## Autoregressive and Moving average models

**Proposition 1.** *Let* $\{\epsilon_t\}$ *be a w.w.n. and* $\{\theta_i\}$ *a sequence of absolutely summable real numbers (i.e.* $\sum_{i=0}^{\infty} |\theta_i| < \infty$*), then*

1. $\forall t$*, the process* $y_{t,n} = \mu + \sum_{i=0}^{n} \theta_i \epsilon_{t-i}$ *is mean square convergent (with respect to* $n$*) and* $\{y_t\}$ *is weakly stationary.*

2. $\forall t$*, we have that*

$$E(y_t) = \mu \quad (1)$$

$$Var(y_t) = \sigma^2 \sum_{i=0}^{\infty} \theta_i^2 \quad (2)$$

$$Cov(y_t, y_{t-j}) = \sigma^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+j} \quad (3)$$

$$\sum_{j=0}^{\infty} |Cov(y_t, y_{t-j})| < \infty \quad (4)$$

3. *Let* $\{\epsilon_t\}$ *be an i.i.d. process then* $\{y_t\}$ *is ergodic and strictly stationary.*

*Proof.* To properly write down the proof, we need to know the two following theorems. The first theorem is an extension of the Cauchy criterion for random variables and it gives us a criterion to test if a random variable converges in mean square (without knowing its limiting random variable). The second theorem shows that if a sequence of random variable converges in mean square then, the first two moments of the limiting distributions coincide.

**Theorem 1.** *Let* $\{X_n\}_n$ *be a sequence of random variables with* $E(X_n^2) < \infty \ \forall n$ *then* $E[(X_m - X_n)^2] \to_{m.s.} 0 \ \forall m, n \to \infty$ *if and only if it exists a random variable* $X$ *such that* $X_n \to_{m.s} X$ *with* $E(X^2) < \infty$*.*

**Theorem 2.** *Let* $\{X_n\}_n$*,*$\{Y_m\}_m$ *be two sequences of random variables with* $E(X_n^2) < \infty \ \forall n$*,* $E(Y_m^2) < \infty \ \forall m$ *and* $X_n \to_{m.s} X$ *with* $E(X^2) < \infty$*,* $Y_m \to_{m.s.} Y$ *with* $E(Y^2) < \infty$ *then*

1. $\lim_{n\to\infty} E(X_n Y_n) = E(XY)$*.*

2. $\lim_{n\to\infty} E(X_n) = E(X)$*.*

3. $\lim_{n\to\infty} E(X_n X) = E(X^2)$*.*

*Proof.* 1.

$$\begin{aligned}
\lim_{n\to\infty} |E(X_nY_n) - E(XY)| &= \lim_{n\to\infty} |E((X_n - X)Y_n + X(Y_n - Y))| \\
&\leq \lim_{n\to\infty} [E(|(X_n - X)Y_n|) + E(|X(Y_n - Y)|)] \\
&\leq \lim_{n\to\infty} [\sqrt{E((X_n - X)^2)E(Y_n^2)} + \sqrt{E(X^2)E((Y_n - Y)^2)}] \\
&\leq 0
\end{aligned}$$

2. $\lim_{n\to\infty} E((X_n - X))^2 \leq \lim_{n\to\infty} E((X_n - X)^2) = 0$ (using Jensen's inequality).

3. $\lim_{n\to\infty} E(X_nX) = E(X^2)$, then $E(X(X_n - X)) \leq \sqrt{E((X_n - X)^2)E(X^2)}$.

□

We are ready to prove the MA($\infty$) proposition (1). Note first that $\sum_{i=1}^{\infty} |\theta_i| < \infty$ by assumption. The sequence of real number satisfies the Cauchy criterion (i.e. $\sum_{i=n+1}^{m} |\theta_i| \to 0$, $\forall m, n \to \infty$). Let us denote by $c$ the limiting value (i.e. $\sum_{i=0}^{\infty} |\theta_i| = c$). Then, we have that $(\sum_{i=0}^{\infty} |\theta_i|)^2 = c^2$ and that $\sum_{i=0}^{\infty} \theta_i^2 < (\sum_{i=0}^{\infty} |\theta_i|)^2 < \infty$. We conclude that $\{\theta_i^2\}$ converges and satisfies the Cauchy's criterion.

