Lecture 9-a
Time Series:
Identification of AR, MA & ARMA Models

Brooks (4th edition): Chapter 6

 1

Review: Times Series

- A time series y_t is a process observed in sequence over time, t = 1, ..., T $\Rightarrow$ $Y_t = \{y_1, y_2, y_3, ..., y_T\}$.

- Main feature of time series: *dependence*.

- Popular models for $E[y_t | I_{t-1}]$:

– AR process: $E_t[y_t | I_{t-1}] = f(y_{t-1}, y_{t-2}, y_{t-3}, ...\}$

Example: AR(1) process, $y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t$.

– MA process: $E_t[y_t | I_{t-1}] = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, ...)$

Example: MA(1) process, $y_t = \mu + \theta_1\, \varepsilon_{t-1} + \varepsilon_t$

– ARMA process: $E_t[y_t | I_{t-1}] = f(y_{t-1}, y_{t-2}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...)$

Review: Times Series – Forecasting

• We want to select an appropriate time series model to forecast y_t. The linear models we consider: AR(p), MA(q) or ARMA(p, q).

• Steps for forecasting:

(1) Identify the appropriate model. That is, determine AR, MA or ARMA and the order of the model -i.e., p, q.

Tools: ACF, PACF, Information Criteria

(2) Estimate the model.

OLS, Method of Moments (complicated).

(3) Test the model.

Make sure errors are WN.

(4) Forecast.

Review: Moving Average Process

• A linear MA(q) model:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t,$$

where

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \ldots + \theta_q L^q$$

• Check stationarity (Constant moments)

• **Mean**

$$E[y_t] = E[\varepsilon_t] + \theta_1 E[\varepsilon_{t-1}] + \theta_2 E[\varepsilon_{t-2}] + \ldots + \theta_q E[\varepsilon_{t-q}] = 0$$

• **Variance**

$$\text{Var}[y_t] = \text{Var}[\varepsilon_t] + \theta_1^2 \text{Var}[\varepsilon_{t-1}] + \theta_2^2 \text{Var}[\varepsilon_{t-2}] + \ldots + \theta_q^2 \text{Var}[\varepsilon_{t-q}]$$
$$= (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\, \sigma^2.$$

• **Covariance**

$$\gamma(q) = \sigma^2 \textstyle\sum_{j=q}^{q} \theta_j \theta_{j-q} \qquad \text{(where } \theta_0 = 1\text{)}$$

$\Rightarrow$ MA(q) is always stationary.

Review: Moving Average Process – Stationarity

$\gamma(q) = \sigma^2 \sum_{j=q}^{q} \theta_j \theta_{j-q}$ (where $\theta_0 = 1$)

In general, for the k *autocovariance:*

$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \theta_{j-k}$ for $|k| \leq q$

$\gamma(k) = 0$ for $|k| > q$

Remark: After lag q, the autocovariances (and autocorrelation functions) are 0.

• It can be shown that for ε_t with the same distribution (say, normal) the autocovariances are non-unique. In this case, we select the MA(q) model that's invertible.

Technical note: An invertible MA(q) is typically required to have roots of the lag polynomial equation $\theta(z) = 0$ greater than one in absolute value ("*outside the unit circle*"). In the MA(1) case, we require $|\theta_1| < 1$.

Review: MA(1) Process – ACF

Example: MA(1) process:

• $\gamma(k)$

$k = 0$ $\gamma(0) = \sigma^2 \sum_{j=0}^{1} \theta_j \theta_{j-0} = \sigma^2 (1 + \theta_1^2)$

$k = 1$ $\gamma(1) = \sigma^2 \sum_{j=1}^{1} \theta_j \theta_{j-1} = \sigma^2 (\theta_1)$

$k > 1$ $\gamma(k) = 0$

$\Rightarrow$ After lag $q = 1$, the autocovariances are 0.

To get the ACF, we divide $\gamma(k)$ by $\gamma(0)$. Then:

$\rho(0) = \gamma(0)/\gamma(0) = 1$

$\rho(1) = \gamma(1)/\gamma(0) = \theta_1\sigma^2 / \sigma^2 (1+ \theta_1^2) = \theta_1/(1+ \theta_1^2)$

$\vdots$

$\rho(k) = \gamma(k)/\gamma(0) = 0$ (for $k > 1$)

Review: MA(1) Process – ACF

Example (continuation):

$$\rho(1) = \theta_1/(1+\theta_1^2)$$

Note that $|\rho(1)| \leq 0.5$.

When $\theta_1 = 0.5 \Rightarrow \rho(1) = 0.4$.
$\theta_1 = -0.9 \Rightarrow \rho(1) = -0.497238$.
$\theta_1 = 2 \Rightarrow \rho(1) = 0.4$. (same $\rho(1)$ for θ_1 & $1/\theta_1$)

If we use the ACF to select a model, we select the *invertible* process with θ_1 = **0.5**.

Review: MA Process – Estimation

• MA processes are more complicated to estimate since we do not observe the errors, ε_t's: Direct estimation is impossible.

• Two indirect ways:
(1) Using method of moments (MM): We match observed moments and solved for the parameters. For example, for an MA(1):

$$\rho(1) = \theta_1/(1+\theta_1^2)$$

$$r_1 = \frac{\hat{\theta}}{(1+\hat{\theta}^2)} \quad \Rightarrow \quad \hat{\theta} = \frac{1 \pm \sqrt{1-4r_1^2}}{2r_1}$$

• A nonlinear solution and difficult to solve.

(2) Using AR(∞) representation: For MA(1) & $|\theta|<1$, find $a \in (-1; 1)$

$$\varepsilon_t(a) = y_t + a\, y_{t-1} + a^2\, y_{t-2} + a^3\, y_{t-3} + \ldots.$$

and look (numerically) for the least-square estimator

$$\hat{\boldsymbol{\theta}} = \arg\min_\theta \{S(\boldsymbol{y}; \boldsymbol{\theta}) = \textstyle\sum_{i=1}^{T} \varepsilon_i(a)^2\} \qquad (a^i = \theta_1{}^i.)$$

Review: Autoregressive (AR) Process

- An AR(p) process is given by:

$$y_t = \mu + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \ldots + \phi_p\, y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

Using the lag operator we write the AR(p) process: $\phi(L)\, y_t = \varepsilon_t$

with $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$

- Stability of AR(p):

We need the roots of $\phi(z) = 0$ to be outside the unit circle.

For the AR(1) process

$$\phi(z) = 1 - \phi_1 z = 0 \qquad \Rightarrow |z| = \frac{1}{|\phi_1|} > 1$$

That is, the AR(1) process is stable if the root of $\phi(z)$ is greater than one ("*the roots lie outside the unit circle*").

Review: AR(1) Process – Stationarity & ACF

- An AR(1) model:

$$y_t = \phi_1\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

Under the stationarity condition $|\phi_1| < 1$, we derived the moments:

$E[y_t] = \mu = 0$ (assuming $\phi_1 \neq 1$)

$Var[y_t] = \gamma(0) = \sigma^2/(1 - \phi_1^2)$ (assuming $|\phi_1| < 1$)

$\gamma(k) = \phi_1^k\, \gamma(0)$

- ACF: $\rho(k) = \dfrac{\gamma(k)}{\gamma(0)} = \phi_1^k$ (ACF decays with k.)

Patterns:
– when $0 < \phi_1 < 1$ $\Rightarrow$ All autocorrelations are positive.
– when $-1 < \phi_1 < 0$ $\Rightarrow$ The sign of $\rho(k)$ shows an alternating sign pattern beginning a negative value.

Review: AR Process – Stationarity & Ergodicity

Theorem: The linear AR(p) process is strictly stationary and ergodic if and only if the roots of $\phi(z)$ are $|z_j|>1$ for all j, where $|z_j|$ is the modulus of the complex number r_j.

Note: If one of the z_j's equals 1, $\phi(L)$ (& y_t) has a *unit root* –i.e., $\phi(1)=0$. This is a special case of *non-stationarity*.

• Inverting $\phi(L)$ produces a process with an infinite sum of ε_{t-j}'s. If this sum does not explode, we say the process is *stable*.

• AR(p) model: $\phi(L)y_t = \mu + \varepsilon_t$,
where

$$\phi(L) = 1 - \phi_1 L^1 - L^2\phi_2 - \ldots - \phi_p L^p$$

Then, $y_t = \phi(L)^{-1}(\mu + \varepsilon_t)$, $\Rightarrow$ an MA(∞) process!

Review: AR Process – Estimation & Properties

• Back to the general AR(p). Define

$$\boldsymbol{x}_t = (1\ y_{t-1}\ y_{t-2} \ldots y_{t-p})$$
$$\boldsymbol{\beta} = (\mu\ \phi_1\ \phi_2\ \ldots\ \phi_p)$$

Then the model can be written as

$$y_t = \boldsymbol{x}_t{}'\boldsymbol{\beta} + \varepsilon_t$$

• The OLS estimator is $\mathbf{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

• Properties:
  – Using the Ergodic Theorem, OLS estimator is consistent.
  – Using the MDS CLT, OLS estimator is asymptotically normal.
    $\Rightarrow$ asymptotic inference is the same.

• The asymptotic covariance matrix is estimated just as in the cross-section case: The sandwich estimator.

