

STAT 720 sp 2019 hw 2

due on Monday, Feb 25th, 2019

Please make use of my R package `tscourse` to complete the homework. You can install the package with the following commands:

```
library(devtools)
devtools::install_github("gregorkb/tscourse")
```

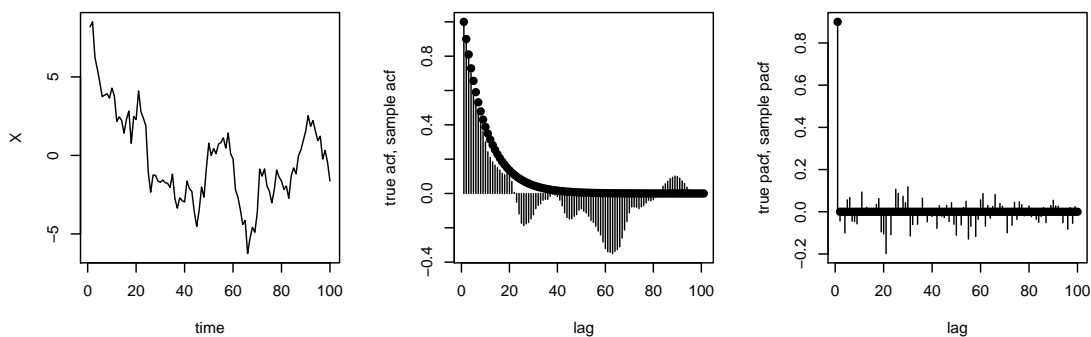
This pulls the package from where it resides in my github repository. You will first need to install the `devtools` package using `install.packages("devtools")`.

1. Consider the AR(1) model

$$X_t = \mu + \phi(X_{t-1} - \mu) + Z_t, \quad t \in \mathbb{Z}, \quad (1)$$

where $\{Z_t, t \in \mathbb{Z}\}$ are independent standard Normal random variables.

- (a) Generate a realization of length $n = 100$ from the above time series with $\phi = 0.9$ and $\mu = 0$. Then make three plots: Plot the simulated time series, the sample autocorrelation function with the true autocorrelations overlaid, and the sample partial autocorrelations with the true partial autocorrelations overlaid. Your plots should resemble the following:



The following R code performs the above:

```
phi <- .9
mu <- 0
sigma <- 1
n <- 100
B <- 1000 # length of burn-in period
X0 <- numeric(B+n)
X0[1] <- 0
```

```

for( i in 2:(B + n))
{
    X0[i] <- mu + phi * X0[i-1] + rnorm(1,0,sigma)
}

X <- X0[-c(1:B)]

gamma <- phi^c(0:n)/(1-phi^2)
rho <- gamma/gamma[1]
alpha <- DL.1step(X,gamma[1],gamma[-1])$alpha

gamma.hat <- sample.acf(X,max.lag=length(X))$gamma.hat
rho.hat <- gamma.hat/gamma.hat[1]
alpha.hat <- DL.1step(X,gamma.hat[1],gamma.hat[-1])$alpha

par(mfrow=c(1,3))

plot(X,type="l",xlab="time")

plot(rho,ylim=range(rho,rho.hat),pch=19,
     ylab="true acf, sample acf",xlab="lag")
points(rho.hat,type="h")

plot(alpha,ylim=range(alpha,alpha.hat),pch=19,
     ylab="true pacf, sample pacf",xlab="lag")
points(alpha.hat,type="h")

```

(b) Give an expression in terms of ϕ for the value σ_∞^2 such that

$$\sqrt{n}(\bar{X}_n - \mu) \rightarrow \text{Normal}(0, \sigma_\infty^2) \quad \text{in distribution}$$

as $n \rightarrow \infty$.

We have

$$\begin{aligned}
 \sigma_\infty^2 &= \sum_{h=-\infty}^{\infty} \gamma(h) = \sum_{h=-\infty}^{\infty} \frac{\phi^{|h|}}{1-\phi^2} \sigma^2 \\
 &= \frac{1}{1-\phi^2} \left(\sum_{h=0}^{\infty} \phi^h + \sum_{h=1}^{\infty} \phi^h \right) \sigma^2 \\
 &= \frac{1}{(1-\phi)(1+\phi)} \left(\frac{1}{1-\phi} + \frac{\phi}{1-\phi} \right) \sigma^2 \\
 &= \frac{1}{(1-\phi)^2} \sigma^2.
 \end{aligned}$$

- (c) Give an expression for a confidence interval for μ which will contain μ with probability 0.95 as $n \rightarrow \infty$.

