Time Series Analysis

CHEA Makara
5th Year Engineering in Majoring Data Science
Department of Applied Mathematics and Statistics, ITC

January 2025

Table of contents

The Method of Moments

- The method of moments is frequently one of the easiest, if not the most efficient, methods for obtaining parameter estimates.
- The simplest example of the method is to estimate a stationary process mean by a sample mean.
- For the autoregressive model, AR(1) case, we have $\rho_1 = \phi$. In the method of moments, ρ_1 is equated to r_1, then

$$r_1 = \hat{\phi}$$

- For AR(2) case, the relationship between the parameter ϕ_1 and ϕ_2 are given by the Yule-Walker equation, that is

$$\rho_1 = \phi_1 + \rho_1\phi_2 \quad \& \quad \rho_2 = \rho_1\phi_1 + \phi_2$$

The Method of Moments Cont.

- In the method of moments, replacing ρ_1 by r_1 and ρ_2 by r_2 then

$$\Longleftrightarrow \begin{cases} r_1 & = \phi_1 + r_1\phi_2 \\ r_2 & = r_1\phi_1 + \phi_2 \end{cases}$$

$$\Longrightarrow \hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2} \quad \& \quad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

- For AR(p) cases proceed, consider the Yulk-Walker equations

$$\begin{cases} r_1 & = \phi_1 + r_1\phi_2 + \cdots + r_{p-1}\phi_p \\ r_2 & = r_1\phi_1 + \phi_2 + r_1\phi_3 + \cdots + r_{p-2}\phi_2 \\ \vdots & \\ r_p & = r_{p-1}\phi_1 + r_{p-2}\phi_2 + r_{p-3}\phi_3 + \cdots + \phi_p \end{cases}$$

The Method of Moments Cont.

- **In case Moving Average**, MA(1) we have

$$\rho_1 = -\frac{\theta}{1+\theta^2}$$

- Replace $\rho_1 \to r_1$, if case $|r_1| < 0.5$ then we get

$$\hat{\theta} = \frac{-1+\sqrt{1-4r_1^2}}{2r_1}$$

- If $r_1 = \pm 0.5$, unique,real solutions exist, namely ∓ 1, but neither is invertible.
- If $|r_1| > 0.5$ no real solution exist, and so the method of moment is fails to estimate value of θ.
- For higher order of MA models, the method of moments quickly gets complicated and this methods are generally produces poor estimated.

The Method of Moments Cont.

- For ARMA(1,1) model case, recall that

$$\rho_k = \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2}\phi^{k-1} \quad \text{for } k \geq 1$$

- Noting that $\phi = \frac{\rho_2}{\rho_1}$, the first estimate ϕas

$$\hat{\phi} = \frac{r_2}{r_1} \quad \text{similarly } r_1 = \frac{(1-\theta\hat{\phi})(\hat{\phi}-\theta)}{1-2\theta\hat{\phi}+\theta^2}$$

Least Squares Estimation

- Consider the AR(1) model with non zero mean, μ, that is

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$$

- Least squares estimation proceeds by minimizing the sum of squares of the differences.

$$(Y_t - \mu) - \phi(Y_{t-1} - \mu)$$

- Since $Y_1, Y_2, \dots, Y_n$ are observed, we can only sum from $t = 2$ to $t = n$.
- Let

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

$$\implies \quad \frac{\partial S_c}{\partial \mu} = \sum_{t=2}^{n} 2[(Y_t - \mu) - \phi(Y_{t-1} - \mu)](-1 + \phi)$$

Least Square Estimation Cont.

$$\text{Set} \quad \frac{\partial S_c}{\partial \mu} = 0 \Longrightarrow \sum_{t=2}^{n} 2[(Y_t - \mu) - \phi(Y_{t-1} - \mu)](-1+\phi) = 0$$

$$\Longrightarrow \sum_{t=2}^{n} Y_t - (n-1)\mu - \phi \sum_{t=2}^{n} Y_{t-1} + \phi(n-1)\mu = 0$$

$$\Longrightarrow \mu(n-1)(-1+\phi) = \phi \sum_{t-2}^{n} Y_{-1} - \sum_{t=2}^{n} Y_t$$

$$\Longrightarrow \mu = \frac{1}{(n-1)(1-\phi)}[\sum_{t=2}^{n} Y_t - \phi \sum_{t=2}^{n} Y_{t-1}]$$

For large n $\frac{1}{n-1}\sum_{t=2}^{n} Y_t \approx \frac{1}{n-1}\sum_{t=2}^{n} Y_{t-1} \approx \bar{Y}, \Longrightarrow \hat{\mu} \approx \frac{1}{1-\phi}(\bar{Y} - \phi\bar{Y}) = \bar{Y}$

Least Square Estimation Cont.