ARMA Process

• A combination of AR(p) and MA(q) processes produces an ARMA(p, q) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

$$= \mu + \sum_{i=1}^{p} \phi_i\, y_{t-i} - \sum_{i=1}^{q} \theta_i L^i \varepsilon_t + \varepsilon_t$$

$$\Rightarrow \phi(L) y_t = \mu + \theta(L)\varepsilon_t$$

• Usually, we insist that $\phi(L) \neq 0$, $\theta(L) \neq 0$ & that the polynomials $\phi(L)$, $\theta(L)$ have no *common factors*. This implies it is not a lower order ARMA model.

ARMA(1,1) – Stationarity & ACF

• For an ARMA(1,1) we have:.

$$y_t = \mu + \phi_1\, y_{t-1} + \theta_1\, \varepsilon_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

• **Moments:** ($\mu = 0$)

$$E[y_t] = \mu \,/\, (1 - \phi_1) = 0 \qquad \text{(assuming } \phi_1 \neq 1)$$

$$Var[y_t] = \sigma^2\, (1 + \theta_1^2) \,/\, (1 - \phi_1^2) \quad \text{(assuming } |\phi_1| < 1)$$

• **Autocovariance function** ($\mu = 0$)

$$\gamma(k) = Cov[y_t,\ y_{t-k}]$$
$$= E[\{\phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t\}\, y_{t-k}]$$
$$= \phi_1 E[y_{t-1}\, y_{t-k}] + \theta_1 E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}]$$
$$= \phi_1\, \gamma(k-1) + \theta_1 E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}]$$

We have a **recursive formula**.

ARMA(1,1) – Stationarity & ACF

- ARMA(1,1): $y_t = \phi_1\, y_{t-1} + \theta_1\, \varepsilon_{t-1} + \varepsilon_t,$

Recursive formula:

$\gamma(k) = \phi_1 \gamma(k-1) + E[\varepsilon_t\, y_{t-k}] + \theta_1 E[\varepsilon_{t-1}\, y_{t-k}]$

$\gamma(0) = \phi_1 \gamma(1) + \sigma^2 + \theta_1(\phi_1 \sigma^2 + \theta_1 \sigma^2)$

$\gamma(1) = \phi_1\, \gamma(0) + \theta_1 \sigma^2$

Two equations for $\gamma(0)$ and $\gamma(1)$. Solving for $\gamma(0)$:

$\gamma(0) = \sigma^2 \dfrac{1 + \theta_1^2 + 2\phi_1\, \theta_1}{1 - \phi_1{}^2}$

$\gamma(1) = \phi_1\, \sigma^2 \dfrac{1 + \theta_1^2 + 2\phi_1\, \theta_1}{1 - \phi_1^2} + \theta_1 \sigma^2 = \sigma^2 \dfrac{(1 + \phi_1\, \theta_1) * (\phi_1 + \theta_1)}{1 - \phi_1^2}$

ARMA(1,1) – Stationarity & ACF

Continuing the process:

$$\begin{aligned}\gamma(2) &= E[y_t\, y_{t-2}] \\ &= E[\{\phi_1 y_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t\}\, y_{t-2}] \\ &= \phi_1 E[y_{t-1}\, y_{t-2}] + \theta_1 E[\varepsilon_{t-1}\, y_{t-2}] + E[\varepsilon_t\, y_{t-2}] \\ &= \phi_1\, \gamma(1)\end{aligned}$$

- In general:

$$\gamma(k) = \phi_1 \gamma(k-1) = \phi_1{}^{k-1} \gamma(1), \quad k > 1$$

$\Rightarrow$ If $|\phi_1|$<1, exponential decay and stationary.

<u>Note</u>: If stationary, ARMA(1,1) and AR(1) show exponential decay. Difficult to distinguish one from the other by looking at the autocovariance functions.

ARMA Process – Representation

- AR Representation: $\Pi(L)(y_t - \mu) = \varepsilon_t \Rightarrow \Pi(L) = \frac{\phi_p(L)}{\theta_q(L)}$

- Pure MA Representation: $(y_t - \mu) = \Psi(L)\varepsilon_t \Rightarrow \Psi(L) = \frac{\theta_q(L)}{\phi_p(L)}$

- Special ARMA(p, q) cases: – p = 0: MA(q)
  – q = 0: AR(p).

ARMA: Stationarity, Causality and Invertibility

Theorem: If ϕ(L) and θ(L) have no common factors, a (unique) *stationary* solution to $\phi(L)y_t = \theta(L)\varepsilon_t$ exists if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$$

This ARMA(p, q) model is *causal* –i.e., AR part can be inverted) if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$$

This ARMA(p, q) model is *invertible* if and only if

$$|z| \leq 1 \Rightarrow \theta(z) = 1 + \theta_1 z - \theta_2 z^2 + \ldots + \theta_p z^p \neq 0.$$

Note: Real data cannot be *exactly* modeled using a finite number of parameters. We choose p, q to create a good approximated model.

ARMA Process

- We defined the ARMA(p, q) model:

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

The mean does not affect the order of the ARMA. Then, if $\mu \neq 0$, we demean the data: $x_t = y_t - \mu$.

Then, $\phi(L)\ x_t = \theta(L)\ \varepsilon_t \quad \Rightarrow\ x_t$ is a *demeaned* ARMA process.

- For the rest of the lecture, we will study:
- Identification of p, q.
- Estimation of ARMA(p, q)

Autocovariance Function (Again)

- For an AR(p) process, WLOG with μ=0 (or demeaned y_t), we get a **recursive formula** to compute $\gamma(k = t - j)$:

$$\gamma(t-j) = \phi_1\gamma(j-1) + \phi_2\gamma(j-2) + ... + \phi_p\gamma(j-p)$$

- The autocovariances, $\gamma(t-j)$, determine a system of equations:

$$\begin{aligned}
\gamma(0) &= E[y_t, y_t] = \phi_1\gamma(1) + \phi_2\gamma(2) + \phi_3\gamma(3) + ... + \phi_p\gamma(p) + \sigma^2 \\
\gamma(1) &= E[y_t, y_{t-1}] = \phi_1\gamma(0) + \phi_2\gamma(1) + \phi_3\gamma(2) + ... + \phi_p\gamma(p-1) \\
\gamma(2) &= E[y_t, y_{t-2}] = \phi_1\gamma(1) + \phi_2\gamma(0) + \phi_3\gamma(1) + ... + \phi_p\gamma(p-2) \\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots
\end{aligned}$$

Using linear algebra, we can write the system as:

$$\mathbf{\Gamma}\, \boldsymbol{\phi} = \boldsymbol{\gamma}$$

where $\mathbf{\Gamma}$ is a pxp matrix of autocovariances, with $\gamma(0)$ on the diagonal; $\boldsymbol{\phi}$ is the px1 vector of AR(p) coefficients; and $\boldsymbol{\gamma}$ is the px1 vector of $\gamma(k)$ autocovariances.

Autocorrelation Function (ACF)

• Now, we define the autocorrelation function (ACF):

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\text{covariance at lag } k}{\text{variance}}$$

The ACF lies between -1 and +1, with $\rho(0) = 1$.

• Dividing the autocovariance system by γ(0), we get:

$$\begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(0) & \cdots & \rho(p-2) \\ \vdots & \vdots & \cdots & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix}$$

Or using linear algebra: $\mathbf{P}\, \boldsymbol{\phi} = \boldsymbol{\rho}$

• These are "*Yule-Walker*" equations, which can be solved numerically.

ACF – Estimation & Correlogram

• **Estimation**:
Easy: Use sample moments to estimate γ(*k*) and plug in formula:

$$r_k = \hat{\rho}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

Then, we plug the $\hat{\rho}_k$ in the Yule-Walker equations and solve for $\boldsymbol{\phi}$:

$$\widehat{\mathbf{P}}\, \boldsymbol{\phi} = \widehat{\boldsymbol{\rho}}$$

• The sample *correlogram* is the plot of the ACF against *k*. As the ACF lies between -1 and +1, the correlogram also lies between these values.

:

ACF – Distribution

• **Distribution**:
For a linear, stationary process, with large T, the distribution of the sample ACF, $r_k = \hat{\rho}_k$ is approximately normal with:

$$\mathbf{r} \xrightarrow{d} \text{N}(\boldsymbol{\rho}, \mathbf{V}/T), \qquad \mathbf{V} \text{ is the covariance matrix.}$$

Under H_0: $\rho_k = 0$ for all $k>1$.

$$\mathbf{r} \xrightarrow{d} \text{N}(\mathbf{0}, \mathbf{I}/T) \qquad \Rightarrow \text{Var}[\text{r}(k)] = 1/T.$$

• Under H_0, the SE[r] = $1/\sqrt{T}$ $\Rightarrow$ **95% C.I.: 0 $\pm$ 1.96 * $1/\sqrt{T}$**

Then, for a white noise sequence, approximately 95% of the sample ACFs should be within the above C.I. limits.

ACF – AR(1)

Example: Sample ACF for an AR(1) process:
Under stationarity:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k \qquad k = 0, 1, 2, \ldots$$

If $|\ \phi_1\ | < 1$, the ACF will show exponential decay.

