Università di Pavia

Forecasting

Eduardo Rossi

Forecast of Y_{t+1} based on a set of variables observed at date t, $\mathbf{X}_t$: $Y^*_{t+1|t}$. The loss function

$$MSE(Y^*_{t+1|t}) = E[Y_{t+1} - Y^*_{t+1|t}]^2$$

The forecast with the smallest MSE is

$$Y^*_{t+1|t} = E[Y_{t+1}|\mathbf{X}_t]$$

Suppose $Y^*_{t+1|t}$ is a linear function of $\mathbf{X}_t$:

$$\widehat{Y}_{t+1|t} = \alpha'\mathbf{X}_t$$

if

$$E[(Y_{t+1} - \alpha'\mathbf{X}_t)\mathbf{X}'_t] = \mathbf{0}'$$

then $\alpha'\mathbf{X}_t$ is the linear projection of Y_{t+1} on $\mathbf{X}_t$.

The LP projection produces the smallest MSE among the class of linear forecasting rule

$$\widehat{P}(Y_{t+1}|\mathbf{X}_t) = \alpha'\mathbf{X}_t$$

$$MSE[\widehat{P}(Y_{t+1}|\mathbf{X}_t)] \geq MSE[E(Y_{t+1}|\mathbf{X}_t)]$$

using

$$E[(Y_{t+1} - \alpha'\mathbf{X}_t)\mathbf{X}_t'] = \mathbf{0}'$$

$$E[Y_{t+1}\mathbf{X}_t'] = \alpha' E[\mathbf{X}_t\mathbf{X}_t']$$

$$\alpha' = E[Y_{t+1}\mathbf{X}_t']E[\mathbf{X}_t\mathbf{X}_t']^{-1}$$

The MSE associated with a LP is given by

$$E[(Y_{t+1} - \alpha'\mathbf{X}_t)^2] = E[(Y_{t+1})^2] - 2E(\alpha'\mathbf{X}_t Y_{t+1}) + E(\alpha'\mathbf{X}_t\mathbf{X}_t\alpha)$$

Replacing α

$$E[(Y_{t+1} - \alpha'\mathbf{X}_t)^2] = E[(Y_{t+1})^2] - 2E(Y_{t+1}\mathbf{X}'_t)[E(\mathbf{X}_t\mathbf{X}'_t)]^{-1}E(\mathbf{X}_t Y_{t+1})$$
$$+E(Y_{t+1}\mathbf{X}'_t)[E(\mathbf{X}_t\mathbf{X}'_t)]^{-1}[E(\mathbf{X}_t\mathbf{X}'_t)][E(\mathbf{X}_t\mathbf{X}'_t)]^{-1}E(\mathbf{X}_t Y_{t+1})$$

$$E[(Y_{t+1} - \alpha'\mathbf{X}_t)^2] = E[(Y_{t+1})^2] - E(Y_{t+1}\mathbf{X}'_t)[E(\mathbf{X}_t\mathbf{X}'_t)]^{-1}E(\mathbf{X}_t Y_{t+1})$$

If $\mathbf{X}_t$ includes a constant term, then

$$\widehat{P}[(aY_{t+1} + b)|\mathbf{X}_t] = a\widehat{P}(Y_{t+1}|\mathbf{X}_t) + b$$

The forecast error is

$$[aY_{t+1} + b] - [a\widehat{P}(Y_{t+1}|\mathbf{X}_t) + b] = a[Y_{t+1} - \widehat{P}(Y_{t+1}|\mathbf{X}_t)]$$

is uncorrelated with $\mathbf{X}_t$ as required of a linear projection.

LP is closely related to OLS regression

$$y_{t+1} = \beta' \mathbf{X}_t + u_t$$

$$\widehat{\beta} = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t' \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t Y_{t+1} \right]$$

$\widehat{\beta}$ is constructed from the sample moments, while α is constructed from population moments.
If $\{\mathbf{X}_t, Y_{t+1}\}$ is covariance stationary and ergodic for second moments, then the sample moments will converge to the population moments as the sample size T goes to infinity

$$\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t' \xrightarrow{p} E[\mathbf{X}_t \mathbf{X}_t']$$

$$\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t y_{t+1} \xrightarrow{p} E[\mathbf{X}_t Y_{t+1}]$$

implying

$$\widehat{\beta} \xrightarrow{p} \alpha$$

$\widehat{\beta}$ is consistent for the LP coefficient.

Forecasting based on lagged ϵ's.

