2. An Introduction to Moving Average Models and ARMA Models

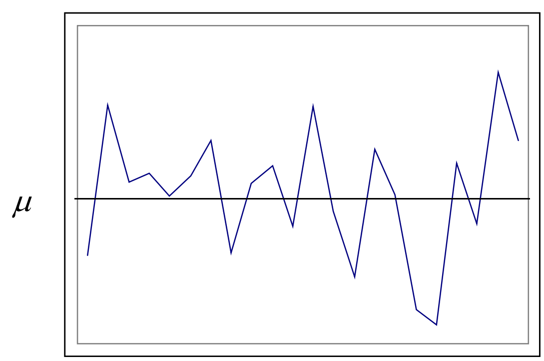
2.1 White Noise
2.2 The MA(1) model
2.3 The MA(q) model .
2.4 Estimation and forecasting of MA models.
2.5 ARMA(p,q) models.

- The Moving Average (MA) models are very different from the Autoregressive models both in terms of how we write them down and think about them as well as the implied dynamics.
- We begin with the simple idea of a white noise process and then build up to the moving average models.

2.1 White Noise

- A white noise time series is simply a mean zero, variance σ^2 series with *all* autocorrelations equal to zero.
- Example: $Y_t=\varepsilon_t$ where ε_t is iid $N(0,\sigma^2)$.
  This is just the iid Normal time series model that we talked about before.
- We could extend this model by giving it a mean: $Y_t=\mu+\varepsilon_t$
- Since ε_t is mean zero the mean of Y is μ.

White Noise for iid $N(\mu,\sigma^2)$



2.2 The MA(1) model

- We could make this model a little more interesting by saying that r_t is determined by two sequantial values of ε_t as in:

$$r_t = \mu + \underbrace{\theta\varepsilon_{t-1}}_{past\ \text{value}} + \varepsilon_t$$

- The model is clearly different from the AR model since we write the time series for r_t as a combination of two random outcomes from a Normal.

- Notice that both Y_t and Y_{t-1} are determined by ε_{t-1}

$$Y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t$$
$$Y_{t-1} = \mu + \theta\varepsilon_{t-2} + \varepsilon_{t-1}$$

- So if ε_{t-1} happens to be very large, and θ is positive, then it is likely that both Y_t and Y_{t-1} will be large.

- Even though ε_t is iid, the Y_t will be correlated!

- For technical reasons (and without loss of generality) we will restrict $|\theta| \le 1$.

- To better understand how the model works, lets think about how we would generate data for this model.
- Consider the model

$$r_t = .5\varepsilon_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ \mathrm{N}(0,1)$$

t	ε_t	ε_{t-1}	r_t
1	0.51	0.83	0.93
2	0.90	0.51	1.15
3	-1.19	0.90	-0.75
4	1.48	-1.19	0.89
5	1.50	1.48	2.24
6	-0.08	1.50	0.67
7	-0.12	-0.08	-0.16
8	-0.73	-0.12	-0.79
9	-0.19	-0.73	-0.56

Generate random outcomes from a N(0,1)

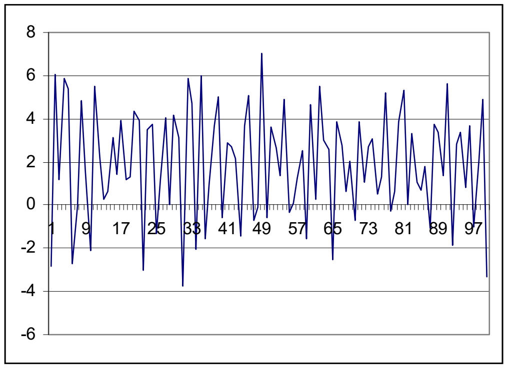
Let's look at some simulated data from an MA(1) model