• Suppose $\phi_1 = 0.4$. Then,

$\rho(0) = \mathbf{1}$
$\rho(1) = \mathbf{0.4}$
$\rho(2) = 0.4^2 = \mathbf{0.16}$
$\rho(3) = 0.4^3 = \mathbf{0.064}$
$\rho(4) = 0.4^4 = \mathbf{0.0256}$
$\vdots$
$\rho(k) = \mathbf{0.4}^k$

ACF – MA(q)

Example: Sample ACF for an MA(q) process:

$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$$

$$\rho(k) = \frac{\sum_{j=k}^{q}\theta_j\theta_{j-k}}{(1+\theta_1{}^2+\theta_2{}^2+\cdots+\theta_q{}^2)} \qquad k \le q$$

$$= 0 \qquad \text{otherwise.}$$

Suppose we have an MA(3). Then, for different k's:

$$\rho(0) = 1$$

$$\rho(1) = \frac{\theta_1 + \theta_2\theta_1 + \theta_3\theta_2}{(1 + \theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)}$$

$$\rho(2) = \frac{\theta_2 + \theta_3\theta_1}{(1 + \theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)}$$

$$\rho(3) = \frac{\theta_3}{(1 + \theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)}$$

$$\rho(k) = 0 \qquad \text{for } |k| > 3.$$

ACF – MA(q=3)

Example (continuation): $y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3}$

Suppose $\theta_1 = 0.5$; $\theta_2 = 0.4$; $\theta_3 = 0.2$. Then,

$$\rho(0) = 1$$

$$\rho(1) = \frac{\theta_1 + \theta_2\theta_1 + \theta_3\theta_2}{(1 + \theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)} = \frac{0.5+0.4*0.5+0.1*0.4}{1 + 0.5^2 + 0.4^2 + 0.1^2} = \mathbf{0.5211}$$

$$\rho(2) = \frac{\theta_2 + \theta_3\theta_1}{(1 + \theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)} = \frac{0.4 + 0.1*0.5}{1 + 0.5^2 + 0.4^2 + 0.1^2} = \mathbf{0.3169}$$

$$\rho(3) = \frac{\theta_3}{(1 + \theta_1{}^2 + \theta_2{}^2 + \theta_3{}^2)} = \frac{0.1}{1 + 0.5^2 + 0.4^2 + 0.1^2} = \mathbf{0.0704}$$

$$\rho(k) = \mathbf{0} \qquad \text{for } |k| > 3.$$

ACF – ARMA(1, 1)

Example: Sample ACF for an ARMA(1,1) process:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

From the autocovariances, we get

$$\gamma(0) = \sigma^2 \frac{1 + \theta_1^{\,2} + 2\phi_1 \theta_1}{1 - \phi_1^{\,2}}$$

$$\gamma(1) = \sigma^2 \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 - \phi_1^{\,2}}$$

$$\gamma(k) = \phi_1 \gamma(k-1) = \phi_1^{\,k-1} \sigma^2 \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 - \phi_1^{\,2}}$$

Then,

$$\rho(k) = \phi_1^{\,k-1} \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 + \theta_1^{\,2} + 2\phi_1 \theta_1}$$

⇒ If $|\phi_1| < 1$, exponential decay. Similar pattern to AR(1).

ACF – ARMA(1, 1)

Example (continuation): Sample ACF for an ARMA(1,1) process:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

The ACF for an ARMA(1,1):

$$\rho(k) = \phi_1^{\,k-1} \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 + \theta_1^{\,2} + 2\phi_1 \theta_1}$$

Suppose $\phi_1 = \mathbf{0.4}$, $\theta_1 = \mathbf{0.5}$. Then,

$$\rho(0) = \mathbf{1}$$

$$\rho(1) = \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5} = \mathbf{0.6545}$$

$$\rho(2) = 0.4 * \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5} = \mathbf{0.2618}$$

$$\rho(3) = 0.4^2 * \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5} = \mathbf{0.0233}$$

⋮

$$\rho(k) = 0.4^{k-1} * \frac{(1 + 0.4 * 0.5) * (0.4 + 0.5)}{1 + 0.5^2 + 2*0.4*0.5}$$

ACF – Example: U.S. Stock Returns

Example: **US Monthly Returns** (1871 – 2020, T = 1,795)

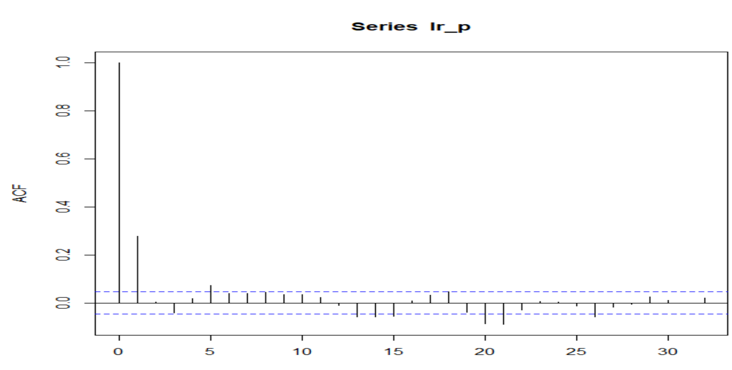
```
Sh_da <- read.csv("C://Financial Econometrics/Shiller_2020data.csv", head=TRUE,
sep=",")
x_P <- Sh_da$P
x_D <- Sh_da$D
T <- length(x_P)
lr_p <- log(x_P[-1]/x_P[-T])
lr_d <- log(x_D[-1]/x_D[-T])
acf_p <- acf(lr_p)                    # acf: R function that estimates the ACF
> acf_p
Autocorrelations of series ‘lr_p’, by lag
```

0	1	2	3	4	5	6	7	8	9	10	11
1.000	0.279	0.004	-0.043	0.017	0.074	0.039	0.039	0.044	0.035	0.034	0.022
12	13	14	15	16	17	18	19	20	21	22	23
-0.010	-0.059	-0.058	-0.056	0.009	0.033	0.047	-0.040	-0.087	-0.090	-0.029	0.005
24	25	26	27	28	29	30	31	32			
0.003	-0.013	-0.058	-0.018	-0.005	0.026	0.011	0.000	0.020			

SE(r_k) = 1/sqrt(T) = 1/sqrt(1,795) = .0236. $\Rightarrow$ 95% C.I.: $\pm$ 2* 0.0236

ACF – Example: U.S. Stock Returns

Example (continuation): Correlogram for **US Monthly Returns** (1871 – 2020)



Note: With the exception of first correlation, correlations are small. However, many are significant, not strange result when T is large.

ACF – Example: U.S. Stock Dividends

Example: **US Monthly Changes in Dividends** (1871 – 2020, T=1,795)

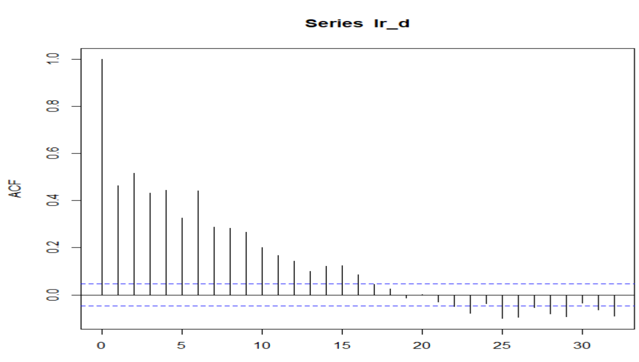
```
acf_d <- acf(lr_d)
> acf_d
Autocorrelations of series 'lr_d', by lag
```

0	1	2	3	4	5	6	7	8	9	10	11
1.000	0.462	0.516	0.432	0.444	0.326	0.442	0.288	0.283	0.265	0.202	0.168
12	13	14	15	16	17	18	19	20	21	22	23
0.142	0.100	0.122	0.123	0.085	0.045	0.026	-0.013	0.001	-0.029	-0.049	-0.077
24	25	26	27	28	29	30	31	32			
-0.038	-0.100	-0.095	-0.055	-0.081	-0.092	-0.034	-0.063	-0.089			

High correlations and significant even after 32 months!

ACF – Example: U.S. Stock Dividends

Example (continuation): Correlogram for **US Monthly Changes in Dividends** (1871 – 2020)



Note: Correlations are positive for almost 1.5 years, then become negative.

ACF – Joint Significance Tests

Recall we compute the Ljung-Box (LB) statistic as:

$$LB = T * (T+2) \sum_{k=1}^{m} \left(\frac{\hat{\rho}_k^2}{(T-k)}\right)$$

The LB test can be used to determine if the first *m* sample ACFs are jointly equal to zero.

Under H_0: $\rho_1 = \rho_2 = \ldots = \rho_m = 0$, $\quad LB \xrightarrow{d} \chi^2_m$

ACF – Joint Significance Tests

Example: LB test with **20 lags** for **US Monthly Returns** and **Changes in Dividends** (1871 – 2020)

```
> Box.test(lr_p, lag=20, type= "Ljung-Box")

        Box-Ljung test

data:  lr_p
X-squared = 208.02, df = 20, p-value < 2.2e-16
```

⇒ Reject H_0 at 5% level. Joint significant first 20 correlations.

```
> Box.test(lr_d, lag=20, type= "Ljung-Box")

        Box-Ljung test

data:  lr_d
X-squared = 2762.7, df = 20, p-value < 2.2e-16
```

⇒ Reject H_0 at 5% level. Joint significant first 20 correlations.

Partial ACF (PACF)

• The ACF gives us a lot of information about the order of the dependence when the series we analyze follows a MA process: The ACF is zero after q lags for an MA(q) process.

