

Please read the relevant pages in your textbook. You should be aware of the following points:

- Autocorrelation is found in many time series.
- Autocorrelation does not cause a bias in your parameter estimates.
- Autocorrelation, however, will create a bias in the standard errors of the estimates- for positive autocorrelation, the estimated standard error will be smaller than the true standard error.
- The result is that one can no longer trust the t-statistics from OLS.

In a time series, autocorrelated errors are those in which the correlation between an error and its lagged values is non-zero. In other words, the current value of the residual is conditioned by previous values.

The figure below shows, on the right, the original data (black dots) and the fitted linear regression line (red line). On the left, the residuals are plotted. In this figure, if a residual is a large positive value, then its immediately preceding neighbor is also likely to be a large positive value; if negative, then its preceding neighbor is also likely to be negative. In cases like this, where the sign and size of adjacent residuals are similar, we have *positive autocorrelation*.

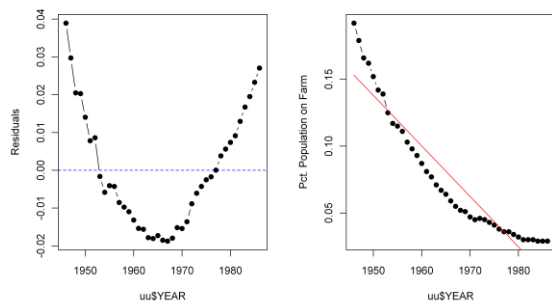


Figure 1: Residuals (left) and model fit (right).

Autocorrelation between a residual and its lagged-one value is *first order autocorrelation*; autocorrelation over two lags is termed *second order*, and so on.

How does one test for temporal autocorrelation?

The **Durbin-Watson statistic** is the most common test for autocorrelation. The null hypothesis for this test is that there is no autocorrelation. In general, values of the Durbin Watson statistic close to 2 allow one to reject

the null hypothesis. There are several important limitations of the Durbin-Watson test:

- The statistic tests only for correlation between the current error and the immediately preceding error (first order autocorrelation).
- The statistic is biased (towards 2, thus falsely showing that there is no autocorrelation) when lagged values of the dependent variable are used as independent variables.
- The test often falls into the "indeterminate" range—i.e., it gives an ambiguous result.

For these reasons, experienced econometricians rarely use the Durbin-Watson test. The test favored by most practitioners is the **Breusch-Godfrey test**, or the Lagrange multiplier test for autocorrelation. The idea behind the test is rather simple:

- Estimate your model, storing your error terms.
- Conduct another regression (the auxiliary regression) in which you regress all of the independent variables, and as many lagged error terms as you think appropriate, on the current value of the error.
- The Breusch-Godfrey statistic is created from elements of this auxiliary regression as follows:
 $(T-p) \cdot R^2$ where T =number of observations, p =number of lagged error terms, and R^2 is the R^2 .
- Your null hypothesis is that this regression will explain very little (i.e., that there is no autocorrelation). The statistic is distributed chi-squared, with p degrees of freedom. Reject H_0 if the p-value of the Breusch-Godfrey statistic is less than 0.05.

The R package *lmtest* contains the procedure *bgtest* which estimates the Breusch-Godfrey test. The null hypothesis is that there is no autocorrelation.

```
library(lmtest)
fm<-lm(FARMPOP~TIME,data=uu)
bgtest(fm, order = 4)
```

If temporal autocorrelation exists, what do we do about it?

The most common solution in older econometric work was to transform the data, using the **Cochrane-Orcutt procedure**. Today, one is taught that autocorrelation is probably the product of model misspecification. On encountering autocorrelation, then, your first reaction should be to try to **respecify your model**. That is, you should hunt for some other way of modeling the relationship between your dependent and independent variables.

It often happens that autocorrelation vanishes when **dynamics** (i.e., lagged values) are introduced in a time series model. By this is meant that lagged values of the dependent or one or more independent variables are included as independent variables.

For the data in Figure 1, the shape of the graph suggests that a **semi-log model** or a **polynomial model** would result in non-autocorrelated residuals.

If respecification is unsuccessful, a corrected estimate of the standard error will correct for the main problem in autocorrelation: unreliable t-statistics. The R package *sandwich* will provide an estimate of the **Newey-West autocorrelation consistent covariance matrix**.

```
library(sandwich)
fm<-lm(FARMPPOP~TIME, data=uu)
coeftest(fm,vcov=NeweyWest(fm, lag=4))
```

MORE ADVICE ON PAPERS

The typical paper begins with an *Introduction*, in which you state what you are doing and why it is important. This is followed by a section giving a *Literature Review*. These two sections constitute the *ex ante* discussion. Next comes a discussion of the *Data and Methods* used, and then a section presenting the *Results*. Your Table 1, showing the descriptive statistics for the variables you use, will come in the data and methods section. Your Table 2, showing the regression results, will come in the section presenting the results. Following the results comes a section usually called *Discussion* where you explain what the results mean—this is the *ex post* discussion. Finally, you wrap everything up with the *Summary and Conclusion* section, where you state what you did and what made it important. This is followed by your *References* section.

You will also need to write an abstract for your paper, which should be placed on the paper's cover page. The cover page requires special attention. For this course, model your cover page after the example given at S:\TEFF\662\R\Title page for term papers.doc .