Let us define $y_{t,n} = \mu + \sum_{i=0}^{n} \theta_i \epsilon_{t-i}$. We first apply theorem 1 stated above:

$$\begin{aligned}
E[(y_{t,m} - y_{t,n})^2] &= E[(\sum_{i=n+1}^{m} \theta_i \epsilon_{t-i})^2] \\
&= \sigma^2 \sum_{i=n+1}^{m} \theta_i^2 \\
&\to 0 \quad \forall m, n \to \infty.
\end{aligned}$$

We conclude that $y_t \to_{m.s.} y_t$ and that $E(y_t^2) < \infty$. However, we do not know the distribution of $y_t$ (we can show that it is unique using Cauchy-Swartz inequality). We now apply theorem

2 to take the limit outside the expectation and we get

$$
\begin{aligned}
E[y_t] &= \lim_{n\to\infty} E(\mu + \sum_{i=0}^{n} \theta_i \epsilon_{t-i}) \\
&= \mu \\
Cov(y_t, y_{t-j}) &= \lim_{n\to\infty} E[(y_{t,n} - \mu)(y_{t-j,n} - \mu)] \\
&= \lim_{n\to\infty} E[(\sum_{i=0}^{n} \theta_i \epsilon_{t-i})(\sum_{i=0}^{n} \theta_i \epsilon_{t-j-i}) \\
&= \lim_{n\to\infty} \sigma^2 \sum_{i=0}^{n} \theta_{i+j}\theta_i \\
V(y_t, y_{t-j}) &= \sigma^2 \sum_{i=0}^{\infty} \theta_i^2.
\end{aligned}
$$

The stochastic process $\{y_t\}$ is thus weakly stationary. We need to show that the covariance coefficients are absolutely summable (i.e., $\sum_{j=0}^{\infty} |Cov(y_t, y_{t-j})| < \infty$).

$$
\begin{aligned}
\sum_{j=0}^{\infty} |Cov(y_t, y_{t-j})| &= \sigma^2 \sum_{j=0}^{\infty} |\sum_{i=0}^{\infty} \theta_{i+j}\theta_i| \\
&\leq \sigma^2 \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} |\theta_{i+j}||\theta_i| \\
&= \sigma^2 \sum_{i=0}^{\infty} |\theta_i| \sum_{j=0}^{\infty} |\theta_{i+j}| \\
&\leq c\sigma^2 \sum_{i=0}^{\infty} |\theta_i| \\
&\leq \sigma^2 c^2 \quad (< \infty).
\end{aligned}
$$

□

**Proposition 2.** *The moments of an AR(1) process:* $y_t = c + \beta y_{t-1} + \epsilon_t$ *with* $\epsilon_t \sim w.w.n(0, \sigma^2)$ *are given by*

$$
\begin{aligned}
E(y_t) &= \frac{c}{1-\beta}, && (5) \\
Var(y_t) &= \frac{\sigma^2}{1-\beta^2}, && (6) \\
Cov(y_t, y_{t-j}) &= \frac{\sigma^2 \beta^j}{1-\beta^2}, && (7) \\
Corr(y_t, y_{t-j}) &= \beta^j. && (8)
\end{aligned}
$$

*Proof.* Let us first remind (and prove) this standard result on sequences of real numbers:

**Theorem 3.** *If $a \in [-1,1]$ then $\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$.*

*Proof.* Note that the sequence of real numbers $\{a^i\}$ converges when $a \in [-1,1]$ since it satisfies the Cauchy's criterion: $\sum_{i=m+1}^{n} a^i \to 0$ when $m,n \to \infty$. We now find the limiting value as follows,

$$\begin{aligned}
\sum_{i=0}^{\infty} a^i &= 1 + \sum_{i=1}^{\infty} a^i, \\
&= 1 + \sum_{i=0}^{\infty} a^{i+1}, \\
&= 1 + a\sum_{i=0}^{\infty} a^i, \\
\sum_{i=0}^{\infty} a^i(1-a) &= 1, \\
\sum_{i=0}^{\infty} a^i &= \frac{1}{1-a}.
\end{aligned}$$

□

By recursive reasoning, we get that $y_t = c\sum_{i=0}^{\infty}\beta^i + \sum_{i=0}^{\infty}\beta^i\epsilon_{t-i}$. The process stands for a MA($\infty$) process. Proposition 1 applies if $\sum_{i=0}^{\infty}|\beta^i| < \infty$. If $|\beta| < 1$ then the series is convergent. In fact, the coefficients are absolutely summable since $\sum_{i=0}^{\infty}|\beta^i| \leq \sum_{i=0}^{\infty}(|\beta|)^i < \frac{1}{1-|\beta|} < \infty$. Applying proposition 1, we get

$$\begin{aligned}
E(y_t) &= c\sum_{i=0}^{\infty}\beta^i \quad (= \frac{c}{1-\beta}), && (9) \\
Var(y_t) &= \sigma^2\sum_{i=0}^{\infty}(\beta^2)^i \quad (= \frac{\sigma^2}{1-\beta^2}), && (10) \\
Cov(y_t, y_{t-j}) &= \sigma^2\sum_{i=0}^{\infty}\beta^i\beta^{i+j} \quad (= \frac{\sigma^2\beta^j}{1-\beta^2}), && (11) \\
Corr(y_t, y_{t-j}) &= \beta^j. && (12)
\end{aligned}$$

