# Forecasting

in Economics, Business, Finance and Beyond

CH 7: Cycles II: The Wold Representation and Its Approximation

## Wold's Theorem

Many different dynamic patterns are consistent with covariance stationarity.

Thus, if we know only that a series is covariance stationary, it's not at all clear what sort of model we might fit to describe its evolution.

The trend and seasonal models that we've studied aren't of use; they're models of specific nonstationary components.

2

## Wold's Theorem

Effectively, what we need now is an appropriate model for what's left after fitting the trend and seasonal components – a model for a covariance stationary residual.

Wold's representation theorem points to the appropriate model.

3

## Wold's Theorem

**Wold's Theorem** says that we can represent any covariance-stationary time series process as an infinite distributed lag of white noise:

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i},$$

$$\varepsilon_t \sim WN(0,\sigma^2),$$

4

# Wold's Theorem

In our statement of Wold's theorem we assumed a zero mean.

That may seem restrictive, but it's not.

The deviation from the mean has a zero mean, by construction.

Working with zero-mean processes therefore involves no loss of generality while facilitating notational economy.

5

# Finite approximations

In short, the correct "model" for any covariance stationary series is some infinite distributed lag of white noise.

We call this the general linear process, "general" because any covariance stationary series can be written that way, and "linear" because the Wold representation expresses the series as a linear function of its innovations (the $\varepsilon_t$'s).

The problem is that we do not have infinitely many data points to use in estimating infinitely many coefficients!

6

# Finite approximations

We want to find a good way to approximate the unknown (and unknowable) true model that drives real world data.

The key to this is to find a parsimonious, yet accurate, approximation.

We use two building blocks of such models: the MA and AR models, as well as their combination.

We start by examining the dynamic patterns generated by the basic types of models.

7

# Moving Average (MA) models

- MA(1)

$$\begin{aligned} y_t &= \varepsilon_t + \theta\varepsilon_{t-1} \\ &= (1+\theta L)\varepsilon_t \\ where\ \varepsilon_t &\sim WN(0,\sigma^2) \end{aligned}$$

Intuition?

Past shocks (*innovations*) in the series feed into the succeeding period.

8

## Properties of an MA(1) series

Uncond. Mean

$$Ey_t = E\left(\varepsilon_t + \theta\varepsilon_{t-1}\right)$$
$$= E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) = 0$$

Uncond. Variance

$$\operatorname{var}(y_t) = \operatorname{var}\left(\varepsilon_t + \theta\varepsilon_{t-1}\right)$$
$$= \operatorname{var}(\varepsilon_t) + \theta^2 \operatorname{var}(\varepsilon_{t-1})$$
$$= \sigma^2 + \theta^2\sigma^2$$
$$= \sigma^2(1+\theta^2)$$

This uses 2 statistics results:

1) for 2 random vars x and y, E(x+y)=E(x)+E(y)

2) If a is a constant:

E(ax) = a E(x)

This uses 2 statistics results:

1) var(x+y)= var(x)+var(y), if they are uncorrelated

2) var(ax) = $a^2$var(x)

9

## Properties of an MA(1) series

Conditional mean (the mean given that we are a particular time period, *t*; i.e. given an "information set" $\Omega_{t-1}$):

$$\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots\}$$

$$E(y_t \mid \Omega_{t-1}) = E\left((\varepsilon_t + \theta\varepsilon_{t-1}) \mid \Omega_{t-1}\right)$$
$$= E(\varepsilon_t \mid \Omega_{t-1}) + \theta E(\varepsilon_{t-1} \mid \Omega_{t-1})$$
$$= 0 \qquad + \theta\varepsilon_{t-1}$$

We read this as "the expected value of $y_{t-1}$ conditional on the information set $\Omega_{t-1}$"

That just means we want the expected value taking into account that we know the values of the past epsilons.

10

## Properties of an MA(1) series

The key insight is that the conditional mean moves over time in response to the evolving information set.

The model captures the dynamics of the process, and the evolving conditional mean is one crucial way of summarizing them.

An important goal of time series modeling, especially for forecasters, is capturing such conditional mean dynamics - the unconditional mean is constant (a requirement of stationarity), but the conditional mean varies in response to the evolving information set.

