Forecasting

in Economics, Business, Finance and Beyond

CH 06: Cycles I: Autoregressions and Wold's Chain Rule

Cycles

- We've already considered models with trend and seasonal components. In this chapter we consider a crucial third component, cycles. When we speak of cycles, we have in mind a much more general notion of cyclicality: any sort of stable, mean-reverting dynamics not captured by trends or seasonality.
- Trend and seasonal dynamics are simple, so we can capture them with simple models. Cyclical dynamics are a bit more complicated

2

Basic Ideas

A realization of a time series is an ordered set,

$$\{\ldots; y_{-2}; y_{-1}; y_0; y_1; y_2; \ldots\}$$

Typically the observations are ordered in time (hence the name time series), but they don't have to be.

In theory, a time series realization begins in the infinite past and continues into the infinite future. This perspective may seem abstract and of limited practical applicability, but it will be useful in deriving certain very important properties of the models we'll be using shortly. In practice, of course, the data we observe is just a finite subset of a realization.

3

Basic Ideas

We'd like a series' mean and its covariance structure (that is, the covariances between current and past values) to be stable over time, in which case we say that the series is **covariance stationary**.

4

Covariance Stationary

The first requirement for a series to be covariance stationary is that the mean of the series be stable over time. Then we can write $E[y_t] = \mu$, for all t.

The second requirement for a series to be covariance stationary is that its covariance structure be stable over time. This is a bit tricky, but tremendously important, and we do it using the **autocovariance function**.

5

Autocovariance

The autocovariance at displacement h is just the covariance between y_t and y_{t-h}. It will depend on h, and it may also depend on t.

$$\gamma(t; h) = \text{cov}(y_t; y_{t-h}) = E[(y_t - \mu)(y_{t-h} - \mu)]$$

If the covariance structure is stable over time, as required by covariance stationarity, then the autocovariances depend only on displacement, h, not on time, t, and we write

$$\gamma(t; h) = \gamma(h)$$

for all t.

6

Autocovariance

The autocovariance function is important because it provides a basic summary of cyclical dynamics in a covariance stationary series.

We graph and examine the autocovariances as a function of h. Typically, we'll consider only non-negative values of h. Symmetry reflects the fact that the autocovariance of a covariance stationary series depends only on displacement; it doesn't matter whether we go forward or backward.

Note also that $\gamma(0) = \text{cov}(y_t; y_t) = \text{var}(y_t)$.

7

Autocovariance

We require that the variance of the series $\gamma(0)$ - the autocovariance at displacement 0, be finite. It can be shown that no autocovariance can be larger in absolute value than $\gamma(0)$, so if $\gamma(0) < 1$, then so too are all the other autocovariances.

It may seem that the requirements for covariance stationarity are quite stringent, but appearances can be deceptive. Although many series are not covariance stationary, it is frequently possible to work with models that give special treatment to nonstationary components such as trend and seasonality, so that the cyclical component that's left over is likely to be covariance stationary.

8

Autocorrelation function (ACF)

$\mathrm{Cov}(X_s, X_t) = \mathrm{E}[(X_s - \mu_s)(X_t - \mu_t)].$

$\rho_X(h) = \dfrac{\mathrm{Cov}(X_{t+h}, X_t)}{\mathrm{Cov}(X_t, X_t)}$

For observations $x_1, \ldots, x_n$ of a time series, the **sample mean** is $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$

The **sample autocovariance function** is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \qquad \text{for } -n < h < n.$$

The **sample autocorrelation function** is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

9

Autocorrelation function (ACF)

The autocorrelation is simply the autocovariance, "normalized," or "standardized," by the product of the standard deviations of *x* and *y*.

The autocorrelation is often more informative and easily interpreted than the autocovariance.

The autocorrelation, moreover, does not depend on the units in which *x* and *y* are measured, whereas the autocovariance does.

We usually work with the autocorrelation function.

10

Partial Autocorrelation Function (PACF)

The **partial autocorrelation function** (**PACF**) gives the partial correlation of a time series with its own lagged values, controlling for the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags.

We denote the PACF at lag h by $\alpha(h)$.

11

Partial Autocorrelation Function (PACF)

To see why we need the PACF, consider the following:

Suppose that $\alpha(1)$ is the only non-zero value of $\alpha(h)$ for $h > 0$. Then X_1 is correlated to X_2, X_2 is correlated to X_3, and so on.

When we check for correlation at lag 2 ($\alpha(2)$) we will get a non-zero estimate because of the lag 1 correlation; X_1 and X_3 will be correlated due to the lag 1 correlation. The PACF adjusts for this "carried-over" correlation.

12

Partial Autocorrelation Function (PACF)

As with the autocorrelations, we often graph the partial autocorrelations as a function of *h* and examine their qualitative shape.

Like the autocorrelation function, the partial autocorrelation function provides a summary of a series' dynamics, but as we'll see, it does so in a different way.

13

Partial Autocorrelation Function (PACF)

All of the covariance stationary processes that we will study have autocorrelation and partial autocorrelation functions that approach zero, one way or another, as the displacement gets large.

The precise decay patterns of autocorrelations and partial autocorrelations of a covariance stationary series, however, depend on the specifics of the series.

14

White Noise

Before we estimate time series models, we need to understand their population properties, assuming that the postulated model is true.

The simplest of all such time series processes is the fundamental building block from which all others are constructed.

Suppose that $X_t = \varepsilon_t$, where $\varepsilon_t \sim (0; \sigma^2)$; in other words, ε_t is uncorrelated over time.

Such a process, with zero mean, constant variance, and no serial correlation, is called zero-mean white noise, or simply white noise.

15

White Noise

For short we often write $X_t \sim \mathrm{WN}(0; \sigma^2)$.

Such a process, with zero mean, constant variance, and no serial correlation, is called zero-mean white noise, or simply white noise.