Series 1 θ=-.9 μ=2 σ=2

Series 2 θ=.9 μ=2 σ=2

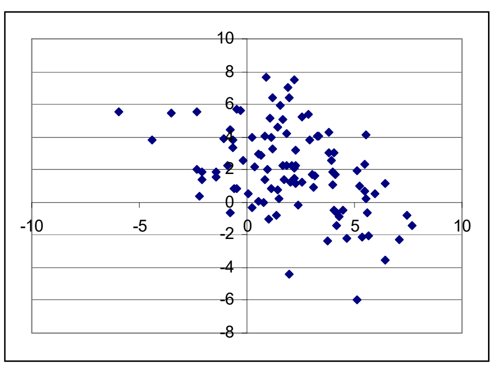
Series 3 θ=.1 μ=2 σ=2

Note that we are interested in θ here. μ just determines the level while σ determines how "spread out" the time series will be.

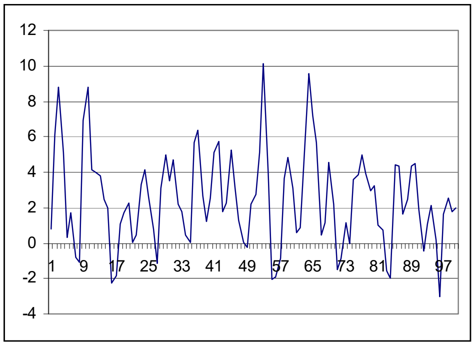


Series 1 θ=-.9 μ=2 σ=2

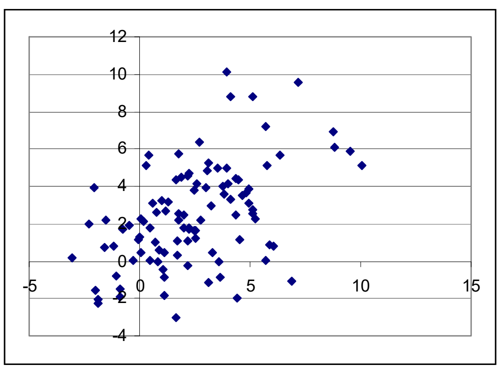
It is a little easier to see the dependence by plotting y_t against y_{t-1}



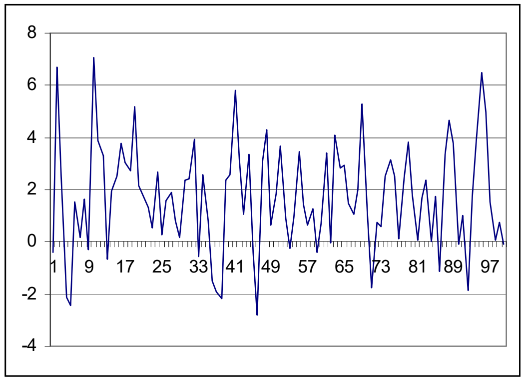
The autocorrelation between y_t and y_{t-1} is about -.5.



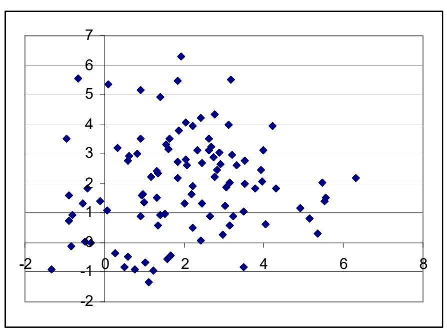
Series 2 $\theta=.9$ $\mu=2$ $\sigma=2$



The autocorrelation between y_t and y_{t-1} is about .5.



Series 3 θ =.1 μ=2 σ=2



The autocorrelation between y_t and y_{t-1} is about .2.