11

## Properties of an MA(1) series

Conditional variance

$$\operatorname{var}(y_t \mid \Omega_{t-1}) = E\left[\left(y_t - E(y_t \mid \Omega_{t-1})\right)^2 \mid \Omega_{t-1}\right]$$
$$= E(\varepsilon_t^2 \mid \Omega_{t-1})$$
$$= E(\varepsilon_t^2) \;=\; \sigma^2$$

To see this, plug in the MA(1) expression for $y_t$ and the expression for the expected value of $y_t$ conditional on the information set (which we just calculated). Notice they differ only by $\varepsilon_t$, the innovation in y that we have not yet observed

12

## Properties of an MA(1) series

Autocovariance function:

$$\gamma(\tau) = E(y_t y_{t-\tau}) = E((\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-\tau} + \theta\varepsilon_{t-\tau-1}))$$
$$= \begin{cases} \theta\sigma^2, & \tau = 1 \\ 0, & otherwise \end{cases}$$

Let's look at how that is derived...

13

## Autocovariance function

Here is the derivation of that:

$$E\left((\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-\tau} + \theta\varepsilon_{t-\tau-1})\right) =$$
$$= E\left(\varepsilon_t\varepsilon_{t-\tau}\right) + E\left(\varepsilon_t\theta\varepsilon_{t-\tau-1}\right) + E\left(\theta\varepsilon_{t-1}\varepsilon_{t-\tau}\right) + E\left(\theta\varepsilon_{t-1}\theta\varepsilon_{t-\tau-1}\right)$$
$$when\ \tau = 1:$$
$$= E\left(\varepsilon_t\varepsilon_{t-1}\right) + \theta E\left(\varepsilon_t\varepsilon_{t-2}\right) + \theta E\left(\varepsilon_{t-1}\varepsilon_{t-1}\right) + \theta^2 E\left(\varepsilon_{t-1}\varepsilon_{t-2}\right)$$
$$= 0 + 0 + \theta\sigma^2 + 0$$

When $\tau > 1$, all terms are zero. Check it!

14

## Properties of an MA(1) series

Autocorrelation function is just this divided by the variance:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \begin{cases} \dfrac{\theta}{1+\theta^2}, & \tau = 1 \\ 0, & otherwise \end{cases}$$

15

## Can $y_t$ be written another way?

Under certain circumstances, we can write an MA process in an *autoregressive representation*

- The series can be written as a function of past lags of itself (plus a current innovation)

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

Why does this form have intuitive appeal?

- Today's value of the variable is related explicitly to past values of the variable

16

## Invertibility

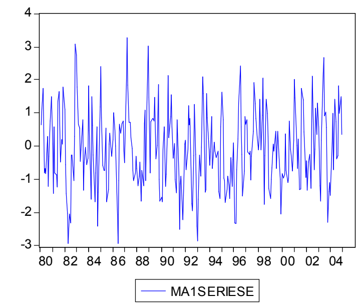
In this case, we say that the MA process is *invertible*

This will occur when $|\theta| < 1$

See the book for a demonstration of how this works and why the key criterion.

17

## What does an MA(1) look like?



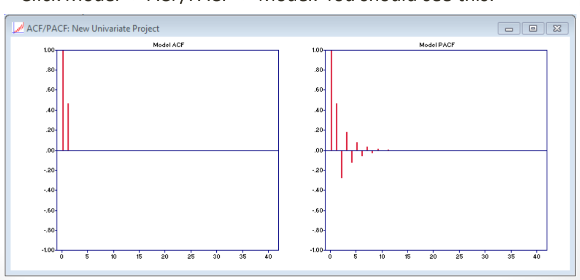

MA 1 seriesa = whitenoise + .7*whitenoise(-1)

18

## What does an MA(1) look like?

Open ITSM with no data. Click Model -> Specify. Enter 1 as the MA order, and 0.7 as the value of Theta(1). Click OK.

