

Econometrics



David Preinerstorfer

Institute for Statistics and Mathematics
Dept. of Finance, Accounting and Statistics

Spring 2026



Last week:

1. Time series
2. Empirical autocovariances, stationarity, residual analysis
3. Models: Stationary processes, AR, and MA models

Plan for today:

1. ARMA models.
2. ARIMA models.
3. ARCH and GARCH models.

A process Y_t containing p AR terms and q MA terms is called an **ARMA(p, q)** process:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \\ + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

- p is the **order** of the AR part; $\phi_1, \dots, \phi_p$ are the **autoregressive parameters** of the model
- q is the **order** of the MA part; $\theta_1, \dots, \theta_q$ are the **moving average parameters** of the model

Representation of MA(1) as AR(∞)

According to representation (69), every stationary AR(1) process has a representation as an MA(∞) process.

Based on recursive substitution of $\varepsilon_{t-j} = Y_{t-j} - \theta\varepsilon_{t-j-1}$, the MA(1) process has an AR(∞) representation:

$$\begin{aligned} Y_t &= \varepsilon_t + \theta\varepsilon_{t-1} = \varepsilon_t + \theta Y_{t-1} - \theta^2\varepsilon_{t-2} \\ &= \varepsilon_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3\varepsilon_{t-3} = \cdots = \varepsilon_t + \sum_{j=1}^{\infty} (-1)^{j-1} \theta^j Y_{t-j}. \end{aligned} \tag{73}$$

The lag operator

The lag operator B moves the index of a variable Z_t back one time unit. Applying it k times, moves the index back k time units:

$$BZ_t = Z_{t-1}, \quad B^2Z_t = B(BZ_t) = BZ_{t-1} = Z_{t-2}, \quad B^kZ_t = Z_{t-k}.$$

Using B , we can rewrite AR processes and MA processes as:

$$\text{AR}(1): \quad (1 - \phi B)Y_t = \varepsilon_t,$$

$$\text{AR}(p): \quad (1 - \phi_1 B - \dots - \phi_p B^p)Y_t = \varepsilon_t,$$

$$\text{MA}(1): \quad Y_t = (1 + \theta B)\varepsilon_t,$$

$$\text{MA}(q): \quad Y_t = (1 + \theta_1 B + \dots + \theta_q B^q)\varepsilon_t.$$

Introducing the characteristic polynomial of the AR(p) and the MA(q) process,

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q,$$

leads to a compact way to write ARMA processes:

$$\text{AR}(p): \quad \phi(B)Y_t = \varepsilon_t,$$

$$\text{MA}(q): \quad Y_t = \theta(B)\varepsilon_t,$$

$$\text{ARMA}(p,q): \quad \phi(B)Y_t = \theta(B)\varepsilon_t.$$

Revisiting the AR(1) process

Define the inversion operator $\theta^\phi(B) = (1 - \phi B)^{-1}$ to be the operator so that $\theta^\phi(1 - \phi B) = 1$ is equal to the identity operator. Then, if $|\phi| < 1$, we obtain:

$$\theta^\phi(B) = (1 - \phi B)^{-1} = 1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \dots \quad (74)$$

If we apply this to invert the AR(1)-process, then we obtain representation (69):

$$Y_t = (1 - \phi B)^{-1}(1 - \phi B)Y_t = \theta^\phi(B)\varepsilon_t.$$

Representation of AR(p) as MA(∞)

To invert an AR(p)-process in a similar way, a factorization of the characteristic polynomial $\phi(B)$ of the AR(p) process is needed:

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p = \prod_{j=1}^p (1 - \lambda_j B), \quad (75)$$

where for each $j = 1, \dots, p$: $|\lambda_j| < 1$. Using the representation involving the p roots $\tilde{\lambda}_1, \dots, \tilde{\lambda}_p$ of $\phi(B)$,

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p = \prod_{j=1}^p (B - \tilde{\lambda}_j),$$

the requirement $|\lambda_j| < 1$ is equivalent to the condition:

all roots $\tilde{\lambda}_j$ of $\phi(B)$ lie outside of the unit circle, since

$$(B - \tilde{\lambda}_j) = -\tilde{\lambda}_j(1 - \lambda_j B), \quad \text{with } \lambda_j = \frac{1}{\tilde{\lambda}_j}.$$

$\lambda_1, \dots, \lambda_p$ are also called the **inverted roots** of $\phi(B)$. The inverted roots λ_j might be complex, but they lie within the unit circle, since $|\lambda_j| < 1$.

