Eco 4306 Economic and Business Forecasting

Lecture 9

Chapter 7: Forecasting with Autoregressive (AR) Processes

Outline

- introduce the autoregressive processes
- autocorrelation function - again helps us understand the past dependence, and help us to predict the dependence between today's information and the future

7.2 Autoregressive Models

- simple linear regression model with cross sectional data

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- suppose we are dealing with time series rather than cross sectional data, so that

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

  and if the explanatory variable is the lagged dependent variable $X_t = Y_{t-1}$ we get

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t$$

- main idea: past is prologue as it determines the present, which in turn sets the stage for future

7.2 Autoregressive Models

- autoregressive (AR) model is a regression model in which the dependent variable and the regressors belong to the same stochastic process, and Y_t is regressed on the lagged values of itself $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$
- stochastic process $\{Y_t\}$ follows an **autoregressive model** of order p, referred as AR(p), if

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$$

  where ε_t is a white noise process
- the order is given by the largest lag in the right-hand side of the model, so a model $Y_t = c + \phi_2 Y_{t-2} + \varepsilon_t$ is an autoregressive process AR(2) even though it has only one regressor in the right-hand side

7.2 Autoregressive Models

- we'll first analyze AR(1) and AR(2), then generalize to an autoregressive process AR(p)
- three questions we want to answer
  1. What does a time series of an AR process look like?
  2. What do the corresponding autocorrelation functions (AC and PAC) look like?
  3. What is the optimal forecast for an AR process?

7.2.1 The AR(1) Process

- consider the AR(1) process

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$$

  for different values of ϕ_1
- ϕ_1 is called the **persistence parameter**, with larger ϕ_1 the series will remain below or above the unconditional mean for longer periods
- AR(1) process is second order weakly stationary if $|\phi_1| < 1$

7.2.1 The AR(1) Process

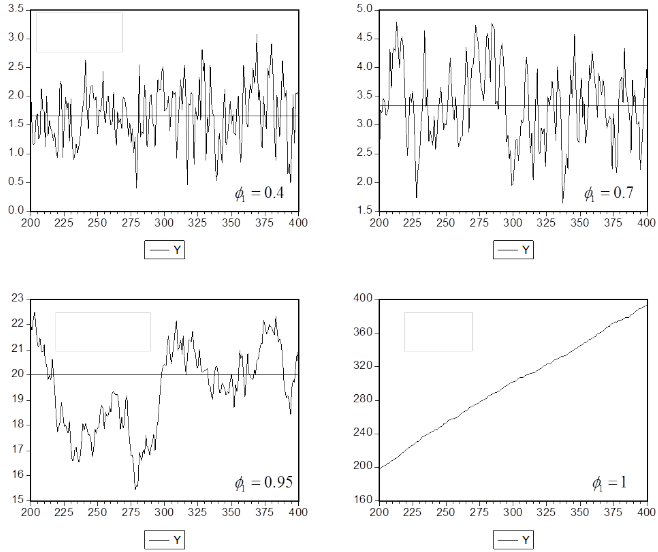
- unconditional population mean, provided that AR(1) is weakly stationary, i.e. if $|\phi_1| < 1$

$$E(Y_t) = E(c + \phi_1 Y_{t-1} + \varepsilon_t) = c + \phi_1 E(Y_{t-1}) = c + \phi_1 E(Y_t) = \frac{c}{1 - \phi_1}$$

- unconditional variance, provided that AR(1) is weakly stationary, i.e. if $|\phi_1| < 1$

$$var(Y_t) = var(c + \phi_1 Y_{t-1} + \varepsilon_t) = \phi_1^2 var(Y_{t-1}) + \sigma_\varepsilon^2 = \phi_1^2 var(Y_t) + \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

7.2.1 The AR(1) Process

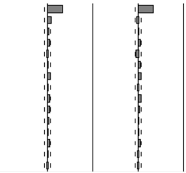


7.2.1 The AR(1) Process

autocorrelation functions of an AR(1) process with $\phi_1 > 0$ have three distinctive features

1. for theoretical autocorrelation (AC) and partial autocorrelation (PAC) functions $\rho_1 = r_1 = \phi_1$ but since sample AC and PAC functions are just estimates of the theoretical ones there is some sampling error
2. AC decreases exponentially toward zero, decay is faster when ϕ_1 is smaller; this exponential decay is given by the formula $\rho_k = \phi_1^k$; e.g. with $\phi_1 = 0.95$ we have $\rho_1 = 0.95, \rho_2 = 0.95^2 = 0.90, \rho_3 = 0.95^3 = 0.86, \dots$
3. PAC is characterized by only one spike: $r_1 \neq 0$, and $r_k = 0$ for $k > 1$

