Forecasting and prediction can be taken as synonyms. Usually, models are estimated to test the theoretical relationship between the dependent variable and one of the independent variables: the real interest therefore centers on the sign and significance of *individual coefficients*. However, when forecasting, interest centers on the *predicted value* of the dependent variable. In time-series studies, data from the present and past are used to estimate a predicted value for one or more time periods in the future. In cross-sectional studies, data from known persons, firms, or regions are used to make guesses about values for persons, firms, or regions for which there is incomplete information.

There are many forecasting methods. *Structural* models take relationships posited by economic theory (e.g., IS-LM models in macroeconomics), estimate the coefficients, then use the estimated coefficients to simulate an economy's progress into the future. *Time-series* models are usually a-theoretic, and estimate future values of a series by using that same series' past and present values. Structural models were developed first, and were shown to be inferior to simple time-series models by researchers during the 1970s. Since that time, the best features of both approaches have been combined in the best-regarded current models such as the Engle-Granger error correction model, and the Harvey approach to time-varying parameters.

No forecast is complete without some effort to assess its accuracy. Evaluating a forecast requires the following steps:

Set aside the last few years from your data set; we call these data the *ex post* forecast period. The remaining data are used to estimate the model parameters—these data are called the *estimation* period.

Use your estimated model to forecast the predicted values in the *ex post* forecast period.

Compare the predicted values with the actual values for the *ex post* forecast period. The accuracy of prediction is termed the *out-of-sample fit*.

A good forecast is more accurate than a random guess. To see if a forecast provides better-than-random information, econometricians will generally compare it to a *random walk*. A random walk assumes that the current value of a time series equals its immediately preceding value plus an error term. A surprising number of economic time series can be shown to be random walks (including stock prices). If a forecast has lower out-of-sample fits than the random walk model, then it should be discarded.

The best forecast will be produced by the model which has a better out-of-sample fit than any other forecast model. However, research has shown that taking the average of several different forecast models often provides the best results. This method, called the "consensus forecast," is used by *The Economist*, for example.