- Similarly, set $\dfrac{\partial S_c}{\partial \phi} = 0$, and replace $\mu \to \bar{Y}$, then

$$\frac{\partial S_c(\phi, \bar{Y})}{\partial \phi} = \sum_{t=2}^{n} 2[(Y_t - \bar{Y}) - \phi(Y_{t-1} - \bar{Y})](Y_{t-1} - \bar{Y}) = 0$$

$$\Longrightarrow \hat{\phi} = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y})^2}$$

- For general AR(p) process, the process consists the same results, for instance

$$\hat{\mu} = \bar{Y}$$

- The case AR(p) process, the condition of least squares estimate of $\phi's$ are obtained by solving the sample Yule-Walker equations.

Least Square Estimator Cont.

- Consider the least square estimation in MA(1) model, $Y_t = e_t - \theta e_{t-1}$

  Recall that the invertibility of MA(1) can be express as

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \cdots - + e_t$$

- Then $S_c(\theta) = \sum (e_t)^2 = \sum [Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \cdots +]^2$
- In this nonlinear process, we need to use the numerical method to minimize $S_c(\theta)$ even in the simple MA with condition $e_0 = 0$

Maximum Likelihood Estimation

- For any observation $Y_1, Y_2, \ldots, Y_n$ time series, the likelihood function L is defined to be the joint probability density o obtaining the data actually observed.
- The maximum likelihood estimators are defined as the values of the parameters for which the data actually observed are the most likely, that is , the values that maximize the likelihood function.
- In AR(1) model, error assumption is that the white noise process, that are independent (uncorrelated), normality distribution random variable with zero mean and common standard deviation.
- The pdf is defined by

$$f(\mu = 0, \sigma_e^2) = (2\pi\sigma_e^2)^{-1/2} exp(-\frac{e_t^2}{2\sigma_e^2}) \quad for -\infty < e_t < \infty$$

- Th joint pdf is

$$(2\pi\sigma_e^2)^{-(n-1)/2} exp(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n} e_t^2)$$

Maximum Likelihood Estimation

- Consider,

$$\begin{cases} Y_2 - \mu &= \phi(Y_1 - \mu) + e_2 \\ Y_3 - \mu &= \phi(Y_2 - \mu) + e_3 \\ \vdots \\ Y_n - \mu &= \phi(Y_{n-1} - \mu) + e_n \end{cases}$$

- If we conditions, $Y_1 = y_1$, then

$$f(y_2, y_3, \ldots, y_n | y_1) = \prod_{t=2}^{n} f(y_t | y_{t-1})$$
$$= (2\pi\sigma_e^2)^{-(n-1)/2} \exp(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}[(y_t - \mu) - \phi(y_{t-1} - \mu)]^2)$$

and clearly, $f(y_1) \sim \mathcal{N}(\mu, \frac{\sigma_e^2}{1-\phi^2})$

Maximum Likelihood Estimation

- The likelihood function,

$$L(\phi, \mu, \sigma_e^2) = (f(y_1)f(y_2, \ldots, y_n|y_1)$$

$$= (2\pi\sigma_Y^2 \exp(-\frac{(Y_1-\mu)^2}{2\sigma_Y^2})) \times (2\pi\sigma_e^2)^{-(n-1)/2} \exp(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}((Y_t-\mu)-\phi(Y_{t-1}-\mu))^2)$$

Substitute $\sigma_Y^2 = \dfrac{\sigma_e^2}{1-\phi^2}$

$$\Leftrightarrow L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2} \exp(-\frac{1}{2\sigma_e^2} S(\phi, \mu)),$$

where $S(\phi, \mu) = \sum_{t=2}^{n}[(Y_t-\mu)-\phi(Y_{t-1}-\mu)]^2 + (1-\phi^2)(Y_1-\mu)$

- The function $S(\phi, \mu)$ is called the unconditional sum-of-squares function.