• If the series we analyze, however, follows an ARMA or AR, the ACF alone tells us little about the orders of dependence: We only observe an exponential decay.

• We introduce a new function that behaves like the ACF of MA models, but for AR models, namely, the partial autocorrelation function (PACF).

• The PACF is similar to the ACF. It measures correlation between observations that are *k* time periods apart, after controlling for correlations at intermediate lags.

Partial ACF

Intuition: Suppose we have an AR(1):

$$y_t = \phi_1 \, y_{t-1} + \varepsilon_t.$$

Then,

$$\gamma(2) = \phi_1^2 \, \gamma(0)$$

The correlation between y_t and y_{t-2} is not zero, as it would be for an MA(1), because y_t is dependent on y_{t-2} through y_{t-1}.

Suppose we break this chain of dependence by removing ("partialing out") the effect y_{t-1}. Then, we consider the correlation between $[y_t - \phi_1 y_{t-1}]$ & $[y_{t-2} - \phi_1 \, y_{t-1}]$ –i.e, the correlation between y_t & y_{t-2} with the linear dependence of each on y_{t-1} removed:

$$\gamma(2) = \text{Cov}(y_t - \phi_1 y_{t-1}, y_{t-2} - \phi_1 \, y_{t-1}) = \text{Cov}(\varepsilon_t, y_{t-2} - \phi_1 \, y_{t-1}) = 0$$

Similarly,

$$\gamma(k) = \text{Cov}(\varepsilon_t, y_{t-k} - \phi_1 \, y_{t-1}) = 0 \quad \text{for all } k > 1.$$

Partial ACF

Definition: The **PACF** of a stationary time series $\{y_t\}$ is

ϕ_{11}= Corr(y_t, y_{t-1}) = ρ(1)

ϕ_{hh} = Corr(y_t – E[y_t | I_{t-1}], y_{t-h} – E[y_{t-h} | I_{t-1}]) for h = 2, 3, ...

This removes the linear effects of y_{t-2}, ..., y_{t-h}.

• The PACF ϕ_{hh} is also the last coefficient in the **best linear prediction** of y_t given y_{t-1}, y_{t-2}, ..., y_{t-h}. (⇒ OLS!)

• Estimation by Yule-Walker equation, using sample estimates:

$$\hat{\boldsymbol{\phi}}_h = [\hat{\boldsymbol{R}}]^{-1}\hat{\boldsymbol{\gamma}}(k) \qquad \Rightarrow \text{a recursive system,}$$

where $\boldsymbol{\phi}_h$ = (ϕ_{h1}, ϕ_{h2} , ..., ϕ_{hh}) and $\boldsymbol{R}$ is the (hxh) correlation matrix.

• OLS is used. Also, a recursive algorithm by Durbin-Levinson.

Partial ACF – AR(p)

Example: AR(p) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$
$$E[y_t|I_{t-1}] = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-h-1}$$
$$E[y_{t-h}|I_{t-1}] = \mu + \phi_1 y_{t-h-1} + \phi_2 y_{t-h-2} + \ldots + \phi_p y_{t-1}$$

Then, $\phi_{hh} = \phi_h$ if $1 \leq h \leq p$

$\quad = 0$ otherwise

⇒ After the p^{th} *PACF,* all remaining PACF are 0 for AR(p) processes.

• The plot of the PACF is called the **partial correlogram**.

Partial ACF – AR(p=2)

Example: We simulate an AR(2) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

```
sim_ar22 <- arima.sim(list(order=c(1,0,0), ar=c(0.5, 0.3)), n=200)       #simulate AR(2) series
plot(sim_ar22, ylab="Simulated Series", main=(expression(AR(2):~~~phi==c(0.5,0.3))))
pacf_ar22 <- pacf(sim_ar22)
```

Print PACF

```
> pacf_ar2

    1      2      3      4      5      6      7      8      9     10     11
 0.558  0.286  0.038  0.103 -0.010  0.009  0.111  0.060 -0.021 -0.076  0.016
    12     13     14     15     16     17     18     19     20     21     22     23
-0.086 -0.139  0.100  0.061 -0.156  0.078 -0.103  0.043 -0.075  0.104  0.024  0.061
```

SE(r_k) ≈ 1/sqrt(200) = .0707. ⇒ 95% C.I.: ± 2* 0.0707

Partial ACF – AR(p=2)

Example: We simulate an AR(2) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 \; y_{t-2} + \varepsilon_t$$

```
sim_ar22 <- arima.sim(list(order=c(1,0,0), ar=c(0.5, 0.3)), n=200)       #simulate AR(2) series
plot(sim_ar22, ylab="Simulated Series", main=(expression(AR(2):~~~phi==c(0.5,0.3))))
pacf_ar22 <- pacf(sim_ar22)
```

Print PACF

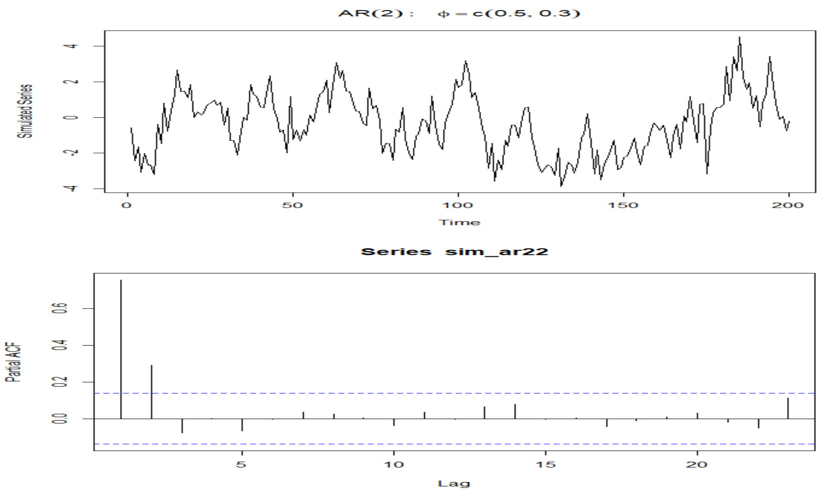
```
> pacf_ar2

    1      2      3      4      5      6      7      8      9     10     11
 0.558  0.286  0.038  0.103 -0.010  0.009  0.111  0.060 -0.021 -0.076  0.016
    12     13     14     15     16     17     18     19     20     21     22     23
-0.086 -0.139  0.100  0.061 -0.156  0.078 -0.103  0.043 -0.075  0.104  0.024  0.061
```

SE(r_k) ≈ 1/sqrt(200) = .0707. ⇒ 95% C.I.: ± 2* 0.0707

Partial ACF – AR(p=2)

Example (continuation): Plot of simulated series and PACF

> plot(sim_ar22, ylab="Simulated Series", main=(expression(AR(2):~~~phi==c(0.5,0.3))))
> pacf_ar2 <- pacf(sim_ar22)



Partial ACF – AR(p=2)

Example (continuation):

Note: The PACF can be calculated by *h* regressions, each one with *h* lags. The *hh* coefficient is the h^{th} order PACF. Using *ar* function:

> ar(sim_ar2, order.max=1, method = "ols")

Coefficients:
1
0.5586

Intercept: -0.008403 (0.0761)

Order selected 1 sigma^2 estimated as 1.152

> ar(sim_ar2, order.max=2, method = "ols")

Coefficients:
1 2
0.3974 **0.2869**

Intercept: -0.009847 (0.07326)

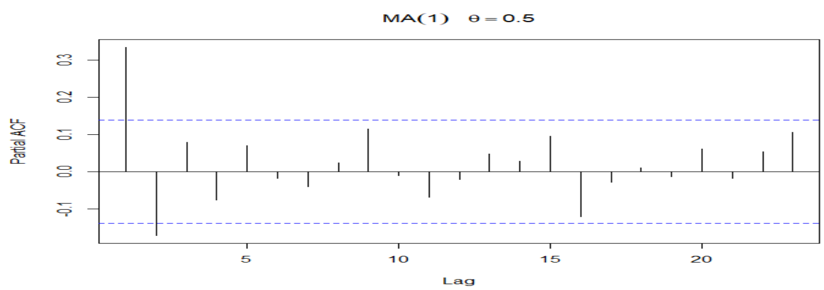
Order selected 2 sigma^2 estimated as 1.063

Partial ACF – MA(q)

• Following the analogy that PACF for AR processes behaves like an ACF for MA processes, we will see exponential decay ("*tails off*") in the partial correlogram for MA process. Similar pattern will also occur for ARMA(p, q) process.

Example: We simulate an MA(1) process with θ_1 = 0.5.

```
sim_ma1 <- arima.sim(list(order=c(0,0,1), ma = 0.5), n=200)
> pacf(sim_ma1)
```

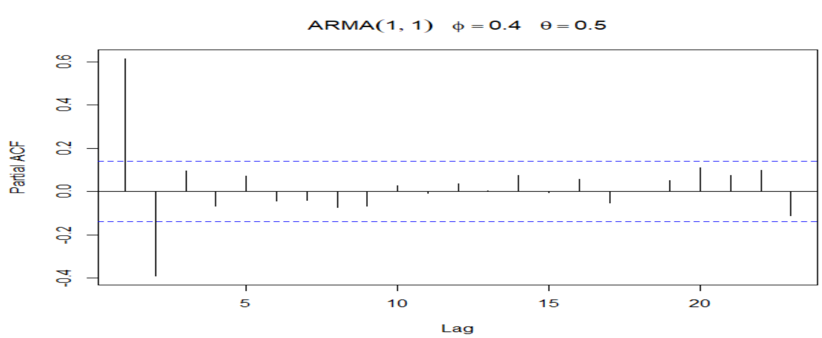


Partial ACF – ARMA(p,q)

• For an ARMA processes, we will see exponential decay ("*tails off*") in the partial correlogram.

