correlograma se requiere las respectivas varianzas:

$$Var(\hat{\rho}_s) \cong \frac{1}{T} (1 + 2(\hat{\rho}_1^2 + \hat{\rho}_2^2 + \dots + \hat{\rho}_{s-1}^2))$$

$$Var(\hat{\phi}_{ss}) \cong \frac{1}{T}$$

En general, $\frac{\hat{\rho}_s}{\sqrt{Var(\hat{\rho}_s)}} \text{ y } \frac{\hat{\phi}_{ss}}{\sqrt{Var(\hat{\phi}_{ss})}} \longrightarrow N(0, 1)$

Detección de la autocorrelación

La AC y PAC no solo se usan para diagnosticar autocorrelación, sino también para identificar el proceso generador de datos de la serie

Si el proceso fuese un AR(1): $u_t = \phi u_{t-1} + \epsilon_t$

AC	PAC
$\hat{\rho}_1 = \frac{\phi \sigma_u^2}{\sigma_u^2} = \phi$	$\phi_{11} = \phi$
$\hat{\rho}_2 = \phi^2$	$\phi_{12} = 0$
\vdots	\vdots
$\hat{\rho}_s = \phi^s$	$\phi_{1s} = 0$

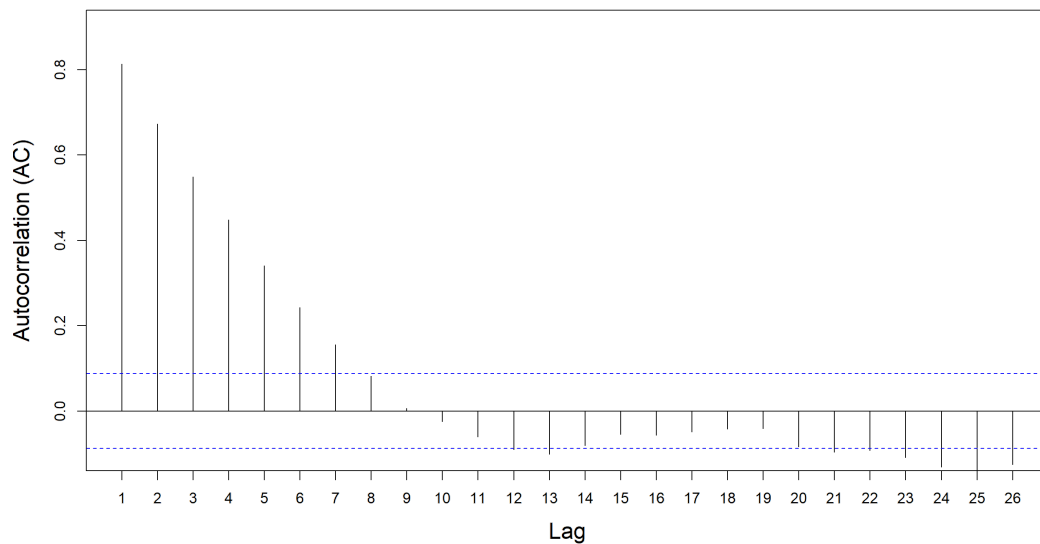
Mientras la AC disminuye lentamente, la PAC se corta abruptamente

Detección de la autocorrelación

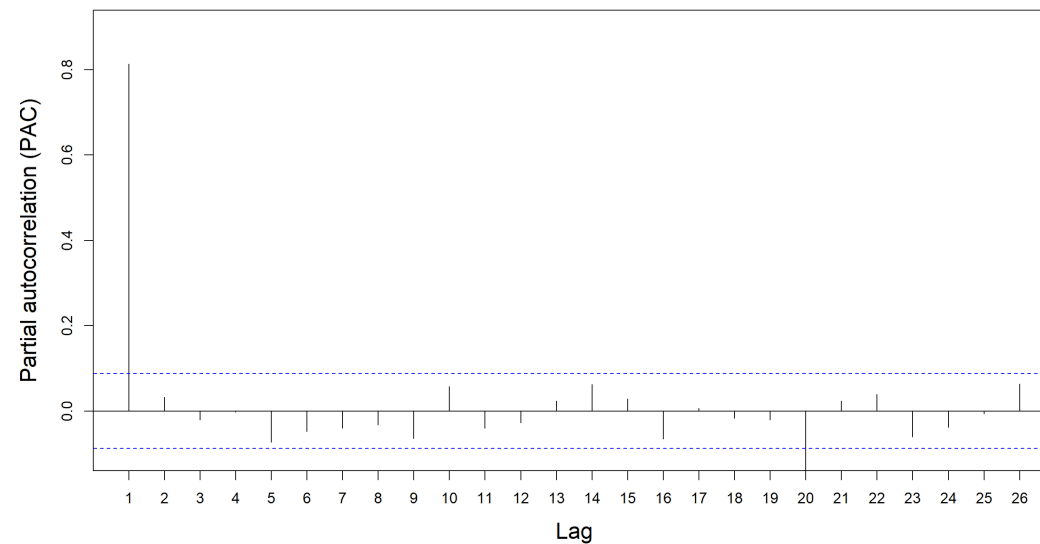
Simulamos un proceso AR(1)

```
ar1 <- arima.sim(list(ar=.8), n=500)
```

```
acf(ar1, 26, xlim=c(1,26), ylim=c(-.1,.9), main = "", ylab = "", xlab = ""  
axis(1, 1:26)  
mtext(side=1, line=3, "Lag",cex=1.5)  
mtext(side=2, line=3, "Autocorrelation (AC)",cex=1.5)
```



```
pacf(ar1, 26, ylim=c(-.1,.9), main = "", ylab = "", xlab = "", xaxt="none"  
axis(1, 1:26)  
mtext(side=1, line=3, "Lag",cex=1.5)  
mtext(side=2, line=3, "Partial autocorrelation (PAC)",cex=1.5)
```