Note that uncorrelated does not necessarily mean independent. If we also have independence, we write $X_t \sim \mathrm{iid}(0; \sigma^2)$.

If the noise in Gaussian and uncorrelated, it is also independent, and we write $X_t \sim \mathrm{iid}\ N(0; \sigma^2)$.

16

Testing the Estimated Noise Sequence

We have looked at techniques to eliminate seasonality and trend from the observed portion of a time series. If the resulting series (the **residuals**) is just iid noise, then the only modeling still to be done is to estimate the mean and variance of the noise components. However, if there is some dependence among the residuals, we may be able to use past residuals to predict future residuals. We need some tools to help us decide whether the residuals are iid noise or whether there is some significant dependence between them that might lead us to an appropriate model for the residuals.

Formally stated, we want to perform tests of the hypothesis H_0: the residual series is iid noise vs. H_a: the residual series has some other structure.

17

Testing the Estimated Noise Sequence

Look at the Sample Autocorrelation Function

Estimators for the autocovariance function are given by

$$\hat{\gamma}(0)=\frac{1}{n}\sum_{t=1}^{n}\left(x_t-\bar{x}\right)^2$$

and

$$\hat{\gamma}(h)=\frac{1}{n}\sum_{t=1}^{n-|h|}\left(x_t-\bar{x}\right)\left(x_{t+|h|}-\bar{x}\right) \qquad -n<h<n$$

It seems reasonable to use these estimators to estimate the autocorrelation function:

$$\hat{\rho}(h)=\frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad \forall\ h\in\mathbb{Z}, \text{ provided } \hat{\gamma}(0)\neq 0$$

18

Testing the Estimated Noise Sequence

Look at the Sample Autocorrelation Function

Note that $\hat{\gamma}(0)=0$ only when there is absolutely no variability in the data.

It can be shown that for large *n* the sample autocorrelations $\hat{\rho}(h)$ h = 1, 2, 3, … , for an iid noise sequence (with finite variance) are approximately iid $N\left(0,1/n\right)$ random variables.

This means that approximately 95% of all values of $\hat{\rho}(h)$ should lie within two standard deviations of 0 (the mean).

Our decision rule would be to reject H_0 if more than 5% of the values of $\hat{\rho}(h)$ lie outside these bounds, or a single value lies far outside these bounds.

We can use a graph in ITSM to check this.

19

Testing the Estimated Noise Sequence

Portmanteau Tests

A portmanteau is a large suitcase that can carry lots of objects.

A portmanteau test is a test that tests several aspects in a single test.

20

Testing the Estimated Noise Sequence

Portmanteau Tests

Version 1. As mentioned before, for large n the sample autocorrelations $\hat{\rho}(h)$, h = 1, 2, 3, … for iid noise are approximately iid $N\left(0,\frac{1}{n}\right)$ random variables, so $\sqrt{n}\hat{\rho}(h)$, h = 1, 2, 3, … for iid noise are approximately iid N(0, 1) (standard normal) random variables. Since the sum of h independent squared standard normal r.v.'s has a χ^2 distribution with h degrees of freedom, we could look at the quantity $Q=n\sum_{j=1}^{h}\left(\hat{\rho}(j)\right)^2$

Large values of Q are evidence against H_0.

21

Testing the Estimated Noise Sequence

Portmanteau Tests

Version 2. ITSM uses a modification of version 1 formulated by Ljung and Box (1978). Replace Q in version 1 with

$$Q_{LB}=n(n+2)\sum_{j=1}^{h}\frac{\left(\hat{\rho}(j)\right)^2}{n-j}$$

The distribution of Q_{LB} is better approximated by a χ^2 (h) distribution than is the distribution of Q in version 1.

22

Testing the Estimated Noise Sequence

Portmanteau Tests

Version 3. A third version of this test was formulated by McLeod and Li (1983). This version tests a slightly different hypothesis. For this version H_0 holds that the data are observations from an iid *normal* sequence. It uses a statistic similar to Version 2 above, with the exception that the each data point is squared before the sample autocorrelations are computed.

23

Testing the Estimated Noise Sequence

The Turning Point Test (Nonparametric)

It seems reasonable that an iid sequence should fluctuate up and down. How fast should this fluctuation occur? In order to answer this we must consider the idea of turning points. We say that there is a turning point in the sequence $\hat{Y}_t$ if either of the two following conditions hold:

$\hat{y}_{t-1}<\hat{y}_t$ and $\hat{y}_{t+1}<\hat{y}_t$ or $\hat{y}_{t-1}>\hat{y}_t$ and $\hat{y}_{t+1}>\hat{y}_t$

It can be shown that the probability that any given point in an iid sequence is a turning point is 2/3.

24

Testing the Estimated Noise Sequence

The Turning Point Test (Nonparametric)

In a sequence of length n there are a possible total of $n - 2$ turning points since the endpoints cannot be checked.

Let T = # turning points in $\hat{Y}_t$. We would then expect to see

$$E[T] = \frac{2(n-2)}{3}$$

turning points. Therefore the quantity T – E[T] should be close to zero. If T – E[T] is a lot larger than zero, it indicates that the sequence is fluctuating faster than we would expect for an iid sequence.

If T – E[T] is a lot smaller than zero, it indicates that there is a positive correlation between neighboring observations.

25

Testing the Estimated Noise Sequence

The Turning Point Test (Nonparametric)

It can also be shown that the variance of T is given by

$$\text{Var}[T] = \frac{16n-29}{90}$$

For large n it can be shown that $T \,\dot{\sim}\, N\left(\frac{2(n-2)}{3}, \frac{16n-29}{90}\right)$

Thus the quantity $\frac{T-E[T]}{\sqrt{V[T]}}$ should then have an approximately standard normal distribution for large n, and we can use it as a test statistic in the usual way.