In light of the previous part, we may use the interval

$$\bar{X}_n \pm 1.96\sigma_\infty n^{-1/2},$$

taking σ_∞ from the previous part.

- (d) For each of the sample sizes $n = 10, 50, 200$, generate 500 length- n realizations of the time series and construct the confidence interval you suggested in the previous part. Compute the proportion of data sets out of the 500 for which the confidence interval contained the true value of μ . Comment on what happens to the coverage probability of the confidence interval as n is increased.

The following code runs the simulations:

```
S <- 500
B <- 1000
n <- 200 # put different values of n here

sigma.inf <- sqrt( 1/(1 - phi)^2 )
X.bar <- lo.ci <- up.ci <- numeric(S)

for(s in 1:S)
{

  X0 <- numeric(B+n)
  X0[1] <- 0

  for( i in 2:(B + n))
  {
    X0[i] <- phi * X0[i-1] + rnorm(1)
  }

  X <- X0[-c(1:B)]

  X.bar[s] <- mean(X)

  lo.ci[s] <- X.bar[s] - 1.96 * sigma.inf/sqrt(n)
  up.ci[s] <- X.bar[s] + 1.96 * sigma.inf/sqrt(n)

}

coverage <- mean( (lo.ci < 0) & (up.ci > 0) )
```

For $n = 10$, I got a coverage of 0.996; for $n = 50$, 0.974; for $n = 200$, 0.946. As n is increased, the coverage approaches the nominal coverage of 95%.

2. Let $\hat{\gamma}_n(\cdot)$ be the sample autocovariance function computed on any sequence of numbers x_1, \dots, x_n .

(a) Show that $\sum_{h=-\infty}^{\infty} \hat{\gamma}_n(h) = 0$.

We have

$$\begin{aligned} \sum_{h=-\infty}^{\infty} \hat{\gamma}_n(h) &= \sum_{h=-(n-1)}^{n-1} \frac{1}{n} \sum_{i=1}^{n-|h|} (X_i - \bar{X}_n)(X_{i+|h|} - \bar{X}_n) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X}_n)(X_j - \bar{X}_n) \\ &= 0. \end{aligned}$$

(b) Suggest a strategy for estimating $\sigma_{\infty}^2 = \sum_{h=-\infty}^{\infty} \gamma(h)$.

The easiest way would be to assume a model for the time series which gives us an expression for $\sum_{h=-\infty}^{\infty} \gamma(h)$ in terms of some parameters. We could estimate these parameters and plug them into this expression to get an estimate of σ_{∞}^2 . In Question 1, for example, we saw that under the AR(1) model the sum $\sum_{h=-\infty}^{\infty} \gamma(h)$ was given by $(1 - \phi)^{-2} \sigma^2$.

3. For some finite $k \geq 1$, let $\boldsymbol{\rho}_k = (\rho(1), \dots, \rho(k))^T$ and $\hat{\boldsymbol{\rho}}_n(k) = (\hat{\rho}_n(1), \dots, \hat{\rho}_n(k))^T$, where $\rho(\cdot)$ is the autocorrelation function of a stationary time series and $\hat{\rho}_n(\cdot)$ is the sample autocorrelation function. Under some conditions on the time series (see Thm 7.2.2. of B&D Theory), we have

$$\sqrt{n}(\hat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k) \rightarrow \text{Normal}(\mathbf{0}, \mathbf{W}_k) \quad \text{in distribution}$$

as $n \rightarrow \infty$, where \mathbf{W}_k is the $k \times k$ matrix with entries given by

$$w_{ij} = \sum_{l=1}^{\infty} [\rho(l+i) + \rho(l-i) - 2\rho(i)\rho(l)][\rho(l+j) + \rho(l-j) - 2\rho(j)\rho(l)]$$

for $1 \leq i, j \leq k$. Now consider the AR(1) model from Question 1, which satisfies the conditions for the above to hold.