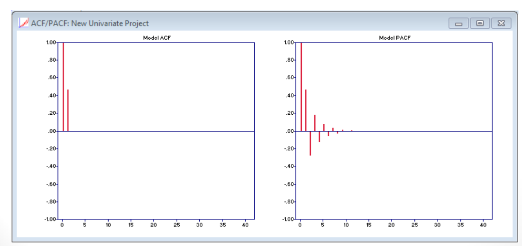
Click Model -> ACF/PACF -> Model. You should see this:




19

## What does an MA(1) look like?

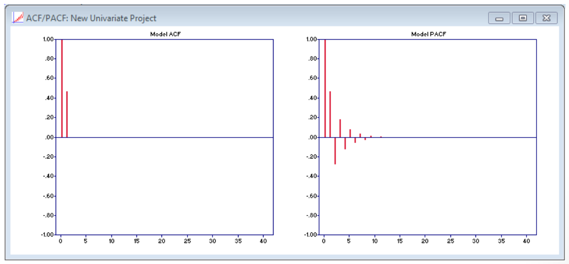
Note that now the ACF has only lag 1 correlation, but the PACF shows decaying correlation with alternate signs. The partial correlations are due to the fact that there is no direct correlation, only indirect correlation through the error terms. For this reason, when we “subtract” the effect of lag 1 correlation, we are removing too much, producing a negative partial correlation at lag 2.




20

## What does an MA(1) look like?

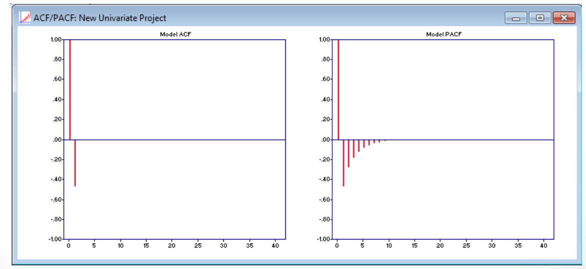
The graphs of the ACF and PACF for any MA(1) process with $0 < \theta < 1$ will be similar to this example.






## What does an MA(1) look like?

Click Model -> Specify. Enter 1 as the MA order, and -0.7 as the value of Theta(1). You should see this:

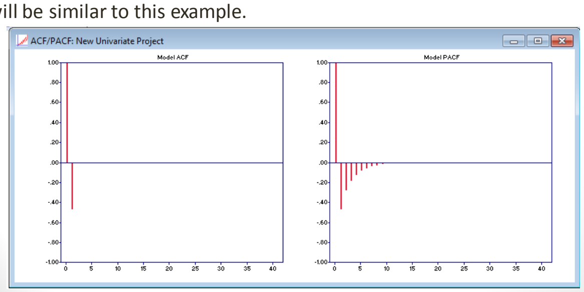





## What does an MA(1) look like?

Note that the ACF again has only lag 1 correlation, and that it has the opposite sign from the previous example. The PACF shows decaying negative correlation.

The graphs of the ACF and PACF for any MA(1) process with $-1 < \theta < 0$ will be similar to this example.






## Modeling Tips

The AR and MA examples illustrate the following two general rules:

- When a purely autoregressive covariance stationary time series involve lags only up to some maximum lag $p > 0$ (an AR($p$) process), all partial autocorrelations beyond lag $p$ are zero, and the autocorrelations decay as the lag increases.
- When a purely moving average covariance stationary time series involve lags only up to some maximum lag $q > 0$ (an MA($q$) process), all autocorrelations beyond lag $q$ are zero, and the partial autocorrelations decay as the lag increases.

The preceding rules are not as helpful as we might expect at first glance. Even a relatively simple series that contains both AR and MA terms can have ACF and PACF graphs of quite different appearances.

# Properties of an MA(1) series

An MA(1) model:

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$
$$= (1+\theta L)\varepsilon_t$$
$$where\ \varepsilon_t = WN(0,\sigma^2)$$

Past shocks (*innovations*) in the series feed into the succeeding period.

25

# Properties of an MA(1) series

An MA(1) has a "short memory"—only last period's shock matters for today

We saw this in the shape of the autocorrelation function:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \begin{cases} \dfrac{\theta}{1+\theta^2}, & \tau = 1 \\ 0, & otherwise \end{cases}$$

There is one significant bar in the autocorrelation graph.

26

# MA(q) series

Higher order MA processes involve additional lags of white noise:

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$$
$$= \varepsilon_t + \theta_1 L\varepsilon_t + \theta_2 L^2\varepsilon_t + \ldots + \theta_q L^q\varepsilon_t$$
$$or,$$
$$y_t = \Theta(L)\varepsilon_t$$
$$where\ \Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$$
$$\varepsilon_t = WN(0,\sigma^2)$$

27

# Autoregressive Models

Review: An AR(1) is:

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

In lag operator form:

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

28

## Properties of AR(1) Model

Key properties of an AR(1):

$$E(y_t) = 0$$

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

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

The variance will only be finite if |phi| < 1. Covariance stationarity requires this.

