

Lecture 5

ARMA Models

Lars A. Lochstoer
UCLA Anderson School of Management

Winter 2025

Outline

- ① Autoregressive Models
- ② Application: Bond Pricing
- ③ Moving Average Models
- ④ ARMA Models
- ⑤ References
- ⑥ Appendix

Autoregressive Models

ARMA Models

- **parsimonious** description of (univariate) time series (mimicking autocorrelation etc.)
- very useful tools for forecasting (and commonly used in industry)
 - ▶ forecasting sales, earnings revenue growth at the firm level or at the industry level
 - ▶ forecasting GPD growth, inflation at the national level

Autoregressive process of order 1

- lagged returns might be useful in predicting returns.
- we consider a model that allows for this:

$$r_{t+1} = \phi_0 + \phi_1 r_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \text{WN}(0, \sigma_\varepsilon^2)$$

- ▶ $\{\varepsilon_t\}$ represents the 'news':

$$\varepsilon_t = r_t - E_{t-1}[r_t]$$

ε_t is what you know about the process at t but not at $t-1$

- ▶ Economists often call ε_t the 'shocks' or 'innovations'.
- this model is referred to as an **AR(1)**

Transition density

Definition

Given an information set \mathcal{F}_t , the **transition density** of a random variable r_{t+1} is the conditional distribution of r_{t+1} given by:

$$r_{t+1} \sim p(r_{t+1} | \mathcal{F}_t; \theta)$$

- The information set \mathcal{F}_t is often (but not always) the history of the process $r_t, r_{t-1}, r_{t-2}, \dots$
- In this case, the transition density is written:

$$r_{t+1} \sim p(r_{t+1} | r_t, r_{t-1}, \dots; \theta)$$

- A transition density is **Markov** if it depends on its finite past.

AR(1) transition density

- Consider the AR(1) model with Gaussian shocks

$$r_{t+1} = \phi_0 + \phi_1 r_t + \varepsilon_{t+1}, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- The transition density is **Markov of order 1**.

$$r_{t+1} \sim p(r_{t+1} | r_t; \theta)$$

the rest of the history r_{t-2}, r_{t-3}, \dots is irrelevant.

- With Gaussian shocks ε_t , the transition density is:

$$r_{t+1} \sim N(\phi_0 + \phi_1 r_t, \sigma_\varepsilon^2)$$

- conditional mean and conditional variance:

$$\begin{aligned} E[r_{t+1} | r_t] &= \phi_0 + \phi_1 r_t, \\ V[r_{t+1} | r_t] &= V[\varepsilon_{t+1}] = \sigma_\varepsilon^2. \end{aligned}$$

Unconditional mean of AR(1)

- assume that the series is covariance-stationary
- compute the unconditional mean μ .
 - take unconditional expectations:

$$E[r_{t+1}] = \phi_0 + \phi_1 E[r_t].$$

- use stationarity: $E[r_{t+1}] = E[r_t] = \mu$:

$$\mu = \phi_0 + \phi_1 \mu,$$

and solving for the unconditional mean:

$$\mu = \frac{\phi_0}{1 - \phi_1}.$$

- mean exists if $\phi_1 \neq 1$ and is zero if $\phi_0 = 0$

Mean Reversion

- if $\phi_1 \neq 1$, we can rewrite the AR(1) process as:

$$r_{t+1} - \mu = \phi_1 (r_t - \mu) + \varepsilon_{t+1}.$$

- suppose $0 < \phi_1 < 1$
 - ▶ when $r_t > \mu$, the process is expected to get **closer** to the mean:

$$E_t[r_{t+1} - \mu] = \phi_1 (r_t - \mu) < (r_t - \mu).$$

- ▶ when $r_t < \mu$, the process is expected to get **closer** to the mean:

$$E_t[r_{t+1} - \mu] = \phi_1 (r_t - \mu) > (r_t - \mu).$$

- the smaller ϕ_1 , the higher the speed of mean reversion

Mean Reversion

- we can rewrite the AR(1) process as:

$$r_{t+2} - \mu = \phi_1^2 (r_t - \mu) + \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}.$$

- suppose $0 < \phi_1 < 1$
 - ▶ when $r_t > \mu$, the process is expected to get **closer** to the mean:

$$E_t[r_{t+2} - \mu] = \phi_1^2 (r_t - \mu) < (r_t - \mu).$$

- ▶ when $r_t < \mu$, the process is expected to get **closer** to the mean:

$$E_t[r_{t+2} - \mu] = \phi_1^2 (r_t - \mu) > (r_t - \mu).$$

Half Life

- we can rewrite the AR(1) process as:

$$r_{t+h} - \mu = \phi^h (r_t - \mu) + \phi^{h-1} \varepsilon_{t+1} + \dots + \varepsilon_{t+h}.$$

- suppose $0 < \phi_1 < 1$
 - ▶ at the **half-life**, the process is expected to cover **1/2** of the distance to the mean:

$$E_t[r_{t+h} - \mu] = \phi_1^h (r_t - \mu) = .5 (r_t - \mu).$$

- the half-life is defined by setting $\phi_1^h = 0.5$ and solving

$$h = \log(0.5) / \log(\phi_1)$$

Variance of AR(1)

Compute the unconditional variance:

- take the expectation of the square of:

$$r_{t+1} - \mu = \phi_1 (r_t - \mu) + \varepsilon_{t+1}.$$

- we obtain the following expression for the unconditional variance:

$$V[r_{t+1}] = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2},$$

provided that $\phi_1^2 < 1$ because the variance has to be positive and bounded

- covariance stationarity requires that

$$-1 < \phi_1 < 1.$$

- in addition, if $-1 < \phi_1 < 1$, we can show that the series is covariance stationary because the mean and variance are finite

Continuous-Time Model

Definition

In a continuous-time model, the log of stock prices, $p_t = \log P_t$, follows an **Ornstein-Uhlenbeck process** if:

$$dp_t = \kappa(\mu_p - p_t)dt + \sigma_p dB_t \quad (1)$$

Continuous-time version of a discrete-time, Gaussian AR(1) process.

Suppose we observe the process (1) at discrete intervals Δt , then this is equivalent to:

$$p_t = \mu + \phi_1(p_{t-1} - \mu) + \sigma \varepsilon_t \quad \varepsilon_t \sim N(0, 1)$$

where

- $\phi_1 = \exp(-\kappa\Delta t)$
- $\mu = \mu_p$
- $\sigma^2 = (1 - \exp(-2\kappa\Delta t)) \frac{\sigma_p^2}{2\kappa}$.

Dynamic Multipliers

- use the expression for the mean of the AR(1) to obtain:

$$r_{t+1} - \mu = \phi_1 (r_t - \mu) + \varepsilon_{t+1}.$$

- by repeated substitution, we get:

$$r_t - \mu = \sum_{i=0}^t \phi_1^i \varepsilon_{t-i} + \phi^{t+1} (r_{-1} - \mu).$$

- value of r_t at t is stated as a function of the **history of shocks** $\{\varepsilon_\tau\}_{\tau=0}^{t-1}$ and its value at time $t = -1$
- effect of shocks die out over time provided that $-1 < \phi_1 < 1$.