If $|\lambda_j| < 1$, the AR(p)-process has a representation similar to (69):

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (76)$$

Proof of (76): invert the AR(p)-process as follows:

$$Y_t = \phi(B)^{-1} \phi(B) Y_t = \prod_{j=1}^p (1 - \lambda_j B)^{-1} \varepsilon_t = \prod_{j=1}^p \theta^{\lambda_j}(B) \varepsilon_t,$$

where $\theta^{\lambda_j}(B)$ is defined in (74).

E.g. for $p = 2$:

$$\begin{aligned} \theta^{\lambda_1}(B) \theta^{\lambda_2}(B) &= (1 + \lambda_1 B + \lambda_1^2 B^2 + \dots)(1 + \lambda_2 B + \lambda_2^2 B^2 + \dots) \\ &= \sum_{j=0}^{\infty} \left(\sum_{k=0}^j \lambda_1^k \lambda_2^{j-k} \right) B^j = \sum_{j=0}^{\infty} \psi_j B^j, \quad \psi_j = \sum_{k=0}^j \lambda_1^k \lambda_2^{j-k}. \end{aligned}$$

Stationarity of AR(p)-processes

It turns out that with $|\lambda_j| < 1$ the absolute summability $\sum_{j=0}^{\infty} |\psi_j| < \infty$ is satisfied. Hence, the expectation of Y_t is zero. Representation (76) can be used to compute $\text{Cov}(Y_t, Y_{t+h})$ and shown that $\text{Cov}(Y_t, Y_{t+h}) \equiv \gamma(h) < \infty$, hence the AR(p)-process is stationary.

If we evaluate polynomial (75) at $B = 1$, then $|\lambda_j| < 1$ implies:

$$1 - \phi_1 - \dots - \phi_p = \prod_{j=1}^p (1 - \lambda_j) > 0 \Leftrightarrow \phi_1 + \dots + \phi_p < 1 \quad (77)$$

Note that (77) is a necessary, but not sufficient condition for stationarity.

Stationarity of AR(2)-processes

The roots of $1 - \phi_1 B - \phi_2 B^2 = 0$ are given by:

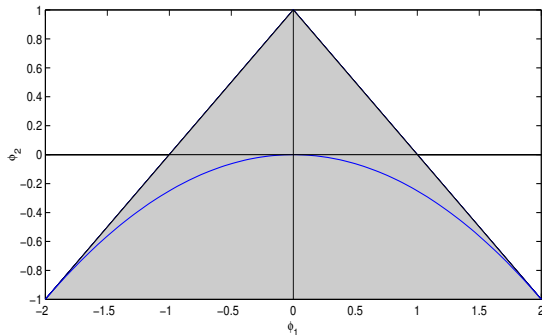
$$|\tilde{\lambda}_{1,2}| = \left| \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \right| > 1$$

The roots are complex, if $\phi_1^2 + 4\phi_2 < 0$.

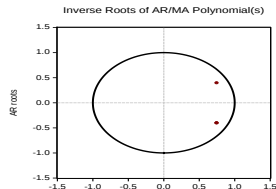
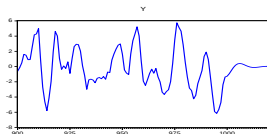
Region of stationarity is characterized by:

$$\phi_1 + \phi_2 < 1, \quad \phi_1 - \phi_2 > -1, \quad \phi_2 > -1$$

If the inverted roots of an AR(2) model are complex, then the corresponding autocorrelation function shows a sinusoidal decay.

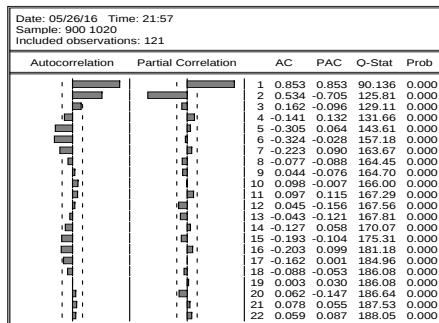


Simulated process with $\phi_1 = 1.5$ and $\phi_2 = -0.7$:



Corresponding empirical ACF:

Correlogram of Y_AR2_ALOIS



Yule-Walker equations

If the stationary condition is satisfied, second order moments can be obtained directly from multiplying (70) with Y_{t-h} and taking expectations. This yields the Yule-Walker equations for the covariances $\gamma(0), \dots, \gamma(p)$:

$$\begin{aligned}h = 0 : \quad & \gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \dots + \phi_p\gamma(p), \\h = 1 : \quad & \gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1) + \dots + \phi_p\gamma(p-1), \\& \dots \quad \dots \\h = p : \quad & \gamma(p) = \phi_1\gamma(p-1) + \phi_2\gamma(p-2) + \dots + \phi_p\gamma(0).\end{aligned}$$

This linear system of $p+1$ equations can be solved explicitly for $\gamma(0), \dots, \gamma(p)$. From these values, autocovariances $\gamma(h)$ for $h > p$ are given by:

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \dots + \phi_p\gamma(h-p).$$

Representation of MA(q) as AR(∞)

If an MA(q) process has a unique AR(∞) representation, it is called invertible.

