

## 2 Linear Regression (theory)

1. Let  $Z = (X', W')'$ . Then,

$$E^*[Y|X, W] = E^*[Y|Z] = E[Y] + \sigma(Y, Z)V(Z)^{-1}(Z - E[Z])$$

Because  $W$  and  $X$  are independent,  $V(Z) = \begin{pmatrix} V(X) & 0 \\ 0 & V(W) \end{pmatrix}$ . So,

$$\begin{aligned} E^*[Y|X, W] &= E[Y] + (\sigma(Y, X) \quad \sigma(Y, W)) \begin{pmatrix} V(X)^{-1} & 0 \\ 0 & V(W)^{-1} \end{pmatrix} \begin{pmatrix} X - E[X] \\ W - E[W] \end{pmatrix} \\ &= E[Y] + (\sigma(Y, X) \quad \sigma(Y, W)) \begin{pmatrix} V(X)^{-1}(X - E[X]) \\ V(W)^{-1}(W - E[W]) \end{pmatrix} \\ &= E[Y] + \sigma(Y, X)V(X)^{-1}(X - E[X]) + \sigma(Y, W)V(W)^{-1}(W - E[W]) \\ &= E[Y] + E^*[Y|X] - E[Y] + E^*[Y|W] - E[Y] \\ &= E^*[Y|X] + E^*[Y|W] - E[Y] \end{aligned}$$

2. Using the formula for the best linear predictor  $\beta_0$

$$\begin{aligned} |\beta_0| &= \frac{|\sigma(X, Y)|}{V(X)} \\ &\leq \frac{\sqrt{V(X)V(Y)}}{V(X)} \\ &= \frac{\sqrt{V(Y)}}{\sqrt{V(X)}} = 1 \end{aligned}$$

In the example of father's height and child's height, the interpretation is that taller fathers may have taller children, but these effects cannot be amplified over generations.

3. From the formula of variance and applying the Law of Iterated Expectations,

$$\begin{aligned} V(Y) &= E[Y^2] - E[Y]^2 \\ &= E[E[Y^2|X]] - E[E[Y|X]]^2 \\ &= E[V(Y|X) + E[Y|X]^2] - E[E[Y|X]]^2 \\ &= E[V(Y|X)] + E[E[Y|X]^2] - [E[Y|X]]^2 \\ &= E[V(Y|X)] + V(E[Y|X]) \end{aligned}$$

4. (a) Using the formula for the best linear predictor,  $E^*[m(Z)|X] = \delta_0 + \xi_0 X$ , and the law of

iterated such that  $E[m(Z)] = E[E[X|Z]] = E[X]$ ,

$$\begin{aligned}
\xi_0 &= \sigma(m(Z), X) V(X)^{-1} \\
&= E \left[ \left( m(Z) - E[m(Z)] \right) \left( X - E[X] \right) \right] \left( E[V(X|Z)] + V(E[X|Z]) \right)^{-1} \\
&= E \left[ \left( E[X|Z] - E[X] \right) \left( X - E[X] \right) \right] \left( E[V(X|Z)] + V(E[X|Z]) \right)^{-1} \\
&= E \left[ \left( E[X|Z] - E[E[X|Z]] \right) \left( X - E[E[X|Z]] \right) \right] \left( E[V(X|Z)] + V(E[X|Z]) \right)^{-1} \\
&= E \left[ \left( E[X|Z] - E[E[X|Z]] \right)^2 \right] \left( E[V(X|Z)] + V(E[X|Z]) \right)^{-1} \\
&= V(E[X|Z]) \left( E[V(X|Z)] + V(E[X|Z]) \right)^{-1}
\end{aligned}$$

By the best linear predictor,  $\delta_0 = E[m(Z)] - \xi_0 E[X] = (1 - \xi_0)E[X]$ .

- (b)  $\xi_0$  is the fraction of total variance of the parents' income that comes from variation of average income between individual neighborhoods, with the rest of the variance due to variation of income within individual neighborhoods. That is, the higher  $\xi_0$  is, the more stratified residential areas are by income.
- (c) First note that  $\rho = \frac{\sigma(A, X)}{\sigma_A \sigma_X}$ , and that the best linear predictor of  $A$  on  $X$  is given by

$$\begin{aligned}
E^*[A|X] &= \mu_A + \sigma(A, X)(\sigma_A^2)^{-1}(X - \mu_x) \\
&= \mu_A + \frac{\rho \sigma_A}{\sigma_X}(X - \mu_x)
\end{aligned}$$

So, taking the linear prediction of  $Y$  with respect to  $X$  and substituting the results from above:

$$\begin{aligned}
E^*[Y|X, m(Z), A] &= \alpha_0 + \beta_0 X + \gamma_0 m(Z) + A \\
E^*[Y|X] &= \alpha_0 + \beta_0 X + \gamma_0 E^*[m(Z)|X] + E^*[A|X] \\
E^*[Y|X] &= \alpha_0 + \beta_0 X + \gamma_0 ((1 - \xi_0)\mu_X + \xi_0 X) + \mu_A + \frac{\rho \sigma_A}{\sigma_X}(X - \mu_x) \\
E^*[Y|X] &= \alpha_0 + \gamma_0(1 - \xi_0)\mu_X + \left( \mu_A - \frac{\rho \sigma_A}{\sigma_X} \mu_x \right) + \left( \beta_0 + \gamma_0 \xi_0 + \rho \frac{\sigma_A}{\sigma_X} \right) X
\end{aligned}$$

- (d) If there is more income stratification in New York than San Francisco,  $\xi_{NYC} > \xi_{SF}$ . So, the intercept for New York will be lower than that of San Francisco, but the coefficient on  $X$  will be higher in New York.