(a) Give an expression for the value of $\rho(h)$ for any $h \geq 1$.

From work done in class, we have that the autocovariance function is given by

$$\gamma(h) = \frac{\phi^{|h|}}{1 - \phi^2} \sigma^2,$$

so that

$$\rho(h) = \gamma(h)/\gamma(0) = \phi^{|h|}.$$

- (b) Find an expression for the asymptotic variance of $\sqrt{n}\hat{\rho}(1)$.

The asymptotic variance is given by w_{11} , which is

$$\begin{aligned} w_{11} &= \sum_{k=1}^{\infty} [\phi^{|k+1|} + \phi^{|k-1|} - 2\phi^k \phi]^2 \\ &= \sum_{k=1}^{\infty} [\phi^{k-1} - \phi^{k+1}]^2 \\ &= \sum_{k=1}^{\infty} [\phi^{2(k-1)} - \phi^{2(k+1)} - 2\phi^{k-1}\phi^{k+1}] \\ &= \sum_{k=1}^{\infty} [\phi^{-2}\phi^{2k} - \phi^2\phi^{2k} - 2\phi^{2k}] \\ &= (\phi^{-2} + \phi^2 - 2) \frac{\phi^2}{1 - \phi^2} \\ &= \frac{(1 + \phi^4 - 2\phi^2)}{1 - \phi^2} \\ &= 1 - \phi^2. \end{aligned}$$

- (c) Suggest a way to construct a 95% confidence interval for the parameter ϕ based on observed data (you will have to estimate the asymptotic variance by plugging in a value).

Since $\rho(1) = \phi$, one way to construct the interval would be

$$\hat{\rho}(1) \pm 1.96 \sqrt{\frac{1 - \hat{\rho}(1)^2}{n}}.$$

- (d) Evaluate the performance of your interval in a simulation study. Generate 500 realizations of length $n = 500$ from the AR(1) model from Question 1 with $\phi = 0.8$ and report the proportion of times your interval contained 0.8.

The following R code runs the simulation:

```
phi <- .8
```

```

S <- 500
B <- 1000
n <- 500

sigma.inf <- sqrt( 1/(1 - phi)^2 )
rho.hat.1 <- lo.ci <- up.ci <- numeric(S)

for(s in 1:S)
{

  X0 <- numeric(B+n)
  X0[1] <- 0

  for( i in 2:(B + n))
  {
    X0[i] <- phi * X0[i-1] + rnorm(1)
  }

  X <- X0[-c(1:B)]

  rho.hat.1[s] <- sample.acf(X,max.lag=10)$rho.hat[2]

  lo.ci[s] <- rho.hat.1[s] - 1.96 * sqrt( 1 - rho.hat.1[s]^2)/sqrt(n)
  up.ci[s] <- rho.hat.1[s] + 1.96 * sqrt( 1 - rho.hat.1[s]^2)/sqrt(n)

}

coverage <- mean( (lo.ci < phi) & (up.ci > phi) )

I got a coverage probability of 0.962.

```

4. Consider the ARMA(1, 1) process defined by

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1} \quad \text{for } t \in \mathbb{Z},$$

where $\{Z_t, t \in \mathbb{Z}\}$ is $\text{WN}(0, \sigma^2)$ and $|\phi| < 1$ and $|\theta| < 1$.

(a) Give the $\text{MA}(\infty)$ representation of the time series $\{X_t, t \in \mathbb{Z}\}$.

