Introduction to the ML Estimation of ARMA processes

Eduardo Rossi

University of Pavia

We consider the AR(p) model:

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \quad t = 1, \ldots, T$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

where $y_0, y_{-1}, \ldots, y_{1-p}$ are given. Notation as a regression model

$$y_t = \mathbf{z}_t' \boldsymbol{\theta} + \varepsilon_t$$

with $\boldsymbol{\theta} = (c, \phi_1, \ldots, \phi_p)'$ and $\mathbf{z}_t = (1, y_{t-1}, \ldots, y_{t-p})'$:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & y_0 & \cdots & y_{1-p} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & y_{T-1} & \cdots & y_{T-p} \end{bmatrix} \begin{bmatrix} c \\ \theta_1 \\ \vdots \\ \theta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

The model is:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

The OLS estimator:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\
&= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\theta} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} \\
&= \boldsymbol{\theta} + \left(\frac{1}{T}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\frac{1}{T}\mathbf{Z}'\boldsymbol{\varepsilon}\right)
\end{aligned}$$

- OLS is no longer linear in y.
- Hence cannot be BLUE. In general OLS in no more unbiased.
- Small sample properties are analytically difficult to derive.

If Y_t is a stable AR(p) process and ε_t is a standard white noise, then the following results hold (Mann and Wald, 1943):

$$\frac{1}{T}(\mathbf{Z}'\mathbf{Z}) \xrightarrow{p} \mathbf{\Gamma}$$

$$\sqrt{T}\left(\frac{1}{T}\mathbf{Z}'\boldsymbol{\varepsilon}\right) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{\Gamma})$$

then consistency and asymptotic normality follows from Cramer's theorem:

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{\Gamma}^{-1})$$

Necessary condition for the consistency of OLS estimator with stochastic (but stationary) regressors is that z_t is asymptotically uncorrelated with ε_t, i.e. plim $\left(\frac{1}{T}\mathbf{Z}'\varepsilon_t\right) = 0$:

$$\begin{aligned} \text{plim}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} &= \text{plim}\left(\frac{1}{T}\mathbf{Z}'\mathbf{Z}\right)^{-1} \text{plim}\left(\frac{1}{T}\mathbf{Z}'\boldsymbol{\varepsilon}\right) \\ &= \boldsymbol{\Gamma}^{-1}\text{plim}\left(\frac{1}{T}\mathbf{Z}'\boldsymbol{\varepsilon}\right) \end{aligned}$$

OLS is no longer consistent under autocorrelation of the regression error as

$$\text{plim}\left(\frac{1}{T}\mathbf{Z}'\boldsymbol{\varepsilon}\right) \neq 0$$

Consider an AR(1) model with first-order autocorrelation of its errors

$$\begin{aligned} Y_t &= \phi Y_{t-1} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

such that $\mathbf{Z}' = [Y_0, \ldots, Y_{T-1}]$. Then

$$E\left[\frac{1}{T}\mathbf{Z}'u\right] = E\left[\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}u_t\right] = \frac{1}{T}\sum_{t=1}^{T} E[Y_{t-1}(\rho(Y_{t-1}-\phi Y_{t-2})+\varepsilon_t)]$$

since

$$u_t = \rho(Y_{t-1} - \phi Y_{t-2}) + \varepsilon_t$$

$$
\begin{aligned}
E\left[\frac{1}{T}\mathbf{Z}'u\right] &= \rho\left(\frac{1}{T}\sum_{t=1}^{T}E[Y_{t-1}^2]\right) - \phi\rho\left(\frac{1}{T}\sum_{t=1}^{T}E[Y_{t-1}Y_{t-2}]\right) + \\
&\quad \left(\frac{1}{T}\sum_{t=1}^{T}E[Y_{t-1}\varepsilon_t]\right) \\
&= \rho\left[\gamma_y(0) - \phi\gamma_y(1)\right]
\end{aligned}
$$

where $\gamma_y(h)$ is the autocovariance function of $\{Y_t\}$ which can be represented as an AR(2) process.

For the Gaussian AR(1) process,

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \quad |\phi| < 1$$

$$\varepsilon_t \sim NID(0, \sigma^2)$$

the joint distribution of

$$\mathbf{Y}_T = (Y_1, \ldots, Y_T)'$$

is

$$\mathbf{Y}_T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

the observations

$$\mathbf{y} \equiv (y_1, y_2, \ldots, y_T)$$

are the single realization of $\mathbf{Y}_T$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix} \sim N(\mu, \mathbf{\Sigma})$$

$$\mu = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} \qquad \mathbf{\Sigma} = \begin{bmatrix} \gamma_0 & \cdots & \gamma_{T-1} \\ \vdots & \ddots & \vdots \\ \gamma_{T-1} & \cdots & \gamma_0 \end{bmatrix}$$