MA(1) model summary

- The MA(1) model is given by:

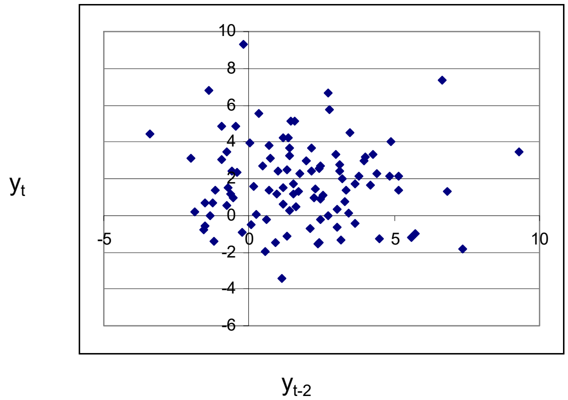
$$Y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t \text{ where } \varepsilon_\mathrm{t} \sim iid\ \mathrm{N}\left(0,\sigma^2\right)$$

- There are three parameters that we need to know, μ, θ, and σ.

2.3 The MA(q) model

- We already saw that for an MA(1) model Y_t is correlated with Y_{t-1}.
- What do you expect for the correlation between Y_t and Y_{t-2}?
- Well, Y_t depends on ε_t and ε_{t-1}
- Y_{t-2} depends on ε_{t-2} and ε_{t-3}
- Since ε_t is iid we should expect that the correlation between Y_t and Y_{t-j} should be zero for j>1.

Series 2 θ=.9 μ=2 σ=2



The autocorrelation between y_t and y_{t-2} is about 0.

- We can allow for richer dependence in Y_t by allowing Y_t to depend on more values of ε_t.

- The MA(q) model says:

$$Y_t = \mu + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

so the value of Y depends on ε_t and q past values of ε.

- An MA(2) model looks like this:

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$
$$\text{where } \varepsilon_t \sim iid \ \mathrm{N}\left(0, \sigma^2\right)$$

- Now Y_t and Y_{t-2} will be correlated since they both depend on ε_{t-2}.

  (Notice that $Y_{t-2} = \mu + \theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4} + \varepsilon_{t-2}$)

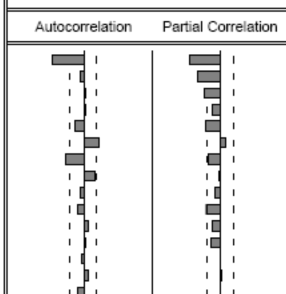
Fact: For an MA(q) model, the first q autocorrelations of Y_t will be non-zero while the $q+j^{th}$ autocorrelations will be (about) zero for all $j>0$.

Fact: For an MA(q) model, the partial autocorrelations should decay slowly.

Here is the ACF and PACF for the first MA model

Correlogram of S1

Date: 07/12/07 Time: 00:06
Sample: 1 100
Included observations: 100



	AC	PAC	Q-Stat	Prob
1	-0.473	-0.473	23.084	0.000
2	-0.046	-0.349	23.309	0.000
3	0.035	-0.234	23.439	0.000
4	0.030	-0.121	23.537	0.000
5	-0.115	-0.229	24.957	0.000
6	0.235	0.098	30.949	0.000
7	-0.277	-0.184	39.368	0.000
8	0.170	-0.014	42.555	0.000
9	-0.053	-0.067	42.874	0.000
10	-0.085	-0.211	43.687	0.000
11	0.074	-0.114	44.319	0.000
12	0.038	-0.140	44.484	0.000
13	-0.039	-0.004	44.662	0.000
14	0.072	0.025	45.272	0.000
15	-0.084	0.000	46.116	0.000

Here is the ACF and PACF for the second MA model

Correlogram of S2

Date: 07/12/07 Time: 00:08
Sample: 1 100
Included observations: 100

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.384	0.384	15.169	0.000
		2	-0.150	-0.348	17.506	0.000
		3	-0.055	0.209	17.820	0.000
		4	-0.088	-0.285	18.635	0.001
		5	-0.176	0.023	21.945	0.001
		6	-0.169	-0.209	25.031	0.000
		7	-0.051	0.108	25.312	0.001
		8	0.064	-0.059	25.765	0.001
		9	0.084	0.091	26.554	0.002
		10	-0.027	-0.197	26.640	0.003
		11	-0.023	0.156	26.700	0.005
		12	0.106	-0.039	28.010	0.006
		13	-0.010	-0.049	28.021	0.009
		14	-0.155	-0.094	30.871	0.006
		15	-0.069	0.030	31.448	0.008