We have

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

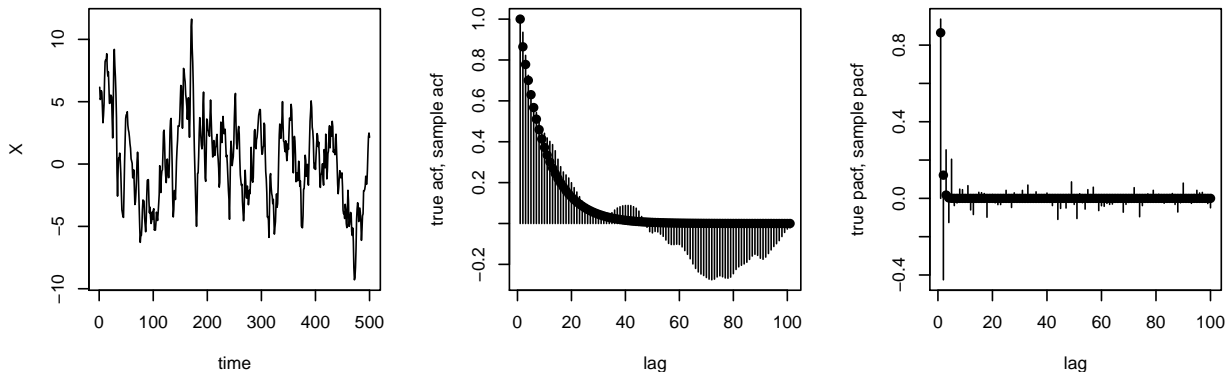
where

$$\psi_j = \begin{cases} \phi^{j-1}(\phi + \theta), & \text{for } j \geq 1 \\ 1, & \text{for } j = 0. \end{cases}$$

(b) Show that the autocovariance function is given by

$$\gamma(h) = \begin{cases} \left[\frac{(\theta + \phi)^2}{1 - \phi^2} + 1 \right] \sigma^2, & \text{if } h = 0 \\ \phi^{|h|-1}(\theta + \phi) \left[1 + \frac{(\theta + \phi)\phi^2}{1 - \phi^2} \right] \sigma^2, & \text{if } |h| \geq 1. \end{cases}$$

(c) Use the $\text{MA}(\infty)$ representation of $\{X_t, t \in \mathbb{Z}\}$ to generate realizations of length $n = 500$ of the time series by truncating to an $\text{MA}(50)$ representation. Do this under your own choices of the parameters ϕ , θ , and σ^2 and then make three plots: Plot a realization of the time series, the sample autocorrelation function with the true autocorrelations overlaid, and the sample partial autocorrelation function with the true partial autocorrelations overlaid. *Report the values of ϕ , θ , and σ^2 that you chose.* My plots with $\phi = 0.9$, $\theta = 0.7$, and $\sigma^2 = 1$ look like the following:



The following R code generates the data and produces the plots:

```
phi <- .9
theta <- .7

B <- 50 # truncate the MA(inf) representation of the time series
n <- 500

psi <- c(1, phi^c(0:(B-1))*(phi + theta))

Z <- rnorm(n+B)
X <- numeric(n)

for(i in 1:n)
{
  ind <- (B+i):i
  X[i] <- sum(psi*Z[ind])
}
```

```

}

max.lag <- 100
gamma <- c((theta + phi)^2/(1-phi^2) + 1,
          phi^c(0:(max.lag-1))*(theta+phi)*(1+(theta+phi)*phi^2/(1-phi^2)))
rho <- gamma/gamma[1]
alpha <- DL.1step(X[1:max.lag],gamma[1],gamma[-1])$alpha

gamma.hat <- sample.acf(X,max.lag=max.lag)$gamma.hat
rho.hat <- gamma.hat/gamma.hat[1]
alpha.hat <- DL.1step(X[1:max.lag],gamma.hat[1],gamma.hat[-1])$alpha

par(mfrow=c(1,3))

plot(X,type="l",xlab="time")

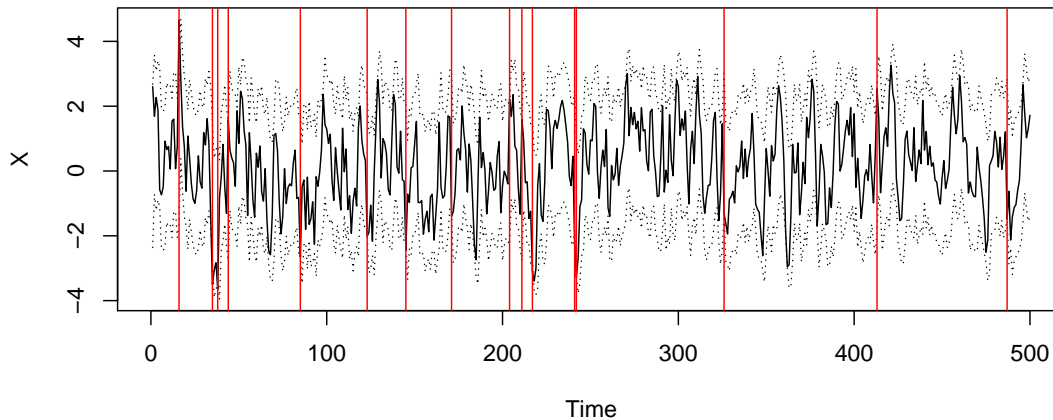
plot(rho,ylim=range(rho,rho.hat),pch=19,
      ylab="true acf, sample acf",xlab="lag")
points(rho.hat,type="h")

plot(alpha,ylim=range(alpha,alpha.hat),pch=19,
      ylab="true pacf, sample pacf",xlab="lag")
points(alpha.hat,type="h")