Example: We simulate an ARMA(1) process with ϕ_1= 0.4 & θ_1= 0.5.

```
sim_arma11 <- arima.sim(list(order=c(1,0,1), ar=0.4, ma=0.5), n=200)
> pacf(sim_arma11)
```



PACF – Example: U.S. Stock Returns

Example: **US Monthly Returns** (1871 – 2020, T=1,795)

pacf_p <- acf(lr_p)   # pacf: R function that estimates the PACF
> pacf_p

Partial autocorrelations of series ‘lr_p’, by lag

1	2	3	4	5	6	7	8	9	10	11
0.278	-0.081	-0.026	0.041	0.058	0.002	0.038	0.032	0.016	0.022	0.009

12	13	14	15	16	17	18	19	20	21	22	23
-0.023	-0.057	-0.032	-0.045	0.027	0.017	0.037	-0.059	-0.051	-0.050	0.005	24

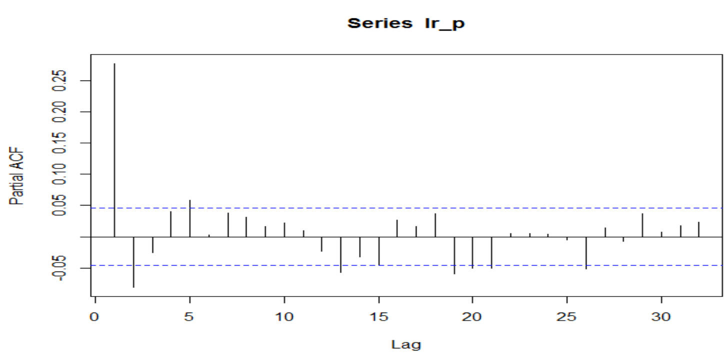
23	24	25	26	27	28	29	30	31	32
0.006	0.004	-0.005	-0.051	0.014	-0.007	0.037	0.008	0.018	0.023

SE(r_k) = 1/sqrt(1,795) = .0236.   ⇒ 95% C.I.: ± 2* 0.0236

PACF – Example: U.S. Stock Returns

Example (continuation): Correlogram for **US Monthly Returns** (1871 – 2020)

> pacf(lr_p)



Note: With the exception of the first partial correlation, partial correlations are small, though, again, some are significant.

PACF – Example: U.S. Stock Dividends

Example: **US Monthly Stock Dividends** (1871 – 2020, T=1,795)

```
pacf_d <- pacf(lr_d)
> pacf_d

Partial autocorrelations of series 'lr_d', by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.462  0.385  0.160  0.150 -0.033  0.189 -0.054 -0.056  0.027 -0.082 -0.019
    12     13     14     15     16     17     18     19     20     21     22     23
-0.063 -0.035  0.067  0.043  0.010 -0.057 -0.046 -0.043 -0.008 -0.031 -0.039
    24     25     26     27     28     29     30     31     32
-0.041  0.050 -0.036 -0.030  0.091  0.006 -0.017  0.044 -0.002 -0.042
```

Higher partial correlations than for stock returns.

ARIMA Models: Identification – Correlations

- Correlation approach.

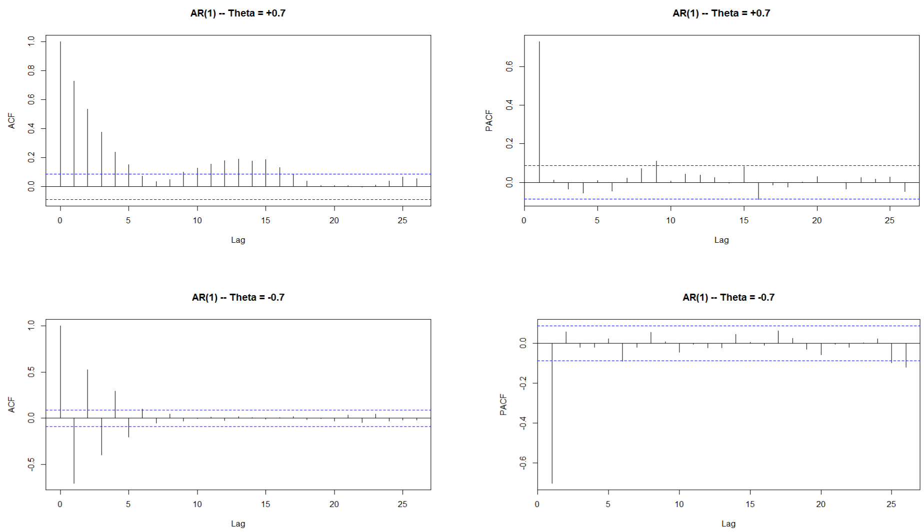
Basic tools: sample ACF and sample PACF.
- ACF identifies order of MA: Non-zero at lag q; zero for lags > q.
- PACF identifies order of AR: Non-zero at lag p; zero for lags > p.
- All other cases, try ARMA(p, q) with $p > 0$ and $q > 0$.

Summary: For $p > 0$ & $q > 0$.

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	0 after lag q	Tails off
PACF	0 after lag p	Tails off	Tails off

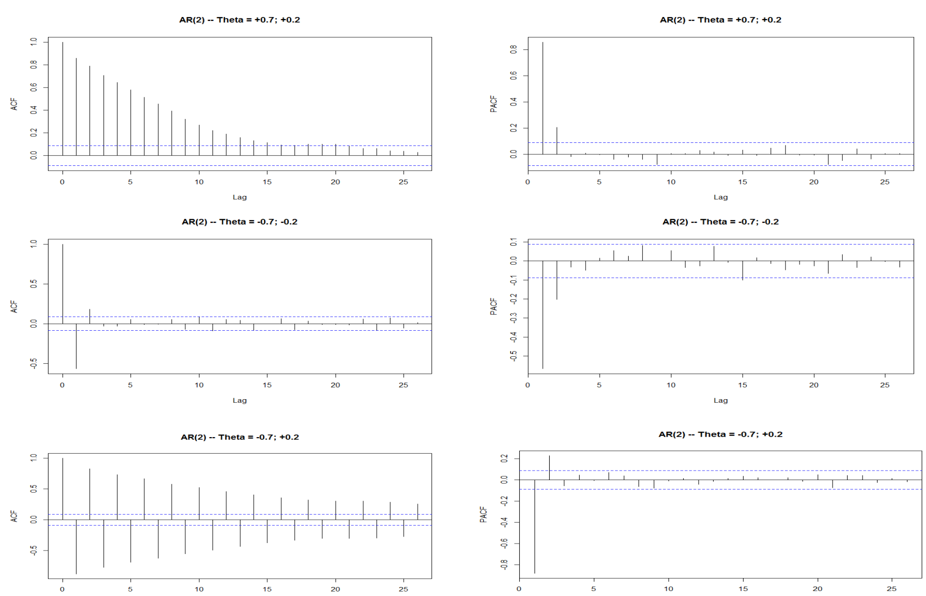
Note: Ideally, "Tails off" is exponential decay. In practice, in these cases, we may see a lot of non-zero values for the ACF and PACF.

ARMA Models: Identification – AR(1)

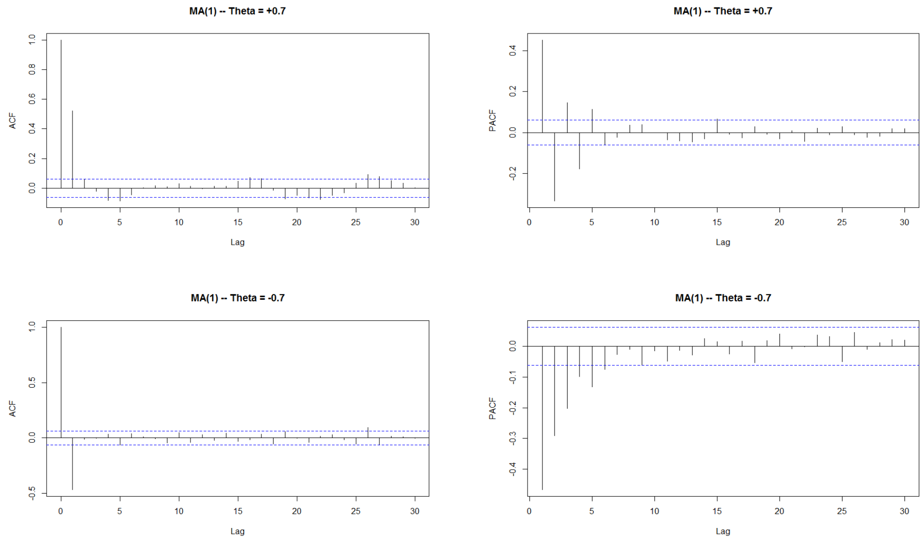


49

ARMA Models: Identification – AR(2)