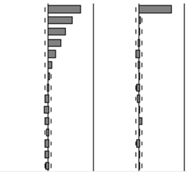
Sample: 2 1000
Included observations: 999



Autocorrelation / Partial Correlation	AC	PAC
1	0.334	0.334
2	0.082	-0.034
3	0.018	0.002
4	0.043	0.044
5	-0.022	-0.056
6	0.006	0.033
7	0.060	0.058
8	0.016	-0.031
9	0.041	0.053
10	0.051	0.024
11	0.028	-0.006
12	0.001	-0.001
13	0.036	0.037
14	0.008	-0.023
15	-0.026	-0.023

$\phi_1 = 0.4$

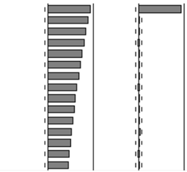
Sample: 2 1000
Included observations: 999



Autocorrelation / Partial Correlation	AC	PAC
1	0.732	0.732
2	0.546	0.021
3	0.400	-0.015
4	0.279	-0.030
5	0.168	-0.058
6	0.085	-0.025
7	0.032	-0.003
8	-0.026	-0.057
9	-0.071	-0.033
10	-0.084	0.011
11	-0.060	0.057
12	-0.048	-0.010
13	-0.060	-0.053
14	-0.063	-0.013
15	-0.054	0.006

$\phi_1 = 0.7$

Sample: 2 1000
Included observations: 999

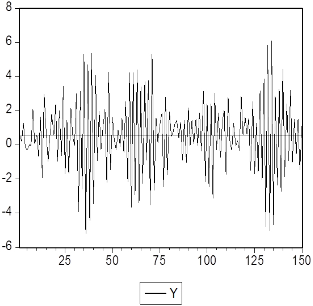


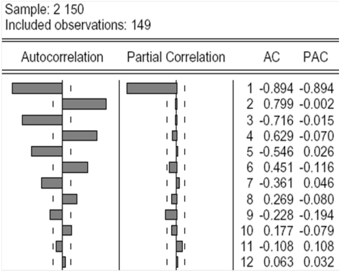
Autocorrelation / Partial Correlation	AC	PAC
1	0.951	0.951
2	0.904	-0.006
3	0.861	0.009
4	0.817	-0.019
5	0.774	-0.016
6	0.734	0.005
7	0.696	-0.002
8	0.658	-0.022
9	0.622	0.013
10	0.590	0.010
11	0.561	0.011
12	0.534	0.017
13	0.508	-0.014
14	0.481	-0.016
15	0.457	0.007

$\phi_1 = 0.95$

7.2.1 The AR(1) Process

- if $\phi_1 < 0$ the autocorrelation functions have the same three properties above
- main difference: negative sign of the persistence parameter, causes the oscillating behavior of AC which switch between positive an negative numbers

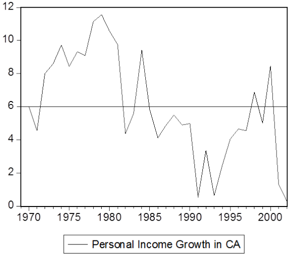


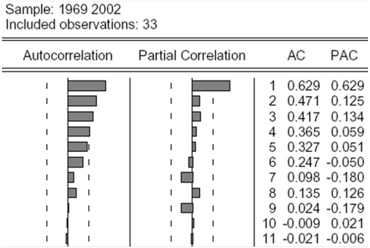

Sample: 2 150
Included observations: 149

Autocorrelation	Partial Correlation		AC	PAC
		1	-0.894	-0.894
		2	0.799	-0.002
		3	-0.716	-0.015
		4	0.629	-0.070
		5	-0.546	0.026
		6	0.451	-0.116
		7	-0.361	0.046
		8	0.269	-0.080
		9	-0.228	-0.194
		10	0.177	-0.079
		11	-0.108	0.108
		12	0.063	0.032

7.2.1 The AR(1) Process

Growth of Per Capita Personal Income Growth in California, 1969-2002





Sample: 1969 2002
Included observations: 33

Lag	AC	PAC
1	0.629	0.629
2	0.471	0.125
3	0.417	0.134
4	0.365	0.059
5	0.327	0.051
6	0.247	-0.050
7	0.098	-0.180
8	0.135	0.126
9	0.024	-0.179
10	-0.009	0.021
11	-0.021	-0.006

