Homework 3

Ze Cheng Zheng

February 2021

1 Introduction

The Monte Carlo analysis is a board class of computer algorithms that rely on repeated random sampling to obtain numerical results. What Granger and Newbold did to replicate this Monte Carlo scheme was that they did a spurious regression, Granger and Newbold regressed two different random walks together and put them through a loop where there is around 100 simulations to see the percentage of times that $H_0(H_0$ = no relationship) is rejected.They showed that out of 100 simulations of random walks, it showed there to be a high amount of times that H_0 was reject showing that there is a relationship, but there was a low R^2 meaning that the has a weak correlation of the linear relationship between the two random walks . In the first test there was 76 times out of 100 simulations that the H_0 was rejected but there was a 0.26 R^2. The H_0 shows that there is a relationship 76 out of the 100 times but our R^2 tells us that we have no very weak relationship. (If R^2 is higher than 0.7 then it is considered a strong relationship).

2 Replication and Results

After doing a single sample regression analysis of independent random walks our R^2 shows there is a low value of 0.18, which means there is a low correlation between the two independent random walks. I then did same exact regression, trying to replicate the Granger and Newbold analysis. I set my seed (beginning point) as 2000, I will have n=100 as the sample size and we are going to repeat the experiment r=100 times. Then I compare with the critical value at 5%, how much of those experiment times is higher than the critical value. This will tell us how many times H_0 will be rejected. The results is that we have a R^2 of 0.58 which tell us that we have a weak value and the data is not at all close to the fitted line. Also though code it shows 0.001 mean with a max of 1.000 telling me that majority or the R^2 is not greater than 0.05, which tells me there is not a relationship. But it also shows that we rejected the H_0 777 times out of the 1000 times of the repeated experiment showing that there is a relationship between the two independent walks. From doing this I have similar results as Granger and Newbold as they also had a low R^2 but had a high percentage of their H_0 being rejected.

Table 1: My Results vs Granger and NewBold's Results

Results	My Results	Granger and Newbold Levels (M=1)
Average R^2	0.58	0.26
Percent of time H_0 is rejected	77.7	76

3 Discussion and Conclusion

The results of my replication shows very similar results to the Granger and Newbold analysis where we both get very high percentages of rejecting the null hypothesis (H_0) which proves that there is a relationship but we also get a very low valued R^2 which proves there to be very little linear relationship between the two variables. This is does not make sense to me seeing as one thing shows that there is a relationship but another saying that there is not a relationship between the two random walks. This tells me that this series in this experiment is strongly auto-correlated, even though there should not be a relationship there is some how a relationship after repeating the experiment several times.

Monte Carlo Progam

Ze Zheng

2/18/2021

```r
#Ze Zheng

#Monte Carlo Simulation

library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──
```

```
## ✓ ggplot2 3.3.2     ✓ purrr   0.3.4
## ✓ tibble  3.0.6     ✓ dplyr   1.0.3
## ✓ tidyr   1.1.2     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.5.0
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

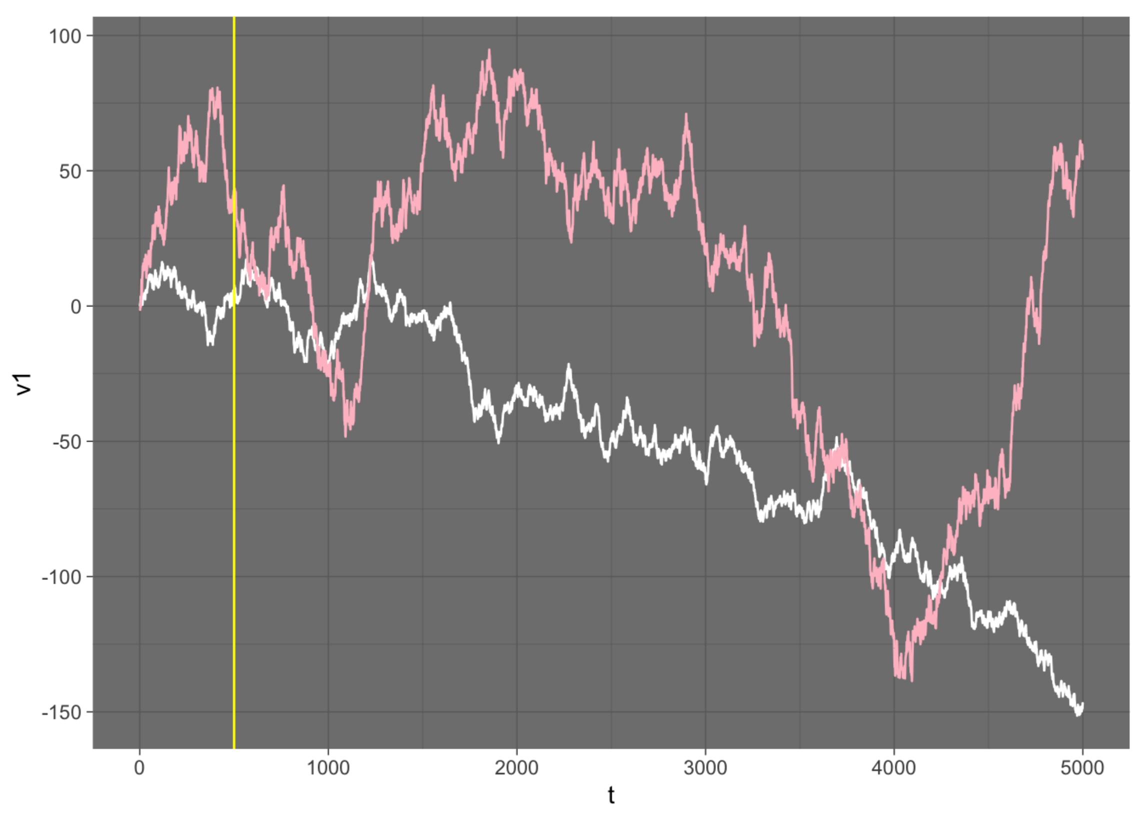
```r
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```r
library(urca)