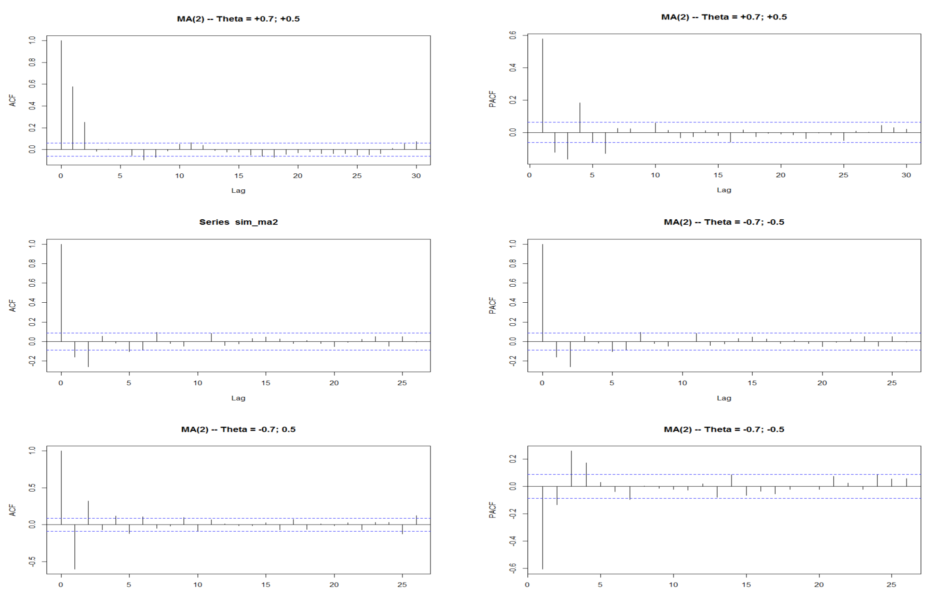


ARMA Models: Identification – MA(1)

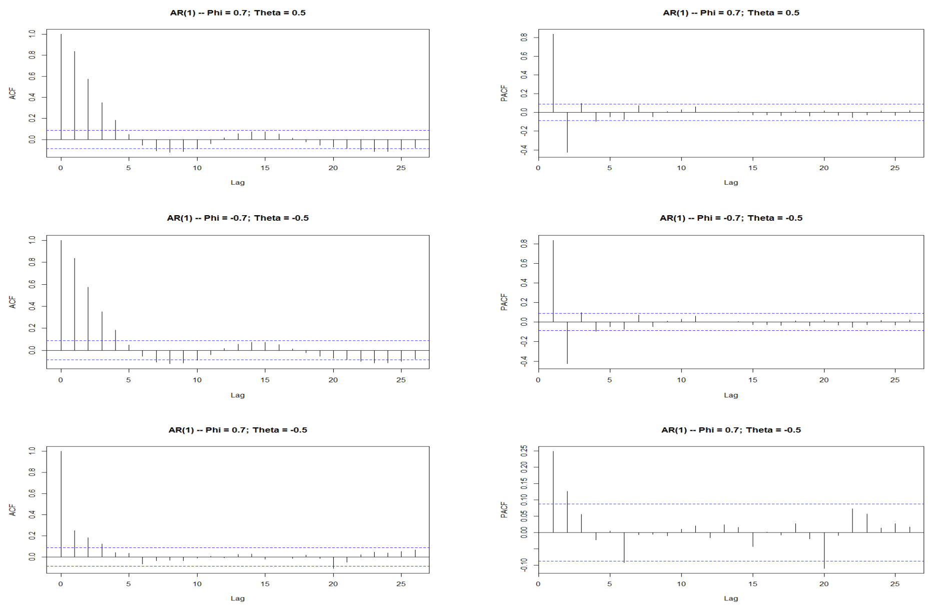


51

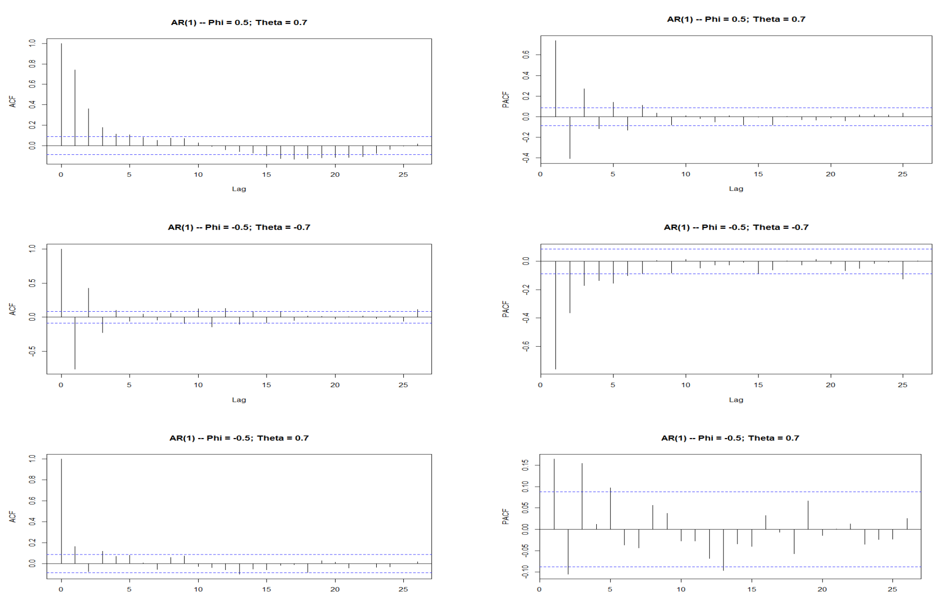
ARMA Models: Identification – MA(2)



ARMA Models: Identification – ARMA(1,1)

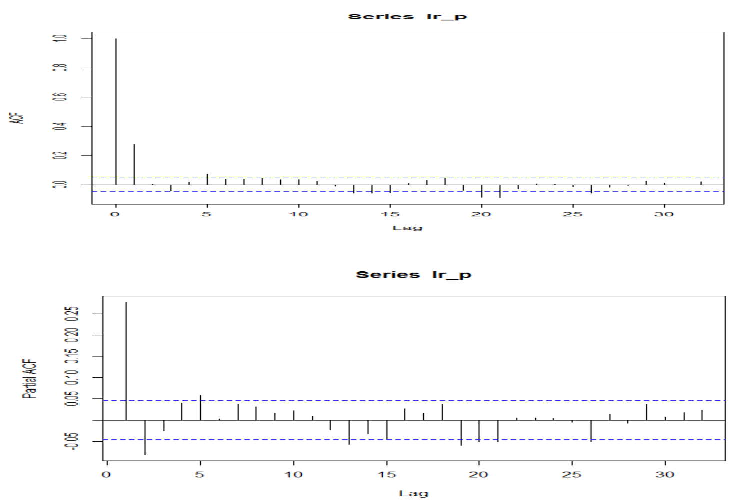


ARMA Models: Identification – ARMA(1,1)



ARMA Models: Identification – ARMA(1,1)

Example: **Monthly US Returns** (1871 - 2020).

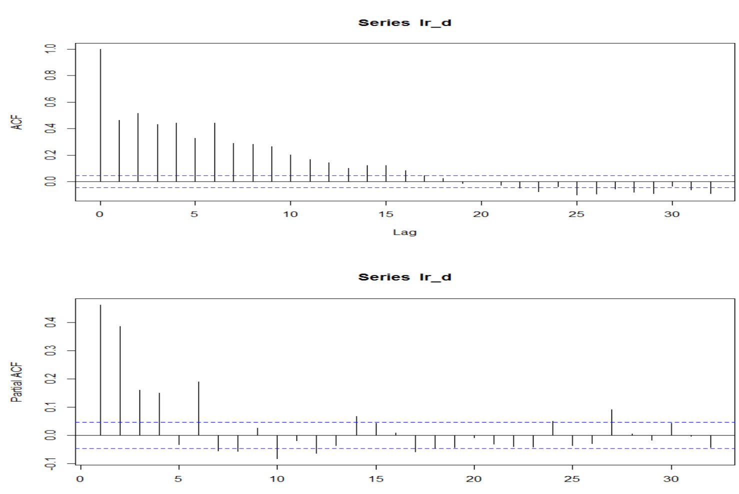


• Note: ARMA(1,1), MA(1), AR(2)?

55

ARMA Models: Identification – ARMA(1,1)

Example: **Monthly Changes in US Dividends** (1871 - 2020).

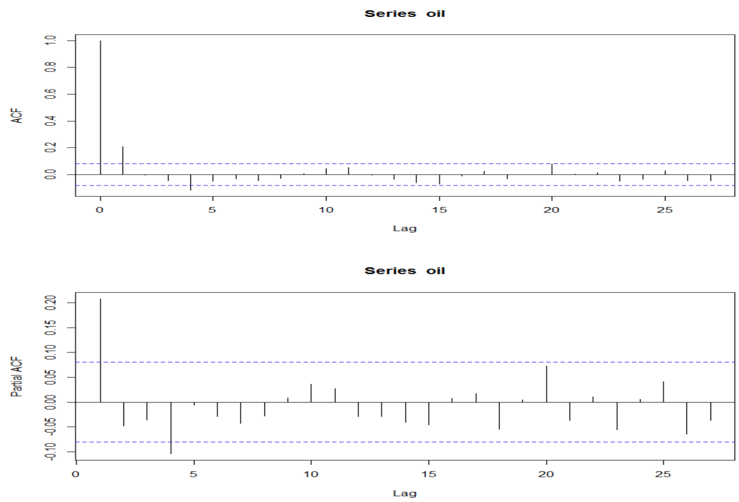


• Note: Not clear: Maybe long a ARMA(p, q) or needs differencing?