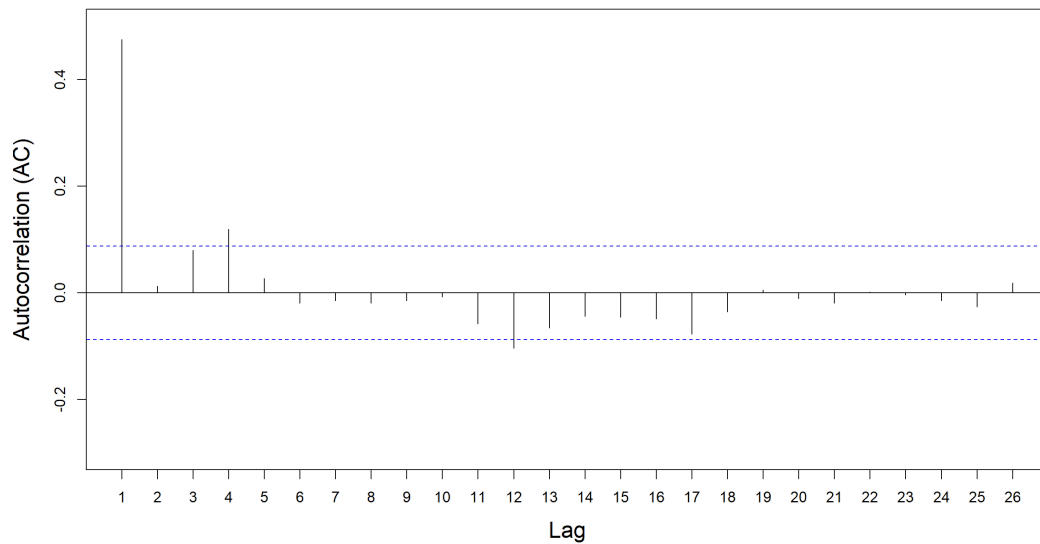
Detección de la autocorrelación

Con un proceso MA(1) sucede todo lo contrario: $u_t = \theta\epsilon_{t-1} + \epsilon_t$

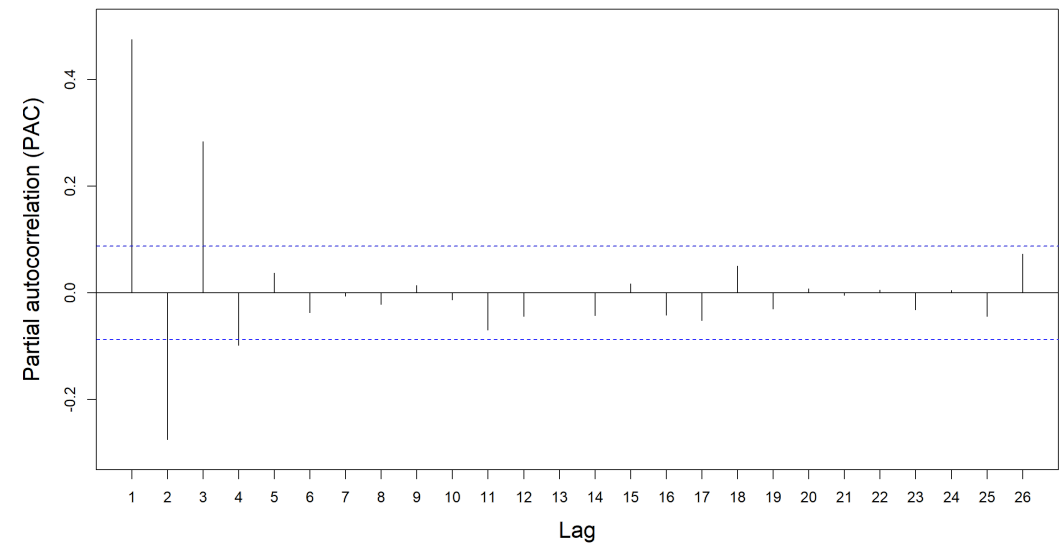
Simulamos un proceso MA(1)

```
ma1 <- arima.sim(list(ma=.7), n=500)
```

```
acf(ma1, 26, xlim=c(1,26), ylim=c(-.3,.5), main = "", ylab = "", xlab = ""  
axis(1, 1:26)  
mtext(side=1, line=3, "Lag",cex=1.5)  
mtext(side=2, line=3, "Autocorrelation (AC)",cex=1.5)
```



```
pacf(ma1, 26, ylim=c(-.3,.5), main = "", ylab = "", xlab = "", xaxt="none"  
axis(1, 1:26)  
mtext(side=1, line=3, "Lag",cex=1.5)  
mtext(side=2, line=3, "Partial autocorrelation (PAC)",cex=1.5)
```