The p.d.f. of the sample $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$ is given by the multivariate normal density

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{T}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \right\}$$

Denoting $\boldsymbol{\Sigma} = \sigma_y^2 \boldsymbol{\Omega}$ with $\boldsymbol{\Omega}_{ij} = \phi^{|i-j|}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \gamma_0 & \cdots & \gamma_{T-1} \\ \vdots & \ddots & \vdots \\ \gamma_{T-1} & \cdots & \gamma_0 \end{bmatrix} = \gamma_0 \begin{bmatrix} 1 & \cdots & \frac{\gamma_{T-1}}{\gamma_0} \\ \vdots & \ddots & \vdots \\ \frac{\gamma_{T-1}}{\gamma_0} & \cdots & 1 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \sigma_y^2 \boldsymbol{\Omega} = \sigma_y^2 \begin{bmatrix} 1 & \cdots & \rho(T-1) \\ \vdots & \ddots & \vdots \\ \rho(T-1) & \cdots & 1 \end{bmatrix}$$

$$\rho(j) = \phi^j$$

Collecting the parameters of the model in $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$, the joint p.d.f. becomes:

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) = (2\pi\sigma_y^2)^{-\frac{T}{2}} |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma_y^2}(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \right\}$$

Collecting the parameters of the model in $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$, the sample log-likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma_y^2) - \frac{1}{2}\log(|\boldsymbol{\Omega}|) - \frac{1}{2\sigma_y^2}(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

Sequential Factorization

The **prediction-error decomposition** uses the fact that the ε_t are independent, identically distributed:

$$f(\varepsilon_2, \ldots, \varepsilon_T) = \prod_{t=2}^{T} f_\varepsilon(\varepsilon_t).$$

and:

$$g_{\mathbf{Y}}(y_T, \ldots, y_1) = \left[\prod_{t=2}^{T} g_{Y_t|Y_{t-1}}(y_t|y_{t-1})\right] \times g_{Y_1}(y_1)$$

by the Markov property. We assume that the marginal density of Y_1 is that of ε_1 with

$$E[Y_1] = \mu = \frac{c}{1-\phi}$$

$$E[(Y_1-\mu)^2] = \sigma_y^2 = \frac{\sigma^2}{1-\phi^2}$$

Since

$$\varepsilon_t = Y_t - (c + \phi Y_{t-1})$$

then

$$g_{Y_t|Y_{t-1}}(y_t|y_{t-1}) = f_\varepsilon(y_t|y_{t-1}) = f_\varepsilon(\varepsilon_t) \quad t = 2, \ldots, T$$

Hence:

$$g_{\mathbf{Y}}(y_T, \ldots, y_1) = \left\{ \left[\prod_{t=2}^{T} f_\varepsilon(y_t | y_{t-1}) \right] f_\varepsilon(y_1) \right\}$$

For $\varepsilon_t \sim NID(0, \sigma^2)$, the log-likelihood is given by:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) \\
&= \sum_{t=2}^{T} \log f_\varepsilon(y_t | y_{t-1}; \boldsymbol{\theta}) + \log f_y(y_1; \boldsymbol{\theta}) \\
&= -\frac{T}{2} \log(2\pi) - \left\{ \frac{T-1}{2} \log(\sigma^2) + \frac{1}{2\sigma^2} \sum_{t=2}^{T} \varepsilon_t \right\} \\
&\quad - \left\{ \frac{1}{2} \log(\sigma_y^2) + \frac{1}{2\sigma_y^2} (y_1 - \mu)^2 \right\}
\end{aligned}$$

where $Y_1 \sim N(\mu, \sigma_y^2)$ with $\mu = \frac{c}{1-\phi^2}$ and $\sigma_y^2 = \frac{\sigma^2}{1-\phi^2}$.

Maximization of the exact log likelihood for an AR(1) process must be accomplished numerically.

Gaussian AR(p):

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma^2)$$

$$\boldsymbol{\theta} = (c, \phi_1, \ldots, \phi_p, \sigma^2)'$$

Exact MLE

Using the prediction-error decomposition, the joint p.d.f is given by:

$$f_{\mathbf{Y}}(y_1, y_2, \ldots, y_T; \boldsymbol{\theta}) = \left[\prod_{t=p+1}^{T} f_{\varepsilon}(y_t|\mathbf{y}_{t-1}; \boldsymbol{\theta}) \right] f_{Y_1,\ldots,Y_p}(y_1, \ldots, y_p; \boldsymbol{\theta})$$

but only the p most recent observations matter

$$f_{\varepsilon}(y_t|\mathbf{y}_{t-1}; \boldsymbol{\theta}) = f_{\varepsilon}(y_t|y_{t-1}, \ldots, y_{t-p}; \boldsymbol{\theta})$$

