Temas selectos de econometría y finanzas (tarea 3)

J. Antonio García Ramirez

1 de octubre de 2018

Ejercicio 1

Escribe una función en R que realice las pruebas de raíces unitarias en el contexto de cointegración considerando los valores críticos de Engle y Yoo (1987), Ouliaris et al. (1989) y MacKinnon (1991). Toma en cuenta las dimensiones de las variables, etc.

Basandome en el paper de Engle y Yoo impplemente la prueba con la siguiente función

```r
#Prueba de cointegracion de Engle y Yoo
Cointegracion.Engle <- function(x,y, significancia='0.95')
{
  # entradas
  #  x (numeric): Vector con la primer serie
  #  y (numeric): Vector con la segunda serie
  #  significancia (numeric): Nivel de significancia tres posibles casos
  #       '0.90', '0.95' y '0.9'
  datos <- as.data.frame(cbind(x,y))
  names(datos) <- c('x', 'y')
  n <- dim(datos)[1]
    # Regresamos la segunda serie en la primera y obtenemos los residuos de las
    # primeras diferencias
  residuos <- lm( y~ x, data=datos)$residuals
  res.z <- diff(residuos)
  res.lag <- residuos[2:n]
      # Consideramos una regresion para un modelo AR(1) con las diferencias de los residuos
  datos2 <- as.data.frame(cbind(res.z,res.lag))
  pos.root <- lm(res.z ~. , data=datos2)
        # extraemos el estadistico t de la prueba, para considerar la significancia
  resumen <- summary(pos.root)$coef
  valor.t <- abs(resumen[, 't value'][2] )
  valor.p <- resumen[, 'Pr(>|t|)'][2]
  print(paste0("Valor del estadistico t:   ", round(valor.t,  2)))
  # Generamos los valores criticos para el modelo tau p,
  # el modelo sin tendencia
  # considerando los tamaños de muestra que sugiere el paper
  tabla <- data.frame(tamanio=c(50,100,200,500,1000),
                      caso.z1=c(3.58,3.51,3.46,3.44,3.43),
                      caso.z2=c(2.93,2.89,2.88,2.87,2.86),
                      caso.z3=c(2.6,2.58,2.57,2.57,2.57))
  significancias <- c('caso.z1', 'caso.z2', 'caso.z3')
  names(significancias) <- c('0.99', '0.95', '0.90')
  valor.paper <- significancias[as.character(significancia)]
  #Comparamos con el valor critico
  if (n<=50)
  {
    if(tabla[1, as.character(valor.paper)]< valor.t){
      print("Series no cointegradas")
```

```r
    } else{ print("Series cointegradas") }
  } else if(n>50 & n<=100)
    {
     if(tabla[2, as.character(valor.paper)] < abs(valor.t))
       {
        print("Series no cointegradas")
     }   else{ print("Series cointegradas")}
  } else if(n>100 & n<=200)
    {
      if(tabla[3, as.character(valor.paper)]< abs(valor.t))
        {
          print("Series no cointegradas")
    }
    else{print("Series cointegradas")}
  } else if(n>200 && n<=500){
    if(tabla[4, as.character(valor.paper)] < abs(valor.t))
      {
        print("Series no cointegradas")
    } else{ print("Series cointegradas") }
  }
  else if(n>500){
    if(tabla[5, as.character(valor.paper)] < abs(valor.t))
      {
        print("Series no cointegradas")
    }
    else{print("Series cointegradas")}
  }
}
```

Y para probarla genero dos series no cointegradas de tamaños 50 y 1000

```r
set.seed(0)
x <- runif(50, 0, 1)
y <- runif(50, 0, 1)
Cointegracion.Engle(x, y, significancia='0.90')
```

```
## [1] "Valor del estadistico t:   6.36"
## [1] "Series no cointegradas"
```

```r
x <- runif(1000, 0, 1)
y <- runif(1000, 0, 1)
```

