This document details mathematical proofs of notions that we will cover in class. These proofs are given for interested students who want to expand their statistical skills.

These proofs will not be part of the final exam.

1 Statistical foundations

Proposition 1. *Linear property of the Expectation:* $E(aX+b) = aE(X)+b$ *when* X *is a random variable defined on support* Ω *and* a *and* b *are constants.*

Proof. $E(aX+b) = \int_\Omega (ax+b)f(x)dx = a\int_\Omega xf(x)dx + b = aE(X)+b$. □

Proposition 2. *Property of the variance:* $V(X) = E[(X-E(X))^2] = E(X^2) - E(X)^2$ *in which* X *is a random variable defined on support* Ω.

Proof.

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

□

Proposition 3. *Scaling property of the variance:* $V(aX) = a^2V(X)$ *in which* X *is a random variable defined on support* Ω *and* a *is a constant.*

Proof.

$$\begin{aligned} V(aX) &= E(a^2X^2) - a^2E(X)^2 \\ &= a^2(E(X^2) - a^2E(X)^2) = a^2V(X). \end{aligned}$$

□

Proposition 4. *Let us consider* T *random variables* $\{Y_t\}_{t=1}^T \sim_{i.i.d.} N(\mu, \sigma^2)$. *Then* $\bar{Y} = \frac{\sum_{t=1}^T Y_t}{T}$ *is an unbiased estimator of* μ *(i.e.* $E(\bar{Y}) = \mu$*).*

Proof.

$$\begin{aligned} E(\bar{Y}) &= E(\sum_{t=1}^{T} Y_t/T) \\ &= \frac{1}{T}\sum_{t=1}^{T} E(Y_t) \\ &= \mu \end{aligned}$$

□

Proposition 5. *Let us assume T random variables $\{Y_t\}_{t=1}^T \sim_{i.i.d.} (\mu, \sigma^2)$. Then $\hat{\sigma}^2 = \frac{\sum_{t=1}^T (Y_t - \bar{Y})^2}{T-1}$ is an unbiased estimator of σ^2.*

Proof.

$$\begin{aligned} E(\hat{\sigma}^2) &= (T-1)^{-1} E(\sum_{t=1}^{T}((Y_t-\mu)+(\mu-\bar{Y}))^2) \\ &= (T-1)^{-1} E(\sum_{t=1}^{T}[(y_t-\mu)^2+(\mu-\bar{Y})^2+2(\mu-\bar{Y})(Y_t-\mu)]) \\ &= (T-1)^{-1}(E[\sum_{t=1}^{T}(y_t-\mu)^2]+E[\sum_{t=1}^{T}(\mu-\bar{Y})^2]-2E[(\mu-\bar{Y})\sum_{t=1}^{T}(\mu-y_t)]) \\ &= (T-1)^{-1}(T\sigma^2+TE[(\mu-\bar{Y})^2]-2TE[(\mu-\bar{Y})(\mu-\bar{Y})]) \\ &= (T-1)^{-1}(T\sigma^2-E[T(\mu-\bar{Y})^2]) = (T-1)^{-1}(T-1)\sigma^2 \end{aligned}$$

where the last equality holds because $V(\bar{Y}) = E[(\mu-\bar{Y})^2] = \frac{\sigma^2}{T}$. In fact $V(\bar{Y}) = \frac{1}{T^2}\sum_{t=1}^T V(Y_t) = \frac{\sigma^2}{T}$. □

Proposition 6. *Law of iterated expectations: $E_X(X) = E_Y(E_X(X|Y))$ in which X and Y are random variables (with support Ω_x and Ω_y).*

Proof.

$$\begin{aligned}
E_X(X) &= \int_{\Omega_x} x f(x) dx \\
&= \int_{\Omega_x} x \int_{\Omega_y} f(x,y) dy dx \\
&= \int_{\Omega_y} \int_{\Omega_x} x f(x,y) dx dy \\
&= \int_{\Omega_y} [\int_{\Omega_x} x f(x|y) dx] f(y) dy \\
&= E_Y(E_X(X|Y)).
\end{aligned}$$

□

Proposition 7. *Let us assume M and N random variables* $\{X_i\}_{i=1}^m$,$\{Y_j\}_{j=1}^n$. *The covariance of* $Z_1 = \sum_{i=1}^m X_i$ *and* $Z_2 = \sum_{j=1}^n Y_i$ *is given by* $Cov(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_i) = \sum_i \sum_j Cov(X_i, Y_j)$.