Dynamic Multipliers

calculate the effect of a change ε_0 on r_t :

$$\frac{\partial[r_t - \mu]}{\partial\varepsilon_0} = \phi_1^t.$$

$$\frac{\partial[r_{t+j} - \mu]}{\partial\varepsilon_t} = \phi_1^j.$$

in a covariance stationary model, *dynamic multiplier* only depends on j , not on t

Again, note that we need $|\phi_1| < 1$ for a stationary (non-explosive) system where shocks die out: $\lim_{j \rightarrow \infty} \phi_1^j = 0$

MA(infinity) representation

- use the expression for the mean of the AR(1) to obtain:

$$r_{t+1} - \mu = \phi_1 (r_t - \mu) + \varepsilon_{t+1}.$$

- by repeated substitution:

$$r_t - \mu = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}.$$

- ▶ **linear** function of past innovations!
- ▶ fits into class of linear time series

Autocovariances of an AR(1)

- take the unconditional expectation:

$$(r_t - \mu)(r_{t-j} - \mu) = \phi_1(r_{t-1} - \mu)(r_{t-j} - \mu) + \varepsilon_t(r_{t-j} - \mu).$$

- this yields:

$$E[(r_t - \mu)(r_{t-j} - \mu)] = \phi_1 E[(r_{t-1} - \mu)(r_{t-j} - \mu)] + E[\varepsilon_t(r_{t-j} - \mu)].$$

- or, using notation from Lecture 2:

$$\begin{aligned}\gamma_j &= \phi_1 \gamma_{j-1}, & j > 0 \\ \gamma_0 &= \phi_1 \gamma_{-1} + \sigma_\varepsilon^2, & j = 0\end{aligned}$$

- note that $\gamma_{-j} = \gamma_j$

Autocorrelation Function

- it immediately implies that the ACF is:

$$\rho_j = \phi_1 \rho_{j-1}, \quad j \geq 0$$

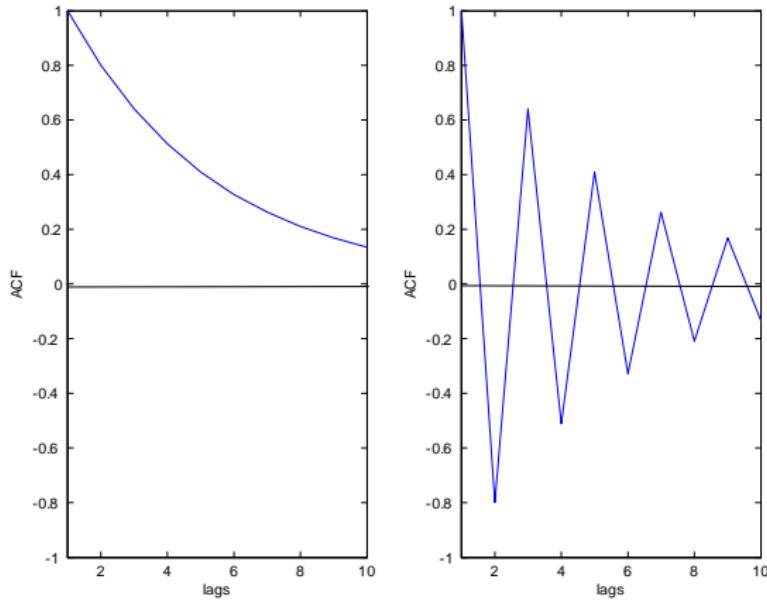
and $\rho_0 = 1$

- combining these two equations imply that:

$$\rho_j = \phi_1^j$$

- ▶ exponential decay at a rate ϕ_1

Autocorrelation Function of an AR(1)



Autocorrelation Function for AR(1). The left panel considers $\phi_1 = 0.8$. The right panel considers $\phi_1 = -0.8$.

AR(p)

Definition

The **AR(p)** model is defined as:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$$

- other lagged returns might be useful in predicting returns
- similar to multiple regression model with p lagged variables as explanatory variables
- the **AR(p)** is **Markov of order p .**

Conditional Moments

- conditional mean and conditional variance:

$$\begin{aligned} E[r_{t+1}|r_t, \dots, r_{t-p+1}] &= \phi_0 + \phi_1 r_t + \dots + \phi_p r_{t-p+1} \\ V[r_{t+1}|r_t, \dots, r_{t-p+1}] &= V[\varepsilon_{t+1}] = \sigma_\varepsilon^2 \end{aligned}$$

- moments conditional on r_t, \dots, r_{t-p+1} are not correlated with $r_{t-i}, i \geq p$

AR(2)

- consider the model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t \quad \varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right)$$

- take unconditional expectations to compute the mean

$$E[r_t] = \phi_0 + \phi_1 E[r_{t-1}] + \phi_2 E[r_{t-2}]$$

- Assuming stationarity and solving for the mean:

$$E[r_t] = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

provided that $\phi_1 + \phi_2 \neq 1$.

- using this expression for μ write the model in deviation from means:

$$r_t - \mu = \phi_1 (r_{t-1} - \mu) + \phi_2 (r_{t-2} - \mu) + \varepsilon_t$$

Autocorrelations of an AR(2)

- take the expectation of :

$$\begin{aligned}(r_t - \mu)(r_{t-j} - \mu) &= \phi_1(r_{t-1} - \mu)(r_{t-j} - \mu) \\ &\quad + \phi_2(r_{t-2} - \mu)(r_{t-j} - \mu) + \varepsilon_t(r_{t-j} - \mu)\end{aligned}$$

- this yields:

$$\begin{aligned}E[(r_t - \mu)(r_{t-j} - \mu)] &= \phi_1 E[(r_{t-1} - \mu)(r_{t-j} - \mu)] \\ &\quad + \phi_2 E[(r_{t-2} - \mu)(r_{t-j} - \mu)] \\ &\quad + E[\varepsilon_t(r_{t-j} - \mu)]\end{aligned}$$

- or, using different notation:

$$\begin{aligned}\gamma_j &= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}, & j > 0 \\ \gamma_0 &= \phi_1 \gamma_{-1} + \phi_2 \gamma_{-2} + \sigma_\varepsilon^2, & j = 0\end{aligned}$$

Autocorrelations of an AR(2)

- the ACF:

$$\begin{aligned}\rho_j &= \phi_1\rho_{j-1} + \phi_2\rho_{j-2}, & j \geq 2 \\ \rho_0 &= \phi_1\rho_{-1} + \phi_2\rho_{-2} + \sigma_\epsilon^2 / \gamma_0, & j = 0\end{aligned}$$

which implies that the ACF of an AR(2) satisfies a second-order difference equation:

$$\begin{aligned}\rho_1 &= \phi_1\rho_0 + \phi_2\rho_1 \\ \rho_j &= \phi_1\rho_{j-1} + \phi_2\rho_{j-2}, & j \geq 2\end{aligned}$$

Roots

Definition

The second-order difference equation for the ACF:

$$(1 - \phi_1 B - \phi_2 B^2) \rho_j = 0,$$

where B is the **back-shift operator**: $B\rho_j = \rho_{j-1}$

Note that we can write the above as:

$$(1 - \omega_1 B) (1 - \omega_2 B) \rho_j = 0$$

- A useful factorization
- Intuitively, the AR(2) is an "AR(1) on top of another AR(1)"
- From AR(1) math, we had that each AR(1) is stationary if its autocorrelation is less than one in absolute value.
- The 'roots' ω_j should satisfy similar property for AR(2) to be stationary

Finding the roots

A simple case:

$$\begin{aligned}1 - \phi_1 B - \phi_2 B^2 &= (1 - \omega_1 B)(1 - \omega_2 B) \\&= 1 - (\omega_1 + \omega_2)B + \omega_1 \omega_2 B^2\end{aligned}$$

and so we solve using the relations:

$$\begin{aligned}\phi_1 &= \omega_1 + \omega_2 \\ \phi_2 &= -\omega_1 \omega_2\end{aligned}$$