En la modelación ARMA, el correlograma es fundamental para identificar p y q

Detección de la autocorrelación

4. La prueba de multiplicadores de Lagrange (LM) o de Breusch-Godfrey (BF)

Para evitar algunos inconvenientes de la prueba d de Durbin-Watson, se propuso una prueba para la autocorrelación que es general porque permite:

- regresoras no estocásticas, como los valores rezagados de la regresada
- esquemas autorregresivos de orden mayor, como el AR(1), AR(2), etc
- MA simples o de orden superior

El test funciona de la siguiente forma: asuma el siguiente modelo con una estructura AR(p) en u_t

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t$$

La hipótesis nula H_0 por demostrar es

$$H_0 = \rho_1 = \rho_2 = \dots = \rho_p = 0$$

Es decir, no existe correlación serial de ningún orden

Detección de la autocorrelación

La prueba LM implica los siguientes pasos:

1. Estime el modelo por MCO y obtenga \hat{u}_t
2. Haga la regresión \hat{u}_t sobre las X_t y $\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-p}$, y obtenga R^2 de esta regresión auxiliar
3. Si el tamaño de la muestra es grande, Breusch y Godfrey demostraron que

$$LM = nR^2 \sim \chi_p^2$$

Se rechaza H_0 si $LM > \chi_{tabla}^2$ o si $pvalue < 1, 5, 10\%$

Detección de la autocorrelación

5. Test de Ljung-Box

Diseñados para procesos AR(p)

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t$$

La hipótesis nula H_0 es

$$H_0 = \rho_1 = \rho_2 = \dots = \rho_p = 0$$

Es decir, no existe correlación serial de ningún orden

El estadístico de prueba es

$$Q = T(T+2) \left[\frac{\hat{\rho}_1^2}{T-1} + \frac{\hat{\rho}_2^2}{T-2} + \dots + \frac{\hat{\rho}_p^2}{T-p} \right] \longrightarrow \chi_p^2$$

Es un test de validez asintótica y su regla de decisión es rechazar H_0 si $Q > \chi_{\alpha}^2$ o si $p\text{value} < 1, 5, 10\%$

Solución a la autocorrelación: el estimador MCG

La idea es transformar el modelo (los datos y la perturbación aleatoria) de tal forma que la perturbación aleatoria del modelo transformado, tenga esfericidad y se puedan aplicar MCO a los datos del modelo transformado

$$\mathbf{Y} = \mathbf{XB} + \mathbf{u}$$

$$E(\mathbf{u}) = \mathbf{0}$$

$$E(\mathbf{X}'\mathbf{u}) = \mathbf{0}$$

$$Cov(\mathbf{u}) = E(\mathbf{uu}') = \sigma_u^2 \mathbf{\Omega}$$

Siendo $\mathbf{\Omega}$ una matriz definida positiva, pues se trata de varianzas

Las matrices definidas positivas pueden descomponerse como:

$$\mathbf{\Omega} = \mathbf{PP}'$$

Siendo \mathbf{P} una matriz no singular (\mathbf{P}^{-1} existe). En el mundo matricial, dadas las propiedades de la inversión de matrices, se da que:

$$\mathbf{\Omega}^{-1} = (\mathbf{PP}')^{-1} = \mathbf{P}'^{-1}\mathbf{P}^{-1} = \mathbf{P}^{-1'}\mathbf{P}^{-1}$$