26

Testing the Estimated Noise Sequence

The Difference-Sign Test for Trend (Nonparametric)

For this test we are testing H_0: there is trend in the data versus H_a: there is no trend.

We count the number S of values of i where $\hat{y}_i > \hat{y}_{i-1};\ i = 2, 3, \ldots, n$

For an iid sequence it can be shown that $E[S] = \frac{n-1}{2}$ and $V[S] = \frac{n+1}{12}$

For large n, $S \,\dot{\sim}\, N(E[S], V[S])$

A large positive value of $S - E[S]$ indicates increasing trend.

A large negative value of $S - E[S]$ indicates decreasing trend.

Caution! This test often fails to detect cyclic trends!

27

Testing the Estimated Noise Sequence

The Rank Test for Linear Trend (Nonparametric)

For this test we are testing H_0: there is linear trend in the data versus H_a: there is no linear trend.

We count the number P of pairs (i, j) where $\hat{y}_j > \hat{y}_i$ and $j > i$, i = 2, 3, …, n.

There are $\binom{n}{2} = \frac{n(n-1)}{2}$ pairs where $j > i$.

For an iid sequence $P\left[\hat{y}_j > \hat{y}_i\right] = \frac{1}{2}$, so $E[P] = \frac{n(n-1)}{4}$.

It can also be shown that for an iid sequence, $V[P] = \frac{n(n-1)(2n+5)}{72}$

28

Testing the Estimated Noise Sequence

The Rank Test for Linear Trend (Nonparametric)

So for large n, $P \dot{\sim} N(E[P], V[P])$.

A large positive value of $P - E[P]$ indicates a positive linear trend.

A large negative value of $P - E[P]$ indicates a negative linear trend.

29

Testing the Estimated Noise Sequence

Fit an Autoregressive Model

We will learn how to do this later in the course!

30

Testing the Estimated Noise Sequence

Look at a Normal Probability (Q-Q) Plot

This is a graphical procedure to see if the residuals can be thought of as observations from a normal distribution.

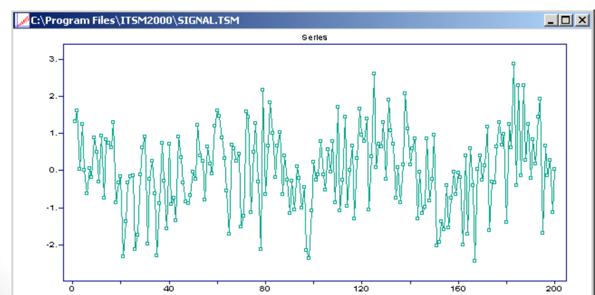
The basic idea for a normal probability plot is that we are going to compare the percentiles in the data set to the corresponding percentiles for a normal distribution with the same mean and variance as the data.

If the residuals are from a normal distribution we expect this plot to show a linear relationship with slope 1 and intercept 0.

31

Using ITSM to Perform These Tests

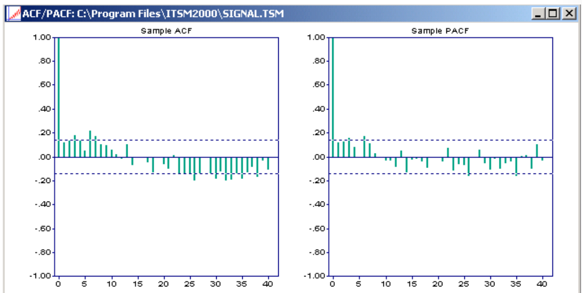
We will use the data set SIGNAL.TSM to show how to use ITSM to test the estimated noise sequence $\hat{Y}_t$. Open the data set. You should see the following graph.



32

Using ITSM to Perform These Tests

To look at the sample autocorrelations click **Statistics** → **ACF/PACF** → **Sample**. You should see the following graph.



The graph on the left gives the autocorrelations and shows the 95% bounds $\pm\frac{1.96}{\sqrt{n}}$

33

Using ITSM to Perform These Tests

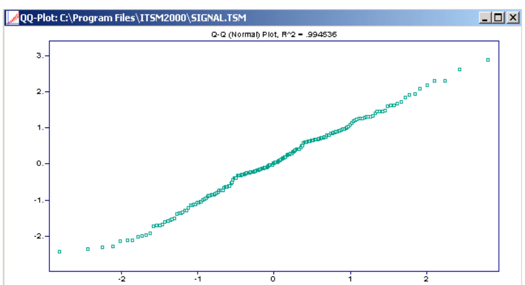
Most of the other tests can be performed by clicking **Statistics** → **Residual Analysis** → **Tests of Randomness**. You will need to enter a value for the maximum lag *h* for the portmanteau tests. You should see the following information window.

```
Tests of randomness: C:\Program Files\ITSM2000\SIGNAL.TSM
============================================
ITSM::(Tests of randomness on residuals)
============================================

Ljung - Box statistic = 51.841 Chi-Square ( 20 ), p-value = .00012

McLeod - Li statistic = 15.987 Chi-Square ( 20 ), p-value = .71746

# Turning points = .13800E+03~AN(.13200E+03,sd = 5.9358), p-value = .31210

# Diff sign points = .10100E+03~AN(99.500,sd = 4.0927), p-value = .71399

Rank test statistic = .10310E+05~AN(.99500E+04,sd = .47315E+03), p-value = .44675

Jarque-Bera test statistic (for normality) = .86432 Chi-Square (2), p-value = .64911

Order of Min AICC YW Model for Residuals = 7
```

34

Using ITSM to Perform These Tests

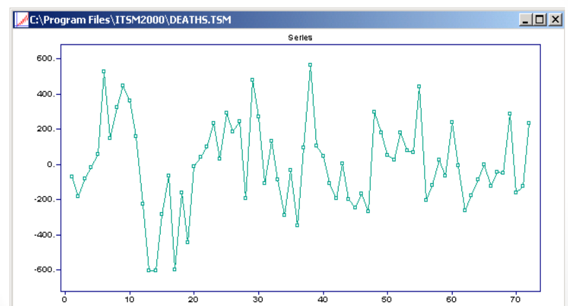
- The Q – Q plot can be obtained by clicking **Statistics** → **QQ Plot (normal)**. You should see the following graph.