The solutions to this are the inverses to the solutions to the second order polynomial in the scalar-valued x :

$$(1 - \phi_1 x - \phi_2 x^2) = 0,$$

- the solutions to this equation are given by:

$$x_1, x_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

- the inverses are the **characteristic roots**: $\omega_1 = x_1^{-1}$ and $\omega_2 = x_2^{-1}$

Roots (real, distinct case)

- two characteristic roots: $\omega_1 = x_1^{-1}$ and $\omega_2 = x_2^{-1}$
- both characteristic roots are real-valued if the discriminant is greater than zero: $\phi_1^2 + 4\phi_2 > 0$
 - ▶ then we can factor the polynomial as:

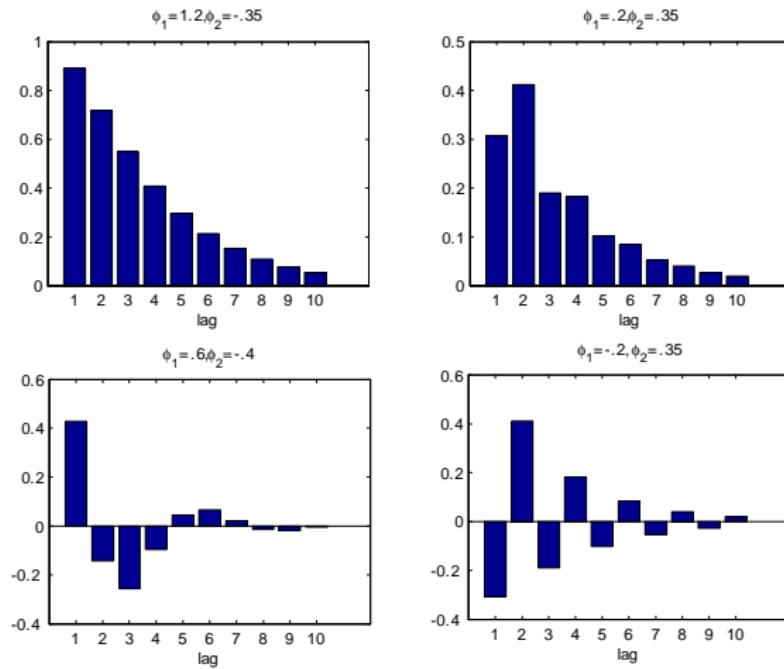
$$(1 - \phi_1 B - \phi_2 B^2) = (1 - \omega_1 B)(1 - \omega_2 B)$$

- ▶ *two AR(1) models on top of each other*
- The ACF will decay like an AR(1) at long lags
 - ▶ Intuition: the effect of the smallest ω dies out more quickly and you are left effectively with just an AR(1)

Roots (complex-valued case)

- two characteristic roots: $\omega_1 = x_1^{-1}$ and $\omega_2 = x_2^{-1}$
- both characteristic roots are complex-valued if the discriminant is negative:
$$\phi_1^2 + 4\phi_2 < 0$$
- Then, $\omega_1 = x_1^{-1}$ and $\omega_2 = x_2^{-1}$ are complex numbers.
- The ACF will look like damped sine and cosine waves.

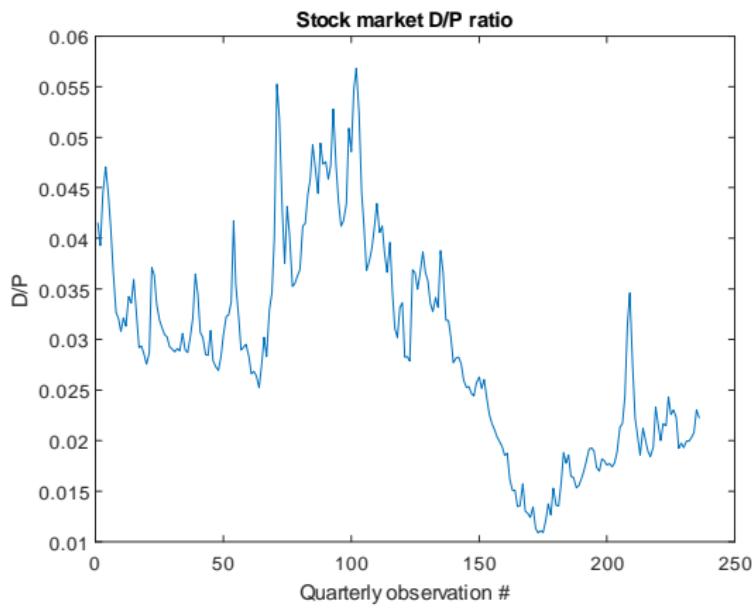
Autocorrelation for AR(2)



Autocorrelation Function for AR(2) processes.

AR(2) Example: The Dividend Price Ratio

- The stock market Dividend to Price ratio is:
 - Sum of last year's dividends to firms in the market divided by current market value
 - A "Valuation Ratio"
 - Very slow-moving (persistent); quarterly postWW2 data for U.S.:



Estimate AR(2) on this variable

```
ARIMA(2,0,0) Model:  
-----  
Conditional Probability Distribution: Gaussian  


| Parameter | Value       | Standard Error | t Statistic |
|-----------|-------------|----------------|-------------|
| Constant  | 0.00123254  | 0.00074679     | 1.65045     |
| AR{1}     | 1.09319     | 0.0527929      | 20.7072     |
| AR{2}     | -0.137308   | 0.051282       | -2.67752    |
| Variance  | 7.84588e-06 | 1.92026e-07    | 40.8583     |


```

- Stationarity test:

$$1 - 1.09319x + 0.13731x^2 = 0$$

- ▶ Roots greater than 1, so stationary despite $\phi_1 = 1.093 > 1$ as $\phi_2 = -0.137$.
- ▶ Unconditional mean:

$$\mu = \frac{0.00123254}{1 - 1.09319 + 0.13731} = 0.0279$$

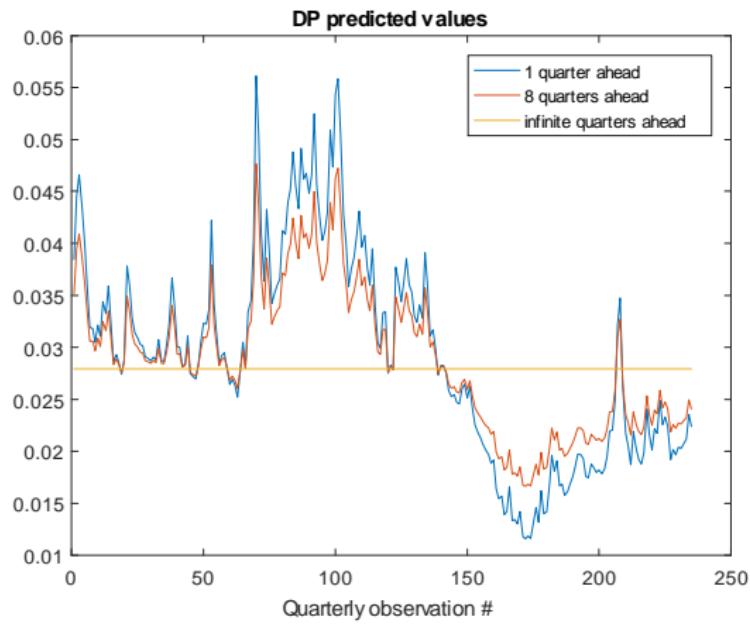
AR(2) DP prediction

Pred_DP1 = uncond_mean + phi1*(DP(2:end)-uncond_mean) + phi2*(DP(1:end-1)-uncond_mean);

Pred_DP2 = uncond_mean + phi1*(Pred_DP1-uncond_mean) + phi2*(DP(2:end)-uncond_mean);

Pred_DP3 = uncond_mean + phi1*(Pred_DP2-uncond_mean) + phi2*(Pred_DP1-uncond_mean);

etc.