□

**Proposition 3.** *The expectation of an AR(p) process given by $y_t = c + \sum_{i=1}^{p}\phi_i y_{t-i} + \epsilon_t$ is equal to $E(y_t) = \frac{c}{1-\sum_{i=1}^{p}\phi_i}$ if the AR(p) process is stationary.*

*Proof.* We propose two different proofs of the proposition.

1. First, because the AR(p) process is stationary, we know that $E(y_t) = E(y_{t-j}) < \infty$ for any $j \in Z$. Taking the expectation of the AR(p) model, we get $E(y_t) = c + \sum_{i=1}^{p} \phi_i E(y_t)$. This gives $E(y_t)(1 - \sum_{i=1}^{p} \phi_i) = c$ which proves the proposition.

2. Using proposition 1, we know that $E(y_t) = c\Psi(1)$ in which $\Psi(L)\phi_p(L) = \Psi(L)(1 - \sum_{i=1}^{p} \phi_i L^i) = 1$. The inverse lag polynomial is given by $\Psi(L) = (1 - \sum_{i=1}^{p} \phi_i L^i)^{-1} = \prod_{i=1}^{p}(1 - \lambda_i L)^{-1}$ where $|\lambda_i| < 1$ are the roots of the characteristic polynomial. From the AR(1) model, we know that $(1-\lambda_i L)^{-1} = \sum_{k=0}^{\infty} \lambda_i^k = \frac{1}{1-\lambda_i}$ if $|\lambda_i| < 1$. We conclude that

$$E(y_t) \equiv c\Psi(1) = c\prod_{i=1}^{p}(\sum_{k=0}^{\infty} \lambda_i^k), \quad (13)$$

$$= c\prod_{i=1}^{p} \frac{1}{1-\lambda_i}, \quad (14)$$

$$= \frac{c}{(1-\lambda_1)\dots(1-\lambda_p)}, \quad (15)$$

$$= \frac{c}{\phi_p(1)}, \quad (16)$$

$$= \frac{c}{1 - \sum_{i=1}^{p} \phi_i}. \quad (17)$$

□

**Proposition 4.** ***Invertibility of a MA(1) model.*** *Let a $M\bar{A}(1)$ process be specified as $y_t = \mu + \bar{\epsilon}_t + \bar{\theta}\bar{\epsilon}_{t-1}$ with $|\bar{\theta}| > 1$ and $\bar{\epsilon}_t \sim i.i.d.(0, \sigma^2)$ then it exists another MA(1) process: $y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$ with $|\theta| < 1$, such that the two first moments of the processes match.*

*Proof.* The moments of the $M\bar{A}(1)$ process are given by:

$$E(y_t) = \mu, \quad (18)$$

$$V(y_t) = \sigma^2(1+\bar{\theta}^2), \quad (19)$$

$$Cov(y_t, y_{t-1}) = \sigma^2\bar{\theta}. \quad (20)$$

Let us set $\theta = \frac{1}{\bar{\theta}}$ and $\epsilon_t = \bar{\theta}\bar{\epsilon}_t$. With this change of variables, the MA(1) process is $y_t = \mu + \bar{\theta}\bar{\epsilon}_t + \bar{\epsilon}_{t-1}$. The first two moments of this MA(1) process are given by: $E(y_t) = \mu$, $V(y_t) = E[(\bar{\theta}\bar{\epsilon}_t + \bar{\epsilon}_{t-1})^2] = \bar{\theta}^2\sigma^2 + \sigma^2 = \sigma^2(1+\bar{\theta}^2)$. The covariance is also equivalent to the

$M\bar{A}(1)$ process since we have

$$Cov(y_t, y_{t-1}) = E[(\bar{\theta}\bar{\epsilon}_t + \bar{\epsilon}_{t-1})(\bar{\theta}\bar{\epsilon}_{t-1} + \bar{\epsilon}_{t-2})], \tag{21}$$

$$= E[\bar{\theta}(\bar{\epsilon}_{t-1})^2] \ \ (= \bar{\theta}\sigma^2). \tag{22}$$