Basandome en el paper de MacKinnon impplemente la prueba con la siguiente función:

```r
Cointegracion.Mckinnon <- function(x,y, significancia='0.95',
                                   tipo = c("none", "trend", "both"))
{
  # entradas
  #  x (numeric): Vector con la primer serie
  #  y (numeric): Vector con la segunda serie
  #  significancia (character): Nivel de significancia tres posibles casos
  #       '0.90', '0.95' y '0.99'
  # tipo (character): Tipo de regresión, sin tendencia ni intercepto 'none',
  #                   tendencia lineal 'trend', tendencia e intercepto 'both'
  datos <- data.frame(x=x, y=y)
  n <- dim(datos)[1]
```

```r
#  consideramos una regresion entre ambas las series y obtenemos los residuos
residuos <- lm( y~ x, data=datos)$residuals
res.z <- diff(residuos)
res.lag <- residuos[2:n]
datos2 <- data.frame(res.z = res.z,res.lag = res.lag)
temp <- c(1, 2, 3)
names(temp) <- c('0.99', '0.95', '0.9')
resultado <- temp[as.character(significancia)]
if (tipo == "none")
{
  #consideramos el caso AR(1) con las diferencias de los residuos sin intercepto
  root <- lm(res.z ~ .-1, data=datos2)
   # extraemos el valor critico
  valor.t <- summary(root)$coef[, 3]
  print(paste("Estadistico t: ",round(valor.t,2)))
   # guardamos los valores criticos
  tabla <- data.frame(tamanio=c(0.99,0.95,0.9),
                      inf= c(-2.565, -1.941,-1.616),
                      beta1 = c(-2.235, -0.268,0.265),
                      beta2= c(-3.627,-3.365,-2.714),
                      beta3= c(0, 31.223,25.364))
  limite <- tabla[resultado,2] + (tabla[resultado,3]/ n)+
    (tabla[resultado,4]/(n^2)) + (tabla[resultado,5]/(n^3))
  if (valor.t > - abs(limite) | valor.t < abs(limite))
  {
    print("Series no cointegradas")
  }   else{print("Series cointegradas")}
}
if (tipo=="trend")
{
  #consideramos el caso AR(1) con las diferencias de los residuos con tendencia
  root <- lm(res.z~. , data=datos2)
  valor.t <- summary(root)$coef[, 3][2]
  print(paste("Valor estadistico t:",round(valor.t,2)))
         # guardamos los valores criticos
  tabla <- data.frame(size=c(0.99,0.95,0.9),
                      inf=c(-3.430,-2.861,-2.566),
                      beta1=c(-6.539,-2.890,-1.538),
                      beta2=c(-16.786,-4.234,-2.809),
                      beta3=c(-79.433,-40.040,0))
  limite <- tabla[resultado,2] + (tabla[resultado,3]/n)+(tabla[resultado,4]/(n^2))+
    (tabla[resultado,5]/(n^3))
  if (valor.t > -abs(limite) | valor.t < abs(limite))
  {
    print("Series no cointegradas")
  } else{print("Series cointegradas")}
}
if (tipo=="both")
{
  #consideramos el caso AR(1) con las diferencias de los residuos con tendencia e intercepto
  datos2$t <- 1:(n-1)
  root <- lm(res.z~. , data=datos2)
  valor.t <- summary(root)$coef[, 3][2]
```

```
    print(paste("Valor estadistico t:",round(valor.t, 2)))
     # guardamos los coeficientes del paper
    tabla <- data.frame(size=c(0.99,0.95,0.9),
                        inf=c(-3.958,-3.410,-3.127),
                        beta1=c(-6.539,-2.890,-1.538),
                        beta2=c(-16.786,-4.234,-2.809),
                        beta3=c(-79.433,-40.040,0))
    limite <- tabla[resultado,2]+ (tabla[resultado,3]/n)+
      (tabla[resultado,4]/(n^2)) + (tabla[resultado,5]/(n^3))
    if (valor.t > -abs(limite) | valor.t <abs(limite)){
      print("Series no cointegradas")
    }    else{print("Series cointegradas")}
  }
}
```

Y para probarla genero dos series no cointegradas de tamaños 50 y 1000

```
set.seed(0)
x <- runif(50, 0, 1)
y <- runif(50, 0, 1)
Cointegracion.Mckinnon(x, y, tipo = 'none')
```

```
## [1] "Estadistico t:  6.42"
## [1] "Series no cointegradas"
```

```
set.seed(0)
x <- runif(1000, 0, 1)
y <- runif(1000, 0, 1)
Cointegracion.Mckinnon(x, y, tipo = 'both')
```

```
## [1] "Valor estadistico t: 32.11"
## [1] "Series no cointegradas"
```