Invertibility depends on the roots $\tilde{\omega}_1, \dots, \tilde{\omega}_q$ of the characteristic polynomial $\theta(B)$ of an MA(q) process, since

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q = \prod_{j=1}^q (1 - \omega_j B),$$

where $\omega_j = 1/\tilde{\omega}_j$ are the inverted roots.

Invertibility holds, if the inverted roots ω_j of $\theta(B)$ lie within the unit circle, i.e.,

$|\omega_j| < 1$, or equivalently, if the roots $\tilde{\omega}_j$ of $\theta(B)$ lie outside the unit circle: $|\tilde{\omega}_j| \geq 1$.

Invertible MA(q) processes are uniquely determined by their ACF.

The ARMA(p,q)-process

$$\phi(B)Y_t = \theta(B)\varepsilon_t,$$

- is stationary, if the AR part, i.e., $\phi(B)Y_t = \tilde{\varepsilon}_t$, is stationary;
- is invertible, if the MA part, i.e., $\tilde{Y}_t = \theta(B)\varepsilon_t$, is invertible.

Advantage of the ARMA(p,q): **parsimony** - low order ARMA models can substitute high-order AR or MA models.

Representations of invertible, stationary ARMA-Processes

- Pure MA(∞) process, also known as *Wold representation*:

$$Y_t = \psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi(B) = \phi(B)^{-1}\theta(B), \quad (78)$$

$$\Rightarrow V(Y_t) = \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2. \quad (79)$$

- Pure AR(∞) process: $\varepsilon_t = \theta(B)^{-1}\phi(B)Y_t$.

A value of the (non-observable) shock ε_t can be found from *past (observable)* values of Y_t . For a noninvertible process, we have to know also all future values of Y_t .

Example

ARMA(1,1)-Processes has 2 unknown parameters ϕ and θ :

$$Y_t = \phi Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t \Rightarrow (1 - \phi B)Y_t = (1 + \theta B)\varepsilon_t$$

MA-representation has infinitely many parameters that converge to 0:

$$\begin{aligned}\phi^{-1}(B)\theta(B) &= (1 - \phi B)^{-1}(1 + \theta B) = \\ &= (1 + \phi B + \phi^2 B^2 + \dots)(1 + \theta B) \\ &= 1 + (\phi + \theta)B + \phi(\phi + \theta)B^2 + \phi^2(\phi + \theta)B^3 + \dots\end{aligned}$$

ARMA(1,1)-Processes has 2 unknown parameters ϕ and θ :

$$Y_t = \phi Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t \Rightarrow (1 - \phi B)Y_t = (1 + \theta B)\varepsilon_t$$

AR-representation has infinitely many parameters that converge to 0:

$$\begin{aligned}\theta^{-1}(B)\phi(B) &= (1 + \theta B)^{-1}(1 - \phi B) \\ &= (1 - \theta B + \theta^2 B^2 - \dots)(1 - \phi B) \\ &= 1 - (\phi + \theta)B + \theta(\phi + \theta)B^2 - \theta^2(\phi + \theta)B^3 + \dots\end{aligned}$$

E.g. for the ARMA(1,1)-Process:

$$Y_t = 0.5Y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t-1} \Rightarrow (1 - 0.5B)Y_t = (1 - 0.2B)\varepsilon_t$$

MA-representation:

$$\begin{aligned}(1 - 0.5B)^{-1}(1 - 0.2B) &= (1 + 0.5B + 0.5^2B^2 + \dots)(1 - 0.2B) \\ &= 1 + 0.3B + 0.3 \cdot 0.5B^2 + 0.3 \cdot 0.5^2B^3 + \dots\end{aligned}$$

AR-representation:

$$\begin{aligned}(1 - 0.2B)^{-1}(1 - 0.5B) &= (1 + 0.2B + 0.2^2B^2 + \dots)(1 - 0.5B) \\ &= 1 - 0.3B - 0.3 \cdot 0.2B^2 - 0.3 \cdot 0.2^2B^3 + \dots\end{aligned}$$