Stationarity

- Recall: The modulus of $z = a + bi$ is $|z| = \sqrt{a^2 + b^2}$. Thus, for real numbers the modulus is simply the absolute value.

Result:

- An AR(1) process is stationary if its characteristic root is less than one, i.e. if $1/x = \phi_1$ is less than one in modulus. This condition implies that $\rho_j = \phi_1^j$ converges to zero as $j \rightarrow \infty$.
- An AR(2) process is stationary if the two characteristic roots ω_1 and ω_2 (the inverses of the solutions to those two equations) are less than one in modulus.

Stationarity of AR(p)

- An AR(p) process is stationary if all p characteristic roots of the below polynomial are less than one in modulus

$$1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0$$

- see chapter 2 in Hamilton (1994) for details.

Partial Autocorrelation Function

Definition

The PACF of a stationary series is defined as $\{\phi_{j,j}\}, j = 1, \dots, n$

$$r_t = \phi_{0,1} + \phi_{1,1} r_{t-1} + v_{1t}$$

$$r_t = \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + v_{2t}$$

$$r_t = \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + v_{3t}$$

...

- These are simple multiple regressions that can be estimated with least squares.
- $\phi_{p,p}$ shows the incremental contribution of r_{t-p} to r_t over an $AR(p-1)$ model

PACF

Definition

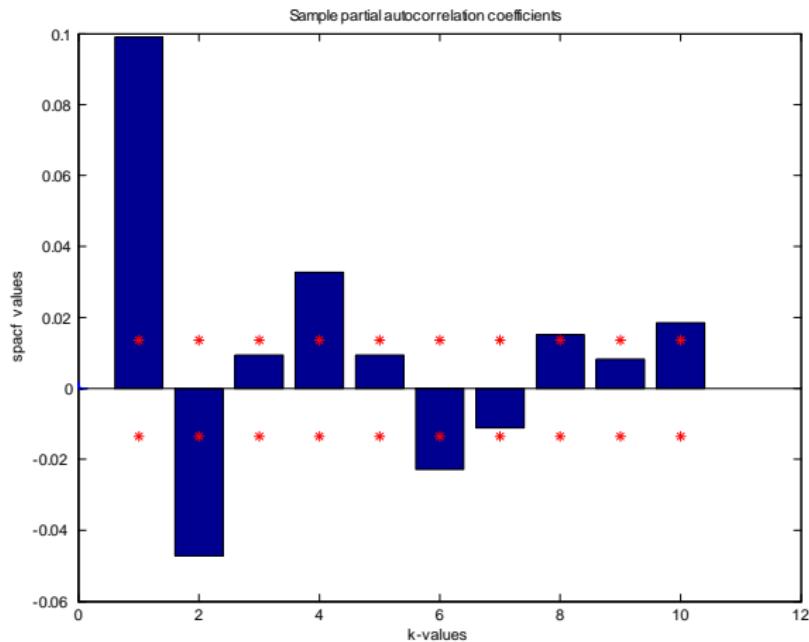
The **sample partial autocorrelations (PACF)** of a time series are defined as
 $\hat{\phi}_{1,1}, \hat{\phi}_{2,2}, \dots, \hat{\phi}_{p,p}, \dots$

Partial Autocorrelation Function

The PACF of an $AR(p)$ satisfies:

- ① $\hat{\phi}_{p,p} \rightarrow \phi_p$ as sample size increases
 - ② $\hat{\phi}_{j,j} \rightarrow 0$ for $j > p$
-
- for an $AR(p)$ series, the sample PACF cuts off after lag p
 - \Rightarrow look at the sample PACF to determine an appropriate value of p

PACF of Daily Log Returns



PACF for Daily log Returns on VW-CRSP Index. Two standard error bands around zero. 1926-2007.

Information Criteria

- information criteria help determine the **optimal lag length**
- the Akaike (1973) information criterion:

$$AIC = -2 \ln(\text{likelihood}) + 2(\text{number of parameters})$$

- the Bayesian information criterion of Schwarz (1978):

$$BIC = -2 \ln(\text{likelihood}) + \ln T(\text{number of parameters})$$

- ▶ the BIC penalty depends on the sample size T
- for different values of p , compute $AIC(p)$ and/or $BIC(p)$ pick the lag length with the minimum AIC/BIC

Manufacturing White Noise

- to check the performance of the AR model you've selected: **check the residuals!!**
- residuals should look like **white noise**
 - ▶ look at the ACF of the residuals
 - ▶ perform Ljung-Box test on residuals
 - ▶ $Q(m) \sim \chi^2(m - p)$ where p is the lag length of the $AR(p)$ model

Forecasting

- suppose we have an AR(p) model
- we want to forecast r_{t+h} using all the info \mathcal{F}_t available at t
- assume we choose the forecast to minimize the **mean square error**:

$$E \left[(y - y_{prediction})^2 \right]$$

- The conditional mean minimizes the mean squared forecast error.
- we will come back to **optimal forecasting** later

1-step ahead forecast error

- the $AR(p)$ model is given by:

$$r_{t+1} = \phi_0 + \phi_1 r_t + \dots + \phi_p r_{t-p+1} + \varepsilon_{t+1}$$

- take the conditional expectation:

$$E_t [r_{t+1}] = \phi_0 + \phi_1 r_t + \dots + \phi_p r_{t-p+1}$$

- the one-step ahead forecast error:

$$v_t(1) = r_{t+1} - \phi_0 - \sum_{i=1}^p \phi_i r_{t-i+1} = \varepsilon_{t+1}$$

- the variance of the one-step ahead forecast error:

$$V[v_t(1)] = \sigma_\varepsilon^2$$

- if ε_t is normally distributed, then the 95 % confidence interval:

$$\pm 1.96 \sigma_\varepsilon$$

2-step ahead forecast error

- the $AR(p)$ model is given by:

$$r_{t+2} = \phi_0 + \phi_1 r_{t+1} + \dots + \phi_p r_{t-p+2} + \varepsilon_{t+2}$$

- we just take the conditional expectation:

$$E_t [r_{t+2}] = \phi_0 + \phi_1 \hat{r}_t(1) + \dots + \phi_p r_{t-p+2}$$

- the two-step ahead forecast error:

$$v_t(2) = \phi_1 v_t(1) + \varepsilon_{t+2} = \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}$$

- the variance of the two-step ahead forecast error:

$$V [v_t(2)] = \sigma_\varepsilon^2 (1 + \phi_1^2)$$

- the variance of the two-step ahead forecast error is larger than the variance of the one-step ahead forecast error

Multi-step ahead forecast error

Result:

The h -step ahead forecast is given by:

$$\hat{r}_t(h) = \phi_0 + \sum_{i=1}^p \phi_i \hat{r}_t(h-i)$$

where $\hat{r}_t(j) = r_{t+j}$ if $j < 0$.