56

ARMA Models: Identification – ARMA(1,1)

Example: **Monthly Log Changes in Oil Prices** (1973 - 2020).

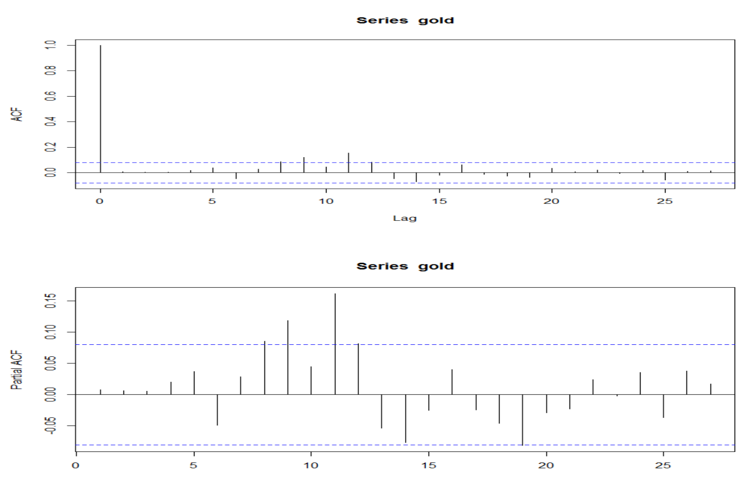


• Note: MA(1), AR(4)?

57

ARMA Models: Identification – ARMA(1,1)

Example: **Monthly Log Changes in Gold** (1973 - 2020).



• Note: No clear ARMA structure.

58

ARMA Model: Identification - IC

- It is difficult to identify an ARMA model using the ACF and PACF. It is common to rely on information criteria (IC).

- IC's are equal to the estimated variance or the log-likelihood function plus a penalty factor, that depends on k. Many IC's:

- Akaike Information Criterion (AIC)
  AIC = -2 * (ln $L - k$) = -2 ln L + 2 * k
  ⇒ if normality AIC = T * ln($\mathbf{e'e}/T$) + 2* k (+constants)

- Bayes-Schwarz Information Criterion (BIC or SBIC)
  BIC = -2 * ln L – ln(T) * k
  ⇒ if normality AIC = T * ln($\mathbf{e'e}/T$) + ln(T) * k (+constants)

- Hannan-Quinn (HQIC)
  HQIC = -2*(ln $L - k$ [ln(ln(T))]
  ⇒ if normality AIC = T * ln($\mathbf{e'e}/T$) + 2 k [ln(ln(T))] (+constants)

ARIMA Model: Identification - IC

- There are modifications of IC to get better finite sample behavior, a popular one is AIC corrected, AICc, statistic:

$$AICc = T\,\widehat{\ln\sigma}^2 + \frac{2k(k+1)}{T-k-1}$$

- AICc converges to AIC as T gets large. Using AICc is not a bad idea.

- For AR(p) models, other AR-specific criteria are possible: Akaike's final prediction error (FPE), Akaike's BIC, Parzen's CAT.

- Hannan and Rissannen's (1982) *minic* (*=Min*imum *IC*): Calculate the BIC for different p*'s* (estimated first) and different q*'s*. Select the best model –i.e., lowest BIC.

Note: Box, Jenkins, and Reinsel (1994) proposed using the AIC above.

ARMA Model: Identification - IC

Example: Monthly US Returns (1871 - 2020) Hannan and Rissannen (1982)'s minic, based on AIC.

Minimum Information Criterion

Lags	**MA 0**	**MA 1**	**MA 2**	**MA 3**	**MA 4**	**MA 5**
AR 0	-6403.59	-6552.94	-6552.69	-6554.27	-6552.88	-6557.37
AR 1	-6545.22	-6552.23	-6551.86	-6552.42	-6552.64	**-6561.48**
AR 2	-6554.76	-6553.28	-6554.85	-6554.35	**-6564.32**	-6559.48
AR 3	-6553.94	-6552.53	-6554.44	-6552.33	-6550.36	-6558.52
AR 4	-6554.98	-6559.83	**-6559.92**	-6558.94	-6554.1	-6558.16
AR 5	**-6558.81**	-6558.65	-6557.45	-6555.78	-6558.66	-6556.06

• Note: Best Model is ARMA(2,4); other potential candidates: ARMA(1,5), ARMA(4,2), ARMA (5,0).

Non-Stationary Time Series Models

• The ACF is as a rough indicator of whether a trend is present in a series. A slow decay in ACF is indicative of highly correlated data, which suggests a true *unit root process*, or a *trend stationary process*.

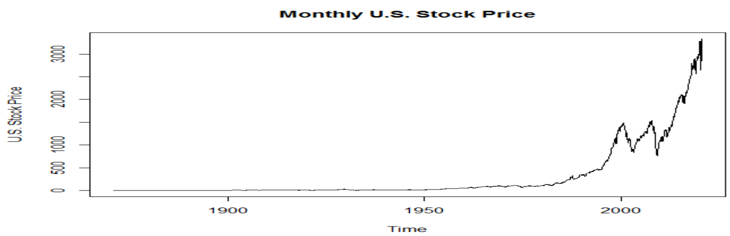
• Formal tests can help to determine whether a system contains a trend and whether the trend is deterministic or stochastic (unit root).

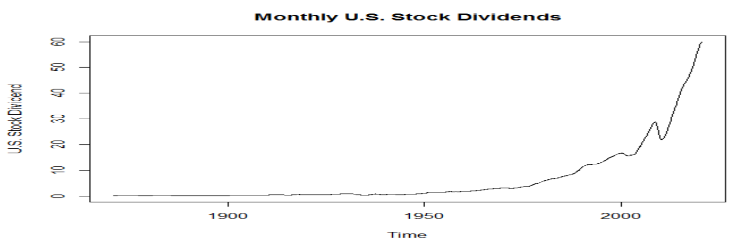
• We will analyze two situations faced in ARMA models:
(1) Deterministic trend – Simple model: $y_t = \alpha + \beta\, t + \varepsilon_t$
– Solution: *Detrending* –i.e., regress y_t on a constant and a time trend, t. Then, keep residuals for further modeling.

(2) Stochastic trend – Simple model: $y_t = \mu + y_{t-1} + \varepsilon_t$.
– Solution: *Differencing* –i.e., apply $\Delta = (1 - L)$ operator to y_t. Then, use Δy_t for further modeling.

Non-Stationary Time Series Models

Example: Plot of US Monthly Prices and Dividends (1871 – 2020)





Non-Stationary Models: Deterministic Trend

- Suppose we have the following model, with a determinist trend:

$$y_t = \alpha + \beta\, t + \varepsilon_t. \qquad \Rightarrow \Delta y_t = y_t - y_{t-1}$$
$$= \beta\, t - \beta\, (t-1) + \varepsilon_t - \varepsilon_{t-1}$$
$$= \beta + \varepsilon_t - \varepsilon_{t-1}$$
$$\Rightarrow \mathrm{E}[\Delta y_t] = \beta$$

- $\{y_t\}$ will show only temporary departures from trend line $\alpha + \beta\, t$. This type of model is called a *trend stationary* (TS) model.

- If a series has a deterministic time trend, then we **detrend** y_t. That is, we remove the influence of t from y_t: We simply regress y_t on an intercept and a time trend (t = *1, 2, …, T*); then, save the residuals:

$$e_t = y_t - \hat{\alpha} - \hat{\beta}\, t \qquad \text{(the residuals are the } \textit{detrended } y_t \text{ series)}$$

- But, we do not necessarily get stationary series by detrending.

Non-Stationary Models: Deterministic Trend

• Many economic series exhibit "exponential trend/growth". They grow over time like an exponential function over time instead of a linear function. In this cases, it is common to work with logs

$$\ln(y_t) = \alpha + \beta\, t + \varepsilon_t. \qquad (\Rightarrow y_t = e^{\alpha + \beta\, t + \varepsilon_t})$$

$\Rightarrow$ The average growth rate is: $E[\Delta \ln(y_t)] = \beta$

• We can have a more general model:

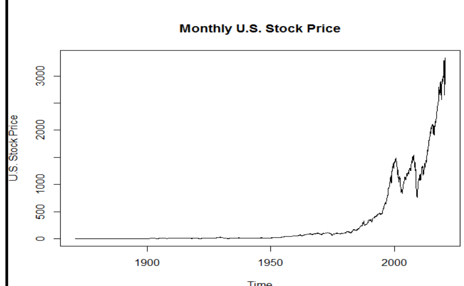
$$y_t = \alpha + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \beta_1 t + \beta_2 t^2 + \ldots + \beta_k t^k + \varepsilon_t.$$

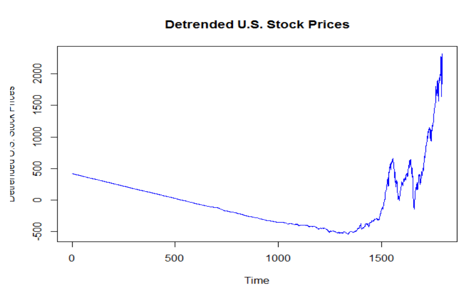
• Estimation of AR(p) with a trend component:
- OLS.
- Frish-Waugh method (a 2-step method):
  (1) Detrend y_t: regress y_t against a constant & a time trend, t. Then, get the residuals (=y_t without the influence of t).
  (2) Use residuals to estimate the AR(p) model.