```

- (d) Generate a realization of the times series of length $n = 500$ under the choices $\phi = 0.5$, $\theta = 0.2$ and $\sigma^2 = 1$. Then use the Durbin-Levinson algorithm to obtain the one-step-ahead predictions for X_1, \dots, X_n . Moreover, construct 95% prediction intervals for the one-step-ahead predictions using the output of the Durbin-Levinson algorithm and make a plot of the time series which shows the one-step-ahead prediction intervals. Report the proportion of times the one-step-ahead prediction interval contained the next value in the series.

My simulation produced the plot below, in which the vertical lines mark the time points at which the value of the time series fell outside of the prediction interval. The prediction interval on my data set contained the next observation 95.6% of the time.



The following R code produces the plot and computes the coverage over X_1, \dots, X_n of the 95% one-step-ahead prediction interval:

```
gamma <- c( (theta + phi)^2/(1-phi^2) + 1,
            phi^c(0:(n-1))*(theta + phi)*(1 + (theta+phi)*phi^2/(1-phi^2)) )

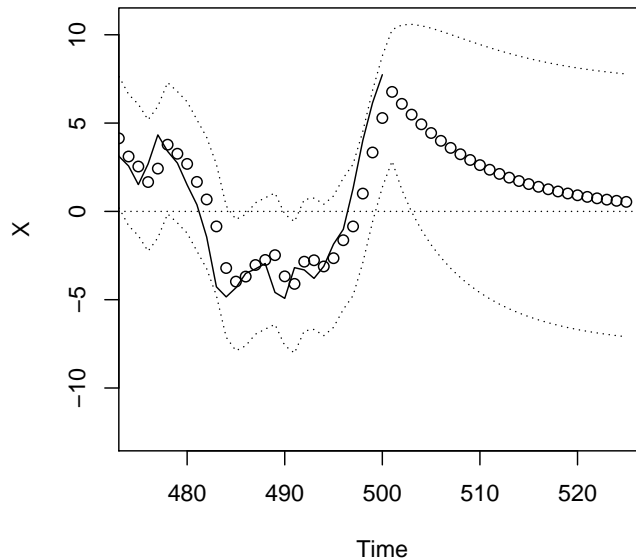
DL.1step.out <- DL.1step(X-mean(X),gamma[1],gamma[-1])
X.pred <- DL.1step.out$X.pred + mean(X)

lo.pred <- X.pred - 1.96 * sqrt(DL.1step.out$v)
up.pred <- X.pred + 1.96 * sqrt(DL.1step.out$v)

coverage <- mean(lo.pred[-(n+1)] < X & X < up.pred[-(n+1)])

plot(X,type="l",ylim=range(lo.pred[-1],up.pred[-1]),xlab="Time")
lines(lo.pred[-(n+1)],lty=3)
lines(up.pred[-(n+1)],lty=3)
abline(v= which(lo.pred[-(n+1)] > X | X > up.pred[-(n+1)]),col="red")
```

- (e) Now use the innovations algorithm to make h -step ahead predictions: Generate a realization of the time series of length $n = 500$ under $\phi = 0.9$, $\theta = 0.7$, and $\sigma^2 = 1$, and get one-step-ahead predictions for X_1, \dots, X_{n+1} as well as predictions of X_{n+2}, \dots, X_{n+25} . Make a plot of the time series with the predicted values and 95% prediction intervals overlaid. Plot only over the range of time points from $n - h, \dots, n + h$, for $h = 25$. My plot looks like this:



The following R code generates the plot.

```

phi <- .9
theta <- .7

B <- 50
n <- 500

psi <- c(1,phi^c(0:(B-1))*(phi + theta))

Z <- rnorm(n+B)
X <- numeric(n)

for(i in 1:n)
{
  ind <- (B+i):i
  X[i] <- sum(psi*Z[ind])
}

h <- 25
gamma <- c( (theta + phi)^2/(1-phi^2)+1,
            phi^c(0:(n+h-2))*(theta+phi)*(1+(theta+phi)*phi^2/(1-phi^2)))

K <- matrix(NA,n+h,n+h)
for(j in 1:(n+h))
  for(i in 1:(n+h))

```

```

      {
          K[i,j] <- gamma[1+abs(i-j)]
      }

innov.hstep.out <- innov.hstep(X-mean(X),h,K)
X.pred <- innov.hstep.out$X.pred + mean(X)

lo.pred <- X.pred - 1.96 * sqrt(innov.hstep.out$v)
up.pred <- X.pred + 1.96 * sqrt(innov.hstep.out$v)

plot(X,type="l",xlim=c(n-h,n+h),ylim=range(lo.pred,up.pred),xlab="Time")
points(innov.hstep.out$X.pred)
abline(h=0,lty=3)
lines(lo.pred,lty=3)
lines(up.pred,lty=3)

```

5. Problem 3.9 from B&D Intro. The data set for parts (b) and (c) is in the R package `datasets` and can be read into R with the command `data(USAccDeaths)`.

For the model

$$Y_t = Z_t + \theta_1 Z_{t-1} + \theta_{12} Z_{t-12},$$

where $\{Z_t, t \in \mathbb{Z}\}$ is $\text{WN}(0, \sigma^2)$, we have

$$\begin{aligned}
 \text{Cov}(Y_t, Y_{t+h}) &= \text{Cov}(Z_t + \theta_1 Z_{t-1} + \theta_{12} Z_{t-12}, Z_{t+h} + \theta_1 Z_{t+h-1} + \theta_{12} Z_{t+h-12}) \\
 &= \mathbb{E}[Z_t Z_{t+h} + \theta_1 Z_t Z_{t+h-1} + \theta_{12} Z_t Z_{t+h-12} \\
 &\quad + \theta_1 Z_{t-1} Z_{t+h} + \theta_1^2 Z_{t-1} Z_{t+h-1} + \theta_1 \theta_{12} Z_{t-1} Z_{t+h-12} \\
 &\quad + \theta_{12} Z_{t-12} Z_{t+h} + \theta_{12} \theta_1 Z_{t-12} Z_{t+h-1} + \theta_{12}^2 Z_{t-12} Z_{t+h-12}] \\
 &= \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_{12}^2), & \text{if } h = 0 \\ \sigma^2 \theta_1, & \text{if } h \pm 1 \\ \sigma^2 \theta_{12} \theta_1, & \text{if } h \pm 11 \\ \sigma^2 \theta_{12}, & \text{if } h \pm 12 \\ 0, & \text{otherwise.} \end{cases}
 \end{aligned}$$

We see that we may write

$$\theta_1 = \frac{\gamma(11)}{\gamma(12)}, \quad \theta_{12} = \frac{\gamma(11)}{\gamma(1)}, \quad \text{and} \quad \sigma^2 = \gamma(1) \frac{\gamma(12)}{\gamma(11)}.$$

The following R code computes estimators $\hat{\theta}_1$, $\hat{\theta}_{12}$ and $\hat{\sigma}^2$ of θ_1 , θ_{12} and σ^2 on the data $\{\nabla \nabla_{12} X_t\}$, where $\{X_t, t = 1, \dots, 72\}$ are the values in the dataset `USAccDeaths`:

```
data(USAccDeaths)
Y <- diff(diff(USAccDeaths,12))
gamma.hat <- sample.acf(Y)$gamma.hat

theta1.hat <- gamma.hat[12]/gamma.hat[13]
theta12.hat <- gamma.hat[12]/gamma.hat[2]
sigma.hat <- sqrt(gamma.hat[2]/theta12.hat)
```

We get $\hat{\theta}_1 = -0.5923191$, $\hat{\theta}_{12} = -0.5627678$ and $\hat{\sigma} = 308.8488$.