The likelihood function for the complete sample is:

$$f_{\mathbf{Y}}(y_1, y_2, \ldots, y_T; \boldsymbol{\theta}) = \left[\prod_{t=p+1}^{T} f_{\varepsilon}(y_t | y_{t-1}, \ldots, y_{t-p}; \boldsymbol{\theta})\right] f_y(y_1, \ldots, y_p; \boldsymbol{\theta})$$

With $\varepsilon_t \sim NID(0, \sigma^2)$

$$f_{\varepsilon}(y_t | y_{t-1}, \ldots, y_{t-p}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_t - c - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p})^2}{2\sigma^2}\right].$$

The first p observations are viewed as the realization of a p-dimensional Gaussian variable with moments:

$$E(\mathbf{Y}_p) = \boldsymbol{\mu}_p$$

$$E\left[(\mathbf{Y}_p - \boldsymbol{\mu}_p)(\mathbf{Y}_p - \boldsymbol{\mu}_p)'\right] = \boldsymbol{\Sigma}_p$$

$$\mathbf{\Sigma}_p = \sigma^2 \mathbf{V}_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix}$$

$$f_y(y_1, \ldots, y_p; \boldsymbol{\theta}) = (2\pi)^{-\frac{p}{2}} |\sigma^{-2}\mathbf{V}_p^{-1}|^{\frac{1}{2}} \exp\left[-\frac{(\mathbf{Y}_p - \boldsymbol{\mu}_p)'\mathbf{V}_p^{-1}(\mathbf{Y}_p - \boldsymbol{\mu}_p)}{2\sigma^2} \right]$$

The log-likelihood is:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \log f_{\mathbf{Y}}(y_1, y_2, \ldots, y_T; \boldsymbol{\theta}) \\
&= \sum_{t=p+1}^{T} \log f_{\varepsilon}(y_t | y_{t-1}, \ldots, y_{t-p}; \boldsymbol{\theta}) + \log f_y(y_1, \ldots, y_p; \boldsymbol{\theta}) \\
&= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) + \frac{1}{2} \log |\mathbf{V}_p^{-1}| \\
&\quad - \frac{1}{2\sigma^2} (\mathbf{Y}_p - \boldsymbol{\mu}_p)' \mathbf{V}_p^{-1} (\mathbf{Y}_p - \boldsymbol{\mu}_p) \\
&\quad - \sum_{t=p+1}^{T} \frac{(y_t - c - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p})^2}{2\sigma^2}
\end{aligned}
$$

The exact MLE follows from:

$$
\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})
$$

Conditional MLE = OLS

Take $\mathbf{y}_p = (y_1, \ldots, y_p)'$ as fixed pre-sample values

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} f_{\mathbf{Y}_{p+1},\ldots,\mathbf{Y}_T|\mathbf{Y}_1,\ldots,\mathbf{Y}_p}(y_{p+1}, \ldots, y_T|\mathbf{y}_p; \boldsymbol{\theta})$$

$$= \arg\max_{\boldsymbol{\theta}} \prod_{t=p+1}^{T} f_{\varepsilon}(y_t|y_{t-1}, \ldots, y_{t-p}; \boldsymbol{\theta})$$

Conditioning on $\mathbf{Y}_p$:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \log f_{\mathbf{Y}_{p+1},\ldots,\mathbf{Y}_T|\mathbf{Y}_1,\ldots,\mathbf{Y}_p}(y_{p+1}, \ldots, y_T|\mathbf{y}_p; \boldsymbol{\theta}) \\
&= \sum_{t=p+1}^{T} \log f_{\varepsilon}(\varepsilon_t|\mathbf{Y}_{t-1}; \boldsymbol{\theta}) \\
&= -\frac{T-p}{2}\log(2\pi) - \frac{T-p}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t+p}^{T}\varepsilon_t^2
\end{aligned}$$

where

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

Thus the MLE of $(c, \phi_1, \ldots, \phi_p)$ results by minimizing the sum of squared residuals:

$$\arg \max_{(c,\phi_1,\ldots,\phi_p)} \mathcal{L}(c, \phi_1, \ldots, \phi_p) = \arg \min_{(c,\phi_1,\ldots,\phi_p)} \sum_{t=p+1}^{T} \varepsilon_t^2(c, \phi_1, \ldots, \phi_p)$$

The conditional ML estimate of σ^2 turns out to be:

$$\widehat{\sigma}^2 = \frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{\varepsilon}_t^2$$

- The ML estimates $\widetilde{\gamma} = (\widetilde{c}, \widetilde{\phi_1}, \ldots, \widetilde{\phi_p})'$ are equivalent to OLS estimates.
- $(\widetilde{c}, \widetilde{\phi_1}, \ldots, \widetilde{\phi_p})$ are consistent estimators if $\{Y_t\}$ is stationary and $\sqrt{T}(\widetilde{\gamma} - \gamma)$ is asymptotically normally distributed.
- The exact ML estimates and the conditional ML estimates have the same large-sample distribution.