Non-Stationary Models: Deterministic Trend

Example: We detrend U.S. Stock Prices

```
T <- length(x_P)                          # length of series
trend <- c(1:T)                           # create trend
det_P <- lm(x_P ~ trend)                  # regression to get detrended e
detrend_P <- det_P$residuals
plot(detrend_P, type="l", col="blue", ylab ="Detrended U.S. Prices", xlab ="Time")
title("Detrended U.S. Stock Prices")
```

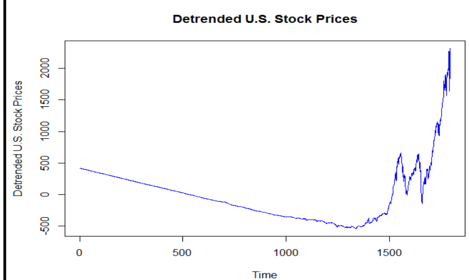


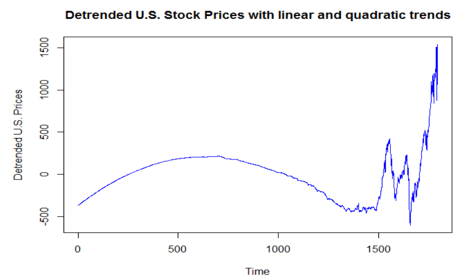


Non-Stationary Models: Deterministic Trend

Example: We detrend U.S. Stock Prices adding a square trend

```
trend2 <- trend^2
det_P <- lm(x_P ~ trend + trend2)          # regression to get detrended e
detrend_P <- det_P$residuals
plot(detrend_P, type="l", col="blue", ylab ="Detrended U.S. Prices", xlab ="Time")
title("Detrended U.S. Stock Prices with linear and quadratic trends")
```

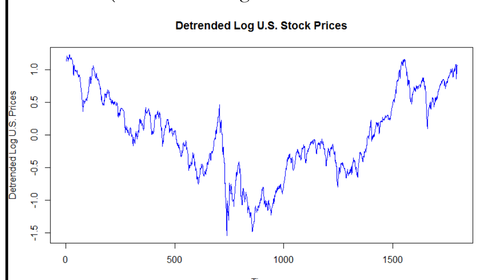


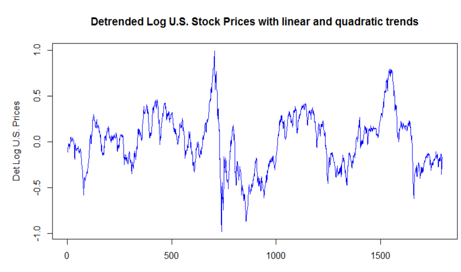


Non-Stationary Models: Deterministic Trend

Example: We detrend Log U.S. Stock Prices adding a squared trend

```
l_P <- log(x_P)
det_lP <- lm(l_P ~ trend)                  # regression to get detrended e
detrend_lP <- det_lP$residuals
plot(detrend_lP, type="l", col="blue", ylab ="Detrended Log U.S. Prices", xlab ="Time")
title("Detrended Log U.S. Stock Prices")

det_lP2 <- lm(l_P ~ trend + trend2)        # regression to get detrended e
det_lP2 <- det_lP2$residuals
plot(det_lP2, type="l", col="blue", ylab ="Det Log U.S. Prices", xlab ="Time")
title("Detrended Log U.S. Stock Prices with linear and quadratic trends")
```





Non-Stationary Models: Stochastic Trend

• The more modern approach is to consider trends in time series as a variable trend.

• A variable trend exists when a trend changes in an unpredictable way. Therefore, it is considered *stochastic*.

• Recall the AR(1) model: $y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$

• As long as $|\phi_1| < 1$, everything is fine, we have a stationary AR(1) process: OLS is consistent, t-stats are asymptotically normal, etc.

• Now consider the special case where $\phi_1 = 1$:

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

Q: Where is the (stochastic) trend? No t term.

Non-Stationary Models: Stochastic Trend

• Let us replace recursively the lag of y_t on the right-hand side:

$$\begin{aligned} y_t &= \mu + y_{t-1} + \varepsilon_t \\ &= \mu + (\mu + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &\ldots \\ &= y_0 + \underbrace{t\,\mu}_{\boxed{\text{Deterministic trend}}} + \textstyle\sum_{j=0}^{t} \varepsilon_{t-j} \end{aligned}$$

• This process is called a "*random walk with drift*": y_t grows with t.

• Each ε_t shock represents a shift in the intercept. All values of $\{\varepsilon_t\}$ have a 1 as coefficient $\Rightarrow$ each shock never vanishes (permanent).

• We remove the trend by **differencing** y_t

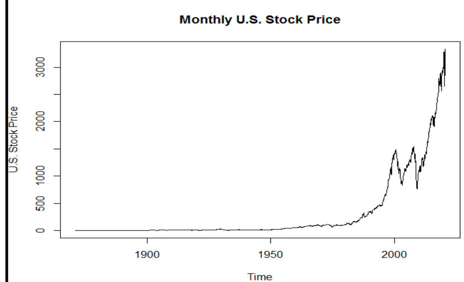
$$\Rightarrow \Delta y_t = (1 - L)\, y_t = \mu + \varepsilon_t$$

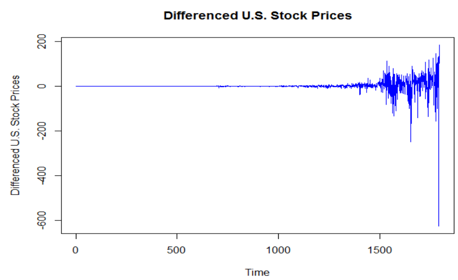
Note: Applying the $(1 - L)$ operator to a time series is called *differencing*

Non-Stationary Models: Stochastic Trend

Example: We difference U.S. Stock Prices, using the *diff* R function:

```
diff_P <- diff(x_P)
> plot(diff_P,type="l", col="blue", ylab ="Differenced U.S. Stock Prices", xlab ="Time")
> title("Differenced U.S. Stock Prices")
```





Non-Stationary Models: Stochastic Trend

• y_t is said to have a *stochastic trend* (ST), since each ε_t shock gives a permanent and random change in the conditional mean of the series.

• For these situations, we use *Autoregressive Integrated Moving Average (ARIMA)* models.

• Q: Deterministic or Stochastic Trend?
They appear similar: Both lead to growth over time. The difference is how we think of ε_t. Should a shock today affect y_{t+1}?

– TS: $y_{t+1} = \mu + \beta\,(t+1) + \varepsilon_{t+1}$ $\quad\Rightarrow \varepsilon_t$ does not affect y_{t+1}.

– ST: $y_{t+1} = \mu + y_t + \varepsilon_{t+1} = \mu + [\mu + y_{t-1} + \varepsilon_t] + \varepsilon_{t+1}$
$= 2 * \mu + y_{t-1} + \varepsilon_t + \varepsilon_{t+1} \quad\Rightarrow \varepsilon_t$ affects y_{t+1}.
(In fact, the shock ε_t has a *permanent* impact.)

ARIMA(p, d, q) Models

• For $p, d, q \geq 0$, we say that a time series $\{y_t\}$ is an *ARIMA* (p, d, q) *process* if $w_t = \Delta^d\, y_t = (1-L)^d\, y_t$ is ARMA(p, q). That is,

$$\phi(L)(1-L)^d\, y_t = \theta(L)\, \varepsilon_t$$

• Applying the $(1-L)$ operator to a time series is called *differencing*.

Notation: If y_t is non-stationary, but $\Delta^d\, y_t$ is stationary, then y_t is *integrated* of order d, or I(d). A time series with *unit root* is I(1). A stationary time series is I(0).

Examples:

Example 1: RW: $y_t = y_{t-1} + \varepsilon_t$.

y_t is non-stationary, but

$$w_t = (1-L)\; y_t = \varepsilon_t \qquad \Rightarrow w_t \sim \text{WN!}$$

Now, y_t ~ ARIMA(0, 1, 0).

ARIMA(p, d, q) Models

Example 2: AR(1) with time trend: $y_t = \mu + \delta\, t + \phi_1\, y_{t-1} + \varepsilon_t$.

y_t is non-stationary, but

$$\begin{aligned} w_t &= (1-L)\, y_t \\ &= \mu + \delta\, t + \phi_1\, y_{t-1} + \varepsilon_t - [\mu + \delta\,(t-1) + \phi_1\; y_{t-2} + \varepsilon_{t-1}]. \\ &= \delta + \phi_1\, w_{t-1} + \varepsilon_t - \varepsilon_{t-1} \qquad \Rightarrow w_t \sim \text{ARIMA(1, 1)}. \end{aligned}$$

Now, y_t ~ ARIMA(1, 1, 1).

• We call both process *first difference stationary*.

Note:

− Example 1: Differencing a series with a unit root in the AR part of the model reduces the AR order.

− Example 2: Differencing can introduce an extra MA structure. We introduced non-invertibility (θ_1=1). This happens when we difference a TS series. Detrending should be used in these cases.