Autocovariance and -correlation function of ARMA-Processes

Use the Wold representation $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ to determine the autocovariance and the autocorrelation function:

$$\begin{aligned}\text{Cov}(Y_t, Y_{t+h}) &= E(Y_t \cdot Y_{t+h}) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \psi_k E(\varepsilon_{t-j} \varepsilon_{t+h-k}) = \\ &= \sum_{j=0}^{\infty} \psi_j \psi_{j+h} E(\varepsilon_{t-j}^2) = \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}.\end{aligned}$$

Hence, $\text{Cov}(Y_t, Y_{t+h}) \equiv \gamma(h)$ for all t and $\rho(h) = \gamma(h)/\sigma^2$.

Example

ARMA(1,1)-Processes has following Wold representation:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad \psi_j = (\phi + \theta)\phi^{j-1}, j \geq 1.$$

Therefore

$$\begin{aligned} \gamma(h) &= \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} = \sigma_{\varepsilon}^2 (1 + (\phi + \theta)^2) \cdot \sum_{j=1}^{\infty} \phi^{j-1} \phi^{j+h-1} \\ &= \sigma_{\varepsilon}^2 \left(1 + (\phi + \theta)^2 \cdot \phi^h \sum_{j=1}^{\infty} (\phi^2)^{j-1} \right) = \sigma_{\varepsilon}^2 (1 + (\phi + \theta)^2) \cdot \frac{\phi^h}{1 - \phi^2} \end{aligned}$$

Adding a mean

The assumption of a zero mean is too restrictive in most cases. To model a stationary time series Y_t with mean $\mu \neq 0$, assume that $Y_t - \mu$ follows an ARMA-process:

$$\phi(B)(Y_t - \mu) = \theta(B)\varepsilon_t. \quad (80)$$

Such a process evidently has mean equal to μ :

$$E(Y_t - \mu) = E(Y_t) - \mu = 0 \Rightarrow E(Y_t) = \mu.$$

Model (80) can be extended to nonstationary time series with mean functions μ_t (see regression models with ARMA errors).

Estimating ARMA models I

An ARMA(p, q) model for an observed time series $y_1, \dots, y_T$,

$$\begin{aligned}
 y_t - \mu = & \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \dots + \phi_p(y_{t-p} - \mu) + \\
 & + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} + \varepsilon_t,
 \end{aligned} \tag{81}$$

could be viewed as a multiple regression model with intercept $\tilde{\beta}_0 = (1 - \phi_1 - \dots - \phi_p)\mu$, where the AR coefficients $\phi_1, \dots, \phi_p$ correspond to the predictors $y_{t-1}, \dots, y_{t-p}$ (i.e., the past p values of the time series y_t) and the MA coefficients $\theta_1, \dots, \theta_q$ correspond to the predictors $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ (i.e., the past q shocks).

Due to stationarity, (77) holds and μ can be recovered from $\tilde{\beta}_0$ through $\mu = \tilde{\beta}_0 / (1 - \phi_1 - \dots - \phi_p)$.

Estimating ARMA models II

OLS estimation for stationary AR models

If no MA terms are present, then OLS estimation could be applied to (81). Only $n = T - p$ observations, i.e., $y_{p+1}, \dots, y_T$, are used as left-hand variables for estimation, because the predictor $\mathbf{x}_t = (y_{t-1}, \dots, y_{t-p})$ is observed only for $t \geq p + 1$.

OLS estimation is **biased**, because assumption **AX** is violated: ε_t is uncorrelated with all past regressors $\mathbf{x}_{t-h}$, $h \geq 0$, but correlated with future regressors $\mathbf{x}_{t+h}$, e.g., for AR(1) with $\mathbf{x}_{t+1} = y_t$:

$$E(\varepsilon_t \cdot \mathbf{x}_{t+1}) = E(\varepsilon_t \cdot y_t) = E(\varepsilon_t(\phi y_{t-1} + \varepsilon_t)) = \sigma_\varepsilon^2.$$

Estimating ARMA models III

Under the weaker assumption $\overline{\mathbf{A}\mathbf{X}}$: $E(\varepsilon_t \cdot \mathbf{x}_t) = 0$, i.e., ε_t is uncorrelated *with the contemporary regressor* $\mathbf{x}_t$, the OLS estimator is asymptotically, i.e., for large T , unbiased and consistent.