29

Byron Gangnes

## AR(p) series

Higher order AR processes involve additional lags of y:

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

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

30

## Approximating the Wold Representation

When building forecasting models, we don't want to pretend that the model we t is true.

Instead, we want to be aware that we're approximating a more complex reality.

In particular, we've seen that the key to successful time series modeling and forecasting is parsimonious, yet accurate, approximation of the Wold representation.

31

## Approximating the Wold Representation

We will consider three approximations:

- Moving average (MA) models
- Autoregressive (AR) models
- Autoregressive moving average (ARMA) models.

The three models differ in their specifics and have different strengths in capturing different sorts of autocorrelation behavior.

We use the sample autocorrelations and partial autocorrelations, in conjunction with the AIC and the SIC, to suggest candidate forecasting models, which we then estimate.

32

## Rational Distributed Lags

The Wold representation points to the crucial importance of models with infinite distributed lags.

Infinite distributed lag models, in turn, are stated in terms of infinite polynomials in the lag operator, which are therefore very important as well.

Infinite distributed lag models are not of immediate practical use, however, because they contain infinitely many parameters, which certainly inhibits practical application!

Fortunately, infinite polynomials in the lag operator needn't contain infinitely many free parameters.

33

## Rational Distributed Lags

The infinite polynomial B(L) may for example be a ratio of finite-order (and perhaps very low-order) polynomials.

Such polynomials are called rational polynomials, and distributed lags constructed from them are called rational distributed lags.

Suppose, for example, that $B(L) = \frac{\Theta\,(L)}{\Phi\,(L)}$,

where the numerator polynomial is of degree *q*, and the denominator polynomial is of degree *p*.

$$\Theta(L) = \sum_{i=0}^{q} \theta_i L^i \qquad \Phi(L) = \sum_{i=0}^{p} \phi_i L^i$$

34

## Rational Distributed Lags

There are not infinitely many free parameters in the B(L) polynomial; instead, there are only p + q parameters (the $\theta$s and the $\phi$s).

If p and q are small, say 0, 1 or 2, then what seems like a hopeless task - estimation of B(L) - may actually be easy.

More realistically, suppose that B(L) is not exactly rational, but is approximately rational.

Then we can approximate the Wold representation using a rational distributed lag.

35

## Rational Distributed Lags

Rational distributed lags produce models of cycles that economize on parameters (they're parsimonious), while nevertheless providing accurate approximations to the Wold representation.

The popular ARMA and ARIMA forecasting models are simply rational approximations to the Wold representation.

The finite-order moving average processes is a natural and obvious approximation to the Wold representation, which is an infinite-order moving average process.

36

## Rational Distributed Lags

Finite-order moving average processes also have direct motivation: the fact that all variation in time series, one way or another, is driven by shocks of various sorts suggests the possibility of modeling time series directly as distributed lags of current and past shocks, that is, as moving average processes.

The structure of the MA(1) process, in which only the first lag of the shock appears on the right, forces it to have a very short memory, and hence weak dynamics, regardless of the parameter value.

Note that the requirements of covariance stationarity are met for any MA(1) process, regardless of the values of its parameters.

37

## Rational Distributed Lags

If $|\theta| < 1$ , then we say that the MA(1) process is invertible.

In that case, we can “invert" the MA(1) process and express the current value of the series not in terms of a current shock and a lagged shock, but rather in terms of a current shock and lagged values of the series.

That's called an autoregressive representation.

An autoregressive representation has a current shock and lagged observable values of the series on the right, whereas a moving average representation has a current shock and lagged unobservable shocks on the right.

38

## Computing the AR representation

The process is $y_t = \varepsilon_t + \theta\varepsilon_{t-1},\ \varepsilon_t \sim WN(0, \sigma^2)$

We can solve for the innovation as $\varepsilon_t = y_t - \theta\varepsilon_{t-1}$.

Lagging by successively more periods gives expressions for the innovations at various dates:

$$\varepsilon_{t-1} = y_{t-1} - \theta\varepsilon_{t-2}$$

$$\varepsilon_{t-2} = y_{t-2} - \theta\varepsilon_{t-3}$$

$$\varepsilon_{t-3} = y_{t-3} - \theta\varepsilon_{t-4}$$

39

## Computing the AR representation

The process is $y_t = \varepsilon_t + \theta\varepsilon_{t-1},\ \varepsilon_t \sim WN(0, \sigma^2)$

We can solve for the innovation as $\varepsilon_t = y_t - \theta\varepsilon_{t-1}$.