□

**Proposition 5.** *Let us assume an ARMA(1,1) process specified as* $y_t = \alpha + \phi y_{t-1} + \theta\epsilon_{t-1} + \epsilon_t$ *with* $|\phi| < 1$*. Then, we have that*

$$E(y_t) = \frac{\alpha}{1-\phi}, \tag{23}$$

$$V(y_t) = \sigma^2 \frac{1+\theta^2+2\theta\phi}{1-\phi^2}, \tag{24}$$

$$\gamma_1 \equiv Cov(y_t, y_{t-1}) = \sigma^2[\frac{\theta - \theta\phi^2 + 1 + \theta^2 + 2\theta\phi}{1-\phi^2}], \tag{25}$$

$$\gamma_j = \phi^{j-1}\gamma_1, \ \textit{for } j > 1. \tag{26}$$

*Proof.* The MA($\infty$) representation of the ARMA(1,1) process is given by

$$y_t = \frac{\alpha}{1-\phi} + \epsilon_t + \sum_{i=1}^{\infty} \phi^{i-1}(\theta+\phi)\epsilon_{t-i}. \tag{27}$$

Using proposition 1 on MA($\infty$) processes, we know that

$$E(y_t) = \frac{\alpha}{1-\phi} \tag{28}$$

$$Var(y_t) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2, \tag{29}$$

$$Cov(y_t, y_{t-j}) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j}, \tag{30}$$

$$\sum_{j=0}^{\infty} |Cov(y_t, y_{t-j})| < \infty, \tag{31}$$

with $\psi_0 = 1$ and $\psi_i = \phi^{i-1}(\theta+\phi)$ for $i > 0$. Using these results, we still have to compute the variance and the covariance of the process.

$$\sum_{i=0}^{\infty}\psi_i^2 = 1+\sum_{i=1}^{\infty}\phi^{2(i-1)}(\theta+\phi)^2 \tag{32}$$

$$= 1+(\theta+\phi)^2+(\theta+\phi)^2\sum_{i=1}^{\infty}(\phi^2)^i \tag{33}$$

$$= 1+(\theta+\phi)^2\sum_{i=0}^{\infty}(\phi^2)^i \tag{34}$$

$$= 1+\frac{(\theta+\phi)^2}{1-\phi^2} \tag{35}$$

$$= \frac{1+\theta^2+2\theta\phi}{1-\phi^2}. \tag{36}$$

We conclude that $Var(y_t) = \sigma^2\sum_{i=0}^{\infty}\psi_i^2 = \sigma^2\frac{1+\theta^2+2\theta\phi}{1-\phi^2}$. Regarding the covariance, we first apply the Yule-Walker result to get that $\gamma_j = \phi\gamma_{j-1} = \phi^{j-1}\gamma_1$. If you do not want to use the Yule-Walker result, we can just prove it as follows. Since $\psi_i = \phi\psi_{i-1}\ \forall i > 1$, we have that

$$\gamma_j \equiv \sigma^2\sum_{i=0}^{\infty}\psi_i\psi_{i+j} = \sigma^2[\phi\sum_{i=0}^{\infty}\psi_i\psi_{i+j-1}]\ \ \forall j>1, \tag{37}$$

$$\gamma_j = \phi\gamma_{j-1} \tag{38}$$

$$\rho_j = \phi\rho_{j-1} \tag{39}$$

$$= \phi^{j-1}\rho_1 \tag{40}$$

We still need to compute the first order covariance. It is given by

$$\gamma_1 \equiv \sigma^2\sum_{i=0}^{\infty}\psi_i\psi_{i+1} = \sigma^2[(\theta+\phi)+\phi\sum_{i=1}^{\infty}\psi_i\psi_i], \tag{41}$$

$$= \sigma^2[\theta+\phi\sum_{i=0}^{\infty}\psi_i^2] \tag{42}$$

$$= \sigma^2[\frac{\theta-\theta\phi^2+1+\theta^2+2\theta\phi}{1-\phi^2}]. \tag{43}$$

□

**Proposition 6.** *Let us assume an AR(p) process specified as* $y_t = \beta_0+\beta_1 y_{t-1}+\ldots+\beta_p y_{t-p}+\epsilon_t$ *with* $\epsilon_t \sim i.i.d.(0,\sigma^2)$. *Let us also assume that the lag polynomial* $\beta_p(L) = (1-\sum_{i=1}^{p}\beta_i L^i)$ *is stable. Then, the OLS estimator given by* $\hat{\beta} = (X'X)^{-1}X'y$ *with* $X =$