Summary of ACF and PACF for AR(p) and MA(q) models

	AR(p) Model	MA(q) Model
ACF	Slowly decaying	Non-zero for first q and (about) zero thereafter
PACF	Non-zero for first p and (about) zero thereafter	Slowly decaying

2.4 Estimation and forecasting of MA models

- The MA model is a model where the errors from the last period help to predict the following outcome: $r_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t$
- Notice that if we observe the past values of the errors, we can forecast the next outcome with: $E(r_t \mid \varepsilon_{t-1}) = \mu + \theta\varepsilon_{t-1} + E(\varepsilon_t \mid \varepsilon_{t-1}) = \underbrace{\mu + \theta\varepsilon_{t-1}}_{\text{period t forecast given t-1 information}}$

  since $E(\varepsilon_t \mid \varepsilon_{t-1}) = 0$

Also, our prediction error will be $\varepsilon_t = r_t - (\mu + \theta\varepsilon_{t-1})$

- So the MA model is a model where your predictions depend on your past forecast mistakes (errors).
- But how do we estimate this thing? We don't know the right hand side variables until we have fit the model…

- Suppose that we know θ and we know the error at time 0 is 0.
- Then $r_1 = \mu + \theta\varepsilon_0 + \varepsilon_1 = \mu + \theta 0 + \varepsilon_1$
  and $\varepsilon_1 = r_1 - \mu$
- We could then use that value of ε_1 to get the prediction for r_2: $r_2 = \mu + \theta\varepsilon_1 + \varepsilon_2$
  and $\varepsilon_2 = r_2 - (\mu + \theta\varepsilon_1)$
- If we new ε_0 (and θ and μ) we could build all the one step ahead forecasts.

- It turns out that the effects of the initial choice of ε effect the forecasts less and less the further away we move from time 0.
- Basically, we can use any reasonable choice for ε_0 and eventually, as we move forward in time, we get the same values of ε's.

- This result also means that we can still estimate μ and θ by least squares.
  - Recall that the ε are also the one step ahead forecast errors.
  - We can search across all values of μ and θ and find the values that minimize the in sample sum of squared errors (after we set ε_0=0).
  - Again, the choice of 0 is not important for large samples.

- We can also use these approximate values of ε to build one step ahead forecasts.

2.5 ARMA(p,q) models

- The world does not always conform to the simple dichotomy of AR and MA models.
- In reality, the world is complex and we more complex models are often needed.
- There is nothing to prevent us from combining AR and MA models.
- This seems to work very well in practice.

ARMA(1,1)

$$r_t = \beta_0 + \beta_1 r_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$$

where $\varepsilon_{\mathrm{t}} \sim iid\ N\left(0, \sigma^2\right)$

and ε_{t} independent of $r_{t-j},\ j > 0$

- An ARMA(p,q) model will have p lags of *r* and q lags of ε.

$$r_t = \beta_0 + \underbrace{\beta_1 r_{t-1} + \cdots \beta_p r_{t-p}}_{p \text{ AR terms}} + \underbrace{\theta_1 \varepsilon_{t-1} + \cdots \theta_q \varepsilon_{t-q}}_{q \text{ MA terms}} + \varepsilon_t$$

- The model becomes difficult to interpret as the dynamics are much more complex.
- It is also difficult to infer *p* and *q* from the ACF and PACF.
- However, we can still use the same approach as before.
  - Look at the ACF and PACF (see if its pure AR or MA).
  - If it is not obviously pure AR or MA fit a low order ARMA model (I like to start with ARMA(1,1)).
  - Check the residual series to see if it is roughly uncorrelated.
  - Add additional AR and MA terms if necessary.

- Some examples in Eviews