La propuesta de los MCG es premultiplicar todo el modelo por \mathbf{P}^{-1}

$$\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{XB} + \mathbf{P}^{-1}\mathbf{u}$$

$$\mathbf{Y}^* = \mathbf{X}^*\mathbf{B} + \mathbf{u}^*$$

Solución a la autocorrelación: el estimador MCG

Si la perturbación \mathbf{u}^* es esférica se puede aplicar MCO al modelo con base en \mathbf{Y}^* y \mathbf{X}^* . Hay que ver los supuestos para \mathbf{u}^*

$$\begin{aligned}E(\mathbf{u}^*) &= E(\mathbf{P}^{-1}\mathbf{u}) = \mathbf{P}^{-1}E(\mathbf{u}) = \mathbf{0} \\Cov(\mathbf{u}^*) &= E((\mathbf{u}^* - E(\mathbf{u}^*))(\mathbf{u}^* - E(\mathbf{u}^*))') = E(\mathbf{u}^*\mathbf{u}^{*'}) \\&= E(\mathbf{P}^{-1}\mathbf{u}\mathbf{u}'\mathbf{P}^{-1'}) = \mathbf{P}^{-1}E(\mathbf{u}\mathbf{u}')\mathbf{P}^{-1'} \\&= \sigma_u^2\mathbf{P}^{-1}\mathbf{\Omega}\mathbf{P}^{-1'} \\&= \sigma_u^2\mathbf{P}^{-1}\mathbf{P}\mathbf{P}'\mathbf{P}^{-1'} \\&= \sigma_u^2\mathbf{I}\end{aligned}$$

Por lo tanto, en el modelo $\mathbf{Y}^* = \mathbf{X}^*\mathbf{B} + \mathbf{u}^*$ se cumple la hipótesis de perturbaciones esféricas y se puede aplicar MCO al modelo transformado, dando como resultado $\hat{\mathbf{B}}_{MCG}$

$$\begin{aligned}\hat{\mathbf{B}}_{MCG} &= (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{Y}^* \\&= ((\mathbf{P}^{-1}\mathbf{X})'(\mathbf{P}^{-1}\mathbf{X}))^{-1}(\mathbf{P}^{-1}\mathbf{X})'\mathbf{P}^{-1}\mathbf{Y} \\&= (\mathbf{X}'\mathbf{P}^{-1'}\mathbf{P}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}^{-1'}\mathbf{P}^{-1}\mathbf{Y} \\&= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{Y}\end{aligned}$$

$\hat{\mathbf{B}}_{MCO}$ son un caso particular cuando $\mathbf{\Omega} = \mathbf{I}$

Es inmediato plantear que en el modelo transformado

$$Cov(\hat{\mathbf{B}}_{MCG}) = \sigma_u^2(\mathbf{X}'\mathbf{\Omega}\mathbf{X})^{-1}$$

Para obtener $\mathbf{\Omega}$ hay que modelar el tipo de situación específica que se quiere resolver, estos es si la autocorrelación es del tipo AR(1), MA(1), etc

Solución a la autocorrelación: el estimador MCG

Descartado un origen de la autocorrelación en la incorrecta especificación del modelo se debe encontrar un proceso generador de datos (PGD) que de cuenta de la autocorrelación. Suponiendo que dicho proceso sea un AR(1), es decir:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{u}$$

$$u_t = \rho u_{t-1} + \epsilon_t$$

con ϵ_t ruido blanco

Es posible demostrar que

$$\mathbf{P}^{-1} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & . & . & 0 \\ -\rho & 1 & 0 & . & . & 0 \\ 0 & -\rho & 1 & . & . & -\rho \\ . & . & . & . & . & . \\ . & . & . & . & 1 & 0 \\ 0 & 0 & 0 & . & -\rho & 1 \end{bmatrix}_{T \times T}$$

Recordando que la solución MCG es transformar el modelo y obtener

$$\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{XB} + \mathbf{P}^{-1}\mathbf{u}$$

Solución a la autocorrelación: el estimador MCG

En matrices se tiene

$$\mathbf{P}^{-1}\mathbf{Y} = \begin{bmatrix} \sqrt{1-\rho^2}Y_1 \\ Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ \vdots \\ Y_t - \rho Y_{t-1} \end{bmatrix} \quad \mathbf{P}^{-1}\mathbf{X} = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}X_{21} & \dots & \sqrt{1-\rho^2}X_{k1} \\ 1-\rho & X_{22} - \rho X_{21} & \dots & X_{k2} - \rho X_{k1} \\ 1-\rho & X_{23} - \rho X_{22} & \dots & X_{k3} - \rho X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1-\rho & X_{2T} - \rho X_{2T-1} & \dots & X_{kT} - \rho X_{kT-1} \end{bmatrix}$$