Asymptotically, the sampling distribution of the OLS estimator is Gaussian. The standard errors of the asymptotic normal distribution depend on the AR parameters, e.g.,

$$\text{AR(1): } \sqrt{n}(\hat{\phi} - \phi) \rightarrow_d \text{Normal}(0, (1 - \phi)^2)$$

$$\text{AR(2): } \sqrt{n}(\hat{\phi}_i - \phi_i) \rightarrow_d \text{Normal}(0, (1 - \phi_i)^2) \quad \text{for } i=1,2$$

Estimating ARMA models IV

Fitting an MA(q) process

In the MA(q) model

$$y_t = \varepsilon_{t-1}(\beta)\theta_1 + \cdots + \varepsilon_{t-q}(\beta)\theta_q + \varepsilon_t,$$

past errors serve as predictors for y_t , however, they depend on the parameters

$\beta = (\theta_1, \theta_2, \dots, \theta_q)$:

$$\varepsilon_t(\beta) = y_t - \theta_1\varepsilon_{t-1}(\beta) - \cdots - \theta_q\varepsilon_{t-q}(\beta).$$

As a consequence, the estimation problem is **nonlinear** and numerical optimization has to be applied.

Non-linear least square estimation for ARMA models

For each $t \geq p + 1$, predict $\hat{y}_t(\beta)$ from past observations and shocks:

$$\hat{y}_t(\beta) = \tilde{\beta}_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1}(\beta) + \cdots + \theta_q \varepsilon_{t-q}(\beta),$$

and compute the corresponding prediction error $\varepsilon_t(\beta)$:

$$\varepsilon_t(\beta) = y_t - \hat{y}_t(\beta).$$

In matrix notation, ε satisfies $\mathbf{A}\mathbf{H}$:

$$\mathbf{y} = \mathbf{X}(\beta)\beta + \varepsilon.$$

$\mathbf{X}(\beta)$ has $1 + p + q$ columns, where the last q columns depend on β .

As for OLS, the aggregated prediction error is determined as the sum of squared residuals SSR:

$$\text{SSR}(\beta) = \varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}(\beta)\beta)'(\mathbf{y} - \mathbf{X}(\beta)\beta).$$

The non-linear least square (NLS) estimator $\hat{\beta}$ of $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \mu)$ is based on minimizing $\text{SSR}(\beta)$. However, the estimation problem is **nonlinear**, because $\mathbf{X}(\beta)$ depends on β .

Taking the first derivatives of $\text{SSR}(\beta)$ does not lead to a linear system of equations to determine β as for OLS. The estimation problem is solved using numerical optimization.

Because $\mathbf{X}(\beta)$ depends on β , the information matrix $\partial^2 \text{SSR}(\beta) / \partial \beta^2$ changes with β .

If ε satisfies **AH** and $\overline{\mathbf{AX}}$, then the NLS estimator is asymptotically unbiased and consistent, asymptotically efficient, and the sampling distribution is asymptotically normal.

To derive the covariance matrix of the sampling distribution of the NLS estimator $\hat{\beta}$, various methods are common, e.g., evaluating the Hessian matrix at $\hat{\beta}$.

Partial autocorrelation function I

Additional information about the dependence structure is given by the correlation of Y_t and Y_{t-h} after **controlling for the effect of the intermediate values** $Y_{t-1}, \dots, Y_{t-h+1}$.

For the AR(1) process $Y_t = \phi Y_{t-1} + \varepsilon_t$, e.g., Y_t and Y_{t-2} are correlated, since $\rho(2) = \phi^2 > 0$. The correlation between Y_t and Y_{t-2} is not 0, because Y_t depends on Y_{t-2} through Y_{t-1} .

However, **conditional on knowing the previous observation** Y_{t-1} , the actual observation Y_t is independent of Y_{t-2} for an AR(1)-Model: the coefficient ϕ_2 of Y_{t-2} in the (overfitting) AR(2)-model is equal to 0:

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

Partial autocorrelation function II

The partial autocorrelation $\pi(h)$ at lag h is equal to the regression coefficient ϕ_{hh} of Y_{t-h} in the regression of Y_t on $Y_{t-1}, \dots, Y_{t-h}$:

$$Y_t = \phi_{h1}Y_{t-1} + \dots + \phi_{hh}Y_{t-h} + \varepsilon_t$$

The partial autocorrelations $\pi(h)$ as a function of the lag h are called partial autocorrelation function, **PACF**.

The PACF of an AR(p) process is zero for lags $h > p$, i.e., it 'cuts off' at lag p , since $\pi(h) = \phi_{hh} = 0$ for an overfitting AR(h) model.

The PACF of an MA(1) process decays exponentially, see AR(∞) representation in (73).