The program on the following page estimates an autoregressive (AR) model, in which current values of a variable are a function of its past values. The estimated coefficients are then used to produce a forecast. Your exercise today is to run the program and produce 5, 15, and 25 year forecasts of per capita GDP for a country of your own choosing. Be sure to evaluate your forecast by comparing fitted values with actual values in an *ex post* forecast period.

```
#--Forecasting (S:\TEFF\662\R\r12a.R)--
rm(list=ls(all=TRUE))
#--Set path to your directory with data and program--
setwd("s:/teff/662/R/")
options(echo=TRUE)
library(foreign)
library(fImport)
library(tseries)
library(AER)

sessionInfo()
class(sunspot.year)
tsp(sunspot.year)

nyi<-length(sunspot.year)
os<-start(sunspot.year)
oe<-end(sunspot.year)
nyo<-55

insamp<-as.ts(sunspot.year[1:(nyi-nyo)])
tsp(insamp)<-c(os[1],oe[1]-nyo,1)

insampar <- ar(insamp,method="ols")
outfit<-predict(insampar, n.ahead=nyo)$pred
insampfit<-sunspot.year-insampar$resid
tsp(insampfit)

cdd<-ts.union(sunspot.year,outfit,insampfit)
plot(cdd,plot.type="single",col=c("red","blue","black"),lty=1:3,lwd=1:3)
legend("topleft", legend = c("actual", "out of samp fit","in-samp fit"),
col=c("red","blue","black"),lty=1:3,lwd=1:3)

MAPEin<-mean(abs(insampar$resid),na.rm=TRUE)
MAPEout<-mean(abs(cdd[,"sunspot.year"]-cdd[,"outfit"]),na.rm=TRUE)
#--can see that insample fit is much better than outsample fit--
cbind(MAPEin,MAPEout)


#--bring in GDP and C from fred at St.Louis Fed--
pcec<-fredImport("PCEC")@data
pcectpi<-fredImport("PCECTPI")@data #--Pers. Cons. Expend.: Chain-type Price Index
gdp<-fredImport("GDP")@data
gdpdef<-fredImport("GDPDEF")@data #--Gross Domestic Product: Implicit Price Deflator
#--check to make sure all have same start and end dates--
str(gdp);str(pcec);str(gdpdef);str(pcectpi)

#--Deflate, create growth rates---
C<-pcec/pcectpi
pcC<-diff(C,4)/lag(C,4)
Q<-gdp/gdpdef
pcQ<-diff(Q,4)/lag(Q,4)

cdd<-cbind(Q,C,pcQ,pcC)
plot(cdd)

plot(cdd[,3:4],plot.type="single",col=c("red","blue"),lty=1:2,lwd=1:2)
legend("topright", legend = c("Pct Ch GDP", "Pct Ch C"),
col=c("red","blue"),lty=1:2,lwd=1:2)


#--forecast growth rate of GDP--
y<-as.ts(pcQ[which(!is.na(pcQ))])
tsp(y)
nyi<-length(y)
nyo<-5

insampar<-ar(y[1:(nyi-nyo)],method="ols")
outfit<-predict(insampar, n.ahead=nyo)$pred
insampfit<-as.ts(y[1:(nyi-nyo)]-insampar$resid)
```

```
cdd<-ts.union(y,outfit,insampfit)
plot(cdd,plot.type="single",col=c("red","blue","black"),lty=1:3,lwd=1:3)
legend("topright", legend = c("actual", "out of samp fit","in-samp fit"),
col=c("red","blue","black"),lty=1:3,lwd=1:3)


#--forecast growth rate of Consumption--
y<-as.ts(pcC[which(!is.na(pcC))])
tsp(y)
nyi<-length(y)
nyo<-25

insampar<-ar(y[1:(nyi-nyo)],method="ols")
outfit<-predict(insampar, n.ahead=nyo)$pred
insampfit<-as.ts(y[1:(nyi-nyo)]-insampar$resid)
str(y);str(outfit);str(insampfit)
class(y);class(outfit);class(insampfit)

cdd<-ts.union(y,outfit,insampfit)
plot(cdd,plot.type="single",col=c("red","blue","black"),lty=1:3,lwd=1:3)
legend("topright", legend = c("actual", "out of samp fit","in-samp fit"),
col=c("red","blue","black"),lty=1:3,lwd=1:3)


#--forecast per capita GDP (Angus Maddison's data)---

#xx<-read.csv("s:/teff/662/R/maddisonPOP.csv") #-population--
xx<-read.csv("s:/teff/662/R/maddisonPCGDP.csv") #-per capita GDP--
nms<-names(xx)[-1]
xx<-(xx[which(xx$year>=1820 & xx$year<=2008),nms])
nms<-names(which(apply(!is.na(xx[,nms]),2,sum)>=50))
xx<-as.ts(xx[,nms])
adf.test(xx[which(!is.na(xx[,"USA"])),"USA"])
tsp(xx)<-c(1820,2008,1)
#xx<-(diff(xx,1)/lag(xx,k=-1)) #--uncomment if you prefer growth rate--
dimnames(xx)[[2]]<-nms

off<-"Germany"
z<-which(!is.na(xx[,off]))
yrs<-seq(tsp(xx)[1],tsp(xx)[2],1)[z]
y<-as.ts(xx[z,off])
nyi<-length(y)
nyo<-25

yin<-as.ts(y[1:(nyi-nyo)])
#insample years
yrs[c(1,(nyi-nyo))]
insampar<-ar(yin,method="ols")
insampar
outfit<-predict(insampar, n.ahead=nyo)$pred
#outsample years
yrs[c((nyi-nyo+1),nyi)]
insampfit<-as.ts(yin-insampar$resid)

cdd<-ts.union(y,outfit,insampfit)
plot(cdd,plot.type="single",col=c("red","blue","black"),lty=1:3,lwd=1:3,main=off)
legend("topleft", legend = c("actual", "out of samp fit","in-samp fit"),
col=c("red","blue","black"),lty=1:3,lwd=1:3)
```