Maximum Likelihood Estimation

- Log-likelihood function

$$\begin{aligned}\mathcal{L}(\phi,\mu.\sigma_e^2) &= \log(L(\phi,\mu,\sigma_e^2)) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_e^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_e^2}S(\phi,\mu)\end{aligned}$$

- For given values of ϕ and μ, $\mathcal{L}(\phi,\mu,\sigma_e^2)$ can be maximized analytically with respect to σ_e^2 in terms of the yet-to-be-determined estimators of ϕ and μ. We obtain

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi},\hat{\mu})}{n}$$

- The estimation of ϕ and μ from the comparison of the unconditional sum of squares function $S(\phi,\mu)$ with conditional $S_c(\phi,\mu)$, reveals

$$S(\phi,\mu) = S_c(\phi,\mu) + (1-\phi^2)(Y_1-\mu)^2$$

Properties of the Estimates

- For the large n, the estimators are approximately unbiased and normally distributed and then

- AR(1): $Var(\hat{\phi}) \approx \dfrac{1-\phi^2}{n}$

- AR(2) :
$$\begin{cases} Var(\hat{\phi}_1) & \approx Var(\hat{\phi}_2) \approx \dfrac{1-\phi_2^2}{n} \\ Corr(\hat{\phi}_1, \hat{\phi}_2) & \approx -\dfrac{\phi_1}{1-\phi_2} = -\rho_1 \end{cases}$$

- Ma(1) : $Var(\hat{\theta}) \approx \dfrac{1-\theta^2}{n}$

- MA(2) : $\begin{cases} Var(\hat{\theta}_1) \approx Var(\hat{\theta}_2) \approx \dfrac{1-\theta_2^2}{n} \\ Corr(\hat{\theta}_1, \hat{\theta}_2) \approx -\dfrac{\theta_1}{1-\theta_2} \end{cases}$

- ARMA(1, 1) :
$$\begin{cases} Var(\hat{\phi}) \approx \left[\dfrac{1-\phi^2}{n}\right] \left[\dfrac{1-\phi\theta}{\phi-\theta}\right]^2 \\ Var(\hat{\theta}) \approx \left[\dfrac{1-\theta^2}{n}\right] \left[\dfrac{1-\phi\theta}{\theta-\phi}\right]^2 \\ Corr(\hat{\phi}, \hat{\theta}) \approx \dfrac{\sqrt{(1-\phi^2)(1-\theta^2)}}{1-\phi\theta} \end{cases}$$

Table of contents

Residual Analysis

- Consider the AR(2) Model with constant term

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_0 + e_t$$

- Estimation ϕ_1, ϕ_2 and θ_0, the residual is defined by

$$\hat{e}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \hat{\theta}_0$$

- For general ARMA model containing moving average terms,then we use inverted, infinite autoregressive form of the model to define residuals. However, the residual is defined by

$$\text{residual} = \text{actual} - \text{predicted}$$

- If the model is correctly specified and the parameter estimates are reasonably close to the true values, then the residuals should nearly the properties of white noise (independent, iid to normal distribution with zero and common standard deviation).

Residual Analysis

- The first diagnostic check is to inspect a plot of residual over time. If the model is adequate, we expect the plot to suggest a rectangular scatter around a zero horizontal level with no trends whatsover.
- Normality of residuals, we consume the quatile-quantile plots for assessing normality.
- In addition, the Shapiro-Wilk normality test and Jarque-Berra test are also applied for test normality.
- Autocorrelation of the residuals, generally speaking, the residuals are approximately normally distributed with zero means; however, for small lags k and j, the variance of $\hat{r}_k$ can be substantially less than $1/n$ and the estimate $\hat{r}_k$ and $\hat{r}_j$ can be highly correlated.
- For the lager lags, the approximate variance $1/n$ does appply, and further $\hat{r}_k$ and $\hat{r}_j$ are approximately uncorrelated.