How to choose the order p and q of the AR and the MA term?

Box/Jenkins: analyze the empirical AC and PAC function

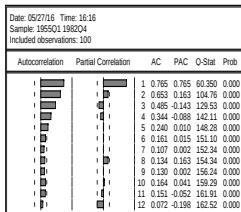
Process	ACF	PACF
MA(q)	cuts off at lag q	decays exponentially
AR(p)	decays exponentially	cuts off at lag p

Can be difficult in practice, if both AR and MA terms are present.

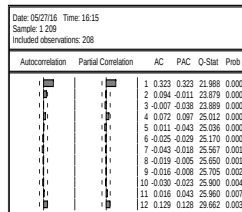
Order Selection from the empirical ACF

left: AR(1), because of $PACF(h)=0$, right: MA(1), because $ACF(h)=0$ for $h > 1$.

Correlogram of SAVE



Correlogram of D(AMEX)



Model Selection by Overfitting

Various null hypothesis for an ARMA(p,q)-model:

$H_0 : \phi_p = 0$	ARMA(p-1,q)-model
$H_0 : \theta_q = 0$	ARMA(p, q-1)-model
$H_0 : \phi_p = 0, \theta_q = 0$	ARMA(p-1, q-1)-model

A single parameter is tested using the t -statistic; r coefficients are tested simultaneously using the Wald statistic W or the F-statistic $F = W/r$.

Asymptotically under $\overline{\mathbf{A}\mathbf{X}}$: $t \rightarrow_d N(0, 1)$, $W \rightarrow_d \chi_r^2$, $F \rightarrow_d F(r, \text{df})$, with $\text{df} = n - (p + q + 1) = T - 2p - q - 1$.

Fit ARMA(p, q) model to time series data from an ARMA(p_0, q_0) model with $\phi_{p_0} \neq 0$ and $\theta_{q_0} \neq 0$ and consider the innovations

$$\varepsilon_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} - \tilde{\beta}_0.$$

- If $p \geq p_0$ and $q \geq q_0$, then ε_t is a zero-mean **white noise process** with constant variance $V(\varepsilon_t) = \sigma_\varepsilon^2$, i.e., $\rho(h) = 0, h > 0$, even if $p > p_0$ or $q > q_0$.
- If either $p < p_0$ or $q < q_0$ (or both), then due to the omitted variable bias, ε_t is not a **white noise process**, but correlated.

Check the empirical ACF of **estimated** residual $\hat{\varepsilon}_t = \varepsilon_t(\hat{\beta})$.

Model Selection Criteria

AIC and SC can be used to select the model order. For time series models, also the **Hannan-Quinn (HQ) criterion** is useful which has a stronger penalty than AIC, but less strong than SC:

$$AIC = \log(\text{SSR}) + 2 \cdot k,$$

$$HQ = \log(\text{SSR}) + 2 \log(\log(n)) \cdot k,$$

$$SC = \log(\text{SSR}) + \log n \cdot k,$$

where $n = T - p$ is the number of observations and $k = p + q + 1$ is the number of parameters (if a constant mean μ is present). Choose the model order that minimizes a particular criterion.

If **AH** is violated, because of correlated errors u_t , then u_t can be modelled by a stationary ARMA model with zero-mean white noise innovations ε_t :

$$\begin{aligned}y_t &= \mathbf{x}_t\boldsymbol{\beta} + u_t, & \phi(B)u_t &= \theta(B)\varepsilon_t, \\ \Leftrightarrow \phi(B)(y_t - \mathbf{x}_t\boldsymbol{\beta}) &= \theta(B)\varepsilon_t.\end{aligned}\tag{82}$$

- Covariance matrix $\boldsymbol{\Omega}$ of ε is derived from the ARMA structure.

Forecasting with ARMA models I

- Make statement about the process Y_t at a future date $t + k$ based on the information available at time t , i.e., $I_t = \{Y_t, Y_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots\}$. k is called the forecasting horizon.
- Point forecasts are based on the conditional expectation $\hat{Y}_{t+k|t} = E(Y_{t+k} | I_t)$ using the model specification. Their uncertainty is measured through $V(\hat{Y}_{t+k|t})$.
- Multi-step (or dynamic) forecasts derive $\hat{Y}_{t+1|t}, \hat{Y}_{t+2|t}, \dots$ for a fixed t . Static forecasts are sequential one step-ahead forecasts $\hat{Y}_{2|1}, \dots, \hat{Y}_{t|t-1}, \dots, \hat{Y}_{T|T-1}$.