35

Using ITSM to Perform These Tests

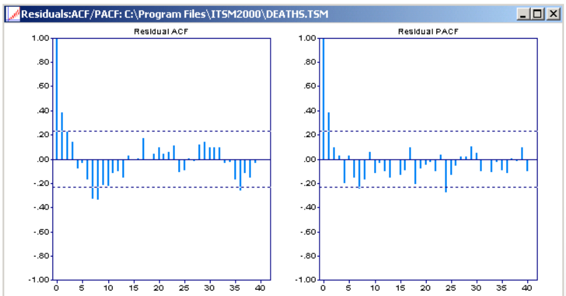
For the estimated noise sequence $\hat{Y}_t$ for the DEATHS.TSM data the ITSM output is as follows:



36

Using ITSM to Perform These Tests

The ACF and PACF:



37

Using ITSM to Perform These Tests

The test results:

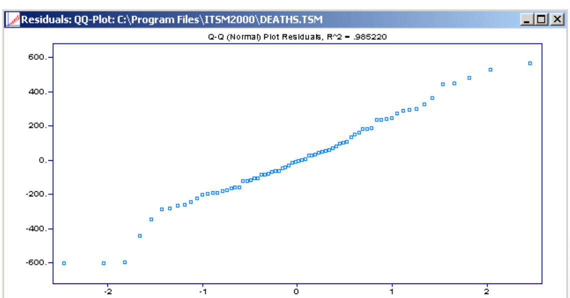
Tests of randomness: C:\Program Files\ITSM2000\DEATHS.TSM

```
===========================================
ITSM::(Tests of randomness on residuals)
===========================================

Ljung - Box statistic = 55.384 Chi-Square ( 20 ), p-value = .00004

McLeod - Li statistic = 15.829 Chi-Square ( 20 ), p-value = .72716

# Turning points = 43.000~AN(46.667,sd = 3.5324), p-value = .29926

# Diff sign points = 35.000~AN(35.500,sd = 2.4664), p-value = .83935

Rank test statistic = .12450E+04~AN(.12780E+04,sd = .10285E+03), p-value = .74833

Jarque-Bera test statistic (for normality) = .13695 Chi-Square (2), p-value = .93382

Order of Min AICC YW Model for Residuals = 1
```

38

Using ITSM to Perform These Tests

The Q-Q plot:



39

Autoregressive Models

A natural starting point for a forecasting model is to use past values of Y (that is, Y_{t-1}, Y_{t-2},…) to forecast Y_t.

An ***autoregression*** is a regression model in which Y_t is regressed against its own lagged values.

The number of lags used as regressors is called the ***order*** of the autoregression.

In a ***first order autoregression***, Y_t is regressed against Y_{t-1}

In a ***p^{th} order autoregression***, Y_t is regressed against Y_{t-1}, Y_{t-2}, …, Y_{t-p}.

40

Autoregressive Models

- Relates the current value of a series to its own past lags. An AR(1) is:

$$y_t = \phi y_{t-1} + \varepsilon_t,$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

How would I write that in lag operator form?

$$(1 - \phi L) y_t = \varepsilon_t$$

We would like to know what its time-series properties are. How can we figure that out?

41

Properties of AR(1) Model

Transform it into an expression involving lags of epsilon by "backward substitution":

$$y_t = \phi y_{t-1} + \varepsilon_t.$$

but $y_{t-1} = \phi y_{t-2} + \varepsilon_{t-1}$

Substituting gives $y_t = \varepsilon_t + \phi\left(\varepsilon_{t-1} + \phi y_{t-2}\right).$

but $y_{t-2} = \phi y_{t-3} + \varepsilon_{t-2}.$

So, $y_t = \varepsilon_t + \phi\left(\varepsilon_{t-1} + \phi\left(\phi y_{t-3} + \varepsilon_{t-2}\right)\right),$

or $y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3 y_{t-3}.$

and so forth...

42

Properties of AR(1) Model

So we can write

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots,$$

Or:

$$y_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} = \left(\sum_{i=0}^{\infty} (\phi L)^i\right)\varepsilon_t,$$

$$y_t = \frac{1}{1 - \phi L}\varepsilon_t$$ As long as $|\phi| < 1$

The last step comes from the fact that the summation is a geometric series.

43

Autoregressive Models

Notice that you can use algebra on the original AR(1) expression in lag operator form to get this same result

$$y_t = \phi y_{t-1} + \varepsilon_t$$

In lag operator form:

$$(1 - \phi L) y_t = \varepsilon_t$$

Divide both sides by the expression in parentheses (the lag polynomial)

$$y_t = \frac{1}{1 - \phi L}\varepsilon_t$$

44

Properties of AR(1) Model

Key properties are :

$$E(y_t) = 0$$

$$E(y_t \mid y_{t-1}) = \phi y_{t-1}$$

$$\text{var}(y_t) = \frac{\sigma^2}{1-\phi^2}$$

The variance will only be finite if |phi| < 1. Covariance stationarity requires this. Intuition, if phi = 1, the series can wander infinitely far away from its starting point, since any shock is permanent.

45

AR(p) series

Higher order AR processes involve additional lags of y:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t, \ or$$

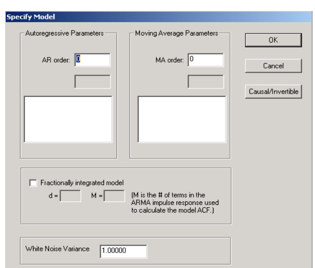
$$\Phi(L) y_t = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) y_t = \varepsilon_t$$

- What do the autocorrelation and partial autocorrelation functions for an AR(p) look like?