Basandome en el paper de Phillips Ouliaris impplemente la prueba con la siguiente función:

```
Cointegracion.Outliaris <-function (x, y, media = TRUE)
{
   # entradas
  #  x (numeric): Vector con la primer serie
  #  y (numeric): Vector con la segunda serie
  # media (numeric): Booleano para quitar tendencia
  library(sandwich)
  datos <- data.frame(x=x, y=y)
   # regresion sin intercepto
  residuos <- lm(x ~ .-1, data=datos)$residuals
  z <- embed(residuos, 2)
  ut <- z[, 1]
  ut1 <- z[, 2]
  n <- dim(z)[1] #cuidado con el lag
  residuos2 <- lm(ut ~ ut1 - 1) #consideramos la regresion sin intercepto
  resumen.residuos2 <- summary(residuos2)
  k <- residuos2$residuals
  k.ssqr <- sum(k^2)/n
  l <- trunc(n/100)
     #varianza de largo plazo Newey-West
  tl.ssqr <- lrvar(k, prewhite = FALSE, type="Newey-West")
```

```r
  alpha <- resumen.residuos2$coefficients[1, 1]
  estadistico <- n*(alpha - 1) - 0.5 * n**2 * (tl.ssqr - k.ssqr)/(sum(ut1^2))
  if (media) #checamos que caso es
  {
    # capturamos los valores limite del paper para tendencia
    tabla <- cbind(c(28.32, 34.17, 41.13, 47.51, 52.17),
                   c(20.49, 26.09,32.06, 37.15, 41.94),
                   c(17.04, 22.19, 27.58, 32.74,37.01))
  } else {
        # capturamos los valores limite del paper para sin tendencia
    tabla <- cbind(c(22.83, 29.27, 36.16, 42.87, 48.52),
                   c(15.64, 21.48,27.85, 33.48, 38.09),
                   c(12.54, 18.18, 23.92, 28.85,33.8))
  }
  tablep <- c(0.01, 0.05, 0.1)
  p.valor <- approx(tabla[ncol(datos) - 1, ], tablep, estadistico, rule = 2)$y
  if( estadistico > p.valor)
  {
    print('Series cointegradas')
  } else print('Series no cointegradas')
  print(paste("Estadistico :", estadistico))
  print(paste("p-value: ",p.valor))
}
```

Y para probarla genero dos series no cointegradas de tamaños 50 y 1000

```r
set.seed(0)
x <- runif(50, 0, 1)
y <- runif(50, 0, 1)
Cointegracion.Outliaris(x, y, media=TRUE)
```

```
## [1] "Series no cointegradas"
## [1] "Estadistico : -11.9314387526224"
## [1] "p-value:  0.1"
```

```r
set.seed(0)
x <- runif(1000, 0, 1)
y <- runif(1000, 0, 1)
Cointegracion.Outliaris(x, y, media=FALSE)
```

```
## [1] "Series no cointegradas"
## [1] "Estadistico : -361.241969099302"
## [1] "p-value:  0.1"
```

Ejercicio 2

Programa una función en R que genere el procedimiento de Sargan y Bhargava (1983).

Basandome en el paper programe el procedimiento para los estadisticos R y R^* de la siguiente forma:

```r
Sargan.Bhargava <- function(x, y, significancia='0.99')
{
    # entradas
  #  x (numeric): Vector con la primer serie
  #  y (numeric): Vector con la segunda serie
```

```r
  # significancia (character): Nivel de significancia puede ser '0.99' o '0.95'
  datos <- data.frame(x=x, y=y)
   # consideramos la regresion de una serie contra la otra
    regresion <- lm(x~y , datos)
  residuos <- diff(resid(regresion))
    #construimos una matriz para los residuos
    n <- dim(datos)[1]
  S <- matrix(0,nrow = n-1,ncol = n-1)
    S[lower.tri(S)] <- 1 # hacer 1 la parte inferior
    diag(S) <- 1
    S <- rbind(rep(0,n-1),S) #agregamos filas
    K <- as.matrix(diff(x))
    I <- diag(n)
    X.star <- cbind(rep(1,n),x)
    Iota <- diag(n-1) - K%*%solve(t(K)%*%K)%*%t(K) #las cuentas del paper
    Iota.cero <- diag(n) - (1/n * rep(1, n) %*% t(rep(1,n)))
    Psi <- I - X.star%*% solve( t(X.star) %*% X.star ) %*% t(X.star)
    R.star<- (t(residuos) %*% Iota %*% residuos) / ( t(residuos) %*%
    R <- (t(residuos)%*%residuos ) / ( t(residuos)%*%t(S) %*%
                                   Iota.cero %*% S %*%residuos )

    # capturamos los valores criticos
    tabla <- matrix(c(1.592, 1.022, 0.747, 0.484, 0.257, 2.404, 1.560, 1.156, 0.755, .404,         2.095
    colnames(tabla) <- c(11,21,31,51,101)
    row.names(tabla)<-c("0.95L","0.95U","0.99L","0.99U")
  posicion.tabla <- which.min(abs(n-as.integer(colnames(tabla))))
    print(paste0("Estadísstico R: ", round(R,2)))
    fila <- match( paste0(significancia,"L"), row.names(tabla))
    fila2 <- match( paste0(significancia,"U"), row.names(tabla))
    if(R < tabla[fila, posicion.tabla] ){
    print("Series cointegradas")
  } else if(R > tabla[fila2, posicion.tabla]){
     print("series no cointegradas")
  }
    print(paste0("Estadístico R*:  ", round(R.star,2)))
    if(R.star < tabla[fila ,posicion.tabla] ){
      print("Series cointegradas")
    } else if(R.star > tabla [fila2, posicion.tabla])
      {
      print("series no cointegradas")
    }
}
```