Se conoce como primeras diferencias **GENERALIZADAS**, cada observación menos ρ veces la anterior

En estricto rigor el MCG lleva a una regresión:

- sin intercepto
- modificando la primera observación por $\sqrt{1-\rho^2}$
- el resto de observaciones como primera diferencia generalizada

En el modelo transformado la primera observación viene generada por:

$$\sqrt{1-\rho^2}Y_1 = \sqrt{1-\rho^2}\beta_1 + \sqrt{1-\rho^2}\beta_2X_{12} + \dots + \sqrt{1-\rho^2}\beta_kX_{k1} + \epsilon_1$$

El resto de las observaciones $t = 2, 3, \dots, T$ viene dado por:

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + \dots + \beta_k(X_{kt} - \rho X_{kt-1}) + \epsilon_t$$

Solución a la autocorrelación: el estimador MCG

En el anterior modelo transformado sólo falta encontrar un valor para ρ . Existen varias formas de hacerlo

- De manera directa a partir del Durbin-Watson

$$d = 2(1 - \hat{\rho}) \implies \hat{\rho} = 1 - \frac{d}{2}$$

- Estimando ρ a partir del residuo:

se estima $Y_t = X_t' \hat{\beta} + \hat{u}_t$

se estima $\hat{u}_t = \hat{\rho} \hat{u}_{t-1} + \epsilon$

$\hat{\rho}$: primero coeficiente de autocorrelación

- Método de Durbin

La idea viene de estimar el modelo en primeras diferencias generalizadas y despejar $\hat{\rho}$. Se estima un modelo en el cual $Y_t = f(Y_{t-1}, X_{2t}, X_{2t-1}, \dots, X_{kt}, X_{kt-1})$, y la propuesta de Durbin es tomar el coeficiente estimado de Y_{t-1} como $\hat{\rho}$

- Malla de Hildreth-Lu

Es un procedimiento de búsqueda típico de los métodos de análisis numérico. La idea es buscar numéricamente el $\hat{\rho}$ que minimiza la SCR del modelo transformado

Solución a la autocorrelación: el estimador MCG

- Procedimiento iterativo de Cochran-Orcutt

En estricto rigor es una solución de MV al problema de Autocorrelación. También busca minimizar la SCR pero con una estrategia diferente

- Se estima el modelo inicial $Y_t = X_t' \hat{\beta} + \hat{u}_t$
- Se obtiene un estimativo de $\hat{\rho}$ (a partir del \hat{d} o del residuo)
- Se transformo el modelo (normalmente excluyendo la primera observación) y se vuelve a obtener $\hat{\hat{\beta}}$ y unos nuevos residuos
- Se repite el proceso hasta la convergencia

$$\hat{\beta} \longrightarrow \hat{u}_t \longrightarrow \hat{\rho} \longrightarrow \hat{\hat{\beta}} \longrightarrow \hat{\hat{u}} \longrightarrow \hat{\hat{\rho}} \longrightarrow \dots$$

Note que en vista de que se utiliza $\hat{\rho}$ en vez de la verdadera ρ , todos estos métodos de estimación se conocen en la bibliografía como métodos MCG factibles (MCGF) o MCG estimados (MCGE).

Solución a la autocorrelación: el estimador MCG

Existen tres procedimientos muy utilizados en el ejercicio empírico para corregir autocorrelación (procedimientos disponibles en softwares como R o Stata):

- el método de Prais–Winsten: procedimiento iterativo, asume estructura $AR(1)$, preserva la 1a obs
- el método Cochrane–Orcutt: procedimiento iterativo, asume estructura $AR(1)$, pierde la 1a obs
- el método Newey-West (corrige heterocedasticidad y autocorrelación)

Ejercicio aplicado en R: relación entre salarios y productividad

En este ejemplo vamos analizar la relación entre salarios y productividad en el sector de negocios de Estados Unidos entre 1960 y 2005. Se tienen datos sobre índices de remuneración real por hora (Y) y producción por hora (X), la base de los índices es 1992=100

