# The Linear Prediction Operator

## Properties + Minimum MSE

STAT 464 / 864 | Fall 2024

Discrete Time Series Analysis

Skyepaphora Griffith, Queen's University

## Minimizing the MSE:

$$E\left[\left(Y-(a_0+\overbrace{a_1W_N+\cdots+a_NW_1}^{a^TW})\right)^2\right]$$

Differentiate with respect to coefficient $a_i$ → set equal to 0 → solve for coefficient

We found $\boxed{a_0 = \mu_Y - a^T\mu_W}$

Plug that in, and we find that **a**,

the vector of remaining coefficients,

solves $\boxed{\Gamma a = \gamma}$

$$\begin{aligned}\gamma &= \mathrm{Cov}(Y, W)\\ &= (\mathrm{Cov}(Y, W_N), \ldots, \mathrm{Cov}(Y, W_1))\\ \Gamma[i,j] &= \mathrm{Cov}(W_{N+1-i}W_{N+1-j})\end{aligned}$$

**Linear Prediction:** best linear predictor of Y, based on W, that minimizes MSE

$$P(Y|W) = \mu_Y - a^T\mu_W + a^TW = \mu_Y + a^T(W-\mu_W)$$

**Note:** this result is theoretical. Values highlighted in blue must be estimated, in practice.

## Minimum MSE

We know our predictor minimized it, but what *is* it, anyway?

$$\mathrm{MSE} = \mathrm{E}\Big[\big(Y - P(Y|W)\big)^2\Big]$$

$$= \mathrm{E}\Big[\Big(Y - \big(\mu_Y + a^T(W - \mu_W)\big)\Big)^2\Big] = \mathrm{E}\Big[\Big((Y - \mu_Y) - a^T(W - \mu_W)\Big)^2\Big]$$

$$= \mathrm{E}\Big[(Y - \mu_Y)^2 - 2(Y - \mu_y)a^T(W - \mu_W) + a^T(W - \mu_W)(W - \mu_W)^T a\Big]$$

$$= \mathrm{E}\big[(Y - \mu_Y)^2\big] - 2a^T\mathrm{E}\big[(Y - \mu_y)(W - \mu_W)\big] + a^T\mathrm{E}\big[(W - \mu_W)(W - \mu_W)^T\big]a$$

$$= \mathrm{Var}(Y) - 2a^T\gamma + a^T\Gamma a = \mathrm{Var}(Y) - 2a^T\gamma + a^T\gamma = \mathrm{Var}(Y) - a^T\gamma$$

# Properties of P( · | W) Overview

$$P(c_0 + c_1 U + c_2 V|W) = c_0 + c_1 P(U|W) + c_2 P(V|W)$$

$$P(Y|W) = \mathrm{E}[Y]$$

$$\mathrm{Cov}\big(W_i, Y - P(Y|W)\big) = 0,$$

$$i = 1, \ldots, N$$

$$P(W_i|W) = W_i$$

LHS

## P1) Linearity

$$\boxed{P(c_0 + c_1U + c_2V|W)} = c_0 + c_1P(U|W) + c_2P(V|W)$$

$$= \mathrm{E}[c_0 + c_1U + c_2V] + a^T(W - \mu_W)$$

$$= \boxed{\mathrm{E}[c_0] + c_1\mathrm{E}[U] + c_2\mathrm{E}[V] + a^T(W - \mu_W)}$$

$a$ satisfies $\Gamma a = \gamma$

$\Gamma$ is the covariance matrix of W,

$\gamma$ is the vector of covariances:

$$\gamma = \mathrm{Cov}(c_0 + c_1U + c_2V, W)$$

$$= c_1\mathrm{Cov}(U, W) + c_2\mathrm{Cov}(V, W)$$

$$= c_1\gamma_U + c_2\gamma_V$$

RHS

## P1) Linearity

$$P(c_0 + c_1 U + c_2 V|W) = \boxed{c_0 + c_1 P(U|W) + c_2 P(V|W)}$$

$$c_0 + c_1\left(\mathrm{E}[U] + a_1^T(W - \mu_W)\right) + c_2\left(\mathrm{E}[V] + a_2^T(W - \mu_W)\right)$$

$$= c_0 + c_1\mathrm{E}[U] + c_2\mathrm{E}[V] + c_1 a_1^T(W - \mu_W) + c_2 a_2^T(W - \mu_W)$$

$$= c_0 + c_1\mathrm{E}[U] + c_2\mathrm{E}[V] + \boxed{(c_1 a_1^T + c_2 a_2^T)(W - \mu_W)}$$

LHS $= \underline{\mathrm{E}[c_0] + c_1\mathrm{E}[U] + c_2\mathrm{E}[V] + \boxed{a^T(W - \mu_W)}}$

$a_1$ satisfies $\Gamma a_1 = \gamma_U$

$a_2$ satisfies $\Gamma a_2 = \gamma_V$

$$\Gamma(c_1 a_1 + c_2 a_2) = \gamma = c_1\gamma_U + c_2\gamma_V$$

# P1) Uncorrelated Y and W

Recall: the best linear predictor is

$$\boxed{\begin{aligned} &\mu_Y + a^T(W - \mu_W) \\ &\Gamma a = \gamma \end{aligned}}$$

If **Y** is uncorrelated with predictors $W_i$ for all i, then $\gamma = 0$

The solution is $a = \vec{0} \implies P(Y|W) = \mu_Y$

# P3) Uncorrelated Predictors and Residuals

Residuals

$$Y - P(Y|W) = Y - \mu_Y - a^T(W - \mu_W)$$

Mean of Residuals

$$\mathrm{E}[Y - P(Y|W)] = \mu_Y - \mu_Y - a^T(\mu_W - \mu_W) = 0$$

Cov[ Residuals, Predictor i ]

$$\mathrm{Cov}(Y - P(Y|W), W_i) = \boxed{\mathrm{E}\big[\big(Y - \mu_Y - a^T(W - \mu_W)\big)W_i\big]}$$

This was set equal to 0 when we optimized the MSE of our LP

# P4) the best LP of a predictor is itself

$$\begin{aligned} P(W_{N+1-i}|W) &= \mathrm{E}[W_{N+1-i}] + \underline{a}^T(W - \mu_W) \\ &= \mathrm{E}[W_{N+1-i}] + (W_{N+1-i} - \mathrm{E}[W_{N+1-i}]) \\ &= W_{N+1-i} \end{aligned}$$

$$\Gamma \underline{a} = \gamma$$

$$\begin{aligned} \Gamma[\cdot, i] &= \left(\mathrm{Cov}(W_{N+1-i}, W_N), \ldots, \mathrm{Cov}(W_{N+1-i}, W_1)\right)^T \\ &= \gamma \end{aligned}$$

i^th component

$$\underline{a} = (0, \ldots, 0, 1, 0, \ldots, 0)^T$$