Proof. We start by showing that $Cov(\sum_{i=1}^m X_i, Y_1) = \sum_{i=1}^m Cov(X_i, Y_1)$.

$$\begin{aligned}
Cov(\sum_{i=1}^m X_i, Y_1) &= E([\sum_{i=1}^m X_i] Y_1) - E([\sum_{i=1}^m X_i]) E(Y_1) \\
&= \sum_{i=1}^m E(X_i Y_1) - E(X_i) E(Y_1), \\
&= \sum_{i=1}^m Cov(X_i, Y_1).
\end{aligned}$$

Since the covariance is symmetric (i.e. $Cov(X,Y) = Cov(Y,X)$), the converse result also holds true, that is $Cov(X_1, \sum_{j=1}^n Y_j) = \sum_{j=1}^n Cov(X_1, Y_j)$.

$$\begin{aligned}
Cov(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_i) &= \sum_{i=1}^m Cov(X_i, \sum_{j=1}^n Y_i), \\
&= \sum_{i=1}^m \sum_{j=1}^n Cov(X_i, Y_j)
\end{aligned}$$

□

Corollary 1. *Let us assume M random variables* $\{X_i\}_{i=1}^m$. *The sum of these random variables is denoted by* $Z = \sum_{i=1}^m X_i$. *Then,* $Var(Z) = \sum_{i=1}^m \sum_{j=1}^m Cov(X_i, X_j)$. *Note also that* $Z = \sum_{i=1}^m X_i = \mathbb{1}_m' X$ *in which* $\mathbb{1}_m = (1, 1, \ldots, 1)' \in \Re^{mx1}$. *So the result implies that* $Var(\mathbb{1}_m' X) = \mathbb{1}_m' Var(X) \mathbb{1}_m$.

Proof. Since $\mathrm{Var}(\sum_{i=1}^{m} X_i) = Cov(\sum_{i=1}^{m} X_i, \sum_{j=1}^{m} X_j)$, we just apply proposition 7 to prove the result. $\square$

Proposition 8. *Let us denote by $f_X(x)$ the density function of a random variable X. Assuming that $Z = X - a$ in which a is a constant, then the density function of Z is given by $f(z) = f_X(z + a)$.*

Proof. The cumulative density function (cdf) of Z is given by

$$P[Z \leq z] \quad = \quad P[X \leq z + a].$$

The probability density function is the derivative of the cdf. Therefore, we have

$$\begin{aligned} f(z) = \frac{dP[Z \leq z]}{dz} &= \frac{dP[X \leq z + a]}{d(z + a)} \frac{d(z + a)}{dz}, \\ &= f_X(z + a). \end{aligned}$$

$\square$

Proposition 9. *Let us denote by $f_X(x)$ the density function of a random variable X. Assuming that $Z = \frac{X-a}{b}$ in which a and b are constant with $b > 0$, then the density function of Z is given by $f(z) = f_X(bz + a)b$.*

Proof. The cumulative density function (cdf) of Z is given by

$$P[Z \leq z] \quad = \quad P[X \leq bz + a].$$

The probability density function is the derivative of the cdf. Therefore, we have

$$\begin{aligned} f(z) = \frac{dP[Z \leq z]}{dz} &= \frac{dP[X \leq bz + a]}{d(bz + a)} \frac{d(bz + a)}{dz}, \\ &= b f_X(bz + a). \end{aligned}$$

$\square$

Corollary 2. *Let us assume that $X \sim N(\mu, \sigma^2)$. Its density function is given by $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)$. Then, the random variable $Z = \frac{X-\mu}{\sigma}$ exhibits a density function given by $f(z) = \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{z^2}{2}\big)$, i.e. $Z \sim N(0, 1)$.*

Proof. By applying proposition 9 with $a = \mu$ and $b = \sigma$, we have that $f(z) = \sigma f_X(\sigma z + \mu)$. It leads to the following simplifications:

$$\begin{aligned} f(z) &= \sigma \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\big), \\ &= \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{z^2}{2}\big). \end{aligned}$$