Making use of these expressions for lagged innovations we can substitute backward in the MA(1) process, yielding

$$y_t = \varepsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - \dots$$

In lag-operator notation, we write the infinite autoregressive representation as

$$\frac{1}{1+\theta L} y_t = \varepsilon_t$$

40

## Computing the AR representation

Note that the back substitution used to obtain the autoregressive representation only makes sense, and in fact a convergent autoregressive representation only exists, if $|\theta| < 1$, because in the back substitution we raise $\theta$ to progressively higher powers.

We can restate the invertibility condition in another way: the inverse of the root of the moving average lag operator polynomial $(1 + \theta L)$ must be less than one in absolute value.

41

## Why all this matters

Autoregressive representations are appealing to forecasters, because one way or another, if a model is to be used for real-world forecasting, it's got to link the present observables to the past history of observables, so that we can extrapolate to form a forecast of future observables based on present and past observables.

Superficially, moving average models don't seem to meet that requirement, because the current value of a series is expressed in terms of current and lagged unobservable shocks, not observable variables.

42

## Why all this matters

But under the invertibility conditions that we've described, moving average processes have equivalent autoregressive representations.

Thus, although we want autoregressive representations for forecasting, we don't have to start with an autoregressive model.

However, we typically restrict ourselves to invertible processes, because for forecasting purposes we want to be able to express current observables as functions of past observables.

43

## Why all this matters

An MA($q$) process has the potential to deliver better approximations to the Wold representation than an MA(1) process, at the cost of more parameters to be estimated.

In the MA(q) case the conditional mean depends on $q$ lags of the innovation.

Thus the MA(q) process has the potential for longer memory as $q$ increases.

The potentially longer memory of the MA(q) process emerges clearly in its autocorrelation function; all autocorrelations beyond displacement $q$ are zero.

44

## 7.2.4 Autoregressive Moving Average (ARMA) Models

## ARMA models

We have looked at autoregressive (AR) and moving average (MA) models.

The two kinds of dynamics can be combined into what is called an *autoregressive moving-average (ARMA) model*.

We like ARMA models because they are usually very parsimonious and can capture fairly sophisticated dynamic behavior.



## Importance of the Wold decomposition

Any stationary process can be written as a linear combination of lagged values of a white noise process (MA(∞) representation).

This implies that if a process is stationary we immediately know how to write a model for it.

Problem: we might need to estimate a lot of parameters (in most cases, an infinite number of them!)

ARMA models: they are an approximation to the Wold representation. This approximation is more parsimonious (=less parameters)

47

## Birth of the ARMA(p,q) models

Under general conditions the infinite lag polynomial of the Wold decomposition can be approximated by the ratio of two finite-lag polynomials:

$$\Psi(L) \approx \frac{\Theta_q(L)}{\Phi_p(L)}$$

Therefore $Z_t = \Psi(L)a_t \approx \frac{\Theta_q(L)}{\Phi_p(L)} a_t$ ,

$$\Phi_p(L)Z_t = \Theta_q(L)a_t$$

$$(1-\phi_1 L - \ldots - \phi_p L^p)Z_t = (1+\theta_1 L + \ldots + \theta_q L^q)a_t$$

$$\underbrace{Z_t - \phi_1 Z_{t-1} - \ldots - \phi_p Z_{t-p}}_{\textbf{AR(p)}} = \underbrace{a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}}_{\textbf{MA(q)}}$$

48

## ARMA (p,q)

A time series $X_t$ is said to be an ARMA(p,q) process if it satisfies

$$\sum_{j=0}^{p}\phi_j X_{t-j} = \sum_{k=0}^{q}\theta_k Z_{t-k}$$

Where $\phi_0 = \theta_0 = 1$, $\phi_p \neq 0$, $\theta_q \neq 0$, $Z_t \sim WN(0,\sigma^2)$,

and the polynomials $1-\phi_1 w - \phi_2 w^2 \ldots - \phi_p w^p$

and $1+\theta_1 w + \theta_2 w^2 \ldots + \theta_q w^q$

have no common factors.

49

## ARMA (p,q)

$\phi(\mathbf{B})X_t$ is called the autoregression polynomial.