- the h -step ahead forecast converges to the unconditional expectation $E(r_t)$ as $h \rightarrow \infty$
- this is referred to as **mean reversion**

Estimation: conditional least squares

- assume we observe or can condition on the first p observations.
- AR(p) model is then a linear regression model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \varepsilon_t, \quad t = p+1, \dots, T$$

- using least squares, the fitted model is

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \dots + \hat{\phi}_p r_{t-p}$$

and the residual is $v_t = r_t - \hat{r}_t$

- the estimated variance of the residuals is:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=p+1}^T v_t^2}{T - 2p - 1}$$

ML Estimation

- alternatively, we could use maximum likelihood.
- the log-likelihood function is:

$$\ln p(r_1, r_2, \dots, r_T; \theta) = \sum_{t=2}^T \ln p(r_t | r_{t-1}, \dots, r_1; \theta) + \ln p(r_1; \theta)$$

- for example, assume Gaussian shocks ε_t then $p(r_t | r_{t-1}, \dots, r_{t-p}; \theta)$ is normal
- the difference between least squares and ML estimation of $(\phi_0, \phi_1, \dots, \phi_p)$ are the initial distributions $p(r_1; \theta), p(r_2 | r_1; \theta) \dots$
- Conditional least squares of an AR(p) drops the first p terms in the likelihood.

Example: ML Estimation of AR(1)

- assume the initial value r_1 comes from the stationary dist.
- unconditional moments:

$$E[r_1] = \frac{\phi_0}{1 - \phi_1}, \quad V[r_1] = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2},$$

- hence, the density $p(r_1; \theta)$ of the first observation r_1 is normal with the above (unconditional) mean and variance
- for $t > 1$, the conditional moments:

$$E[r_t | r_{t-1}] = \phi_0 + \phi_1 r_{t-1}, \quad V[r_t | r_{t-1}] = \sigma_\varepsilon^2$$

- hence, the conditional density $p(r_t | r_{t-1}; \theta)$ is normal with the above (conditional) mean and (conditional) variance

Example: ML Estimation of AR(1)

- the log-likelihood function is:

$$\begin{aligned}\ln p(r_1, r_2, \dots, r_T; \theta) &= \sum_{t=2}^T \ln p(r_t | r_{t-1}, \dots, r_1; \theta) + \ln p(r_1; \theta) \\ &= -\frac{1}{2} \sum_{t=2}^T \left(\ln(2\pi) + \ln(\sigma_\varepsilon^2) + \frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\sigma_\varepsilon^2} \right) \\ &\quad + \ln p(r_1; \theta)\end{aligned}$$

- choose parameters $\theta = (\phi_0, \phi_1, \sigma_\varepsilon^2)$ to maximize the log-likelihood function
 $p(r_1; \theta)$ is typically chosen to be the stationary distribution

$$r_1 \sim N\left(\frac{\phi_0}{1 - \phi_1}, \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}\right)$$

Exact vs. Conditional ML

- the conditional ML estimator drops the initial condition
- exact log-likelihood function:

$$\ln p(r_1, r_2, \dots, r_T; \theta) = \sum_{t=2}^T \ln p(r_t | r_{t-1}, \dots, r_1; \theta) + \ln p(r_1 | \theta)$$

- conditional log-likelihood function:

$$\ln p(r_{p+1}, \dots, r_T; \theta) = \sum_{t=p+1}^T \ln p(r_t | r_{t-1}, \dots, r_1; \theta)$$

- the conditional log-likelihood ‘conditions’ on the first data point and drops the first p terms.
- Conditional ML is the same as least squares. The solution can be calculated analytically.

Summary: AR(p) models

- *dynamic model*, e.g. AR(p):

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \varepsilon_t$$

- ▶ constant determines mean through: $\mu = \frac{\phi_0}{1-\phi_1-\phi_2-\dots-\phi_p}$
 - ▶ coefficients $(\phi_1, \phi_2, \dots, \phi_p)$ must satisfy stationarity restrictions for a well-specified model:
 - ▶ objective: parsimonious model of dynamics of r_t
-
- For AR(p) models, you can maximize the conditional MLE in closed-form...conditional least squares....but there is no guarantee that it will satisfy the stationarity restrictions.
 - Calculating the full MLE requires numerical optimization.

Application: Bond Pricing

Bond Notation

- an n -period zero coupon bond pays one dollar n periods from now
- notation:
 - ▶ $P_t^{(n)}$ denotes the price of an n -period zero-coupon bond.
 - ▶ $p_t^{(n)} = \log(P_t^{(n)})$ denotes the log price
 - ▶ the yield of an n -period zero-coupon bond is:

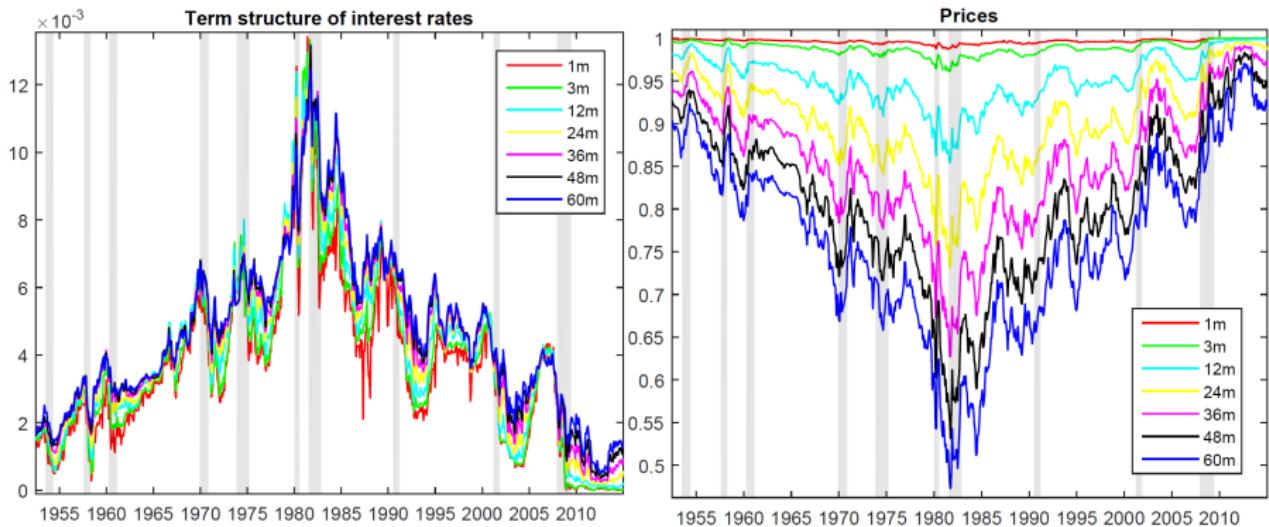
$$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}$$

- ▶ the holding period return is:

$$hpr_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}$$

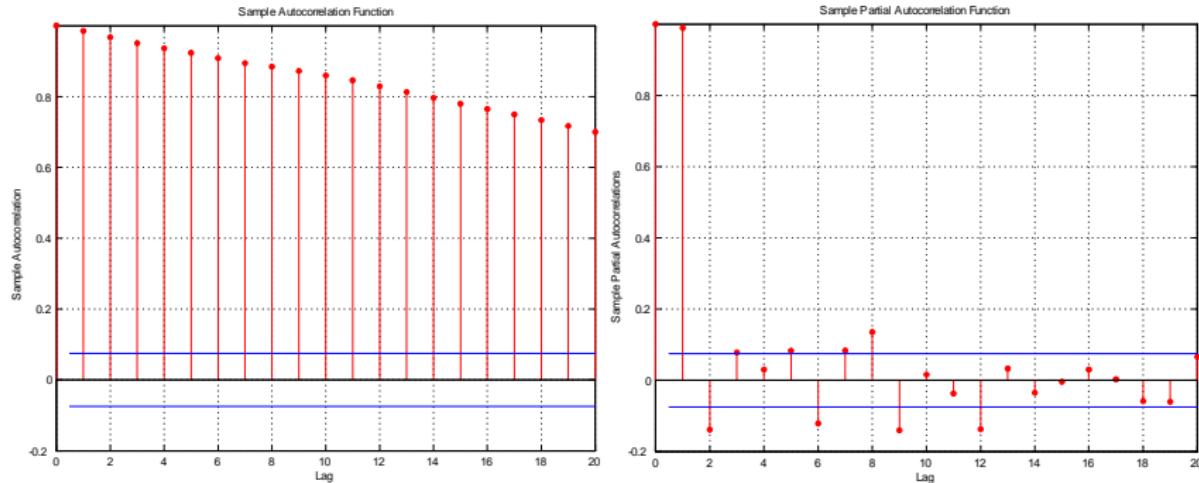
- ▶ the short term interest rate $y_t^{(1)}$ is given special notation r_t

Term Structure of Interest Rates



CRSP Fama-Bliss Zero-Coupon Bond Data. Sample: 1952.6-2014.12. Monthly data. Left: yields. Right: prices

ACF and PACF of 1 month yield



CRSP Fama-Bliss Zero-Coupon Bond Data. Sample: 1952.6-2014.12. Monthly data. Yield on one month zero-coupon bond $y_t^{(1)}$.