□

1.1 Risk minimization

Proposition 10. ***Mean-variance portfolio:*** *Given a universe of N asset returns $X = (X_1, X_2, \ldots, X_N)'$, we want to find the portfolio that exhibits the smallest variance. We denote the covariance matrix of the N asset returns by $Var(X) = \Sigma$. In practice, we solve the following optimization program: $\hat{\omega} = (\hat{\omega}_1, ..., \hat{\omega}_N)' = argMin_\omega V[X_p]$ such that $\sum_{i=1}^N \omega_i = 1$ and where $X_p = \sum_i^N \omega_i X_i = \omega' X$. Then the optimal portfolio is given by $\hat{\omega} = \frac{\Sigma^{-1}\mathbb{1}}{\mathbb{1}'\Sigma^{-1}\mathbb{1}}$.*

Proof. We denote the optimization function, in which we have already added the Lagrangian to account for the constraint, by $L = \omega'\Sigma\omega - \lambda(\omega'\mathbb{1} - 1)$. First, note that the optimization function is convex with respect to ω since $\frac{d^2 L}{d\omega d\omega} = \mathrm{Var}(X)$ which is definite semi-positive (any covariance matrix is definite semi-positive). Therefore, if it exists a solution, it is a global minimum. Let us find it by taking the derivatives and equaling them to zero:

$$\begin{aligned} \frac{dL}{d\omega} &= 2\Sigma\omega - \mathbb{1}\lambda, \\ \hat{\omega} &= \lambda 2\Sigma^{-1}\mathbb{1}. \\ \frac{dL}{d\lambda} &= 1 - \omega'\mathbb{1} \\ \mathbb{1}'\hat{\omega} &= 1 \\ \hat{\lambda} 2\mathbb{1}'\Sigma^{-1}\mathbb{1} &= 1 \\ \hat{\lambda} &= \frac{1}{2\mathbb{1}'\Sigma^{-1}\mathbb{1}}. \end{aligned}$$

By combining the value of $\hat{\lambda}$ and $\hat{\omega}$, we find that $\hat{\omega} = \hat{\lambda} 2\Sigma^{-1}\mathbb{1} = \frac{\Sigma^{-1}\mathbb{1}}{\mathbb{1}'\Sigma^{-1}\mathbb{1}}$. □

2 Simple linear regression

In this section, we consider a simple linear regression given by

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t.$$

The sum of squared residuals (SSR), i.e. $SSR(\hat{\beta}_1, \hat{\beta}_2) = \sum_{t=1}^{T} \hat{\epsilon}_t^2 = \sum_{t=1}^{T}(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2$, is the standard criterion used to derive the estimators of β_1 and β_2.

Proposition 11. *Minimizing the SSR criterion leads to the OLS estimators given by*

$$\begin{aligned}
\hat{\beta}_1 &= \bar{y} - \hat{\beta}_2 \bar{x}, \\
\hat{\beta}_2 &= \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}.
\end{aligned}$$

Proof. The SSR can be expand as follows:

$$SSR(\beta_1, \beta_2) = \sum_{t=1}^{T}(y_t - \beta_1)^2 + \beta_2^2 \sum_{t=1}^{T} x_t^2 - 2\beta_2 \sum_{t=1}^{T}(y_t - \beta_1)x_t.$$

The SSR is a convex function with respect to β_2. Let us take the derivative with respect to this parameter:

$$\begin{aligned}
\frac{d}{d\beta_2} SSR(\beta_1, \beta_2) &= 2\beta_2 \sum_{t=1}^{T} x_t^2 - 2\sum_{t=1}^{T}(y_t - \beta_1)x_t \quad (= 0), \\
\hat{\beta}_2 &= \frac{\sum_{t=1}^{T}(y_t - \hat{\beta}_1)x_t}{\sum_{t=1}^{T} x_t^2}.
\end{aligned}$$

To find out the estimator, we need an analytical expression for $\hat{\beta}_1$. Using the same expanding strategy as above, we simplify the SSR function as follows,

$$\begin{aligned}
\hat{\beta} &= \text{argmin} \sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2, \\
SSR(\beta_1, \beta_2) &= \sum_{t=1}^{T}(y_t - \beta_2 x_t)^2 + T\beta_1^2 - 2\beta_1 \sum_{t=1}^{T}(y_t - \beta_2 x_t), \\
\frac{d}{d\beta_2} SSR(\beta_1, \beta_2) &= 2T\beta_1 - 2\sum_{t=1}^{T}(y_t - \beta_2 x_t), \\
\hat{\beta}_1 &= \frac{\sum_{t=1}^{T}(y_t - \hat{\beta}_2 x_t)}{T}, \\
&= \bar{y} - \hat{\beta}_2 \bar{x}.
\end{aligned}$$