u1 <- rnorm(5000,0,1)
u2 <- rnorm(5000,0,2)
v1 <- u1
v2 <- u2

t<-(1:5000)

#Random Walk  1
for (i in 2:5000) v1[i]<-v1[i-1]+u1[i]

#Random Walk  2

for (i in 2:5000) v2[i]<-v2[i-1]+u2[i]
ggplot()+
  geom_line(aes(x=t,y=v1),col="white")+
  geom_line(aes(x=t,y=v2),col="pink")+
  theme_dark()+
  geom_vline(xintercept = 500,color="yellow")
```




```r
# Random Walk 1 and Random Walk 2 Regression (w1~w2)
reg1<-lm(v1~v2)

# Stargazer Regression Result
stargazer(reg1, report="vcs", type="text",
          omit.stat = c("ser", "f"),
          ci=TRUE, omit.table.layout = "n")
```

```
##
## ================================
##              Dependent variable:
##              -------------------
##                      v1
## --------------------------------
## v2                  0.378
##                (0.360, 0.397)
##
## Constant           -50.446
##              (-51.499, -49.393)
##
## --------------------------------
## Observations        5,000
## R2                  0.246
## Adjusted R2         0.246
## ================================
```

```r
summary(reg1)
```

```
##
## Call:
## lm(formula = v1 ~ v2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -123.318  -17.888   -4.814   30.353   70.851
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -50.446087   0.537132  -93.92   <2e-16 ***
## v2            0.378368   0.009364   40.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.71 on 4998 degrees of freedom
## Multiple R-squared:  0.2462, Adjusted R-squared:  0.2461
## F-statistic:  1633 on 1 and 4998 DF,  p-value: < 2.2e-16
```

```r
sum= summary(reg1)
#Generating f-stats
fstat<-sum$fstatistic
fstat # value, numf, dendf find from fstats
```

```
##   value   numdf   dendf
## 1632.79    1.00 4998.00
```

```r
# f-stats only for value
fstat<-sum$fstatistic["value"]
fstat
```

```
##   value
## 1632.79
```

```r
fstat<-sum$r.squared
fstat # R^2
```

```
## [1] 0.2462437
```

```r
fstat<-sum$adj.r.squared
fstat # adjusted R^2
```

```
## [1] 0.2460929
```

```r
#Clears environment so that previous code does not conflict with new code
rm(list=ls())

# Taken From Applied Econometrics
#Econ 508 - Fall 2014
#from UIUC
#Monte Carlo
set.seed(2000)
r <- 1000
n <- 100
A <- array(0,c(1,1000))
rsq<-numeric(r) # a spot to save R-squared

#The Loop
for(i in 1:r){
    u<-rnorm(n)
    v<-rnorm(n)
    y<-rep(0,n)
    x<-rep(0,n)
    for(j in 2:n){
      x[j] <- x[j-1] + u[j]
      y[j] <- y[j-1] + v[j]
    }
    A[1,i] <- round(summary(lm(y~x))$coef["x","t value"],2)
}


#Generate the number of rejections out of 100 simulations
B<-as.matrix(abs(A)>abs(qt(0.05,n-1)))
apply(B,1,sum)
```

```
## [1] 777
```

```r
reg2<-(lm(y~x))
rsq[j]<-summary(reg2)$r.squared

# The result
reg2<-(lm(y~x))
stargazer(reg2, report="vcs", type="text",
          omit.stat = c("ser", "f"),
          ci=TRUE, omit.table.layout = "n")
```

```
##
## ================================
##              Dependent variable:
##              -------------------
##                       y
## --------------------------------
## x                   1.393
##                (1.159, 1.628)
##
## Constant            0.638
##                (-0.395, 1.671)
##
## --------------------------------
## Observations         100
## R2                  0.580
## Adjusted R2         0.576
## ================================
```

```r
#Summary of the regression and fstats
sum=summary(reg2)

#Generate Fstats
fstat<-sum$fstatistic
fstat # value, numf, dendf find from fstats
```

```
##    value    numdf    dendf
## 135.6108   1.0000  98.0000
```

```r
rej<-ifelse(rsq >= 0.05,1,0)# If R^2 is greater or = 0.05 (True of False)
summary(rej)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   0.001   0.000   1.000
```