7.2.1 The AR(1) Process

- recall: under quadratic loss function the optimal point forecast is conditional mean, $f_{t,h} = \mu_{t+h|t} = E(Y_{t+h}|I_t)$

h	$\mu_{t+h\|t}$	$\sigma^2_{t+h\|t}$
1	$c + \phi_1 y_t$	σ^2_ε
2	$(1+\phi_1)c + \phi_1^2 y_t$	$(1+\phi_1^2)\sigma^2_\varepsilon$
$\vdots$		
s	$(1+\phi_1+\phi_1^2+\ldots+\phi_1^{s-1})c + \phi_1^s y_t$	$(1+\phi_1^2+\phi_1^4+\ldots+\phi_1^{2(s-1)})\sigma^2_\varepsilon$

- note that as $s \to \infty$ the forecast converges to the unconditional mean

$$f_{t,s} = (1+\phi_1+\phi_1^2+\phi_1^3+\ldots)c = \frac{c}{1-\phi_1}$$
$$\sigma^2_{t+s|t} = (1+\phi_1^2+\phi_1^4+\phi_1^6+\ldots) = \frac{\sigma^2_\varepsilon}{1-\phi_1^2}$$

- forecasting with an AR(1) is limited by the short memory of the process - in the long run the forecast converges to the unconditional mean

7.2.2 The AR(2) Process

- consider the AR(2) process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

- unconditional population mean, provided that AR(2) is weakly stationary

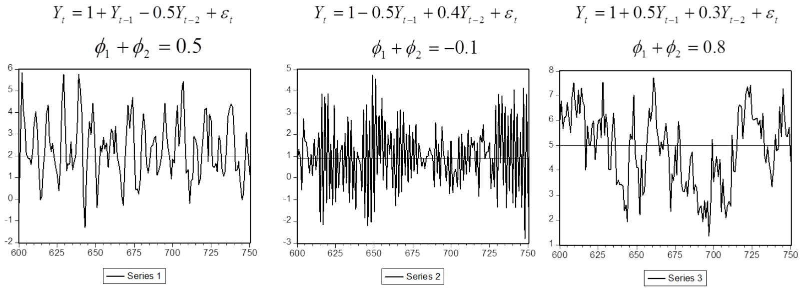
$$\begin{aligned} E(Y_t) &= E(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t) = c + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) \\ &= c + \phi_1 E(Y_t) + \phi_2 E(Y_t) = \frac{c}{1 - \phi_1 - \phi_2} \end{aligned}$$

- unconditional variance, provided that AR(2) is weakly stationary

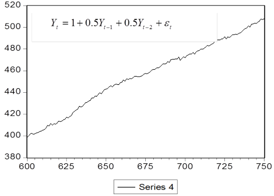
$$\begin{aligned} var(Y_t) &= var(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t) = \phi_1^2 var(Y_{t-1}) + \phi_2^2 var(Y_{t-2}) + \sigma_\varepsilon^2 \\ &= \phi_1^2 var(Y_t) + \phi_2^2 var(Y_t) + \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2 - \phi_2^2} \end{aligned}$$

7.2.2 The AR(2) Process

- larger values of $\phi_1 + \phi_2$ imply smoother time series



- if $\phi_1 + \phi_2 = 1$ time series becomes non-stationary

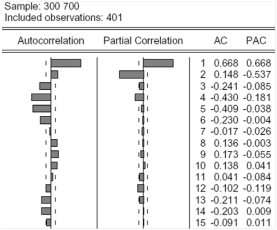


7.2.2 The AR(2) Process

autocorrelation functions of an AR(2) process have three distinctive features

1. for theoretical autocorrelation (AC) and partial autocorrelation (PAC) functions $\rho_1 = r_1$ and $r_2 = \phi_2$ but since sample AC and PAC functions are just estimates of the theoretical ones there is some sampling error
2. AC decreases toward zero, either in wave-like pattern, in oscillating pattern, or in exponentially decaying pattern
3. PAC is characterized by only two non-zero spikes: $r_1 \neq 0$, $r_2 \neq 0$, and $r_k = 0$ for $k > 2$

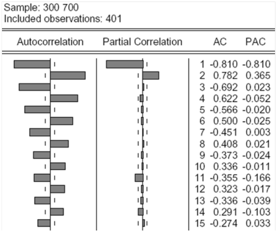
$Y_t = 1 + Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$