Plugging our expression of $\hat{\beta}_1$ into the expression of $\hat{\beta}_2$ leads to:

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{t=1}^{T}(y_t - (\bar{y} - \hat{\beta}_2\bar{x}))x_t}{\sum_{t=1}^{T} x_t^2}, \\
\hat{\beta}_2[\sum_{t=1}^{T} x_t^2] &= [\sum_{t=1}^{T}(y_t - \bar{y})x_t] + \hat{\beta}_2 T\bar{x}^2, \\
\hat{\beta}_2 &= \frac{\sum_{t=1}^{T}(y_t - \bar{y})x_t}{(\sum_{t=1}^{T} x_t^2) - T\bar{x}^2}.
\end{aligned}$$

Note that $\sum_{t=1}^{T}(x_t - \bar{x})^2 = \sum_{t=1}^{T} x_t^2 + T\bar{x}^2 - 2T\bar{x}^2 = \sum_{t=1}^{T} x_t^2 - T\bar{x}^2$. The OLS estimator is then equal to

$$\hat{\beta}_2 = \frac{\sum_{t=1}^{T}(y_t - \bar{y})x_t}{\sum_{t=1}^{T}(x_t - \bar{x})^2}.$$

In addition, note that

$$\begin{aligned}
\sum_{t=1}^{T}(y_t - \bar{y})\bar{x} &= \bar{x}[\sum_{t=1}^{T} y_t - T\bar{y}], \\
&= 0.
\end{aligned}$$

So we can add and substract $\bar{x}$ and get:

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{t=1}^{T}(y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} + \frac{\sum_{t=1}^{T}(y_t - \bar{y})\bar{x}}{\sum_{t=1}^{T}(x_t - \bar{x})^2}, \\
&= \frac{\sum_{t=1}^{T}(y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}.
\end{aligned}$$

□

Proposition 12. *Assuming that the linear regression is just about a constant, $y_t = \beta_1 + \epsilon_t$, the OLS estimator is the sample mean, that is*

$$\hat{\beta}_1 = \bar{y}.$$

Proof. There are multiple ways to prove this result.

1. The OLS estimator $\hat{\beta}_1$ of the simple linear regression is given by $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}$ (see proposition 11). By setting $x_t = 0$ for all t since there is no explanatory variable, we get $\hat{\beta}_1 = \bar{y}$.

2. By adding and subtracting $\bar{y}$ to the SSR function, we get

$$\begin{aligned} SSR(\beta_1) &= \sum_{t=1}^{T}(y_t - \bar{y} + \bar{y} - \beta_1)^2, \\ &= \sum_{t=1}^{T}[(y_t - \bar{y})^2 + (\bar{y} - \beta_1)^2 + 2(y_t - \bar{y})(\bar{y} - \beta_1)], \\ &= \sum_{t=1}^{T}(y_t - \bar{y})^2 + \sum_{t=1}^{T}(\bar{y} - \beta_1)^2 + 2(\bar{y} - \beta_1)\sum_{t=1}^{T}(y_t - \bar{y}), \\ &= \sum_{t=1}^{T}(y_t - \bar{y})^2 + \sum_{t=1}^{T}(\bar{y} - \beta_1)^2. \end{aligned}$$

The last expression is minimized when $\beta_1 = \bar{y}$.

3. To find the OLS estimator, we can take the derivative of the SSR criterion and set it to zero. It also leads to $\beta_1 = \bar{y}$.

□

Proposition 13. *The SSR criterion is sensitive to the scale of the dependent variable.*

Proof. We consider a linear regression that is multiplied by a constant k:

$$\begin{aligned} \tilde{y}_t \equiv k y_t &= k\beta_1 + k\beta_2 x_t + k\epsilon_t, \\ &= \tilde{\beta}_1 + \tilde{\beta}_2 x_t + \tilde{\epsilon}_t. \end{aligned}$$

The SSR criterion is given by,

$$\begin{aligned} SSR(\tilde{\beta}_1, \tilde{\beta}_2) &= \sum_{t=1}^{T}\tilde{\epsilon}_t^2, \\ &= \sum_{t=1}^{T}(\tilde{y}_t - \tilde{\beta}_1 - \tilde{\beta}_2 x_t)^2, \\ &= k^2\sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2, \\ &= k^2 SSR(\beta_1, \beta_2). \end{aligned}$$

□

Proposition 14. *Coefficient of determination in a simple regression is equal to the squared correlation: $R^2 = \rho_{xy}$.*