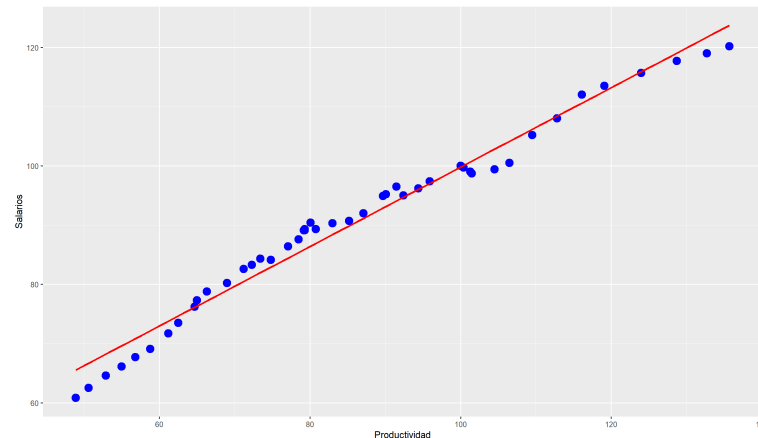
- Código en R

```
library(tidyverse); library(Hmisc); library(forecast); library(lmtest); library(prais); library(orcutt); library(sandwich); library(gujarati)

data("Table12_4")
data <- Table12_4 |> mutate(year = as.numeric(as.character(Year)),
                             y = as.numeric(as.character(Y)), x = as.numeric(as.character(X))) |> select(year, y, x)
```

Graficamos salarios contra productividad

```
ggplot(data, aes(x=x, y=y)) +
  geom_point(color = "blue", size = 4) +
  labs(x="Productividad", y="Salarios") +
  geom_smooth(method = "lm", se = FALSE, formula= y~x, color="red")
```



Ejercicio aplicado en R: relación entre salarios y productividad

Estimamos un modelo de regresión en logs

```
modelo <- lm(log(y) ~ log(x), data=data)
summary(modelo)
```

```
Call:
lm(formula = log(y) ~ log(x), data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.041164	-0.017041	0.001037	0.018077	0.038719

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.60668	0.05471	29.37	<2e-16 ***
log(x)	0.65222	0.01235	52.80	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

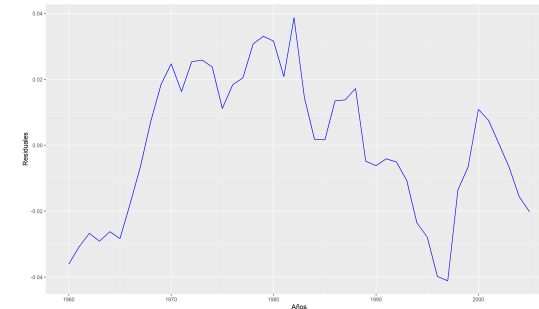
Residual standard error: 0.02208 on 44 degrees of freedom
Multiple R-squared: 0.9845, Adjusted R-squared: 0.9841
F-statistic: 2788 on 1 and 44 DF, p-value: < 2.2e-16

Detección de la autocorrelación

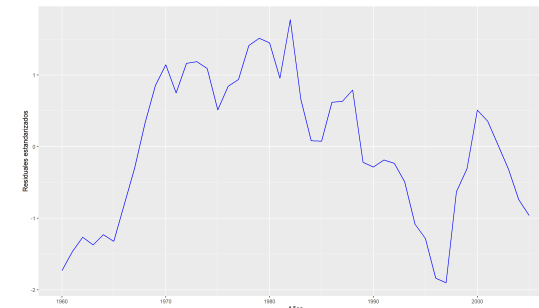
Graficando los residuales

```
data <- data |> mutate(u = modelo$residuals, u_e = rstandard(modelo))
```

```
ggplot(data, aes(x=year, y=u)) +  
  geom_line(color = "blue") + labs(x="Años", y="Residuales")
```



```
ggplot(data, aes(x=year, y=u_e)) +  
  geom_line(color = "blue") + labs(x="Años", y="Residuales estandarizados")
```

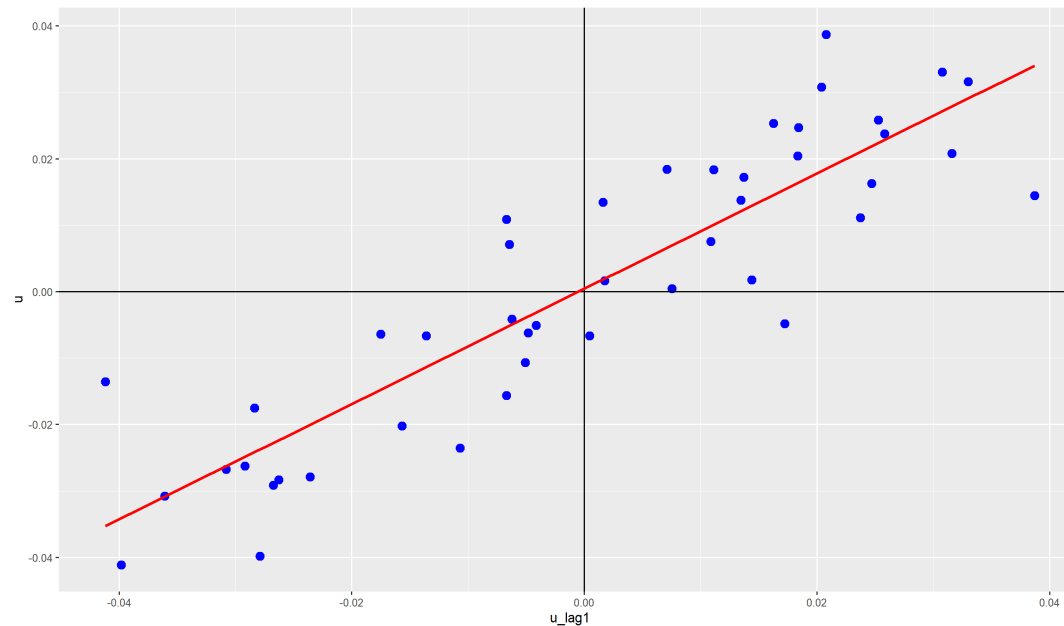


Observamos que tanto u como u_e presentan un patrón, lo que indica que tal vez los residuales no sean aleatorios

Ejercicio aplicado en R: relación entre salarios y productividad

Graficando u_t contra u_{t-1}