Y para probarla genero dos series no cointegradas de tamaños 50 y 1000

```r
set.seed(0)
x <- runif(50, 0, 1)
y <- runif(50, 0, 1)
Sargan.Bhargava(x, y, significancia = '0.99')
```

```
## [1] "Estadísstico R: 2.01"
## [1] "series no cointegradas"
## [1] "Estadístico R*:  1.68"
## [1] "series no cointegradas"
```

```
set.seed(0)
x <- runif(1000, 0, 1)
y <- runif(1000, 0, 1)
Sargan.Bhargava(x, y,significancia = '0.95')
```

```
## [1] "Estadísstico R: 1.94"
## [1] "series no cointegradas"
## [1] "Estadístico R*:  2.03"
## [1] "series no cointegradas"
```

Ejercicio 3

Con los procedimientos anteriores y los visto en clase, piensa en un problema que sea de tu interés, i.e, donde sea evidente relacionar variables no estacionarias y concluye si existe evidencia de que existen relaciones de largo plazo.

Utilizaré dos series que considere de mi interés, correspondientes a las ventas de 50 articulos en diez tiendas diferentes, los datos con los que trabajaré en el proyecto final, disponibles en kaggle

Considere el número de articulos que se vendieron de los 2 primeros articulos de la primer tienda.

Notemos que las series no son estacionales, como lo indica el test aumentado de Dickey-Fuller

```
datos <- read.csv('train.csv')
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
names(datos)
```

```
## [1] "date"  "store" "item"  "sales"
```

```
cuentas <- datos%>% select(store, item)%>%group_by(store,item) %>% summarise(total=n())
item1 <- datos[1:1826, ]
item2 <- datos[1827:(2*1826), ]
adf.test(x)
```