Proof. Note that the estimators of the simple regression are given by

$$\begin{aligned}\hat{\beta}_1 &= \bar{y} - \hat{\beta}_2 \bar{x}, \\ \hat{\beta}_2 &= \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2}.\end{aligned}$$

In addition, note that the squared of the empirical correlation can be expanded as

$$\begin{aligned}\hat{\text{Corr}}(X_t, Y_t) \equiv \rho_{xy} &= \frac{1}{T} \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})^2}\sqrt{\frac{1}{T}\sum_{t=1}^T (y_t - \bar{y})^2}}, \\ \rho_{xy}^2 &= \frac{1}{T^2} \frac{\left(\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})\right)^2}{\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})^2 \frac{1}{T}\sum_{t=1}^T (y_t - \bar{y})^2}, \\ &= \hat{\beta}_2 \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}\end{aligned}$$

The coefficient of determination is given by

$$R^2 = \frac{\sum_{t=1}^T (y_t - \bar{y})^2 - \sum_{t=1}^T \hat{\epsilon}_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

We now develop the SSR to get the following simplification:

$$\begin{aligned}\sum_{t=1}^T \hat{\epsilon}_t^2 &= \sum_{t=1}^T (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2 \\ &= \sum_{t=1}^T (y_t - \bar{y} - \hat{\beta}_2(x_t - \bar{x}))^2, \\ &= \sum_{t=1}^T (y_t - \bar{y})^2 + \hat{\beta}_2^2 (x_t - \bar{x})^2 - 2(y_t - \bar{y})\hat{\beta}_2(x_t - \bar{x}), \\ &= \sum_{t=1}^T (y_t - \bar{y})^2 + \hat{\beta}_2 \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}) - 2\hat{\beta}_2 \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}), \\ &= \sum_{t=1}^T (y_t - \bar{y})^2 - \hat{\beta}_2 \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}).\end{aligned}$$

By plugging the last expression into the R^2 formula, the coefficient of determination is equal to

$$\begin{aligned}R^2 &= \frac{\hat{\beta}_2 \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \\ &= \rho_{xy}^2.\end{aligned}$$

□

Proposition 15. *OLS estimators are unbiased when no-multicollinearity and strict exogeneity (i.e. $E(\epsilon_t|x_1, ..., x_T) = 0$) hold*

Proof. We remind that the OLS estimators are given by

$$\begin{aligned}
\hat{\beta}_1 &= \bar{y} - \hat{\beta}_2 \bar{x}, \\
\hat{\beta}_2 &= \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2}.
\end{aligned}$$

First, let us assume that $E(\hat{\beta}_2|x_1, ..., x_T) = \beta_2$ which means that $E(\hat{\beta}_2) = \beta_2$ since $E(\hat{\beta}_2) = E_x[E(\hat{\beta}_2|x_1, ..., x_T)]$. In such a case, it is easy to show that $E(\hat{\beta}_1|x_1, ..., x_T) = \beta_1$:

$$\begin{aligned}
E(\hat{\beta}_1|x_1, ..., x_T) &= E(\bar{y}|x_1, ..., x_T) - \beta_2 \bar{x}, \\
&= \frac{1}{T} \sum_{t=1}^T E(\beta_1 + \beta_2 x_t + \epsilon_t | x_1, ..., x_T) - \beta_2 \bar{x}, \\
&= \beta_1 + \beta_2 \bar{x} - \beta_2 \bar{x}, \\
&= \beta_1.
\end{aligned}$$

We need to prove unbiasedness of $\hat{\beta}_2$.

$$\begin{aligned}
E(\hat{\beta}_2|x_1, ..., x_T) &= E(\frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2} | x_1, ..., x_T), \\
&= \frac{\sum_{t=1}^T (x_t - \bar{x}) E((y_t - \bar{y})|x_1, ..., x_T)}{\sum_{t=1}^T (x_t - \bar{x})^2}, \\
&= \frac{\sum_{t=1}^T (x_t - \bar{x})(\beta_1 + \beta_2 x_t - \beta_1 - \beta_2 \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}, \\
&= \frac{\sum_{t=1}^T (x_t - \bar{x}) \beta_2 (x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}, \\
&= \beta_2 \frac{\sum_{t=1}^T (x_t - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2}, \\
&= \beta_2
\end{aligned}$$

□

Proposition 16. *Var*$(\hat{\beta}_2|x_1, \ldots, x_T) = \frac{\sigma^2}{\sum_{t=1}^T (x_t - \bar{x})^2}$.