46

Model ACF/PACF

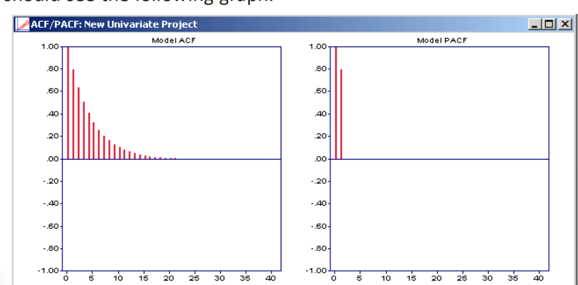
Open ITSM and click **File** → **Project** → **New**. Choose Univariate and click the OK button. This opens a project with no data. Click **Model** → **Specify**. You should see the dialog box below:



47

Model ACF/PACF

Enter 1 as the AR order and set the value of PHI(1) = 0.8. Click OK. The model we are looking at is $X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$. You should see the following graph:

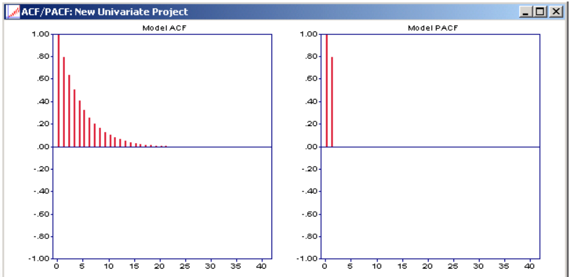


48

Model ACF/PACF

$X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.

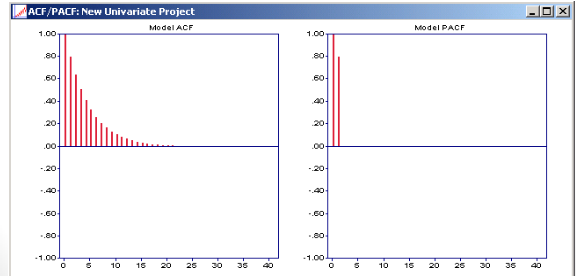
Note that the graph of the ACF shows the expected correlation of 0.8 at lag 1. This is the correlation between X_t and X_{t-1}.



49

Model ACF/PACF

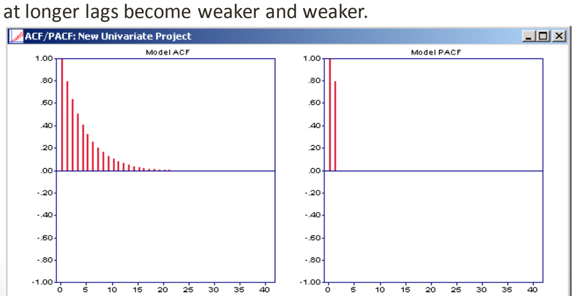
$X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$. The lag2 value gives the correlation between X_t and X_{t-2}. We do not have a lag 2 term in our model, but since there is lag 1 correlation, X_{t-2} is correlated with X_{t-1}, which is in turn correlated with X_t. So X_{t-2} is correlated with X_t.



50

Model ACF/PACF

$X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$. However, since the correlation at lag 1 is not perfect (= 1), the correlation between X_t and X_{t-2} is not as strong as the correlation between X_t and X_{t-1}. Thus the correlations at longer lags become weaker and weaker.

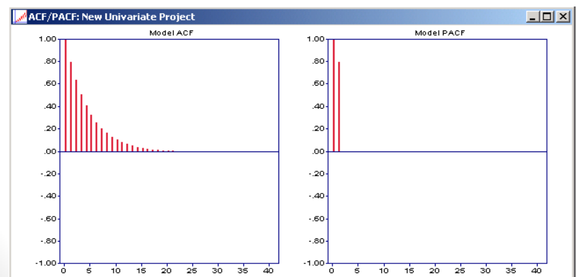


51

Model ACF/PACF

$X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.

Now look at the graph of the PACF. The lag 1 correlation is the same as for the ACF, but at any lag longer than 1 the PACF is zero.

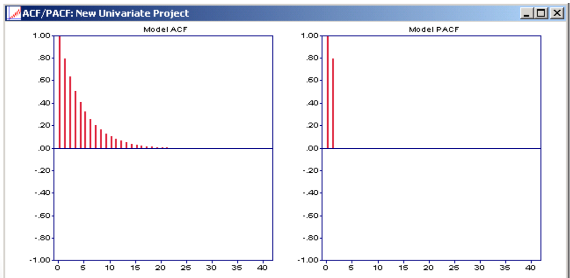


52

Model ACF/PACF

$X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.

This is because the PACF gives the correlation at lag *h* after the effects of correlation at shorter lags has been removed.

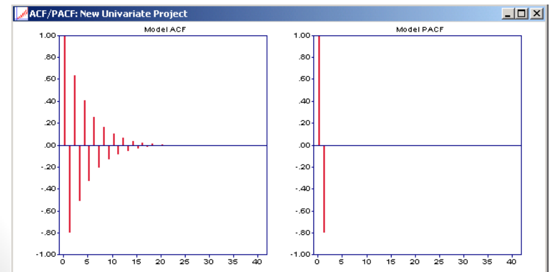


53

Model ACF/PACF

$X_t = -0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.

Note what happens when we change the value of ϕ to -0.8 instead of 0.8. Similar damping pattern, in the ACF, but alternating between positive and negative correlations.