$\theta(\mathbf{B})Z_t$ is called the moving average polynomial.

A stationary ARMA solution exists if the autoregression polynomial and the moving average polynomial have no common roots, and is unique iff the autoregressive polynomial has no roots on the unit circle (we are allowing complex roots).

The **UNIT CIRCLE** is the set of all complex numbers z = a + bi where $|z| \leq 1$.

The ARMA process is causal if the autoregression polynomial has no roots on the unit circle.

The ARMA process is invertible if the moving average polynomial has no roots on the unit circle.

50

## ARMA (p,q)

As with autoregressions and moving averages, ARMA processes have a fixed unconditional mean but a time-varying conditional mean.

In contrast to pure moving average or pure autoregressive processes, however, neither the autocorrelation nor partial autocorrelation functions of ARMA processes cut off at any particular displacement.

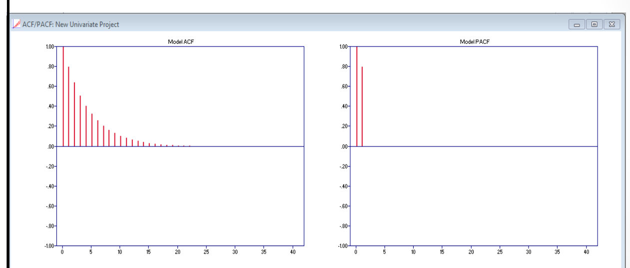
Instead, each damps gradually, with the precise pattern depending on the process.

In general, we can infer the history of $\varepsilon$ from the history of y, and the history of y from the history of $\varepsilon$.

51

## The ACF/PACF of ARMA models

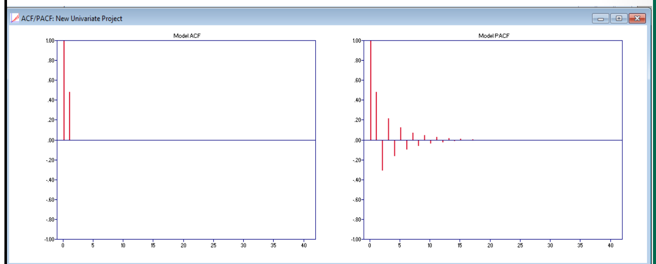
Use ITSM to look at the ACF/PACF of an AR(1) model with $\phi_1 = 0.8$.



52

## The ACF/PACF of ARMA models

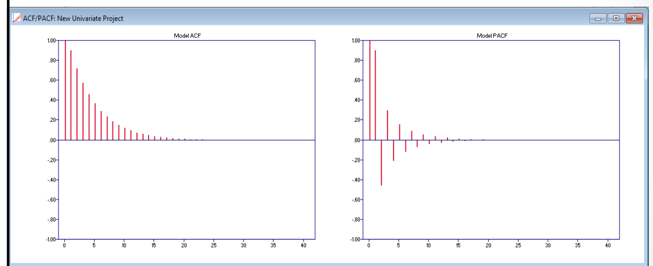
Use ITSM to look at the ACF/PACF of an MA(1) model with $\theta_1 = 0.8$.




53

## The ACF/PACF of ARMA models

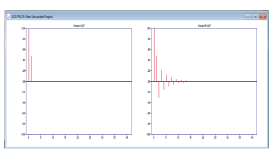
Use ITSM to look at the ACF/PACF of an ARMA(1,1) model with $\phi_1 = 0.8$ and $\theta_1 = 0.8$.

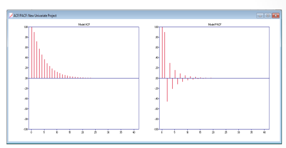


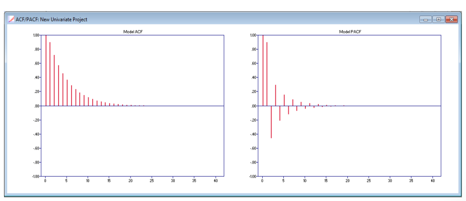

54

## The ACF/PACF of ARMA models

Note that we have the PACF of the AR(1) model and the ACF of the MA(1) model!





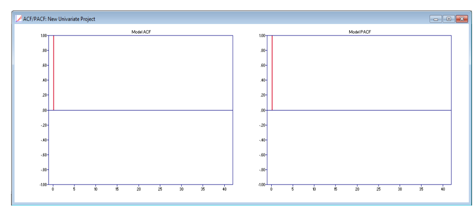


55

## The ACF/PACF of ARMA models

Use ITSM to look at the ACF/PACF of an ARMA(1,1) model with $\phi_1 = 0.8$ and $\theta_1 = -0.8$.