Proof. First, we express the OLS estimator as a weighted sum of the dependent variable as follows,

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2} \\
&= \frac{\sum_{t=1}^T (x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2} (y_t - \frac{1}{T}\sum_{t=1}^T y_t), \\
&= \frac{\sum_{t=1}^T (x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2} y_t - \frac{1}{T}\sum_{t=1}^T y_t \frac{\sum_{t=1}^T (x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2} \\
&= \sum_{t=1}^T [\frac{(x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2} - \frac{1}{T}\frac{\sum_{t=1}^T (x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}] y_t, \\
&= \sum_{t=1}^T \underbrace{[\frac{T(x_t - \bar{x}) - \sum_{t=1}^T (x_t - \bar{x})}{T\sum_{t=1}^T (x_t - \bar{x})^2}]}_{\omega_t} y_t \\
&= \sum_{t=1}^T \underbrace{[\frac{x_t^*}{\sum_{t=1}^T (x_t^*)^2}]}_{\omega_t} y_t \\
&= \sum_{t=1}^T \omega_t y_t,
\end{aligned}$$

in which $x_t^* = x_t - \bar{x}$ and note that $\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x}) = 0$. Since $\mathrm{Cov}(y_t, y_j | x_1, \ldots, x_T) = 0$ (because the error terms are uncorrelated) and $\mathrm{Var}(y_t|x_t) = \sigma^2$ (because of constant variance of the error term), we have that

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_2 | x_1, \ldots, x_T) &= \sum_{t=1}^T \omega_t^2 \mathrm{Var}(y_t | x_1, \ldots, x_T), \\
&= \sigma^2 \sum_{t=1}^T \omega_t^2.
\end{aligned}$$

We now expand the squared of the weight function to proof the result. By setting $\tilde{x}^2 = \sum_{t=1}^T (x_t^*)^2$, it leads to

$$\begin{aligned}
\sum_{t=1}^T \omega_t^2 &= \sum_{t=1}^T [\frac{x_t^*}{\sum_{t=1}^T (x_t^*)^2}]^2, \\
&= \frac{1}{\sum_{t=1}^T (x_t^*)^2}
\end{aligned}$$

The variance of the OLS estimator is thus given by

$$\mathrm{Var}(\hat{\beta}_2 | x_1, \ldots, x_T) = \frac{\sigma^2}{\sum_{t=1}^T (x_t^*)^2}.$$

$\square$

Proposition 17. *OLS Estimators are weighted averages of the error terms.*

Note that the estimators of the simple regression are given by

$$\begin{aligned}\hat{\beta}_1 &= \bar{y} - \hat{\beta}_2 \bar{x}, \\ \hat{\beta}_2 &= \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2}.\end{aligned}$$

We have the following results:

$$\begin{aligned}\bar{y} &= \beta_1 + \beta_2 \bar{x} + \frac{1}{T}\sum_{t=1}^T \epsilon_t, \\ y_t - \bar{y} &= \beta_2 (x_t - \bar{x}) + [\epsilon_t - \frac{1}{T}\sum_{t=1}^T \epsilon_t]\end{aligned}$$

Now, let us remind that $\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x}) = 0$. The OLS estimators are thus equal to

$$\begin{aligned}\hat{\beta}_2 &= \beta_2 + \sum_{t=1}^T \frac{(x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}[\epsilon_t - \frac{1}{T}\sum_{t=1}^T \epsilon_t] \\ &= \beta_2 + \frac{1}{\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})^2}\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})\epsilon_t, \\ \hat{\beta}_1 &= \beta_1 - \frac{\bar{x}}{\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})^2}\frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})\epsilon_t + \frac{1}{T}\sum_{t=1}^T \epsilon_t,\end{aligned}$$

We conclude that the central limit theorem applies if $\{z_t\}$ with $z_t = (x_t - \bar{x})\epsilon_t$ and $\{\epsilon_t\}$ are i.i.d. random variables with finite first two moments.

Proposition 18. *The slope estimator is given by*

$$\hat{\beta}_2 | X \sim N(\beta_2, \frac{\sigma^2}{\sum_t (x_t - \bar{x})^2}) \tag{1}$$

If we replace the true variance by its estimator, then the distribution of $\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^T (x_t - \bar{x})^2}}} | X \sim t(T-2)$.