Asymptotically equivalent

- MLE of the mean-adjusted model

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t$$

  where $\mu = (1 - \phi_1 - \ldots - \phi_p)^{-1}c$.

- OLS of $(\phi_1, \ldots, \phi_p)$ in the mean adjusted model, where

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

- Yule-Walker estimation of $(\phi_1, \ldots, \phi_p)$

$$\begin{bmatrix} \widehat{\phi}_1 \\ \vdots \\ \widehat{\phi}_p \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_0 & \cdots & \widehat{\gamma}_{p-1} \\ \vdots & \ddots & \cdots \\ \widehat{\gamma}_{p-1} & \cdots & \widehat{\gamma}_0 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\gamma}_1 \\ \vdots \\ \widehat{\gamma}_p \end{bmatrix}$$

where

$$\widehat{\gamma}_h = (T-h)^{-1} \sum_{t=h+1}^{T} (y_t - \overline{y})(y_{t-h} - \overline{y})$$

and

$$\widehat{\mu} = \overline{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

Gaussian MA(q):

$$Y_t = \mu + \varepsilon_t + \boldsymbol{\theta}_1 \varepsilon_{t-1} + \ldots + \boldsymbol{\theta}_q \varepsilon_{t-q}$$

$$\varepsilon_t \sim NID(0, \sigma^2)$$

Conditional MLE = NLLS

Conditioning on $\underline{\varepsilon}_0 = (\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{1-q})' = \mathbf{0}$, we can iterate on:

$$\varepsilon_t = Y_t - (\boldsymbol{\theta}_1 \varepsilon_{t-1} + \ldots + \boldsymbol{\theta}_q \varepsilon_{t-q})$$

for $t = 1, \ldots, T$. The conditional likelihood is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \log f_{\mathbf{Y}_T | \underline{\varepsilon}_0 = \mathbf{0}}(\mathbf{y}_T | \underline{\varepsilon}_0 = \mathbf{0}; \boldsymbol{\theta}) \\ &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \end{aligned}$$

where $\boldsymbol{\theta} = (\mu, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_q, \sigma^2)$.

- The MLE of $(\mu, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_q)$ results by minimizing the sum of squared residuals.
- Analytical expressions for MLE are usually not available due to highly non-linear FOCs.
- MLE requires to apply numerical optimization techniques.

- Conditioning requires invertibility, i.e. the roots of

$$1 + \boldsymbol{\theta}_1 z + \boldsymbol{\theta}_2 z^2 + \ldots + \boldsymbol{\theta}_q z^q = 0$$

lie outside the unit circle. For MA(1) process:

$$\varepsilon_t = Y_t - \mu - \boldsymbol{\theta}_1 \varepsilon_{t-1} = (-\boldsymbol{\theta}_1)^t \varepsilon_0 + \sum_{j=1}^{t} (-\boldsymbol{\theta}_1)^j [Y_{t-j} - \mu]$$

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t + \boldsymbol{\theta}_1 \varepsilon_{t-1} + \ldots + \boldsymbol{\theta}_q \varepsilon_{t-q}$$

$$\varepsilon_t \sim NID(0, \sigma^2)$$

Conditional MLE = NLLS

Conditioning on $\mathbf{Y}_0 = (Y_0, Y_{-1}, \ldots, Y_{-p+1})$ and $\underline{\varepsilon}_0 = (\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-q+1})' = \mathbf{0}$, the sequence $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$ can be calculated from $\{Y_1, Y_2, \ldots, Y_T\}$ by iterating on:

$$\varepsilon_t = Y_t - (c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p}) - (\boldsymbol{\theta}_1 \varepsilon_{t-1} + \ldots + \boldsymbol{\theta}_q \varepsilon_{t-q})$$

for $t = 1, \ldots, T$.

The conditional log-likelihood is:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \log f_{\mathbf{Y}_T|\mathbf{Y}_0,\underline{\varepsilon}_0}(\mathbf{y}_T|\mathbf{y}_0,\underline{\varepsilon}_0) \\ &= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma^2}\end{aligned}$$

- One option is to set the initial values equal to their expected values:

$$Y_s = (1-\phi_1-\ldots-\phi_p)^{-1}c \quad s=0,-1,\ldots,-p+1$$

$$\varepsilon_s = 0 \quad s=0,-1,\ldots,-q+1$$

- Box and Jenkins (1976) recommended setting ε's to zero but y's equal to their actual values. The iteration is started at date $t = p + 1$, with $Y_1, Y_2, \ldots, Y_p$ set to the observed values and

$$\varepsilon_p = \varepsilon_{p-1} = \ldots = \varepsilon_{p-q+1} = 0$$