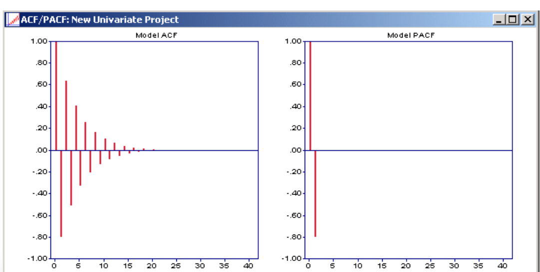


54

Model ACF/PACF

$X_t = -0.8X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.

The graphs of the ACF and PACF for any AR(1) process with $0 < \phi < 1$ will be similar to these examples.



55

MORE ON AR(1) PROCESSES

Consider the AR(1) process $X_t = \phi X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.

If $|\phi| > 1$ we get the unpleasant result that the correlations do not damp out; instead, they resonate.

This means we do not have a covariance stationary time series.

In effect, the only covariance stationary solution involves predicting X_t using future observations rather than past observations! The time series is not **causal**.

Fortunately, it turns out that every AR(1) process with $|\phi| > 1$ can be expressed as an AR(1) process with $|\phi| < 1$, so nothing is "lost" by only considering causal AR(1) processes.

56

MORE ON AR(1) PROCESSES

Consider the AR(1) process $X_t = \phi X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.
If we multiply both sides by X_{t+h} we get

$$\begin{aligned} & X_t X_{t+h} = \phi X_{t-1} X_{t+h} + Z_t X_{t+h} \quad \forall\, t,h \in \mathbb{Z} \\ \Rightarrow\quad & \mathrm{E}\left[X_t X_{t+h}\right] = \phi E\left[X_{t-1} X_{t+h}\right] + E\left[Z_t X_{t+h}\right] \\ \Rightarrow\quad & \gamma(h) = \phi\gamma(h+1) + E\left[Z_t X_{t+h}\right] \\ \Rightarrow\quad & \gamma(h) = \phi\gamma(h+1) + E\left[Z_t X_s\right] \quad \forall\, s,t,h \in \mathbb{Z};\ s<t \end{aligned}$$

Since $E\left[Z_t X_s\right] = 0$, we obtain $\gamma(h) = \phi\gamma(h+1) \quad \forall\, h \in \mathbb{Z}$

Def: The system of equations above are called the **YULE - WALKER** equations.

57

Model ACF/PACF

Let's discuss the AR(p) autocorrelation function in a bit greater depth.

The key insight is that, in spite of the fact that its qualitative behavior (gradual damping) matches that of the AR(1) autocorrelation function, it can nevertheless display a richer variety of patterns, depending on the order and parameters of the process.

It can, for example, have damped monotonic decay, as in the AR(1) case with a positive coecient, but it can also have damped oscillation in ways that AR(1) can't have.

58

Model ACF/PACF

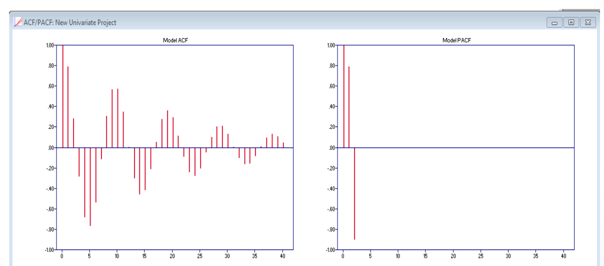
In higher-order autoregressive models, however, the autocorrelations can oscillate with much richer patterns reminiscent of cycles in the more traditional sense.

This occurs when some roots of the autoregressive lag operator polynomial are complex.

59

Model ACF/PACF

$X_t = 1.5\, X_{t-1} - 0.9\, X_{t-2} + Z_t$, $Z_t \sim WN(0, \sigma^2)$. *(Complex roots)*



60

Forecasting Causal Models

Wold's Chain Rule

A very simple recursive method for computing optimal h-step-ahead point forecasts, for any desired h, is available for autoregressions. The recursive method, called the chain rule of forecasting, is best learned by example.

Consider the AR(1) process $X_t = \phi X_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$.

First we construct the optimal 1-step-ahead forecast, and then we construct the optimal 2-step-ahead forecast, which depends on the optimal 1-step-aheadforecast, which we've already constructed.

Then we construct the optimal 3-step-ahead forecast, which depends on the already-computed 2-step-ahead forecast, which we've already constructed, and so on.

61

Forecasting Causal Models

Wold's Chain Rule

The future innovation (error term) is always replaced by its' expected value of 0.

To construct the 1-step-ahead forecast, we write out the process for time T + 1:

$$X_t = \phi X_{t-1} + Z_t$$

Projecting 1 step ahead: $X_{t+1} = \phi X_t + Z_{t} = X_{t+1} = \phi X_t + 0$

62

Forecasting Causal Models

Wold's Chain Rule

To construct the 2-step-ahead forecast, we just repeat this process, plugging in our prediction of .

Continuing in this way, we can recursively build up forecasts for any and all future periods.

Hence the name “chain rule of forecasting."

63

Example: Canadian Employment

In order to gain an appreciation of AR(p) models, we will look at the CAEMP data from the text. (page 161)

The file CAEMP.TSM contains the Canadian employment index, seasonally adjusted, 1961:1-1994:4 (136 observations)

Download the file from Canvas. Save it to your ITSM folder. Open it in ITSM.

64

Example: Canadian Employment

In order to gain an appreciation of AR(p) models, we will look at the CAEMP data from the text. (page 178)

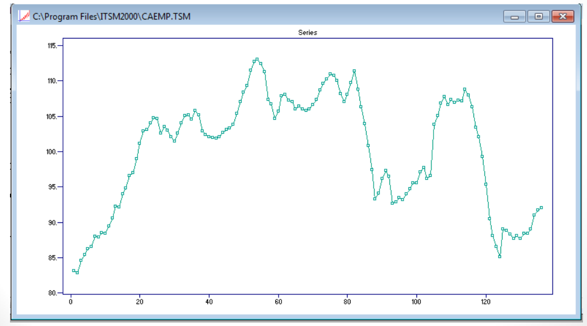
The file CAEMP.TSM contains the Canadian employment index, seasonally adjusted, 1961:1-1994:4 (136 observations)

Download the file from Canvas. Save it to your ITSM folder. Open it in ITSM.