Proof. First, we remind that $Z \equiv (T-2)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum_t \hat{\epsilon}_t^2}{\sigma^2} \sim \chi^2(T-2)$. The conditional distribution of $\hat{\beta}_2$ given $\hat{\sigma}^2$ leads to

$$\hat{\beta}_2 | X \sim N(\beta_2, \frac{\sigma^2}{\sum_{t=1}^T (x_t - \bar{x})^2}), \tag{2}$$

$$\hat{\beta}_2 | X, \hat{\sigma}^2 \sim N(\beta_2, \frac{\sigma^2 \hat{\sigma}^2}{\hat{\sigma}^2 \sum_{t=1}^T (x_t - \bar{x})^2}), \tag{3}$$

$$\frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}} | X, \hat{\sigma}^2 \sim N(0, \frac{\sigma^2}{\hat{\sigma}^2}\frac{(T-2)}{(T-2)\sum_{t=1}^T (x_t - \bar{x})^2}). \tag{4}$$

To simplify the notation, let us denote $\frac{1}{\sum_{t=1}^{T}(x_t-\bar{x})^2}=f(x)$. We have that

$$\begin{aligned}
\frac{\hat{\beta}_2-\beta_2}{\sqrt{\hat{\sigma}^2}}|X,Z &\sim N(0,Z^{-1}(T-2)f(x)), &(5)\\
\frac{\hat{\beta}_2-\beta_2}{\sqrt{\hat{\sigma}^2 f(x)}}|X,Z &\sim N(0,Z^{-1}\underbrace{(T-2)}_{v}). &(6)
\end{aligned}$$

Note that $Z\sim\chi^2(v)$ and its density is given by $f(z)=\frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})}z^{\frac{v}{2}-1}\exp(-\frac{z}{2})$ in which v stands for the degree of freedom (here $v=T-2$). Before integrating out the chi-square, let us remind a very important trick to integrate densities out. Since the chi-square has a proper pdf, it integrates to one (i.e. $\int f(z)dz=1$). We conclude that

$$\int z^{\frac{v}{2}-1}\exp(-\frac{z}{2})dz = 2^{\frac{v}{2}}\Gamma(\frac{v}{2}).$$

We will use this result to integrate out the kernel of a Gamma distribution: $\int z^{\alpha-1}\exp(-z\beta)dz=\beta^{-\alpha}\Gamma(\alpha)$. Integrating out Z, the distribution of $Y|X\equiv\frac{\hat{\beta}_2-\beta_2}{\sqrt{\hat{\sigma}^2 f(x)}}|X$ is given by

$$\begin{aligned}
f(y|X) &= \int f(y|X,z)f(z)dz,\\
&= (2\pi v)^{-\frac{1}{2}}\frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})}\int z^{\frac{v+1}{2}-1}\exp(-\frac{zy^2}{2v}-\frac{z}{2})dz,\\
&= (2\pi v)^{-\frac{1}{2}}\frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})}\int z^{\frac{v+1}{2}-1}\exp(-z\frac{[\frac{y^2}{v}+1]}{2})dz,\\
&= (2\pi v)^{-\frac{1}{2}}\frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})}(\frac{[\frac{y^2}{v}+1]}{2})^{-\frac{v+1}{2}}\Gamma(\frac{v+1}{2}),\\
&= (\pi v)^{-\frac{1}{2}}\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})}([\frac{y^2}{v}+1])^{-\frac{v+1}{2}}
\end{aligned}$$

The final expression is the density of a student distribution with degree of freedom equal to $v=T-2$. So, we have proven that $Y|X\sim t(T-2)$ in which $Y=\frac{\hat{\beta}_2-\beta_2}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^{T}(x_t-\bar{x})^2}}}$.

□

3 Multiple linear regression

In this section, we consider a multiple linear regression given by

$$\begin{aligned}
y_t &= \sum_{i=1}^{K}\beta_i x_{t,i}+\epsilon_t,\\
&= x_t'\beta+\epsilon_t,
\end{aligned}$$

in which $\beta = (\beta_1, \ldots, \beta_K)'$ and $x_t = (x_{t,1}, \ldots, x_{t,K})'$. The sum of squared residuals (SSR), i.e. $SSR(\hat{\beta}) = \sum_{t=1}^{T} \hat{\epsilon}_t^2 = \sum_{t=1}^{T} (y_t - x_t'\hat{\beta})^2$, is the standard criterion used to derive the estimators of β. Note that we can write the linear regression in a matrix expression as follows

$$y = X\beta + \epsilon,$$

where $X = \begin{pmatrix} x_1' \\ x_2' \\ \ldots \\ x_T' \end{pmatrix} = \begin{pmatrix} x_{1,1} & \ldots & x_{1,K} \\ x_{2,1} & \ldots & x_{2,K} \\ \ldots & & \\ x_{T,1} & \ldots & x_{T,K} \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_T \end{pmatrix}$ and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \ldots \\ \epsilon_T \end{pmatrix}$.