The Ljung-Box Test

- In addition to looking at residual correlations at individual lags, it is useful to have a test that takes into account their magnitudes as a group.
- The chi-square distribution for Q statistic is based on the limit theorem as $n \to \infty$, where $Q = n(\hat{r}_1^2 + \hat{r}_2^2 + \cdots + \hat{r}_k^2),$. Nevertheless, the Ljug and Box subsequently discovered that even for $n = 100$, the approximation is not satisfactory. The modified Box-Pierce, or Ljung-Box, statistic is given by

$$Q_* = n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \cdots + \frac{\hat{r}_k^2}{n-K}\right)$$

  Notice that since $\dfrac{(n+2)}{(n-k)} > 1$ for every $k \geq 1$, we have $Q_* > Q$
- Ljung Box test statistic judge the size of p-values. For instance, everything looks very good if the p-value is coveraged by horizontal dashed line ($\alpha = 5\%$)

Overfitting and Parameter Redundancy

- After specifying and fitting what we believe to be an adequate model, we fit a slightly more general model, that is, a model "close by" that contains the original model as a special case.
- For Example, if ab AR(2) model seems appropriates, we might overfit with and AR(3) model. The original AR(2) model would be confirmed if:
  1. the estimate of the additional parameterm ϕ_3, is not significantly different from zero
  2. the estimates for the parameters is common, ϕ_1 and ϕ_2, do not change significantly from their original estimates.
- For any ARMA(p,q) model can be considered as special case of ARMA model with additional parameters equal to zero. However, when generalize ARMA models, we must be aware of the problem of parameter redundancy or lack of indentifiability.

Parameter Redundancy Cont.

- Consider ARMA(1,2) model:

$$Y_t = \phi Y_{t-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, \quad t \leftrightarrow t-1 \qquad (*)$$

$$Y_{t-1} = \phi Y_{t-2} + e_{t-1} - \theta_1 e_{t-2} - \theta_2 e_{t-3} \text{ (multiply constant c)} \qquad (**)$$

$$\text{Take } (*) - (**), \text{ then}$$

$$Y_t - (\phi + c) Y_{t-1} + \phi c Y_{t-2} = e_t - (\theta_1 + c) e_{t-1} - (\theta_2 - \theta_1 c) e_{t-2} + c\theta_2 e_{t-3}$$

- The apparently defines an ARMA(2,3) process. and clearly

$$\Phi(x) Y_t = \Theta(x) e_t$$

$$(1 - (\phi + c)x + \phi c x^2) Y_t = (1 - (\theta_1 + c)x - (\theta_2 - c\theta_1)x^2 + c\theta_2 x^3) e_t$$

$$\Longleftrightarrow (1 - \theta_1 x - \theta_2 x^2)(1 - cx) = 1 - (\theta_1 + c)x - (\theta_2 - c\theta_1)x^2 + c\theta_2 x^3$$

Parameter Redundancy Cont.

- Thus the AR and MA characteristic polynomilas in ARMA(2,3) process have a common factor pf $(1 - cx)$.Even though the Y_t does satify the ARMA(2,3) model, clearly the parameters in that model are not unique-the constant c is completely arbitrary. We say that we have **parameter redundancy** in the ARMA(2,3) model.
- The implications for fitting and overfitting models are as follows
    1. Specify the original model carefully. If a simple model seems at all promising, check it out before trying a more complicated model.
    2. when overfitting, do not increase the orders of both AR and MA parts of the model simultaneously.
    3. Extend the model is directions suggested by the analysis of the residuals. For example, if after fitting model MA(1), substantial correlation remains at lag 2 in the residuals, try an MA(2), not ARMA(1,1).

Table of contents

Minimum Mean Square Error Forecasting

- We would like to forecast the value of Y_{t+l} that will occur ℓ time units into the future.
- We call time t **forecast origin** and ℓ is the **lead time** for the forecast, and denote the forecast itself ad $\hat{Y}_t(\ell)$.
- The minimum mean square error forecast is given by

$$\hat{Y}_t(\ell) = E[Y_{t+\ell}|Y_1, Y_2, \ldots, Y_t]$$

Deterministic Trends

- Consider, $Y_t = \mu_t + X_t$, where $X_t \sim (\mu = 0, \sigma^2 = \gamma_0)$
- Forecasting l time ahead,

$$
\begin{aligned}
\hat{Y}_t(\ell) &= E[\mu_{t+\ell} + X_{t+\ell} | Y_1, Y_2, \ldots, Y_t] \\
&= \underbrace{E[\mu_{t+\ell} | Y_1, Y_2, \ldots, Y_t]}_{\mu_{t+l} \text{ is the constant term}} + \underbrace{E[X_{t+\ell} | Y_1, Y_2, \ldots, Y_t]}_{\text{Independence}} \\
&= \mu_{t+\ell} + \underbrace{E[X_{t+\ell}]}_{0} \\
\hat{Y}_t(\ell) &= \mu_{t+\ell}
\end{aligned}
$$