Infinite MA:

$$(Y_t - \mu) = \psi(L)\epsilon_t$$
$$\epsilon_t \sim WN(0, \sigma^2)$$

$\psi_0 = 1$, $\sum_{j=0}^{\infty} |\psi_j| < \infty$. An infinite number of obs on ϵ through date t: $\{\epsilon_t, \epsilon_{t-1}, \ldots\}$. We know the values of μ and $\{\psi_1, \psi_2, \ldots\}$

$$Y_{t+s} = \mu + \epsilon_{t+s} + \psi_1\epsilon_{t+s-1} + \psi_2\epsilon_{t+s-2} + \ldots + \psi_s\epsilon_t + \psi_{s+1}\epsilon_{t-1} + \ldots$$

The *optimal linear forecast* is:

$$\widehat{E}[Y_{t+s}|\epsilon_t, \epsilon_{t-1}, \ldots] = \mu + \psi_s \epsilon_t + \psi_{s+1}\epsilon_{t-1} + \ldots$$

where $\widehat{E}[Y_{t+s}|\mathbf{X}_t] \equiv P(Y_{t+s}|1, \mathbf{X}_t)$.

The unknown future ϵ's are set to their expected value of zero. The forecast error is

$$Y_{t+s} - \widehat{E}[Y_{t+s}|\epsilon_t, \epsilon_{t-1}, \ldots] = \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + \psi_2 \epsilon_{t+s-2} + \ldots + \psi_{s-1}\epsilon_{t+1}$$

$$E[(Y_{t+s} - \widehat{E}[Y_{t+s}|\epsilon_t, \epsilon_{t-1}, \ldots])^2] = (1 + \psi_1^2 + \ldots + \psi_{s-1}^2)\sigma^2$$

when $s \to \infty$ the MSE converges to the unconditional variance $\sigma^2 \sum_{j=0}^{\infty} \psi_j^2$.

MA(q):

$$\psi(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$$

$$Y_{t+s} = \mu + \epsilon_{t+s} + \theta_1 \epsilon_{t+s-1} + \ldots + \theta_{t+s-q}\epsilon_{t+s-q}$$

The optimal linear forecast is

$$\widehat{E}[Y_{t+s}|\epsilon_t, \epsilon_{t-1}, \ldots] = \begin{cases} \mu + \theta_s \epsilon_t + \theta_{s+1}\epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q+s} & s = 1, \ldots, q \\ \mu & s = q+1, \ldots \end{cases}$$

MSE:

$$
\begin{array}{ll}
\sigma^2 & s = 1 \\
(1 + \theta_1 + \ldots + \theta_{s-1}^2)\sigma^2 & s = 2, 3, \ldots, q \\
(1 + \theta_1^2 + \ldots + \theta_q^2)\sigma^2 & s = q + 1, q + 2, \ldots
\end{array}
$$

The MSE increases with the forecast horizon up until $s = q$. For $s > q$ the forecast is the unconditional mean and the MSE is the unconditional variance of the series.

Compact lag operator

$$\frac{\psi(L)}{L^s} = L^{-s} + \psi_1 L^{1-s} + \psi_2 L^{2-s} + \ldots + \psi_{s-1} L^{-1} + \psi_s L^0 + \psi_{s+1} L^1 + \psi_{s+2} L^2 + \ldots$$

the *annihilation operator* replaces negative powers of L by zero

$$\left[\frac{\psi(L)}{L^s}\right]_+ = \psi_s L^0 + \psi_{s+1} L^1 + \psi_{s+2} L^2 + \ldots$$

$$\widehat{E}[Y_{t+s}|\epsilon_t, \epsilon_{t-1}, \ldots] = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ \epsilon_t$$

Forecasting based on lagged Y's.

In the usual forecasting situation we have obs on lagged Y's. Suppose the infinite MA process has an Infinite AR representation

$$\eta(L)(Y_t - \mu) = \epsilon_t$$

$$\eta(L) = \sum_{j=0}^{\infty} \eta_j L^j, \ \ \eta_0 = 1 \text{ and } \sum_{j=0}^{\infty} |\eta_j| < \infty$$

$$\eta(L) = [\psi(L)]^{-1}.$$

A c.s. AR(p)

$$(1 - \phi_1 L - \phi_2 L^2 + \ldots + \phi_p L^p)(Y_t - \mu) = \epsilon_t$$

$$\phi(L)(Y_t - \mu) = \epsilon_t$$

satisfies

$$\begin{aligned} \eta(L) &= \phi(L) \\ \psi(L) &= [\phi(L)]^{-1} \end{aligned}$$

For an MA(q):

$$Y_t - \mu = (1 + \theta_1 L + \ldots + \theta_q L^q)\epsilon_t$$

$$Y_t - \mu = \theta(L)\epsilon_t$$

$$\begin{aligned} \psi(L) &= \theta(L) \\ \eta(L) &= [\theta(L)]^{-1} \end{aligned}$$

provided that is based on an invertible representation.