65

Example: Canadian Employment

You should see this graph:



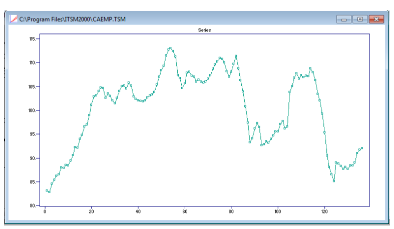
66

Example: Canadian Employment

The series displays no trend, and of course it displays no seasonality because it's seasonally adjusted.

It does, however, appear highly serially correlated.

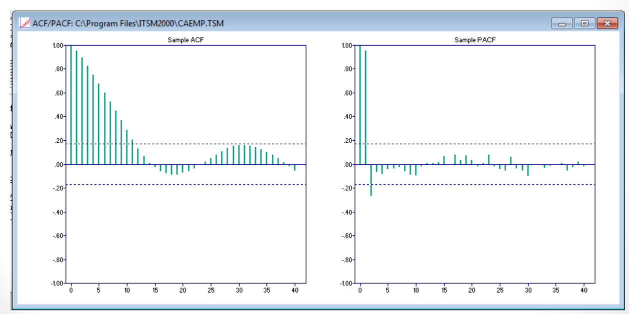
It evolves in a slow, persistent fashion - high in business cycle booms and low in recessions.



67

Example: Canadian Employment

Click Statistics -> ACF/PACF -> Sample. You should see this graph:

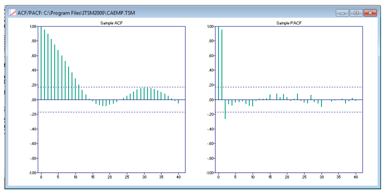


68

Example: Canadian Employment

The sample autocorrelations are very large relative to their standard errors and display slow one-sided decay.

The sample partial autocorrelations, in contrast, are large relative to their standard errors at first (particularly for the 1-quarter displacement) but are statistically negligible beyond displacement 2. This suggests an AR(2) model.



69

Example: Canadian Employment

The option Model -> Estimation -> Preliminary in ITSM contains fast (but not the most efficient) model-fitting algorithms.

They are useful for suggesting the most promising models for the data, but should be followed by maximum likelihood estimation using Model -> Estimation -> Max likelihood.

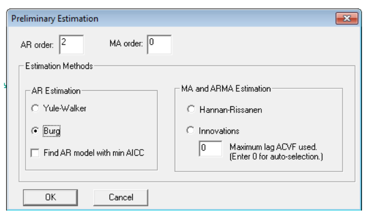
For pure AR models the preliminary estimation option offers a choice between the Burg and Yule–Walker estimates. (Burg is frequently better.)

You can also check the box Find AR model with min AICC to allow the program to fit AR models of orders 0, 1, . . . , 27 and select the one with smallest AICC value.

70

Example: Canadian Employment

Click Model -> Estimation -> Preliminary. The program will ask if you want to subtract the mean. Click Yes. You should see this dialog box:

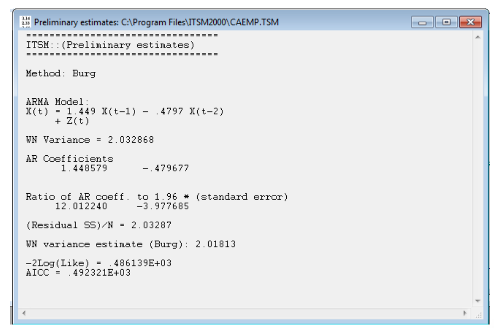


Enter 2 as the AR order. Click the Burg button for preliminary estimation method. Click OK.

71

Example: Canadian Employment

You should see this information box:

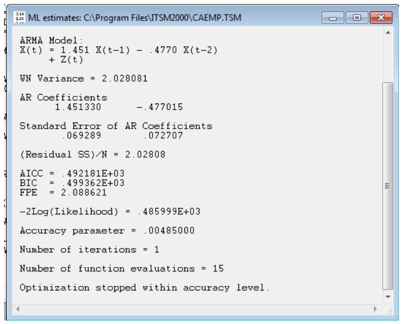

Preliminary estimates: C:\Program Files\ITSM2000\CAEMP.TSM

```
=================================
ITSM::(Preliminary estimates)
=================================

Method: Burg


ARMA Model:
X(t) = 1.449 X(t-1) - .4797 X(t-2)
     + Z(t)

WN Variance = 2.032868

AR Coefficients
      1.448579      -.479677


Ratio of AR coeff. to 1.96 * (standard error)
     12.012240     -3.977685

(Residual SS)/N = 2.03287

WN variance estimate (Burg): 2.01813

-2Log(Like) = .486139E+03
AICC = .492321E+03
```

72

Example: Canadian Employment

Click Model -> Estimation -> Maximum Likelihood. You should see this information box: (Note the results are different!)