Note that since we have the same roots (a NO-NO!) the AR and MA components have cancelled each other. We have a white noise model!

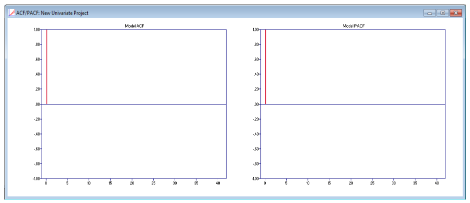


56

## The ACF/PACF of ARMA models

Use ITSM to look at the ACF/PACF of an ARMA(1,1) model with $\phi_1$ = 0.8 and $\theta_1$ = -0.8.

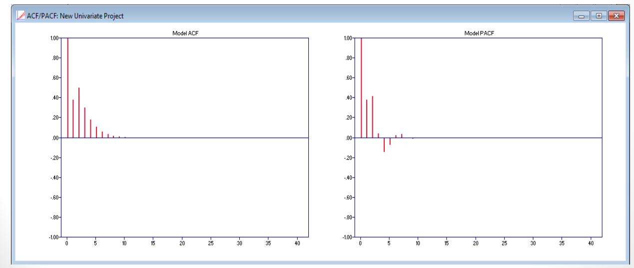
Note that since we have the same roots (a NO-NO!) the AR and MA components have cancelled each other. We have a white noise model!





## The ACF/PACF of ARMA models

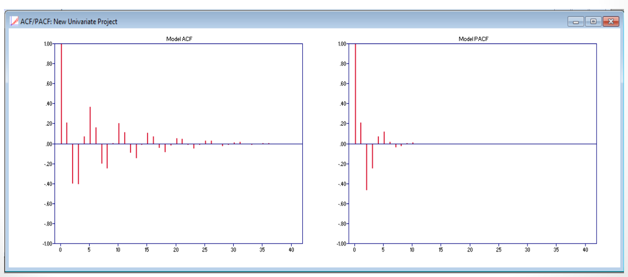
Use ITSM to look at the ACF/PACF of an ARMA(1,2) model with $\phi_1$ = 0.6, $\theta_1$ = -0.4, and $\theta_2$ = 0.4. Click the Causal/Invertible button to see that this model is both causal and invertible.





## The ACF/PACF of ARMA models

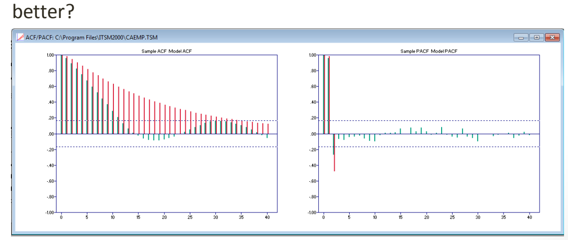
Use ITSM to look at the ACF/PACF of an ARMA(1,2) model with $\phi_1$ = 0.6, $\phi_2$ = -0.8, $\theta_1$ = -0.4, and $\theta_2$ = 0.4. Click the Causal/Invertible button to see that this model is both causal and invertible.





# 7.5 Canadian Employment

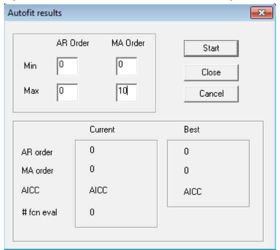
In earlier work we fit an AR(2) model to the Canadian Employment data. Recall that our model ACF/PACF did not match the data as well as we might have liked. Can we do better?

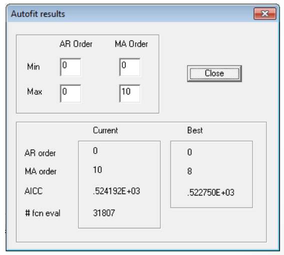


Let's play around with it some more.

# 7.5 Canadian Employment

In ITSM, open the CAEMP data and click Model -> Estimation -> Autofit. Let's try a pure MA model. Change the max AR order to 0, the max MA order to 10, and click Start.