Forecasting with ARMA models II

Dynamic AR(1) model forecasts

Dynamic forecasting of a stationary, non-zero mean AR(1) process

$$Y_{t+k} = \mu(1 - \phi) + \phi Y_{t+k-1} + \varepsilon_{t+k}$$

exploits $E(\varepsilon_{t+k} | I_t) = 0$:

$$\begin{aligned} \text{1-step} \quad & \hat{Y}_{t+1|t} = \mu(1 - \phi) + \phi Y_t \\ \text{2 steps} \quad & \hat{Y}_{t+2|t} = \mu(1 - \phi) + \phi \hat{Y}_{t+1|t} \\ & = \mu(1 - \phi)(1 + \phi) + \phi^2 Y_t \\ \text{k steps} \quad & \hat{Y}_{t+k|t} = \mu(1 - \phi)(1 + \phi + \dots + \phi^{k-1}) + \phi^k Y_t. \end{aligned}$$

Forecasting with ARMA models III

The forecast is a weighted average of the last observation Y_t and the long-run level μ :

$$\hat{Y}_{t+k|t} = \mu(1 - \phi) \frac{1 - \phi^k}{1 - \phi} + \phi^k Y_t = \mu(1 - \phi^k) + \phi^k Y_t,$$

$$\lim_{k \rightarrow \infty} \hat{Y}_{t+k|t} = \mu. \quad (83)$$

- With increasing forecasting horizon k , the weight ϕ^k converges to 0
- The closer $|\phi|$ is to one, the more persistent is the last value of Y_t .

Forecasting with ARMA models IV

MA model forecasts

Dynamic forecasting of an MA(q) process

$$Y_{t+k} = \mu + \theta_1 \varepsilon_{t+k-1} + \cdots + \theta_q \varepsilon_{t+k-q} + \varepsilon_{t+k}$$

$$\text{1-step} \quad \hat{Y}_{t+1|t} = \mu + \theta_1 \varepsilon_t + \cdots + \theta_q \varepsilon_{t+1-q}$$

$$\text{2-steps} \quad \hat{Y}_{t+2|t} = \mu + \theta_1 E(\varepsilon_{t+1} | I_t) + \theta_2 \varepsilon_t + \cdots + \theta_q \varepsilon_{t+2-q} =$$

$$\mu + \theta_2 \varepsilon_t + \cdots + \theta_q \varepsilon_{t+2-q}.$$

$$\text{k-steps} \quad \hat{Y}_{t+k|t} = \mu, \quad \text{for } k > q.$$

The forecast $\hat{Y}_{t+k|t}$ is equal to μ for $k > q$. For an estimated MA model, ε_{t-j} is substituted by the estimated shocks $\hat{\varepsilon}_{t-j}$ for $j \geq 0$.

ARMA model forecasts

Dynamic forecasts from an AR(p) or ARMA(p,q) model show a more complicated pattern.

Once the contribution from the MA part has vanished, the forecasts are driven by the AR part.

The Wold representation of $Y_t - \mu$, given by (78), implies convergence to the long run level μ , i.e., **mean reversion** for a stationary, invertible ARMA process:

$$\lim_{k \rightarrow \infty} \hat{Y}_{t+k|t} = \mu.$$

Hence, ARMA-process are useful mainly for short-term forecasting.

Forecasting with ARMA models VI

Properties of ARMA Forecasts

The point forecast $\hat{Y}_{t+1|t}$ is unbiased under the correct model specification:

$$E(Y_{t+k} - \hat{Y}_{t+1|t} \mid I_t) = 0.$$

Variance of the one-step-ahead forecast: all predictors are deterministic given I_t , hence (ignoring parameter uncertainty):

$$V(Y_{t+1} \mid I_t) = \sigma_\varepsilon^2.$$

The variance of the multi-step-ahead forecast results from the Wold representation (78):

$$V(Y_{t+k} \mid I_t) = \sigma_\varepsilon^2 \sum_{j=0}^{k-1} \psi_j^2.$$

Note that the variance increases with every forecasting step and converges to the unconditional variance σ^2 :

$$\lim_{k \rightarrow \infty} V(Y_{t+k} | I_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2 = V(Y_t) = \sigma^2.$$

E.g. AR(1)-process, where $\psi_j = \phi^j$:

$$V(Y_{t+k} | I_t) = \sigma_\varepsilon^2 (1 + \phi^2),$$

$$V(Y_{t+k} | I_t) = \sigma_\varepsilon^2 (1 + \phi^2 + \dots + \phi^{2(k-1)}) = \sigma_\varepsilon^2 \frac{1 - (\phi^2)^k}{1 - \phi^2} = \sigma^2 (1 - \phi^{2k}).$$