```
data <- data |> mutate(u_lag1 = lag(u, 1))  
  
ggplot(data, aes(x=u_lag1, y=u)) +  
  geom_point(color = "blue", size = 3) +  
  geom_hline(yintercept = 0) + geom_vline(xintercept = 0) +  
  geom_smooth(method = "lm", se = FALSE, formula= y~x, color="red")
```



Ejercicio aplicado en R: relación entre salarios y productividad

Test de Durbin-Watson

```
dwtest(modelo)
```

Durbin-Watson test

```
data: modelo
DW = 0.21756, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

Test LM o de Breusch-Godfrey (BF)

```
bgtest(modelo, order=1)
```

Breusch-Godfrey test for serial correlation of order up to 1

```
data: modelo
LM test = 34.02, df = 1, p-value = 5.456e-09
```

```
coeftest(bgtest(modelo, order=1))
```

z test of coefficients:

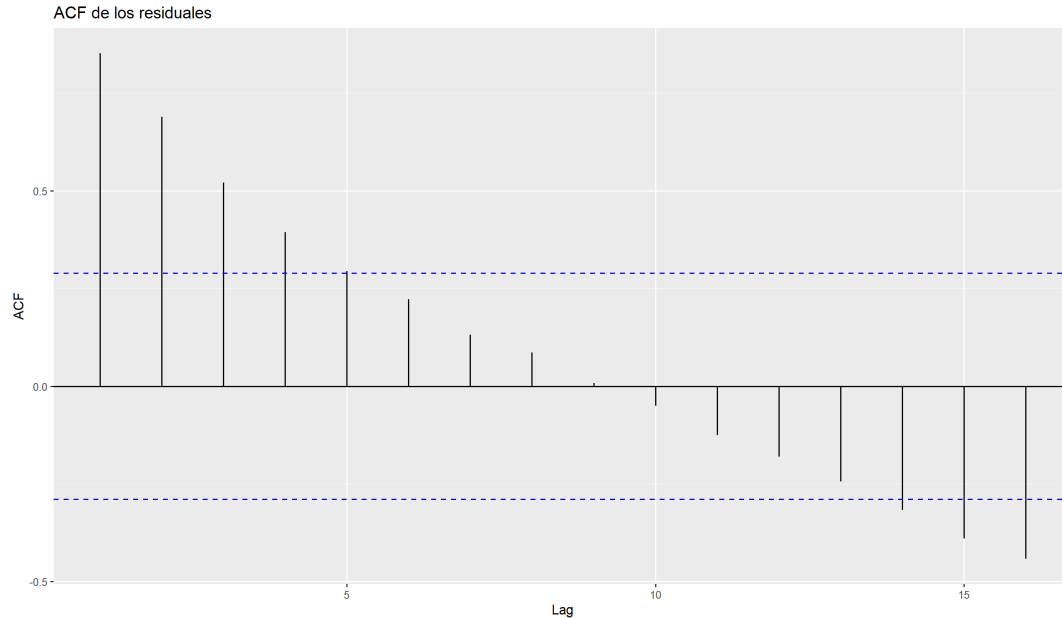
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0062584	0.0282482	0.2216	0.8247
log(x)	-0.0015019	0.0063783	-0.2355	0.8138
lag(resid)_1	0.8686522	0.0786108	11.0500	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