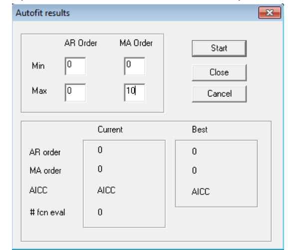



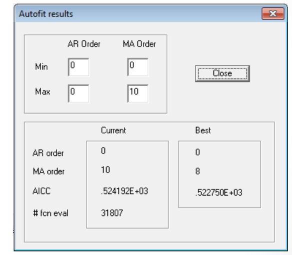

Note that the best model found is an MA(8).

61

# 7.5 Canadian Employment

In ITSM, open the CAEMP data and click Model -> Estimation -> Autofit. Let's try a pure MA model. Change the max AR order to 0, the max MA order to 10, and click Start.



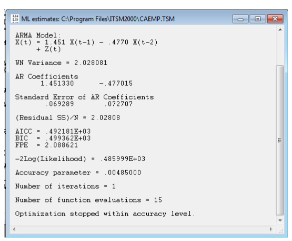



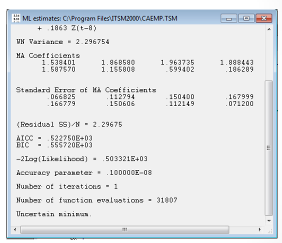

Note that the best model found is an MA(8).

62

# 7.5 Canadian Employment

If we close the estimation dialog box, we see the model information.



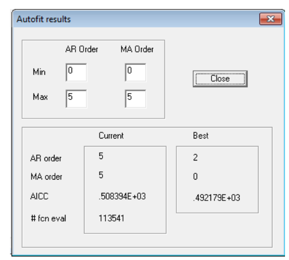




Both the AICC and BIC are higher for the MA(8) model, so we prefer the AR(2) model.

63

# 7.5 Canadian Employment

Let's try an ARMA model. In Autofit, use max orders of 5 for both the AR and MA orders and click Start.

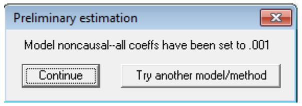



The best model found is just our AR(2) model.

64

# 7.5 Canadian Employment

The text suggests an ARMA(3,1) model. Lets give it a try. Click Model -> Estimation -> Preliminary. Enter 3 and 1 as the AR and MA orders. Choose the Innovations method for MA and ARMA Estimation. Click OK.

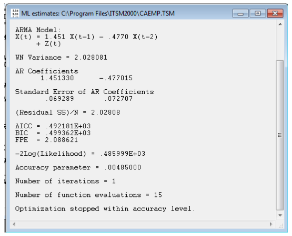


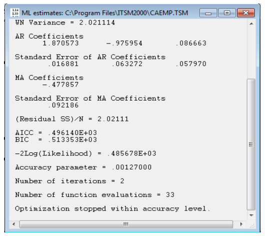

Well, that's not good! Click Try another model/method, choose the Hannan-Rissanen method, and click OK.

65

# 7.5 Canadian Employment

That's more like it! Now click Model -> Estimation -> Max likelihood and click OK in the resulting dialog box.



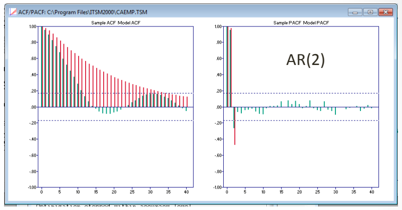

The AICC is slightly worse, and the BIC is even "worser". Let's check the sample/model ACF/PACF.

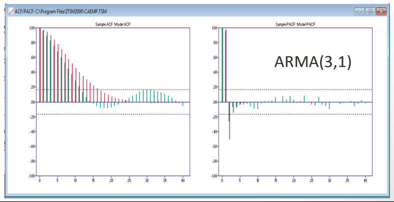
66

# 7.5 Canadian Employment

I'm not sure that things have improved much!

The Q-Q plot of the residuals and the Tests of Randomness for the residuals were also very similar for the two models.




67

# 7.5 Canadian Employment

The author likes the AR(2) model for these reasons:

The AR(2) model is more parsimonious than the ARMA(3,1).

The AICC and BIC both "select" the AR(2) model, and the ACF/PACF "fits" better over longer lags.

His third reason is that one of the roots of the AR equation is almost the same as the root of the MA equation, therefore canceling to an AR(2) model (remember our example?). Note that these are roots, not coefficients. This might be why the innovations method of preliminary estimation failed, but I am not sure.

68