- For the linear trend case, $\mu_t = \beta_0 + \beta_1 t$, then the forecast is

$$\hat{Y}_t(\ell) = \beta_0 + \beta_1(t + \ell)$$

ARIMA Forecasting AR Case

- In AR(1) process with a nonzero mean that satisfies

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t \quad \text{then the forecast one step ahead}$$

$$\begin{aligned} Y_{t+1} - \mu &= \phi(Y_t - \mu) + e_{t+1} \\ \Longrightarrow \hat{Y}_t(1) - \mu &= \phi[E(Y_t|Y_1, Y_2, \ldots, Y_t) - \mu] + E(e_{t+1}|Y_1, Y_2, \ldots, Y_t) \\ &= \phi(Y_t - \mu) - \underbrace{E[e_{t+1}]}_{0} \\ \hat{Y}_t(l) &= \mu + \phi(Y_t - \mu) \end{aligned}$$

- In general lead time ℓ, we can see that

$$\hat{Y}_t(\ell) = \mu + \phi[\hat{Y}_t(\ell - 1) - \mu] \quad \text{for } \ell \geq 1$$

where $\hat{Y}_t(\ell - 1) = E[Y_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t]$, and for $l \geq 1$, e_{t+l} is independent of $\{Y_t\}$

ARIMA Forecasting (AR Case) Cont.

- If we are backward on l in the AR(1) forecast, we have

$$\begin{aligned}
\hat{Y}_t(l) &= \phi[\hat{Y}_t(\ell-1)-\mu]+\mu \\
&= \phi[\phi(\hat{Y}_t(\ell-2)-\mu)]+\mu \\
&\vdots \\
\hat{Y}_t(\ell) &= \phi^{\ell-1}[\hat{Y}_t(1)-\mu]+\mu \\
\Longleftrightarrow \hat{Y}_t(\ell) &= \phi^{\ell}[Y_t-\mu]+\mu
\end{aligned}$$

- In generally, since $|\phi|<1$ then for large l leading time, we shall that

$$\hat{Y}_t(\ell)\approx\mu$$

ARIMA Forecasting (AR Case) Cont.

- Consider the one step ahead for predicting error, $e_t(1)$

$$
\begin{aligned}
e_t(1) &= Y_{t+1} - \hat{Y}_t(1) \\
&= [\mu + (\phi(Y_t - \mu) + e_{t+1}] - [\mu + \phi(Y_t - \mu)] \\
e_t(1) &= e_{t+1}
\end{aligned}
$$

- The white noise process $\{e_t\}$ can be reinterpreted as a sequence of one step ahead forecast errors and clearly $Var(e_t(1)) = \sigma_e^2$
- To investigate the properties of the forecast errors for longer leads, consider

$$
e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell) = \underbrace{Y_{t+\ell}}_{MA(\infty)} - [\mu + \phi^\ell(Y_t - \mu)]
$$

ARIMA Forecasting (AR Case) Cont.

- We have

$$e_t(\ell) = \underbrace{Y_{t+\ell}}_{\text{general linear process}} - \mu - \phi^\ell(Y_t - \mu)$$

$$= e_{t+\ell} - \phi e_{t+\ell-1} + \cdots + \phi^{\ell-1} e_{t+1} + \phi^\ell e_t$$

$$+ \cdots - \phi^l(e_t + \phi e_{t-1} + \dots)$$

$$\text{so that } e_t(\ell) = e_{t+\ell} + \phi e_{t+\ell-1} + \cdots + \phi^{\ell-1} e_{t+1}$$

  which can also be written as $e_t(\ell) = e_{t+\ell} + \psi_1 e_{t+\ell-1} + \psi_2 e_{t+\ell-2} + \cdots + \psi_{\ell-1} e_{t+1}$
- with $E[e_t(\ell)] = 0$ and $Var(e_t(\ell)) = \sigma_e^2(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)$
- In particular, for AR(1) case, $Var(e_t(l)) = \sigma_e^2[\dfrac{1-\phi^{2l}}{1-\phi^2}]$. If ℓ is large, $Var(e_t(\ell)) \approx \dfrac{\sigma_e^2}{1-\phi^2}$
- Noted that $Var(e_t(\ell)) = \underbrace{Var(Y_t)}_{AR(1)} = \gamma_0; \quad l \to \infty$