ARMA(p,q) can be represented as an AR(∞) with

$$\psi(L) = \frac{\theta(L)}{\phi(L)}$$

provided that the roots of $\phi(z)$ and $\theta(z)$ lie outside the unit circle.

When the restrictions are satisfied obs on $\{Y_t, Y_{t-1}, Y_{t-2,\ldots}\}$ will be sufficient to construct $\{\epsilon_t, \epsilon_{t-1}, \ldots\}$.

For example for an AR(1):

$$(1-\phi L)(Y_t-\mu)=\epsilon_t$$

given ϕ and μ and Y_t,Y_{t-1}, the value of ϵ_t can be constructed from

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

For an invertible MA(1):

$$(1+\theta L)^{-1}(Y_t-\mu)=\epsilon_t$$

given an infinite number of obs on Y, we can compute:

$$\epsilon_t=(Y_t-\mu)-\theta(Y_{t-1}-\mu)+\theta^2(Y_{t-2}-\mu)-\theta^3(Y_{t-3}-\mu)+\ldots$$

Under the conditions

$$\widehat{E}[Y_{t+s}|Y_t, Y_{t-1}, \ldots] = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ \eta(L)(Y_t - \mu)$$

the forecast of Y_{t+s} as a function of lagged Y's.
Using $\eta(L) = [\psi(L)]^{-1}$

$$\widehat{E}[Y_{t+s}|Y_t, Y_{t-1}, \ldots] = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ [\psi(L)]^{-1}(Y_t - \mu)$$

Wiener-Kolmogorov prediction formula.

For example for an AR(1):

$$(1-\phi L)(Y_t-\mu)=\epsilon_t$$

$$\psi(L)=\frac{1}{1-\phi L}=1+\phi L+\phi^2 L^2+\ldots+\phi^s L^s+\ldots$$

the annihilation operator is:

$$\left[\frac{\psi(L)}{L^s}\right]_+=\phi^s+\phi^{s+1}L^1+\phi^{s+2}L^2+\ldots=\frac{\phi^s}{1-\phi L}$$

$$\widehat{E}[Y_{t+s}|Y_t,Y_{t-1},\ldots]=\mu+\left[\frac{\psi(L)}{L^s}\right]_+\eta(L)(Y_t-\mu)=\mu+\frac{\phi^s}{1-\phi L}(1-\phi L)(Y_t-\mu)$$

where $\epsilon_t=(1-\phi L)(Y_t-\mu)$.

$$\widehat{E}[Y_{t+s}|Y_t,Y_{t-1},\ldots]=\mu+\phi^s(Y_t-\mu)$$

the forecast decays geometrically from $(Y_t-\mu)$ toward μ as s increases.

Given that $\psi_j = \phi^j$, from the MSE of a MA(∞), we have that the MSE s-period-ahead forecast error is:

$$[1 + \phi^2 + \ldots + \phi^{2(s-1)}]\sigma^2$$

as $s \to \infty$

$$MSE = \frac{\sigma^2}{1 - \phi^2}$$

Stationary AR(p) process

$$Y_{t+s} - \mu = f_{11}^{(s)}(Y_t - \mu) + f_{12}^{(s)}(Y_{t-1} - \mu) + \ldots + f_{1p}^{(s)}(Y_{t-p+1} - \mu) + \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + \cdots + \psi_{s-1}\epsilon_{t+1}$$

$$\psi_j = f_{11}^{(j)}$$

the optimal s-period-ahead forecast is

$$\widehat{Y}_{t+s|t} = \mu + f_{11}^{(s)}(Y_t - \mu) + \ldots + f_{1p}^{(s)}(Y_{t-p+1} - \mu)$$

forecast error

$$Y_{t+s|t} - \widehat{Y}_{t+s|t} = \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + \cdots + \psi_{s-1}\epsilon_{t+1}$$