- ACF is persistent. PACF drops off after 1 month.

Bond Pricing: Expectations Hypothesis

- discrete time models of bond pricing
- examine the simplest possible model: a single factor model, investors risk-neutral with respect to interest rate risks
- the single factor g_t follows an AR(1):

$$g_{t+1} = (1 - \phi)\mu + \phi g_t + \sigma \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, 1)$$

- In "exponential-affine" term structure models the log risk-free rate is linear in factors

$$r_t = \delta_0 + \delta'_1 g_t$$

- Since only one factor in our example, might as well let log short rate be the factor $r_t = y_t^{(1)} = g_t$ (ie, $\delta_0 = 0$ and $\delta_1 = 1$)

Pricing of zero-coupon default-free bonds

- A zero-coupon bond only pays off principal (normalize this to \$1), no coupons.
- If investors are risk-neutral, they always discount with the risk-free rate.
- so, the price of a 1-period zero-coupon bond would be

$$P_t^{(1)} = e^{-r_t} \times \$1$$

- the price of a 2-period zero-coupon bond would be:

$$P_t^{(2)} = E_t[e^{-r_t} P_{t+1}^{(1)}] = E_t[e^{-r_t} e^{-r_{t+1}}]$$

- Price of n -period bond

$$P_t^{(n)} = E_t[\exp\left(-\sum_{j=0}^{n-1} r_{t+j}\right)]$$

Pricing (cont'd)

- We need to take expectations of AR(1) variable in exponential

- Use the fact:

$$E[e^x] = e^{\mu + \frac{1}{2}\sigma^2}$$

if $x \sim N(\mu, \sigma^2)$

- Consider 2-period bond

$$\begin{aligned} P_t^{(2)} &= e^{-r_t} E_t [e^{-r_{t+1}}] \\ &= e^{-r_t} E_t \left[e^{-((1-\phi)\mu + \phi r_t + \sigma \varepsilon_{t+1})} \right] \\ &= e^{-r_t} e^{-(1-\phi)\mu - \phi r_t} E_t [e^{-\sigma \varepsilon_{t+1}}] \\ &= e^{-2\mu - (1+\phi)(r_t - \mu)} E_t [e^{-\sigma \varepsilon_{t+1}}] \\ &= e^{-2\mu - (1+\phi)(r_t - \mu) + \frac{1}{2}\sigma^2}. \end{aligned}$$

Price of an n -period bond

The Price of an n -period zero-coupon bond is:

$$P_t^{(n)} = E_t \left[e^{-r_t} P_{t+1}^{(n-1)} \right]$$

Solving recursively, we get:

$$P_t^{(n)} = \exp(a_n + b_n r_t)$$

where

$$\begin{aligned} a_n &= a_{n-1} + b_{n-1} (1 - \phi) \mu + \frac{1}{2} \sigma^2 b_{n-1}^2, \\ b_n &= b_{n-1} \phi - 1, \end{aligned}$$

with initial conditions $a_1 = 0$ and $b_1 = -1$.

- This implies yields are:

$$y_t^{(n)} = \tilde{a}_n + \tilde{b}_n g_t \quad \tilde{a}_n = -\frac{1}{n} a_n \quad \tilde{b}_n = -\frac{1}{n} b_n$$

- a_n and b_n are difference equations fit through the *cross-section* of yields.

Estimation

- monthly U.S. zero coupon Fama-Bliss data from CRSP.
- 1,3,12,24,36,48,60 month yields
- Model estimated by maximum likelihood (add noise to yield observations)
- short rate $r_t = g_t$ is the one-month yield $y_t^{(1)}$

Parameter	Model	Sample moment	Model-implied
μ	$mean(y_t^{(1)})$	0.00355	0.00362
$\frac{\sigma^2}{1-\phi^2}$	$var(y_t^{(1)})$	5.976e-06	8.368e-06
ϕ	$\rho_{12}(y_t^{(1)})$	0.9756	0.9895

- All yields perfectly conditionally correlated due to 1-factor structure
- High current interest rates expected to revert to mean so long-term rates lower than short-term rates.
- Average slope of yield curve is slightly below zero (since investors' risk-neutral and there is a convexity term), contrary to data

Multiple Factors

- the Vasicek (1977) adds "risk prices" to the this bare-bones model, but is still a simple, one-factor model
- this model cannot capture the slope or curvature of yields, only the level of interest rates.
- we need a richer model with more factors, where we let g_t be a vector of factors \Rightarrow vector autoregressive process
 - ▶ PCA of yields indicated we need three factors to explain yields.
- Fixed Income class will discuss more on this

Moving Average Models

AR(infinity)

- in theory the true data generating process could be an AR(∞):

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \varepsilon_t$$

- implementation:
 - infinite** number of parameters
- solution: constrain parameters

$$x_t = \phi_0 - \theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} - \dots + \varepsilon_t$$

where $\phi_i = -\theta_1^i, i \geq 1$

AR(infinity) to MA(1)

- solution: constrain parameters

$$x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \theta_1^3 x_{t-3} + \dots = \phi_0 + \varepsilon_t$$

- this can be restated as an MA(1) model:

$$x_t = \phi_0 (1 - \theta_1) + (1 - \theta_1 B) \varepsilon_t$$

- ▶ MA(1) is a 'cheap' version of an AR(∞).
- general form of MA(1) model is:

$$x_t = \mu + (1 - \theta_1 B) \varepsilon_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

MA(q)

Definition

A **moving average process of order q** or MA(q) model is:

$$x_t = \mu + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t,$$

where $q > 0$

Stationarity

- consider the MA(1) model:

$$x_t = \mu + (1 - \theta_1 B) \varepsilon_t.$$

- compute the variance of an MA(1) model:

$$V[\mu + (1 - \theta_1 B) \varepsilon_t] = (1 + \theta_1^2) \sigma_\varepsilon^2.$$

- compute the variance of an MA(q) model:

$$V[\mu + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_\varepsilon^2.$$

Computing Autocovariances for MA(1)

- assume the unconditional mean $\mu = 0$
- pre-multiply the MA(1) model by r_{t-j} :

$$r_{t-j} r_t = r_{t-j} \varepsilon_t - \theta_1 r_{t-j} \varepsilon_{t-1}$$

- take expectations
- compute the auto-covariance of an MA(1) model:

$$\gamma_1 = -\theta_1 \sigma_\varepsilon^2, \quad \gamma_j = 0, \quad j > 1$$

- this implies the autocorrelations are:

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}, \quad \rho_j = 0, \quad j > 1$$

► **the ACF is cut off after 1 lag!**

Computing Autocovariances for MA(2)