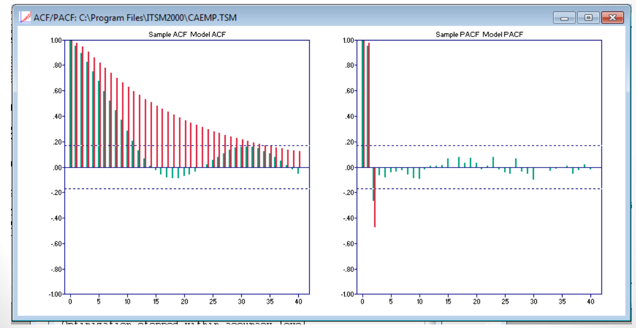



```
ARMA Model:
X(t) = 1.451 X(t-1) - .4770 X(t-2)
     + Z(t)

WN Variance = 2.028081

AR Coefficients
    1.451330      -.477015

Standard Error of AR Coefficients
    .069289       .072707

(Residual SS)/N = 2.02808

AICC = .492181E+03
BIC  = .499362E+03
FPE  = 2.088621

-2Log(Likelihood) = .485999E+03

Accuracy parameter = .00485000

Number of iterations = 1

Number of function evaluations = 15

Optimization stopped within accuracy level.
```

73

Example: Canadian Employment

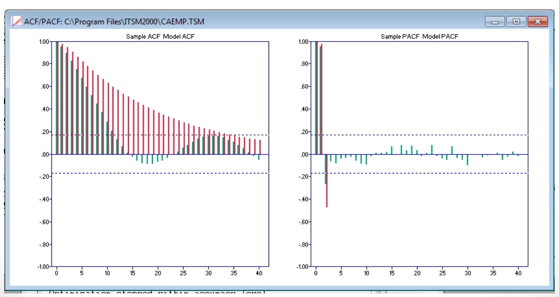
Click Statistics -> ACF/PACF -> Sample/Model. You should see this graph:



74

Example: Canadian Employment

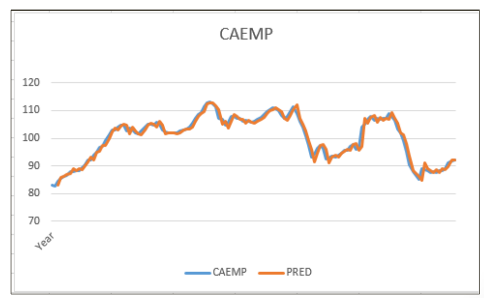
Not horrible, but we are overestimating the magnitude of the lag 2 correlation. This is due to accommodating the cyclic pattern in the data that is NOT in our model. We will address this later in the course!



75

Example: Canadian Employment

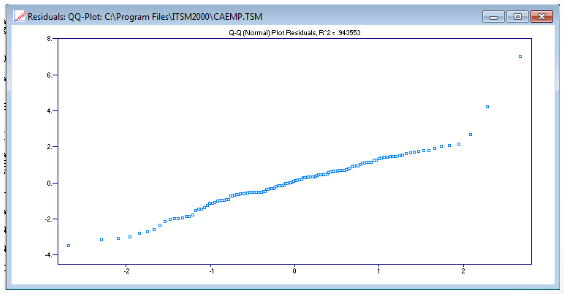
Using Excel to graph the data and the model yields the following: (Note how the model lags the data, but responds fairly quickly to trend changes!



76

Example: Canadian Employment

Click Statistics -> Residual Analysis -> QQ-Plot (normal). You should see the following:

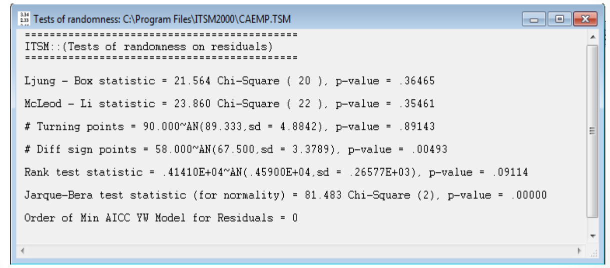


Not great, but not horrible!

77

Example: Canadian Employment

Click Statistics -> Residual Analysis -> Tests of Randomness. You should see the following:

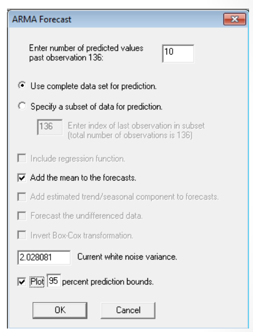


All go the right way. YAY! Also The min AICC YWorder is 0. YAY!

78

Example: Canadian Employment

To predict future values, click Forecasting -> ARMA. You should see this dialog box:



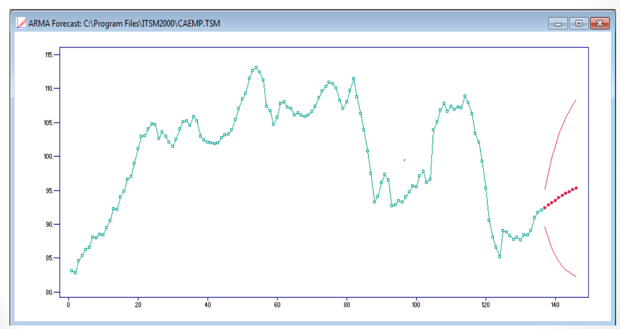
Click the Plot 95% prediction bounds checkbox.

Click OK.

79

Example: Canadian Employment

You should see the following:



80

Example: Canadian Employment

Click the icon in the upper left corner and select info. You should see this :

INFO: ARMA Forecast: C:\Program Files\ITSM2000\CAEMP.TSM

```
=====================
ITSM::(ARMA Forecast)
=====================
```

Step	Prediction	sqrt(MSE)	Approximate 95 Percent Prediction Bounds Lower	Upper
1	92.38863	1.42411	89.59743	95.17983
2	92.76804	2.50997	87.84859	97.68749
3	93.14043	3.41819	86.44090	99.83995
4	93.49989	4.16611	85.33448	.10167E+03
5	93.84397	4.78308	84.46931	.10322E+03
6	94.17186	5.29620	83.79150	.10455E+03
7	94.48362	5.72721	83.25849	.10571E+03
8	94.77966	6.09277	82.83806	.10672E+03
9	95.06062	6.40552	82.50603	.10762E+03
10	95.32715	6.67510	82.24420	.10841E+03

81