To calculate the optimal forecast we use a recursion. Start with the forecast $\widehat{Y}_{t+1|t}$

$$\widehat{Y}_{t+1|t} - \mu = \phi_1(Y_t - \mu) + \ldots + \phi_p(Y_{t-p+1} - \mu)$$

$\widehat{Y}_{t+2|t+1}$:

$$\widehat{Y}_{t+2|t+1} - \mu = \phi_1(Y_{t+1} - \mu) + \ldots + \phi_p(Y_{t-p+2} - \mu)$$

Law of Iterated Projections: Forecast $\widehat{Y}_{t+2|t+1}$ projected on date t information set then we obtain $\widehat{Y}_{t+2|t}$:

$$\widehat{Y}_{t+2|t} - \mu = \phi_1(\widehat{Y}_{t+1|t} - \mu) + \ldots + \phi_p(Y_{t-p+2} - \mu)$$

substituting $\widehat{Y}_{t+1|t}$

$$\begin{aligned}\widehat{Y}_{t+2|t} - \mu &= \phi_1[\phi_1(Y_t - \mu) + \ldots + \phi_p(Y_{t-p+1} - \mu)] + \\ &\quad \phi_2(Y_t - \mu) + \ldots + \phi_p(Y_{t-p+2} - \mu)\end{aligned}$$

$$\begin{aligned}\widehat{Y}_{t+2|t} - \mu &= (\phi_1^2 + \phi_2)(Y_t - \mu) + (\phi_1\phi_2 + \phi_3)(Y_{t-1} - \mu) + \ldots + \\ &\quad (\phi_1\phi_{p-1} + \phi_p)(Y_{t-p+2} - \mu) + \phi_1\phi_p(Y_{t-p+1} - \mu)\end{aligned}$$

The s-period-ahead forecast of an AR(p) process can be obtained by iterating on

$$\widehat{Y}_{t+j|t} - \mu = \phi_1(\widehat{Y}_{t+j-1|t} - \mu) + \ldots + \phi_p(\widehat{Y}_{t+j-p|t} - \mu)$$

Invertible MA(1)

$$(Y_t - \mu) = (1 + \theta L)\epsilon_t$$

with $|\theta| < 1$. Wiener-Kolmogorov formula

$$\widehat{Y}_{t+s|t} = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ (1 + \theta L)^{-1}(Y_t - \mu)$$

Forecast $s = 1$

$$\left[\frac{(1 + \theta L)}{L^1}\right]_+ = \theta$$

$$\begin{aligned}
\widehat{Y}_{t+1|t} &= \mu + \frac{\theta}{1 + \theta L}(Y_t - \mu) \\
&= \mu + \theta(Y_t - \mu) - \theta^2(Y_{t-1} - \mu) + \theta^3(Y_{t-2} - \mu) + \ldots
\end{aligned}$$

Alternatively

$$\epsilon_t = (1+\theta L)^{-1}(Y_t - \mu)$$

in practice

$$\widehat{\epsilon}_t = (Y_t - \mu) - \theta\widehat{\epsilon}_{t-1}.$$

For $s = 2, 3, \ldots$

$$\left[\frac{(1+\theta L)}{L^s}\right]_+ = 0$$

$$\widehat{Y}_{t+s|t} = \mu$$

$$(Y_t - \mu) = \theta(L)\epsilon_t$$

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

$$\widehat{Y}_{t+s|t} = \mu + \left[\frac{1 + \theta_1 L + \ldots + \theta_q L^q}{L^s}\right]_+ \frac{1}{\theta(L)}(Y_t - \mu)$$

$$\left[\frac{1 + \theta_1 L + \ldots + \theta_q L^q}{L^s}\right]_+ = \begin{cases} 1 + \theta_s L + \theta_{s+1} L^2 + \ldots + \theta_q L^{q-s} & s = 1, \ldots, q \\ 0 & s = q+1, \ldots \end{cases}$$

For

$$\widehat{Y}_{t+s|t} = \mu + (\theta_s + \theta_{s+1} L + \ldots + \theta_q L^{q-s})\widehat{\epsilon}_t$$

$$\widehat{\epsilon}_t = (Y_t - \mu) - \theta_1 \widehat{\epsilon}_{t-1} - \ldots - \theta_q \widehat{\epsilon}_{t-q}$$

$$(1-\phi L)(Y_t-\mu) = (1+\theta L)\epsilon_t$$

Stationarity: $|\phi| < 1$. Invertibility: $|\theta| < 1$.