Interval Forecasts

An interval forecast is a predictive interval which covers the future value with probability $1 - \alpha$. Based on the assumption of normality of $Y_{t+k} - \hat{Y}_{t+k|t}$, predictive intervals can be derived from the expectation $\hat{Y}_{t+k|t}$ and the variance $V(Y_{t+k} | I_t)$:

$$\left[\hat{Y}_{t+k|t} - 1.96 \times \sqrt{V(Y_{t+k} | I_t)}, \hat{Y}_{t+k|t} + 1.96 \times \sqrt{V(Y_{t+k} | I_t)} \right].$$

All unknown AR and MA parameters, as well as the variance σ_ε^2 are substituted by estimates. This predictive interval does not account for parameter uncertainty.

ARIMA models

Time series Y_t follows an ARIMA(p, d, q) model, for some natural number d , if the time series $(I - B)^d Y_t$ follows an ARMA(p, q) model.

Note that $(I - B)$ applied to a time series affects differencing the time series once.

Hence, $(I - B)^d$ applied to a time series affects differencing the time series d times.

Differencing a time series d times removes polynomial time trends of order up to d (use induction).

Estimation strategy: difference, and then use methodology for ARMA models. Aggregate again for prediction.

For example, a random walk is an ARIMA(0,1,0) process.

If a time series is not weakly stationary, but becomes stationary after differencing it d times, we call it integrated of order d .

A time series that is integrated of order 1 is said to have a unit root.

Unit root tests (e.g., `adf.test` in R) can be applied to test for the presence of a unit root.

ARCH Model (Autoregressive Conditional Heteroskedasticity)

- ▶ Developed by Robert Engle in the 80s to model time-varying volatility in financial returns.
- ▶ Model for the variance of the error term:

$$x_t = \sigma_t \epsilon_t, \quad \epsilon_t \text{ iid and } N(0, 1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_q x_{t-q}^2$$

- ▶ Normality assumption not needed.
- ▶ Captures volatility clustering: large changes tend to follow large changes.
- ▶ Requires $\alpha_0 > 0$, $\alpha_i \geq 0$ for stability.

Easy to verify: $E(x_t \mid x_{t-1}, x_{t-2}, \dots) = 0$ and $E(x_t^2 \mid x_{t-1}, x_{t-2}, \dots) = \sigma_t^2$, hence "ARCH".

GARCH Model (Generalized ARCH)

- ▶ Proposed by Tim Bollerslev in the 80s as a generalization of ARCH.
- ▶ Adds lagged conditional variance terms:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i x_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

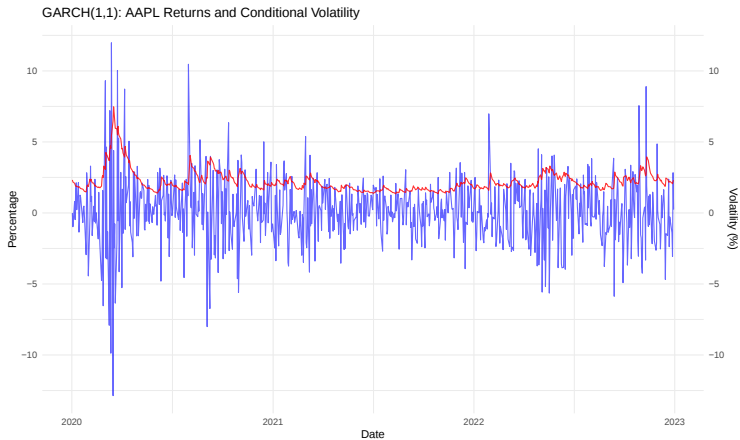
- ▶ GARCH(1,1) is most common:

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

- ▶ Captures both short-term shocks and long-term persistence in volatility.

Again: $E(x_t \mid x_{t-1}, x_{t-2}, \dots) = 0$ and $E(x_t^2 \mid x_{t-1}, x_{t-2}, \dots) = \sigma_t^2$.

In R via: fGarch package, garchFit function delivers



Models can be combined:

1. First fit a suitable ARIMA model.
2. Then, fit a GARCH model to the ARIMA residuals.

Results in ARIMA-GARCH models, which model conditional expectations and conditional volatilities.

Conditional expectation is forecast through the ARIMA part.

Conditional volatility is taken care for using the GARCH part.

Results in forecast intervals that incorporate both types of information.

More info: Financial Time Series Analysis, elective, Fall.