ARIMA Forecasting- MA Case

- To forecast MA(1) case with nonzero mean $Y_t = \mu + e_t - \theta e_{t-1}$
- Forecasting 1 time ahead,

$$
\begin{aligned}
Y_{t+1} &= \mu + e_{t+1} - \theta e_t \\
\Longleftrightarrow \hat{Y}_t(1) &= \mu + \underbrace{E(e_t + 1)}_{0} - \theta E[e_t | Y_1, Y_2, \ldots, Y_t] \\
&= \mu - \theta e_t
\end{aligned}
$$

- Forecasting one step ahead for error ,

$$
\begin{aligned}
e_t(1) &= Y_{t+1} - \hat{Y}_t(1) \\
&= (\mu + e_{t+1} - \theta e_t) - (\mu - \theta e_t) \\
e_t(1) &= e_{t+1}
\end{aligned}
$$

- For longer lead times, we have

$$\hat{Y}_t(\ell) = \mu + E[e_{t+\ell} | Y_1, Y_2, \ldots, Y_t] - \theta E[e_{t+\ell-1} | Y_1, \ldots, Y_t] = \mu \text{ for } \ell > 1$$

The Random Walk with Drift

- Consider the random walk with drift $Y_t = Y_{t-1} + \theta_0 + e_t$
- Here

$$
\begin{aligned}
Y_{t+1} &= Y_t + \theta_0 + e_{t+1} \\
\implies \hat{Y}_t(1) &= E[Y_t | Y_1, \ldots, Y_t] + \theta_0 + E[e_{t+1} | Y_1, Y_2, \ldots, Y_t] \\
\hat{Y}_t(1) &= Y_t + \theta_0
\end{aligned}
$$

- In particular, $\hat{Y}_t(\ell) = \hat{Y}_t(\ell - 1) + \theta_0$ for $l \geq 1$
- Iterating backward on l,

$$
\begin{aligned}
\hat{Y}_t(\ell) &= \hat{Y}_t(\ell - 2) + \theta_0 + \theta_0 = \hat{Y}_t(\ell - 2) + 2\theta_0 \\
&= \hat{Y}_t(\ell - 3) + 3\theta_0 \\
&\vdots \\
\hat{Y}_t(\ell) &= Y_t + \ell\theta_0 \quad \text{for } l \geq 1.
\end{aligned}
$$

Random Walk Cont.

- Forecasting 1 step ahead for error

$$\begin{aligned} e_t(1) &= Y_{t+1} - \hat{Y}_t(1) \\ &= Y_t + \theta_0 + e_{t-1} - (Y_t + \theta_0) \\ &= e_{t+1} \end{aligned}$$

- In particular,

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \hat{Y}_t(\ell) \\ &= (Y_t + \ell\theta_0 + e_{t+1} + \cdots + e_{t+\ell}) - (Y - t + \ell\theta_0) \\ &= e_{t+1} + e_{t+2} + \cdots + e_{t+\ell} \end{aligned}$$

- This model is performed with $\psi_j = 1, \forall j$then

$$Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 = \ell\sigma_e^2$$

ARMA(p,q) Forecasting

- Consider the AMRA$(1,1)$ model with drift, we have

$$\begin{aligned}
Y_t &= \theta_0 + \phi Y_{t-1} + e_t - \theta e_{t-1} \\
Y_{t+1} &= +\theta_0 + \phi Y_t + e_{t+1} - \theta e_t \\
\Longrightarrow \hat{Y}_t(1) &= \theta_0 + \phi E[Y_t | Y_1, Y_2, \ldots, Y_t] + E[e_{t+1} | Y_1, \ldots, Y_t] - \theta E[e_t | Y_1, Y_2, \ldots, Y_t] \\
\hat{Y}_t(1) &= \theta_0 + \phi Y_t - \theta e_t
\end{aligned}$$

Forecast error

$$\begin{aligned}
\hat{e}_t(1) &= Y_{t+1} - \hat{Y}_t(1) \\
&= \theta_0 + \phi Y_t + e_{t+1} - \theta e_t - (\theta_0 + \phi Y_t - \theta e_t) \\
\hat{e}_t(1) &= e_{t+1}
\end{aligned}$$

ARMA(p,q) Forecasting Cont.

