Econ 361: Advanced Econometrics

Prediction

Covariance and Correlation

$$\begin{aligned}
\text{Cov}(X,Y) &\equiv \int_{\Omega_Y}\int_{\Omega_X}(x-E[X])(y-E[Y])\,f(x,y)\,dx\,dy \\
&= \int_{\Omega_Y}\int_{\Omega_X}(xy-xE[Y]-E[X]y+E[X]E[Y])\,f(x,y)\,dx\,dy \\
&= E[XY]-E[X]E[Y]
\end{aligned}$$

If X and Y are statistically independent of each other ...

$$
\begin{aligned}
\text{Cov}(X,Y) &= \int_{\Omega_Y}\int_{\Omega_X} (x - E[X])(y - E[Y])\ \underbrace{f(x)f(y)}_{f(x,y)}\ dx\,dy \\
&= \left(\int_{\Omega_Y} (y - E[Y])\ f(y)\ dy\right)\ \left(\int_{\Omega_X} (x - E[X])\ f(x)\ dx\right) \\
&= \left(\underbrace{\int_{\Omega_Y} y\ f(y)\ dy}_{=E[Y]} - E[Y]\right)\ \left(\underbrace{\int_{\Omega_X} x\ f(x)\ dx}_{=E[X]} - E[X]\right) \\
&= 0
\end{aligned}
$$

Statistical independence implies zero covariance

As such, some people will use **covariance** to try to distill the information one random variable has about the other. The more the covariance differs from zero, the more ”statistically informative” one random variable has about the other.

Caveat: zero covariance does not necessarily imply statistical independence ...

Correlation: a "normalized" version of Covariance

$$\text{Correlation}(X,Y) \quad \equiv \quad \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\ \text{Var}(Y)}}$$

By Cauchy-Schwartz Inequality, we can show that

$$-1 \ \leq \ \text{Correlation}(X,Y) \ \leq \ +1$$

So, we sometimes refer to zero covariance as **"uncorrelated"**

And when two random variables are uncorrelated with each other, we often infer (at times incorrectly) that the two are not statistically informative of each other ...

Correlation and Linear Regression

Linear Regression builds on this intuition of correlation as a prediction tool.

It expands the intuition by allowing for **partial** correlation: the correlation between two random variables controlling for the influence of other random variables

Analogous to the ***ceteris paribus*** concept introduced in Introductory Economics, or the concept of ***partial derivatives*** introduced in multivariable calculus – both parallels we will explore further later this semester

Best Predictor and Best Linear Predictor

More formally, we will develop the intuition of correlation as a prediction tool using the "best predictor" and "best linear predictor" framework where "best" is with respect to some **loss function** – technically "risk function" (but most refer to the loss function underlying the risk function)

Specifically, the loss function for which the intuition seems to hold best is that of **mean squared error (MSE)**

Best Predictor of Y given X

DEFINITION: the **best predictor (BP)** of Y given X for a given loss function is the predictor that minimizes the risk function associated with that loss function

$$BP(Y|X) \quad = \quad \text{argmin}_{\hat{Y}(X)} \, E[\, LF(\hat{Y}(X)) \,]$$

$BP(Y|X)$ for the **MSE** loss function is the function $\hat{Y}(X)$ that minimizes

$$E[\, LF(\hat{Y}(X)) \,] \quad = \quad E[\quad \underbrace{\left(\hat{Y}(X) - Y\right)^2}_{LF(\hat{Y}(X)) \text{ under MSE}} \quad]$$

Best Predictor of Y given X under MSE

$BP(Y|X)$ for the **MSE** loss function is the function $\hat{Y}(X)$ that minimizes

$$E[\,LF(\hat{Y}(X))\,] \quad = \quad E[\quad \underbrace{\left(\hat{Y}(X) - Y\right)^2}_{LF(\hat{Y}(X)) \text{ under MSE}} \quad]$$

Can show that the function $\hat{Y}(X)$ that solves the above minimization problem is

$$E[Y|X]$$

Best *Linear* Predictor of Y given X under MSE

What if we constrained the allowable candidate $\hat{Y}(X)$ to just linear functions of X? So, the underlying "best" optimization problem is now a **constrained** minimization problem?

This leads to the Best **Linear** Predictor of Y given X

For the MSE loss function, we can show that Best Linear Predictor is

$$\text{BLP}_{MSE}(Y|X) \quad = \quad a^* + b^* X$$

where

$$\begin{aligned} a^* &= E[Y] - b^* E[X] \;=\; \mu_Y - b^* \mu_X \\ b^* &= \frac{E[XY] - E[X]\,E[Y]}{E[X^2] - (E[X])^2} \;=\; \frac{\sigma_{XY}}{\sigma_X^2} \end{aligned}$$

Covariance and BLP$_{MSE}(Y|X)$

Note: when Cov$(X, Y) = 0$, $b^* = 0$ and BLP$_{MSE}(Y|X) = E[Y]$

In short, when X and Y are uncorrelated, the BLP$_{MSE}(Y|X)$ is no more informative than simply predicting the mean of Y ...

In order for BLP$_{MSE}(Y|X)$ to be more informative, need X and Y to be correlated

But all of this is under the mean squared error (MSE) loss function ...