$$\widehat{Y}_{t+s|t} = \mu + \left[\frac{1+\theta L}{(1-\phi L)L^s}\right]_+ \frac{1-\phi L}{1+\theta L}(Y_t-\mu)$$

$$\frac{1}{(1-\phi L)} = 1+\phi L+\phi^2 L^2+\ldots$$

$$\begin{aligned}
\left[\frac{1+\theta L}{(1-\phi L)L^s}\right]_+ &= \left[\frac{1}{(1-\phi L)L^s}+\frac{\theta L}{(1-\phi L)L^s}\right]_+ \\
&= \left[\frac{(1+\phi L+\phi^2L^2+\ldots)}{L^s}+\frac{\theta L(1+\phi L+\phi^2L^2+\ldots)}{L^s}\right]_+ \\
&= (\phi^s+\phi^{s+1}L+\phi^{s+2}L^2+\ldots)+ \\
&\quad \theta(\phi^{s-1}+\phi^s L+\phi^{s+1}L^2+\ldots)
\end{aligned}$$

$$\begin{aligned}\left[\frac{1+\theta L}{(1-\phi L)L^s}\right]_+ &= \phi^s(1+\phi L+\phi^2L^2+\ldots)+\\ &\quad \theta\phi^{s-1}(1+\phi L+\phi^2L^2+\ldots)\\ &= (\phi^s+\theta\phi^{s-1})(1+\phi L+\phi^2L^2+\ldots)\\ &= \frac{\phi^s+\theta\phi^{s-1}}{1-\phi L}\end{aligned}$$

$$\begin{aligned}\widehat{Y}_{t+s|t} &= \mu + \left[\frac{1+\theta L}{(1-\phi L)L^s}\right]_+ \frac{1-\phi L}{1+\theta L}(Y_t-\mu)\\ &= \mu + \frac{\phi^s+\theta\phi^{s-1}}{1-\phi L}\frac{1-\phi L}{1+\theta L}(Y_t-\mu)\\ &= \mu + \frac{\phi^s+\theta\phi^{s-1}}{1+\theta L}(Y_t-\mu)\end{aligned}$$

For $s = 2, 3, \ldots$ the forecast

$$\widehat{Y}_{t+s|t} - \mu = \phi(\widehat{Y}_{t+s-1|t} - \mu)$$

the forecast decays geometrically at the rate ϕ toward the unconditional mean μ. The one-period-ahead forecast (s=1) is given by

$$\widehat{Y}_{t+1|t} = \mu + \frac{\phi+\theta}{1+\theta L}(Y_t-\mu)$$

$$\widehat{Y}_{t+1|t} = \mu + \frac{\phi(1+\theta L) + \theta(1-\phi L)}{1+\theta L}(Y_t - \mu)$$
$$= \mu + \phi(Y_t - \mu) + \frac{1-\phi L}{1+\theta L}(Y_t - \mu)$$

$$\widehat{\epsilon}_t = \frac{1-\phi L}{1+\theta L}(Y_t - \mu) = (Y_t - \mu) - \phi(Y_{t-1} - \mu) - \theta\widehat{\epsilon}_{t-1}$$

$$\widehat{\epsilon}_t = Y_t - \widehat{Y}_{t|t-1}$$

$s = 2,$

$$\begin{aligned}
\widehat{Y}_{t+2|t} &= \mu + \frac{\phi^2 + \theta\phi}{1+\theta L}(Y_t - \mu) \\
&= \mu + \phi\frac{\phi + \theta}{1+\theta L}(Y_t - \mu) \\
&= \mu + \phi(\phi+\theta)(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \ldots)(Y_t - \mu) \\
&= \mu + \phi(\phi+\theta)(Y_t - \mu) - \phi(\phi+\theta)\theta(Y_{t-1} - \mu) + \ldots
\end{aligned}$$

ARMA(p,q):

$$\phi(L)(Y_t - \mu) = \theta(L)\epsilon_t$$

$$\widehat{Y}_{t+1|t} - \mu = \phi_1(Y_t - \mu) + \ldots + \phi_p(Y_{t-p+1} - \mu) + \theta_1\widehat{\epsilon}_t + \ldots + \theta_q\widehat{\epsilon}_{t-q+1}$$

$$\widehat{\epsilon}_t = Y_t - Y_{t|t-1} \qquad \widehat{Y}_{\tau|t} = Y_\tau \qquad \tau \leq t$$

$$\widehat{Y}_{t+s|t} - \mu = \begin{cases} \phi_1(\widehat{Y}_{t+s-1|t} - \mu) + \ldots + \phi_p(Y_{t+s-p|t} - \mu) + \theta_1\widehat{\epsilon}_t + \ldots + \theta_q\widehat{\epsilon}_{t+s-q} \\ \text{for } \; s = 1, \ldots, q \\ \phi_1(\widehat{Y}_{t+s-1|t} - \mu) + \ldots + \phi_p(Y_{t+s-p|t} - \mu) \\ \text{for } \; s = q+1, \ldots \end{cases}$$

