3.12 More on surprises [extra material not in handout]

We can also use the properties of the predictive error, ε, to show two features:

1. **The variance of any random variable X is equal to the variance of the predictive error, $\varepsilon = X - E(X)$**

   $$var(X) = \sigma_X^2 = E\left(\varepsilon^2\right) = var(\varepsilon)$$

   which should be fairly obvious since, using the definition of ε

   $$var(X) = E\left[(X - E(X))^2\right] = E\left(\varepsilon^2\right) = E\left[(\varepsilon - E(\varepsilon))^2\right] = E\left[(\varepsilon - 0)^2\right]$$

   but if this isn't obvious, we can also write (again using the definition of ε)

   $$\begin{aligned} X &= E(X) + \varepsilon \\ &\Rightarrow var(X) = var(\varepsilon) \end{aligned}$$

   which we shall see shortly is a special case of one of CW's rules

2. **The expectation $E(X)$ is the best (least squares) forecast of X.**

To see this consider some general forecast, of X, denoted $\widehat{X}$, and let the error of this forecast be $u = X - \widehat{X}$. Then we have (again, expanding inside the expectations operator before taking expectations) the (true) mean squared

error:

$$
\begin{aligned}
E\left(u^2\right) &= E\left[\left(X-\widehat{X}\right)^2\right] \\
&= E\left[\left(X-E\left(X\right)-\left(\widehat{X}-E\left(X\right)\right)\right)^2\right] \\
&= E\left[\left(\varepsilon-\left(\widehat{X}-E\left(X\right)\right)\right)^2\right] \\
&= E\left[\varepsilon^2+\left(\widehat{X}-E\left(X\right)\right)^2+2\varepsilon\left(\widehat{X}-E\left(X\right)\right)\right] \\
&= E\left(\varepsilon^2\right)+\left(\widehat{X}-E\left(X\right)\right)^2+2E\left(\varepsilon\right)\left(\widehat{X}-E\left(X\right)\right) \\
&= \sigma_X^2+\left(\widehat{X}-E\left(X\right)\right)^2
\end{aligned}
$$

ie the (true) mean squared error $E\left(u^2\right)$ is equal to the variance plus the squared "bias", which must be non-negative. Clearly this is minimised if $\widehat{X}=E\left(X\right)$.

- Note that we again used the technique of expanding the term in brackets *before* we took expectations.

5.1 Towards mean-variance utility: [not in handout]

Generalise the example a little

$$\begin{aligned} R &= R_L \text{ with probability } \pi = \frac{1}{2} \\ R &= R_H > R_L \text{ with probability } 1 - \pi = \frac{1}{2} \end{aligned}$$

then we have

$$E(R) \equiv \mu = \frac{R_L + R_H}{2}$$

which we can rewrite, taking μ over to the rhs, as

$$0 = \frac{R_L - \mu}{2} + \frac{R_H - \mu}{2}$$

implying

$$\begin{aligned} R_L - \mu &= -(R_H - \mu) \\ \varepsilon_L &= -\varepsilon_H \end{aligned}$$

hence

$$\begin{aligned} \sigma_R^2 &= E\left(\varepsilon^2\right) = \frac{1}{2}\varepsilon_L^2 + \frac{1}{2}\varepsilon_H^2 \\ &= \varepsilon_H^2 = (R_H - \mu)^2 \\ &\Rightarrow \sigma_X = R_H - \mu = -(R_L - \mu) \end{aligned}$$