Sample: 300 700
Included observations: 401

	AC	PAC
1	0.668	0.668
2	0.148	-0.537
3	-0.241	-0.085
4	-0.430	-0.181
5	-0.409	-0.038
6	-0.230	-0.004
7	-0.017	-0.026
8	0.136	-0.003
9	0.173	-0.055
10	0.138	0.041
11	0.041	-0.084
12	-0.102	-0.119
13	-0.211	-0.074
14	-0.203	0.009
15	-0.091	0.011

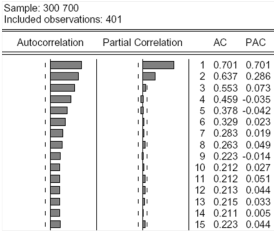
$Y_t = 1 - 0.5Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t$



Sample: 300 700
Included observations: 401

	AC	PAC
1	-0.810	-0.810
2	0.782	0.365
3	-0.692	0.023
4	0.622	-0.052
5	-0.566	-0.020
6	0.500	-0.025
7	-0.451	0.003
8	0.408	0.021
9	-0.373	-0.024
10	0.336	-0.011
11	-0.355	-0.166
12	0.323	-0.017
13	-0.336	-0.039
14	0.291	-0.103
15	-0.274	0.033

$Y_t = 1 + 0.5Y_{t-1} + 0.3Y_{t-2} + \varepsilon_t$



Sample: 300 700
Included observations: 401

	AC	PAC
1	0.701	0.701
2	0.637	0.286
3	0.553	0.073
4	0.459	-0.035
5	0.378	-0.042
6	0.329	0.023
7	0.283	0.019
8	0.263	0.049
9	0.223	-0.014
10	0.212	0.027
11	0.212	0.051
12	0.213	0.044
13	0.215	0.033
14	0.211	0.005
15	0.223	0.044

7.2.2 The AR(2) Process

- recall: under quadratic loss function the optimal point forecast is conditional mean, $f_{t,h} = \mu_{t+h|t} = E(Y_{t+h}|I_t)$

h	$\mu_{t+h\|t}$	$\sigma^2_{t+h\|t}$
1	$c + \phi_1 y_t + \phi_2 y_{t-1}$	σ^2_ε
2	$c + \phi_1 f_{t,1} + \phi_2 y_t$	$(1 + \phi_1^2)\sigma^2_\varepsilon$
$\vdots$		
s	$c + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2}$	$\sigma^2_\varepsilon + \phi_1^2 \sigma^2_{t+s-1\|t} + \phi_2^2 \sigma^2_{t+s-2\|t} + 2\phi_1\phi_2 cov(e_{t,s-1}, e_{t,s-2})$

- just like in the case of AR(1), as $s \to \infty$ the forecast $f_{t,s}$ converges to the unconditional mean, and the variance of the forecast error $e_{t,s}$ converges to the unconditional variance of the process
- forecasting with an AR(2) is again limited by the short memory of the process

7.2.3 The AR(p) Process

- consider the AR(p) process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$$

- unconditional population mean, provided that AR(p) is weakly stationary

$$E(Y_t) = \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$$

- unconditional variance, provided that AR(p) is weakly stationary

$$var(Y_t) = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2 - \phi_2^2 - \ldots - \phi_p^2}$$

7.2.3 The AR(p) Process

autocorrelation functions of an AR(p) process have following features

1. $\rho_1 = r_1$
2. AC decreases toward zero, either in wave-like pattern, in oscillating pattern, or in exponentially decaying pattern
3. PAC has only p non-zero spikes: $r_k \neq 0$ if $k \leq p$, and $r_k = 0$ for $k > p$

7.2.3 The AR(p) Process

- under quadratic loss function the optimal point forecast is conditional mean, $f_{t,h} = \mu_{t+h|t} = E(Y_{t+h}|I_t)$ and we have

h	$\mu_{t+h\|t}$	$\sigma^2_{t+h\|t}$
1	$c + \phi_1 y_t + \phi_2 y_{t-1} + \ldots + \phi_p y_{t-p}$	σ^2_ε
2	$c + \phi_1 f_{t,1} + \phi_2 y_t + \ldots + \phi_p y_{t-p+1}$	$(1 + \phi_1^2)\sigma^2_\varepsilon$
$\vdots$		
s	$c + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2} + \ldots + \phi_p f_{t,s-p}$	$\sigma^2_\varepsilon + \sum_{i=1}^{p} \phi_i^2 \sigma^2_{t+s-i\|t} + 2 \sum_{i=1}^{p} \sum_{j=i+1}^{p}$

- just like in the case of AR(1) and AR(2), as $s \to \infty$ the forecast $f_{t,s}$ converges to the unconditional mean, and the variance of the forecast error $e_{t,s}$ converges to the unconditional variance of the process