$\{Y_t, Y_{t-1}, \ldots, Y_{t-m+1}\}$ observations. Presample ϵ's all equal to 0.
Approximation

$$\widehat{E}[Y_{t+s}|Y_t, Y_{t-1}, \ldots] \cong \widehat{E}[Y_{t+s}|Y_t, \ldots, Y_{t-m+1}, \epsilon_{t-m} = 0, \epsilon_{t-m-1} = 0, \ldots]$$

MA(q):

$$\widehat{\epsilon}_{t-m} = \widehat{\epsilon}_{t-m-1} = \ldots = \widehat{\epsilon}_{t-m-q+1} = 0$$

$$\widehat{\epsilon}_{t-m+1} = Y_{t-m+1} - \mu$$

$$\widehat{\epsilon}_{t-m+2} = Y_{t-m+2} - \mu - \theta_1 \widehat{\epsilon}_{t-m+1}$$

$$\widehat{\epsilon}_{t-m+3} = Y_{t-m+3} - \mu - \theta_1 \widehat{\epsilon}_{t-m+2} - \theta_2 \widehat{\epsilon}_{t-m+1}$$

The values are to be replaced in

$$\widehat{Y}_{t+s|t} = \mu + (\theta_s + \theta_{s+1} L + \theta_{s+2} L^2 + \ldots + \theta_q L^{q-s}) \widehat{\epsilon}_t$$

For $s = q = 1$:

$$\widehat{Y}_{t+s|t} = \mu + \theta(Y_t - \mu) - \theta^2(Y_{t-1} - \mu) + \ldots + (-1)^{m-1}\theta^m(Y_{t-m+1} - \mu)$$

truncated infinite AR. For m and $|\theta|$ small we have a good approximation. For $|\theta| \cong 1$ the approximation may be poorer.

Exact projection of Y_{t+1} on its most recent values

$$\mathbf{X}_t = \begin{bmatrix} 1 \\ Y_t \\ \vdots \\ Y_{t-m+1} \end{bmatrix}$$

Linear Forecast

$$\alpha^{(m)'}\mathbf{X}_t = \alpha_0^m + \alpha_1^m Y_t + \ldots + \alpha_m^m Y_{t-m+1}$$

If Y_t is c.s.

$$E[Y_t Y_{t-j}] = \gamma_j + \mu^2$$

$$\mathbf{X}_t = [1, Y_t, \ldots, Y_{t-m+1}]'$$

implies

$$\alpha^{(m)\prime} = \begin{bmatrix} \mu & (\gamma_1+\mu^2) & \ldots & (\gamma_m+\mu^2) \end{bmatrix} \times \begin{bmatrix} 1 & \mu & \ldots & \mu \\ \mu & (\gamma_0+\mu^2) & \ldots & (\gamma_{m-1}+\mu^2) \\ \vdots & \vdots & & \vdots \\ \mu & (\gamma_{m-1}+\mu^2) & \ldots & (\gamma_0+\mu^2) \end{bmatrix}^{-1}$$

when a constant term is included in $\mathbf{X}_t$ it is more convenient to express variables in deviations from the mean.

Calculate the projection of $(Y_{t+1} - \mu)$ on $(Y_t - \mu), (Y_{t-1} - \mu), \ldots, (Y_{t-m+1} - \mu)$

$$\alpha^{(m)} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \vdots & \vdots & & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_m \end{bmatrix}$$

s-period-ahead forecast

$$\widehat{Y}_{t+s|t} = \mu + \alpha_1^{(m,s)}(Y_t - \mu) + \ldots + \alpha_m^{(m,s)}(Y_{t-m+s} - \mu)$$

$$\begin{bmatrix} \alpha_1^{(m,s)} \\ \vdots \\ \alpha_m^{(m,s)} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \vdots & \vdots & & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_s \\ \vdots \\ \gamma_{s+m-1} \end{bmatrix}$$

Inversion of an $(m \times m)$ matrix. Two algorithms:

1. Kalman Filter to compute finite-sample forecast.
2. Triangular Factorization.