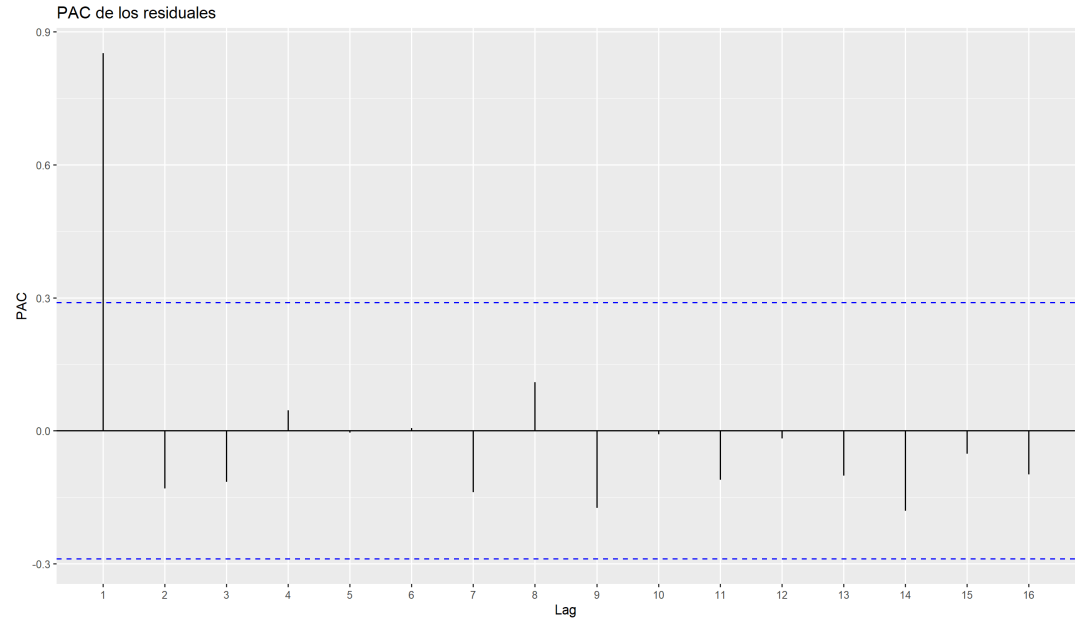
Ejercicio aplicado en R: relación entre salarios y productividad

Test basados en el correlograma

```
ggAcf(data$u,main="ACF de los residuales")
```



```
ggPacf(data$u,main="PAC de los residuales", ylab="PAC")
```



Se detecta un problema de autocorrelación en los residuales y es del tipo AR(1)

Ejercicio aplicado en R: relación entre salarios y productividad

Test de Ljung-Box

```
Box.test(data$u, lag = 1, type = "Ljung-Box")
```

Box-Ljung test

data: data\$u
X-squared = 35.566, df = 1, p-value = 2.465e-09

Ejercicio aplicado en R: relación entre salarios y productividad

Corrección de la autocorrelación

El método de Prais-Winsten: corrige AR(1)

```
pw <- prais_winsten(log(y)~log(x), data = data, index = data$year)
```

```
Iteration 0: rho = 0
Iteration 1: rho = 0.8679
Iteration 2: rho = 0.8912
Iteration 3: rho = 0.8953
Iteration 4: rho = 0.8962
Iteration 5: rho = 0.8964
Iteration 6: rho = 0.8964
Iteration 7: rho = 0.8964
Iteration 8: rho = 0.8964
Iteration 9: rho = 0.8964
```

```
summary(pw)
```

```
Call:
prais_winsten(formula = log(y) ~ log(x), data = data, index = data$year)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.03546 -0.01079  0.00693  0.02630  0.04666
```

```
AR(1) coefficient rho after 9 iterations: 0.8964
```

```
Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.56421    0.13458   11.62 5.3e-15 ***
log(x)       0.66009    0.03035   21.75 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.01014 on 44 degrees of freedom
Multiple R-squared:  0.9974,    Adjusted R-squared:  0.9974
```

El método Cochrane-Orcutt: corrige AR(1)

```
coch <- cochrane.orcutt(modelo)
summary(coch)
```

```
Call:
```

```
lm(formula = log(y) ~ log(x), data = data)
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.955389    0.192416  10.162 5.283e-13 ***
log(x)       0.576819    0.041914  13.762 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.0096 on 43 degrees of freedom
Multiple R-squared:  0.815 , Adjusted R-squared:  0.8107
F-statistic: 189.4 on 1 and 43 DF, p-value: < 2.349e-17
```

```
Durbin-Watson statistic
```

```
(original):    0.21756 , p-value: 5.098e-19
(transformed): 1.70388 , p-value: 1.253e-01
```

El método Newey-West: corrige con orden de rezago más alto

```
coeftest(modelo,vcov=NeweyWest(modelo,lag=2,verbose=T))
```

```
Lag truncation parameter chosen: 2
```

```
t test of coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.606680    0.187208  8.5823 5.975e-11 ***
log(x)       0.652216    0.043447  15.0118 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```