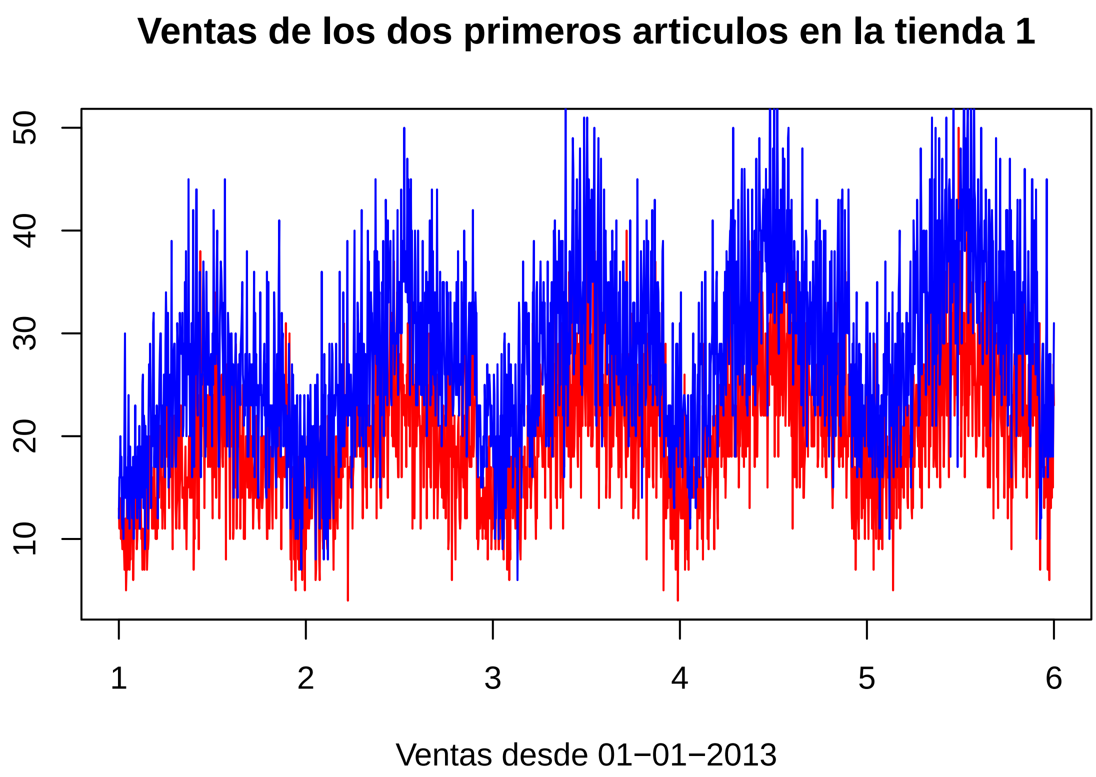
```
## Warning in adf.test(x): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -9.8728, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

```
adf.test.custom(x)
```

```
## Lag optimo
```

```
##          1
```

```
#adf.test(y)
x <- ts(item1$sales, start = 1, frequency = 365)
y <- ts(item2$sales, start = 1, frequency = 365)
ts.plot(x, col='red', xlab='Ventas desde 01-01-2013', ylab='', main='Ventas de los dos primeros articul
lines(y, col='blue')
```



Es de notar que las series se comportan de manera parecida, comportándose a la alza y a la baja en los mismos intervalos. Veamos si las ventas para los dos artículos diferentes en la misma tienda se comportan como series cointegradas.

Uilizando los test implementados en esta tarea no encontramos indicaciones de que las series sean cointegradas, lo cual juega encontra de la tesis de que la cointegración entre las series pueda ser útil para mejorar el pronostico de venta.

```
Cointegracion.Engle(x, y)
```

```
## [1] "Valor del estadistico t:   36.83"
## [1] "Series no cointegradas"
```

```
Cointegracion.Mckinnon(x, y, tipo='none')
```

```
## [1] "Estadistico t:  36.84"
## [1] "Series no cointegradas"
```

```
Cointegracion.Outliaris(x, y, media=TRUE)
```

```
## [1] "Series no cointegradas"
## [1] "Estadistico : -917.579177408375"
```

```
## [1] "p-value:  0.1"
```

```r
Sargan.Bhargava(x, y)
```

```
## [1] "Estadísstico R: 1.85"
## [1] "series no cointegradas"
## [1] "Estadístico R*:  1.44"
## [1] "series no cointegradas"
```

Sin embargo al considerar las primeras diferencias vemos que las series son estacionales, sin embargo los test implementados tampoco arrojan información que apoye el supuesto de cointegración