- Two times ahead forecasting

$$Y_{t+2} = \theta_0 + \phi Y_{t+1} + e_{t+2} - \theta e_{t+1}$$

$$E[Y_{t+2}|Y_1, Y_2, \ldots, Y_t] = \theta_0 + \phi E[Y_{t+1}|Y_1, \ldots, Y_t] + 0 - \theta \underbrace{E[e_{t+1}|Y_1, \ldots, Y_t]}_{0}$$

$$\Longrightarrow \hat{Y}_t(2) = \theta_0 + \phi \hat{Y}_t(1)$$

- Generally, $\hat{Y}_t(\ell) = \phi \hat{Y}_t(\ell - 1) + \theta_0$ for $\ell > 2$.
- For $\ell > q$, where $\ell = 1, 2, 3, \ldots, q$, the autoregressive portion of the difference equation takes over, and we have

$$\hat{Y}_t(\ell) = \phi_1 \hat{Y}_1(\ell - 1) + \phi_2 \hat{Y}_2(\ell - 2) + \cdots + \phi_p \hat{Y}_t(\ell - p) + \theta_0 \text{ for } \ell > q$$

Nonstationary Models

- Consider the ARIMA (1,1,1) model case,

$$\begin{aligned} Y_t - Y_{t-1} &= \theta_0 + \phi(Y_{t-1} - Y_{t-2}) + e_t - \theta e_{t-1} \\ Y_t &= \theta_0 + (1+\theta)Y_{t-1} - \phi Y_{t-2} + e_t - \theta e_{t-1} \\ \implies \hat{Y}_t(1) &= \theta_0 + (1+\theta)Y_t - \phi Y_{t-1} - \theta e_t \\ \hat{Y}_t(2) &= \theta_0 + (1+\theta)\hat{Y}_t(1) - \phi Y_t \end{aligned}$$

Generally, $\hat{Y}_t(\ell) = (1+\phi)\hat{Y}_t(\ell-1) - \phi\hat{Y}_t(\ell-2) + \theta_0$

- and $e_t(\ell) = e_{t+\ell} + \psi_1 e_{t+\ell-1} + \psi_2 e_{t+l-2} + \cdots + \psi_{\ell-1}e_{t+1}$ for $\ell \geq 1$.

with $E[e_t(\ell)] = 0; \quad Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2$ for $\ell \geq 1$

Prediction Limits

- Deterministic Trends model with a white noise stochastic component $\{X_t\}$, we recall that

$$\hat{Y}_t(\ell) = \mu_{t+\ell}$$

and

$$Var(e_t(\ell)) = Var(X_{t+\ell}) = \gamma_0$$

- For a given confidence level $1-\alpha$, we could use a standard normal percentile , $z_{1-\alpha/2}$, that is

$$P\left[-z_{1-\alpha/2} < \frac{Y_{t+\ell} - \hat{Y}_t(\ell)}{\sqrt{Var(e_t(\ell))}} < z_{1-\alpha/2}\right] = 1-\alpha$$

For $(1-\alpha)100\%$ confident that the future observation $Y_{t+\ell}$ consists within the predicting limits

$$\hat{Y}_t(\ell) \pm Z_{\alpha/2}\sqrt{Var(e_t(\ell))}$$

Updating ARIMA Forecasts

- For general updating equation, we shall see that

$$\hat{Y}_{t+1}(\ell) = \hat{Y}_t(\ell + 1) + \psi_\ell[Y_{t+1} - \hat{Y}_t(1)]$$

Notice that $[Y_{t+1} - \hat{Y}_t(1)]$ is the actual forecast error at time $t + 1$ once Y_{t+1} has been observed.

Forecast Weights and Exponentially Weighted Moving Averages

- For ARIMA models, we return to the inverted form of any invertible ARIMA process namely,

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots + e_t$$

  , then

$$\hat{Y}_t(1) = \pi_1 Y_t + \pi_2 Y_{t-1} + \dots$$

  We assume the t is sufficiently large ans/or that the π-weights die out sufficiently quickly so that $\pi_t, \pi_{t+1}, \dots$ are all negligible.

- The forecasts is called an **exponentially weighted moving average (EWMA)**

Forecasting Transformed Series

- **Differencing**
  Suppose we are interested in forecasting a series whose model involves a first difference to achieve stationary.
  There are two methods of forecasting can be considered:
  1. forecasting the original nonstationary series,
  2. Forecasting the stationary difference series $W_t = Y_t - Y_{t-1}$ and then "undoing: the difference by summing to obtain the forecast in original terms.