$\begin{pmatrix} 1 & y_p & \dots & y_1 \\ 1 & y_{p+1} & \dots & y_2 \\ \dots & & \dots & \\ 1 & y_{T-1} & \dots & y_{T-p-1} \end{pmatrix}$, $y = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \dots \\ y_T \end{pmatrix}$ *is consistent. In addition, it is also asymptotically normal with*

$$\sqrt{T}(\hat{\beta} - \beta) \quad \to_d \quad N(0, \sigma^2 E(x_t x_t')),$$

*where* $x_t = (1, y_{t-1}, \dots, y_{t-p})'$.

*Proof.* To prove this proposition, we just need to check all the linear regression assumptions of the asymptotic framework.

1. **Linearity:** The process is linear in $y_t$.

2. $\{y_t\}$ is a stationary and ergodic process due to proposition 1.

3. **Exogeneity:** Weak exogeneity implies that $E(\epsilon_t|y_{t-1}, \dots, y_{t-p}) = 0$. Since $\epsilon_t$ is i.i.d. and $y_{t-i} = f(\epsilon_{t-i}, \epsilon_{t-i-1}, ...)$ using the $MA(\infty)$ representation, we have that $E(\epsilon_t|y_{t-1}, \dots, y_{t-p}) = E(\epsilon_t|f(\epsilon_{t-1}, \epsilon_{t-2}, ...), \dots, f(\epsilon_{t-p}, \epsilon_{t-p-1}, ...)) = E(\epsilon_t) = 0$.

4. **Invertibility:** $E(x_t x_t')$ exists and is not singular:

$$E(x_t x_t') = E(\begin{pmatrix} 1 & y_{t-1} & y_{t-2} & \dots & y_{t-p} \\ y_{t-1} & y_{t-1}^2 & y_{t-1}y_{t-2} & \dots & y_{t-1}y_{t-p} \\ & & & & \\ y_{t-p} & y_{t-p}y_{t-1} & y_{t-p}y_{t-2} & \dots & y_{t-p}^2 \end{pmatrix}) \quad (44)$$

$$= \begin{pmatrix} 1 & E(y_{t-1}) & E(y_{t-2}) & \dots & E(y_{t-p}) \\ E(y_{t-1}) & E(y_{t-1}^2) & E(y_{t-1}y_{t-2}) & \dots & E(y_{t-1}y_{t-p}) \\ & & & & \\ E(y_{t-p}) & E(y_{t-p}y_{t-1}) & E(y_{t-p}y_{t-2}) & \dots & E(y_{t-p}^2) \end{pmatrix} \quad (45)$$

$$= \begin{pmatrix} 1 & \mu & \mu & \dots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \dots & \gamma_{p-1} + \mu^2 \\ & & & & \\ \mu & \gamma_{p-1} + \mu^2 & \gamma_{p-2} + \mu^2 & \dots & \gamma_0 + \mu^2 \end{pmatrix} \quad (46)$$

$$(47)$$

The determinant of $E(x_t x_t')$ can be shown to be strictly positive. We show it for an AR(1) process which corresponds to the matrix $V_{AR1} = \begin{pmatrix} 1 & \mu \\ \mu & \gamma_0 + \mu^2 \end{pmatrix}$. The determinant is given by $|V_{AR1}| = \gamma_0 = V(y_t) > 0$. Thus, we conclude that $E(x_t x_t')$ is invertible.

5. **Martingale difference sequence:** $\{g_t\} = \{\epsilon_t x_t\}$ should be a martingale difference sequence. Since $\epsilon_t$ is i.i.d., we know that $E(\epsilon_t | \epsilon_{t-1}, ..., y_{t-2}\epsilon_{t-1}, ...) = 0$. In addition, $E(y_{t-i}\epsilon_t | \epsilon_{t-1}, ..., y_{t-2}\epsilon_{t-1}, ...) = y_{t-i} E(\epsilon_t | \epsilon_{t-1}, ..., y_{t-2}\epsilon_{t-1}, ...) = 0$ because $y_{t-i} = f(\epsilon_{t-i}, \epsilon_{t-i-1}, ...)$ for $i > 0$.

6. **Homoskedasticity:** Using the fact that $\epsilon_t$ is i.i.d., we have that $E(\epsilon_t^2 | x_t) = E(\epsilon_t^2) = \sigma^2$.

□