```r
x1 <- diff(x, lag=12)
y1 <- diff(y, lag=12)
adf.test(x1)
```

```
## Warning in adf.test(x1): p-value smaller than printed p-value
```

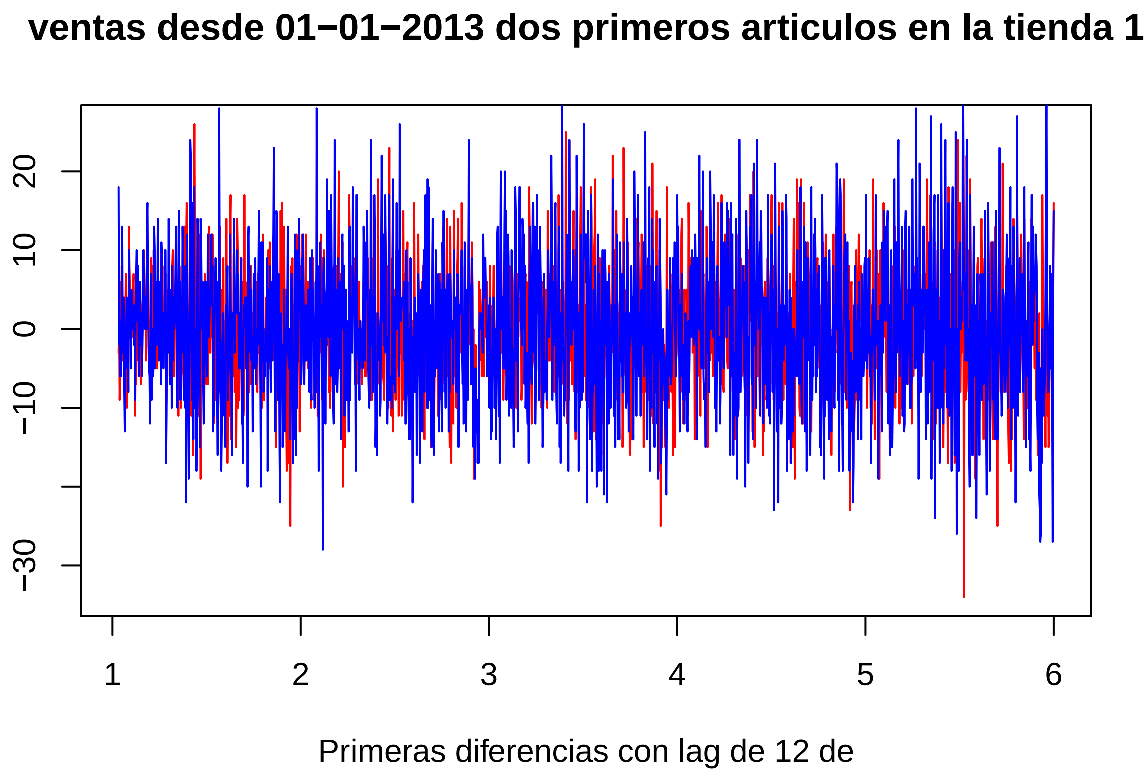
```
##
##  Augmented Dickey-Fuller Test
##
## data:  x1
## Dickey-Fuller = -16.03, Lag order = 12, p-value = 0.01
## alternative hypothesis: stationary
```

```r
adf.test(y1)
```

```
## Warning in adf.test(y1): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y1
## Dickey-Fuller = -15.488, Lag order = 12, p-value = 0.01
## alternative hypothesis: stationary
```

```r
plot(x1, col='red', xlab='Primeras diferencias con lag de 12 de', ylab='', main='ventas desde 01-01-2013
lines(y1, col='blue')
```




```r
Cointegracion.Engle(x1, y1)
```

```
## [1] "Valor del estadistico t:   38.2"
## [1] "Series no cointegradas"
```

```r
Cointegracion.Mckinnon(x1, y1, tipo='none')
```

```
## [1] "Estadistico t:  38.21"
## [1] "Series no cointegradas"
```

```r
Cointegracion.Outliaris(x1, y1, media=TRUE)
```

```
## [1] "Series no cointegradas"
## [1] "Estadistico : -825.846362036764"
## [1] "p-value:  0.1"
```

```r
Sargan.Bhargava(x1, y1)
```

```
## [1] "Estadísstico R: 1.91"
## [1] "series no cointegradas"
## [1] "Estadístico R*:  1.75"
## [1] "series no cointegradas"
```

Como podemos verificiar la regresión del segundo articulo con las primeras diferencias del primer articulo menciona que el coeficiente de la regresión es significativo sin embargo el ajuste es pobre pues el R^2 ajustado es de 0.12

```r
summary(lm(y1~ x1))
```

```
##
```

```
## Call:
## lm(formula = y1 ~ x1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.865  -5.753   0.108   5.913  33.413
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.03121    0.20782    0.15    0.881
## x1           0.42595    0.02695   15.80   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.851 on 1812 degrees of freedom
## Multiple R-squared:  0.1211, Adjusted R-squared:  0.1206
## F-statistic: 249.7 on 1 and 1812 DF,  p-value: < 2.2e-16
```

Concluimos con el comentario acerca de que **no tenemos elementos para afirmar la relación a largo plazo de las ventas para estos dos articulos**, sin embargo valdrá la pena considerar más elementos de las series como en primer lugar determinar la estacionalidad de las mismas y seguir con los test implementados pues como pudimos observar las series no son estacionales.