HOMEWORK

Your assignment is to construct a consumption function, using data downloaded from the St. Louis Fed. The specification is entirely up to you. Follow these basic steps:

- 1) Try a number of specifications, containing lags of both the independent variables and the dependent variable until you find one that effectively eliminates autocorrelation.
- 2) Carry out a Chow test to determine if there is a structural break. [If there is, start over, with a different model for each of the two sub-periods]
- 3) Calculate the long-run multiplier giving the effect of income on consumption. Explain in words exactly what it means.

```

#---R program for Autocorrelated time series (S:\TEFF\662\R\r08a.R)
rm(list=ls(all=TRUE))
setwd("s:/TEFF/662/R/")
library(foreign)
library(fImport)
library(tseries)
library(AER)
library(car)
library(fEcofin)
library(sandwich)
library(lmtest)

#--bring in GDP and C from fred at St.Louis Fed--
pcec<-fredImport("PCEC")@data
gdp<-fredImport("GDP")@data
#--make sure they have the same start and end date--
str(gdp)
str(pcec)
class(gdp)

#--convert to ts format-
C<-as.ts(pcec)
Y<-as.ts(gdp)
#--augmented Dickey-Fuller test--H0: series non-stationary--
adf.test(C)
adf.test(Y)
#--take first difference--
C<-diff(as.ts(pcec),1)
Y<-diff(as.ts(gdp),1)
adf.test(C)
adf.test(Y)
#--make some lags (you'll probably want to make even more)--
C1<-lag(C,-1)
C2<-lag(C,-2)
Y1<-lag(Y,-1)
Y2<-lag(Y,-2)
ww<-ts.intersect(C,C1,C2,Y,Y1,Y2)
ww[1:11,]

cf<-lm(C~Y+Y1+Y2+C1+C2,data=ww)
summary(cf)
#--Breusch-Godfrey test--H0: no autocorrelation--
bgtest(cf,order=4)
#--calculate long run multiplier--
u<-coef(cf)
sum(u[2:4])/(1-sum(u[5:6]))

```

```

#--Monte-Carlo for autocorrelated time series (S:\TEFF\662\R\r08mc.R)
rm(list=ls(all=TRUE))
#--Set path to your own directory--
setwd("S:/teff/662/R/")
options(echo=TRUE)
library(foreign)
library(car)
library(tseries)

estcoef<-NULL
corvals<-NULL
nobs<-50
truecoef<-6
for (i in 1:5000){
x<-matrix(1:nobs,nobs,1)
zerr<-ts(rnorm(nobs+4))
aerr<-0.4*lag(zerr,-1)+0.3*lag(zerr,-2)+0.2*lag(zerr,-3)+0.1*lag(zerr,-4)
kk<-ts.intersect(aerr,zerr)
aerr<-as.matrix(kk[,1])
err<-as.matrix(kk[,2])
aerr<-12*(aerr-mean(aerr))/sd(aerr)
err<-12*(err-mean(err))/sd(err)
y1<-x*truecoef+err
y2<-x*truecoef+aerr
ec<-cbind(err[1:(nobs-1)],err[2:nobs])
aec<-cbind(aerr[1:(nobs-1)],aerr[2:nobs])
corvals<-rbind(corvals,cbind(cor(ec)[1,2],cor(aec)[1,2]))
z1<-summary(lm(y1~x))
cf1<-z1$coefficients[2,1]
se1<-z1$coefficients[2,2]
z2<-summary(lm(y2~x))
cf2<-z2$coefficients[2,1]
se2<-z2$coefficients[2,2]
estcoef<-rbind(estcoef,cbind(cf1,se1,cf2,se2))
}
estcoef<-data.frame(estcoef)
names(estcoef)<-c("NoacBeta","NoacSE","acBeta","acSE")

summary(corvals)
#--take mean and standard deviation of results--
mns<-data.frame(t(apply(estcoef,2,mean)))
sds<-data.frame(t(apply(estcoef,2,sd)))
#--compare mean of SE with SD of estimated coefficient (should be equal)--
cbind(mns$NoacSE,sds$NoacBeta)
cbind(mns$acSE,sds$acBeta)

#--plot histograms of results--
layout(matrix(1:2,1,2))
hist(estcoef$NoacBeta,breaks=50)
lines(density(estcoef$NoacBeta),col="blue")
abline(v=truecoef,col="red",lty=2,lwd=2)
hist(estcoef$acBeta,breaks=50)
lines(density(estcoef$acBeta),col="blue")
abline(v=truecoef,col="red",lty=2,lwd=2)
layout(1)

#--plot histograms of results--
layout(matrix(1:2,1,2))
hist(estcoef$NoacSE,breaks=30,xlim=c(0.05,.25))
mtext("true value: red line;\n mean estimated value: blue line")
abline(v=sd(estcoef$NoacBeta),col="red",lty=2,lwd=2)
abline(v=mean(estcoef$NoacSE),col="blue",lty=3,lwd=2)
hist(estcoef$acSE,breaks=30,xlim=c(0.05,.25))
mtext("true value: red line; mean estimated value: blue line")
abline(v=sd(estcoef$acBeta),col="red",lty=2,lwd=2)
abline(v=mean(estcoef$acSE),col="blue",lty=3,lwd=2)
layout(1)

```