- the same argument implies that the autocorrelations of an MA(2) are:

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \quad \rho_j = 0, \ j > 2$$

- ▶ the ACF is cut off after 2

Forecasting with MA(1)

- consider an MA(1) model:

$$r_{t+1} = \mu + \varepsilon_{t+1} - \theta_1 \varepsilon_t$$

- take conditional expectations:

$$\hat{r}_t(1) = E_t [r_{t+1}] = \mu - \theta_1 \varepsilon_t$$

$$\hat{r}_t(2) = E_t [r_{t+2}] = \mu$$

- the one-step ahead forecast error is given by:

$$v_t(1) = r_{t+1} - \hat{r}_t(1) = \varepsilon_{t+1}$$

- ▶ the variance of the one-step ahead forecast error is σ_ε^2

Forecasting with MA(1)

- consider an MA(1) model:

$$r_{t+1} = \mu + \varepsilon_{t+1} - \theta_1 \varepsilon_t$$

- take conditional expectations:

$$\hat{r}_t(1) = E_t [r_{t+1}] = \mu - \theta_1 \varepsilon_t$$

$$\hat{r}_t(2) = E_t [r_{t+2}] = \mu$$

- the two-step ahead forecast error is given by:

$$v_t(2) = r_{t+2} - \hat{r}_t(2) = \varepsilon_{t+2} - \theta_1 \varepsilon_{t+1}$$

- ▶ the variance of the two-step ahead forecast error is $(1 + \theta_1^2)\sigma_\varepsilon^2$
- ▶ this is the unconditional variance

Forecasting with MA(2)

- consider an MA(2) model:

$$r_{t+1} = \mu + \varepsilon_{t+1} - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1}$$

- take conditional expectations:

$$\hat{r}_t(1) = E_t [r_{t+1}] = \mu - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1}$$

$$\hat{r}_t(2) = E_t [r_{t+2}] = \mu - \theta_2 \varepsilon_t$$

$$\hat{r}_t(3) = E_t [r_{t+3}] = \mu$$

- the one-step ahead forecast error is given by:

$$v_t(1) = r_{t+1} - \hat{r}_t(1) = \varepsilon_{t+1}$$

- ▶ the variance of the one-step ahead forecast error is σ_ε^2

Forecasting with MA(2)

- consider an MA(2) model:

$$r_{t+1} = \mu + \varepsilon_{t+1} - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1}$$

- take conditional expectations:

$$\hat{r}_t(1) = E_t [r_{t+1}] = \mu - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1}$$

$$\hat{r}_t(2) = E_t [r_{t+2}] = \mu - \theta_2 \varepsilon_t$$

$$\hat{r}_t(3) = E_t [r_{t+3}] = \mu$$

- the two-step ahead forecast error is given by:

$$v_t(2) = r_{t+2} - \hat{r}_t(2) = \varepsilon_{t+2} - \theta_1 \varepsilon_{t+1}$$

- ▶ the variance of the two-step ahead forecast error is $(1 + \theta_1^2)\sigma_\varepsilon^2$
- ▶ this is smaller than the unconditional variance

Forecasting with MA(2)

- consider an MA(2) model:

$$r_{t+1} = \mu + \varepsilon_{t+1} - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1}$$

- take conditional expectations:

$$\hat{r}_t(1) = E_t [r_{t+1}] = \mu - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1}$$

$$\hat{r}_t(2) = E_t [r_{t+2}] = \mu - \theta_2 \varepsilon_t$$

$$\hat{r}_t(3) = E_t [r_{t+3}] = \mu$$

- the three-step ahead forecast error is given by:

$$v_t(3) = r_{t+3} - \hat{r}_t(3) = \varepsilon_{t+3} - \theta_1 \varepsilon_{t+2} - \theta_2 \varepsilon_{t+1}$$

- ▶ the variance of the three-step ahead forecast error is $(1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2$
- ▶ this is the unconditional variance

Maximum Likelihood

- MA(q) models can't be estimated using (conditional) least squares because the parameters are a non-linear function of the data
- MA(q) models are commonly estimated using Maximum Likelihood
- this involves assuming a parametric distribution for the shocks ε_t .
- Often, we assume ε_t are normally distributed.

ML Estimation of MA(1)

- conditional moments:

$$\begin{aligned}V[r_t|r_{t-1}] &= \sigma_\varepsilon^2, \\E[r_t|r_{t-1}] &= \mu - \theta\varepsilon_{t-1}\end{aligned}$$

- hence, the density $p(r_t|\mathcal{F}_{t-1};\theta)$ of the first observation is normal with the above (conditional) mean and variance
- suppose we assume that $\varepsilon_0 = 0$.
- then $\varepsilon_1 = r_1 - \mu$
- then $\varepsilon_2 = r_2 - \mu - \theta_1\varepsilon_1$
- we can recursively calculate the sequence $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_t\}$

ML Estimation of MA(1)

- hence, the log-likelihood function is:

$$\begin{aligned}\ln p(r_1, r_2, \dots, r_T; \theta) &= \sum_{t=2}^T \ln p(r_t | r_{t-1}, \dots, r_1; \theta) + \ln p(r_1; \theta) \\ &= -\frac{1}{2} \sum_{t=1}^T \left(\ln(2\pi) + \ln(\sigma_\varepsilon^2) + \frac{(-\varepsilon_t)^2}{\sigma_\varepsilon^2} \right) \\ &\quad + \ln p(r_1; \theta)\end{aligned}$$

- choose parameters $\theta = (\mu, \theta_1, \sigma_\varepsilon^2)$ to maximize the log-likelihood function

ACF and PACF

- *ACF* is useful for determining MA lag length:
 - ▶ autocorrelations are cut off at q for an $MA(q)$: $ACF(k) = 0$ for $k > q$
- *PACF* is useful for determining AR lag length
 - ▶ partial autocorrelations are cut off at p for an $AR(p)$: $PACF(k) = 0$ for $k > p$

ARMA Models

ARMA(p,q)

- certain processes can only be described by AR or MA models if we include lots of lags
 - ▶ unappealing (need to estimate lots of parameters)
- natural solution: ARMA(p, q) processes

ARMA(p,q)

- consider an ARMA(1, 1) model:

$$r_t - \phi_1 r_{t-1} = \phi_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

with $\theta_1 \neq \phi_1$

- the unconditional mean of an ARMA(1, 1) has the same expression as an AR(1)

$$E[r_t] = \frac{\phi_0}{1 - \phi_1}$$

- we can re-write the process as:

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

- take expectations of $[r_t - \mu]^2$ to compute the variance:

$$V[r_t] = \phi_1^2 V[r_{t-1}] + \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 - 2\phi_1\theta_1 E[\varepsilon_{t-1}(r_{t-1} - \mu)]$$

ARMA(1,1)

- this reduces to:

$$V[r_t] = \phi_1^2 V[r_t] + \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 - 2\phi_1\theta_1\sigma_\varepsilon^2$$

- collecting terms, we get:

$$V[r_t] = \sigma_\varepsilon^2 \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}$$

- obviously, we need $\phi_1^2 < 1$

► same stationarity requirement as for AR(1)

ACF of ARMA(1,1)

- to compute the auto-covariances:

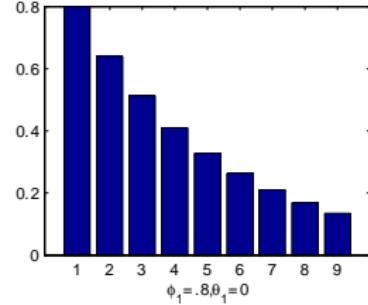
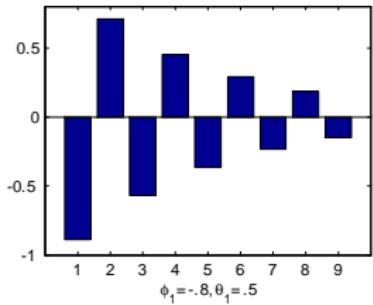
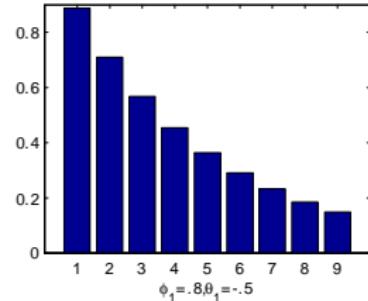
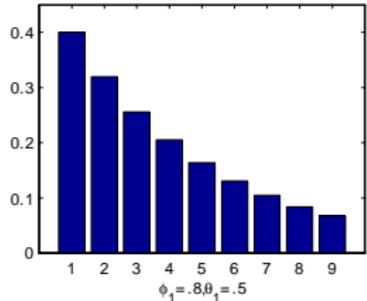
$$\begin{aligned} E[(r_t - \mu)(r_{t-j} - \mu)] &= \phi_1 E[(r_{t-1} - \mu)(r_{t-j} - \mu)] \\ &\quad + E[\varepsilon_t(r_{t-j} - \mu)] \\ &\quad - \theta_1 E[\varepsilon_{t-1}(r_{t-j} - \mu)] \end{aligned}$$

- for $j = 1$, we get: $\gamma_1 = \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2$
- this implies that the ACF is given by:

$$\rho_1 = \phi_1 - \theta_1 \frac{\sigma_\varepsilon^2}{\gamma_0}$$

$$\rho_j = \phi_1 \rho_{j-1}, \quad j > 1$$

Autocorrelation for ARMA(1,1)



Autocorrelation Function for ARMA(1,1) processes.

PACF of ARMA(1,1)

- PACF does not die out at some lag
- slow decay (as is the case for MA models)

ARMA(p,q)

- consider an ARMA(p, q) model:

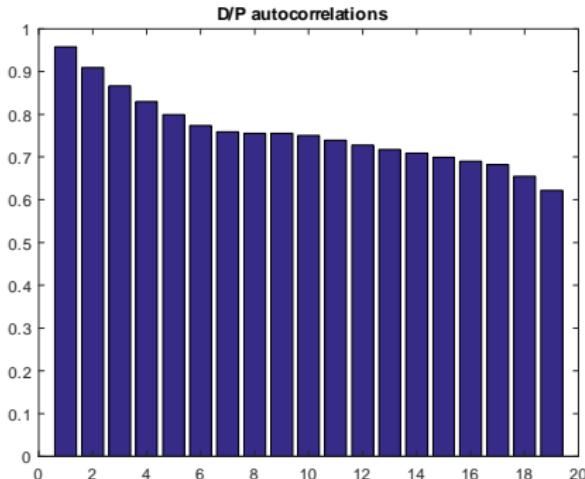
$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \quad \varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$$

- using the backshift operator

$$(1 - \phi_1 B - \dots - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

D/P autocorrelation function

- Revisiting the D/P ratio
 - ▶ Sample autocorrelation function:



- 'Drop off' for about first 4 lags, then stable...
 - ▶ Indicates a 4 lags of MA might be a good representation + 1 lag AR

D/P as ARMA(1,4)

```
ARIMA(1,0,4) Model:  
-----  
Conditional Probability Distribution: Gaussian  
  
Parameter      Value       Standard Error      t Statistic  
-----  
Constant      0.000484308   0.000550722   0.879406  
AR{1}          0.980784     0.0158269    61.9696  
MA{1}          0.103568     0.0573102    1.80715  
MA{2}          -0.172191    0.0611522   -2.81577  
MA{3}          -0.148333    0.0632631   -2.34469  
MA{4}          -0.106098    0.0582711   -1.82076  
Variance      7.50796e-06   1.64215e-07   45.7203
```

- Forecast by getting sample series of residuals, then plug in as needed for forecasts

$$\mu = \frac{0.000484}{1 - 0.9807}$$

$$\begin{aligned}E_t [DP_{t+1}] &= \mu + 0.98(DP_t - \mu) + 0.10\varepsilon_t - 0.17\varepsilon_{t-1} - 0.14\varepsilon_{t-2} - 0.11\varepsilon_{t-3}, \\E_t [DP_{t+2}] &= E_t [E_{t+1} [DP_{t+2}]] \\&= E_t [\mu + 0.98(DP_{t+1} - \mu) + 0.10\varepsilon_{t+1} - 0.17\varepsilon_t - 0.14\varepsilon_{t-1} - 0.11\varepsilon_{t-2}] \\&= \mu + 0.98(E_t [DP_{t+1}] - \mu) - 0.17\varepsilon_t - 0.14\varepsilon_{t-1} - 0.11\varepsilon_{t-2}\end{aligned}$$

etc.

MA representation

- start from this expression:

$$(1 - \phi_1 B - \dots - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

- re-arranging this expression delivers an *MA* representation:

$$r_t = \frac{\phi_0}{(1 - \phi_1 B - \dots - \phi_p B^p)} + \frac{(1 - \theta_1 B - \dots - \theta_q B^q)}{(1 - \phi_1 B - \dots - \phi_p B^p)} \varepsilon_t$$

- more succinctly:

$$r_t = \mu + \psi(B) \varepsilon_t$$

- stationarity: the solutions of $(1 - \phi_1 x - \dots - \phi_p x^p) = 0$ should lie outside of the unit circle

Impulse-Response Function

- consider the MA representation:

$$r_t = \mu + \psi(B)\varepsilon_t$$

- this can be written out as:

$$r_t = \mu + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \dots$$

where $\{\psi_i\}$ is the *impulse response* function of the ARMA model.

- the coefficients $\{\psi_i\}$ are functions of the parameters $\{\phi_i\}$ and $\{\theta_i\}$
- the impulse response function shows the effect today of a shock k periods ago:

$$\frac{\partial r_t}{\partial \varepsilon_{t-k}} = \psi_k$$

Forecasting

- consider the MA representation:

$$r_t = \mu + \psi(B)\varepsilon_t$$

- the h -period ahead forecast :

$$\hat{r}_t(h) = \mu + \psi_h\varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + \dots$$

- the h -period ahead forecast error can be stated as:

$$v_t(h) = \varepsilon_{t+h} + \psi_1\varepsilon_{t+h-1} + \psi_2\varepsilon_{t+h-2} + \dots + \psi_{h-1}\varepsilon_{t+1}$$

- the variance of the h -step ahead forecast error is:

$$V[v_t(h)] = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \sigma_\varepsilon^2$$

Variance of Forecast Error

- the variance of the h -step ahead forecast error is:

$$V[v_t(h)] = \left(1 + \psi_1^2 + \psi_2^2 + \dots + \psi_2^h\right) \sigma_\varepsilon^2$$

- non-decreasing function of forecast horizon
- variance of forecast error converges to variance of process

$$V[v_t(h)] \rightarrow V[r_t]$$

as $h \rightarrow \infty$

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov and F. Csaki (Eds.), 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, pp. 267-281. Budapest: Akademiai Kiado.
- Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50, 745-787.
- Box, G. E. P. and G. Jenkins (1970). *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- Ljung, G. M. and G. E. P. Box (1978). On a measure of a lack of fit in time series models. *Biometrika* 65(2), 297-303.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6(2), 461-464.
- Vasicek, O. A. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5(2), 177-188.