Proposition 19. *The OLS estimator is given by* $\hat{\beta} = (X'X)^{-1}X'y$.

Proof. We first remind that $\frac{d(a'\beta)}{d\beta} = a$ and that $\frac{d(\beta' A\beta)}{d\beta} = 2A\beta$ in which A is a matrix and a is a vector. We want to minimize the sum of squared residuals given by $\text{Argmin}_{\beta}\epsilon'\epsilon = \sum_{t=1}^{T} \epsilon_t^2$. It leads to

$$\begin{aligned} \epsilon'\epsilon &= (y - X\beta)'(y - X\beta), \\ &= y'y + \beta'(X'X)\beta - 2y'X\beta \\ \frac{d\epsilon'\epsilon}{d\beta} &= 2(X'X)\beta - 2X'y \\ \hat{\beta} &= (X'X)^{-1}X'y. \end{aligned}$$

The solution minimizes the SSR function since $\frac{d^2 e'e}{d\beta^2} = (X'X)$ is a definite positive matrix. Note that the proof implies that $X'y - (X'X)\hat{\beta} = X'(y - X\hat{\beta}) = X'e = 0$. For instance, if the regression exhibits a constant (i.e. the first explanatory variable is fixed $x_{t,1} = 1$ for all t), we have that $\sum_{t=1}^{T} e_t = 0$. So, the error term average is equal to zero. □

Proposition 20. *If the no-multicollinearity assumption and the strict exogeneity assumption hold, then the OLS estimator are unbiased.*

Proof. Note that the OLS estimators can be equivalently expressed as

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y, \\ &= (X'X)^{-1}X'(X\beta + \epsilon), \\ &= \beta + (X'X)^{-1}X'\epsilon. \end{aligned}$$

The strict exogeneity assumption implies that $E(\epsilon|X) = 0$. It means that any random variable such as $Z = f(X)\epsilon$, in which $f(X)$ is a function of the explanatory variable, has an expectation equal to zero because $E(Z|X) = E(f(X)\epsilon|X) = f(X)E(\epsilon|X) = 0$. For the OLS estimator, we apply this property with $f(X) = (X'X)^{-1}X'$ and we have that

$$\begin{aligned} E(\hat{\beta}|X) &= \beta + E((X'X)^{-1}X'\epsilon|X), \\ &= \beta. \end{aligned}$$

□

Proposition 21. *The ridge estimator, given by $\hat{\beta}_R = (X'X + \lambda I_K)^{-1}X'y$, minimizes the following penalized criterion $\hat{\beta} = arg\ min_\beta \sum_{t=1}^T \epsilon_t^2 + \lambda \sum_{i=1}^K \beta_i^2$.*

Proof. We first remind that $\frac{d(a'\beta)}{d\beta} = a$ and that $\frac{d(\beta' A \beta)}{d\beta} = 2A\beta$ in which A is a matrix and a is a vector. We want to minimize the penalized sum of squared residuals that can be simplified as

$$\begin{aligned} PSSR(\beta) &= \sum_{t=1}^T \epsilon_t^2 + \lambda \sum_{i=1}^K \beta_i^2, \\ &= \epsilon'\epsilon + \lambda\beta'\beta, \\ &= (y - X\beta)'(y - X\beta) + \lambda\beta'\beta, \\ &= y'y + \beta'(X'X + \lambda I_K)\beta - 2y'X\beta. \end{aligned}$$

Taking the derivative with respect to β and equalling it to zero leads to the ridge estimator:

$$\begin{aligned} \frac{dPSSR(\beta)}{d\beta} &= 2(X'X + \lambda I_K)\beta - 2X'y \\ \hat{\beta}_R &= (X'X + \lambda I_K)^{-1}X'y. \end{aligned}$$

The ridge estimator minimizes the PSSR function since $\frac{d^2 PSSR(\beta)}{d\beta^2} = (X'X + \lambda I_K)$ is a definite positive matrix. □

4 Limited dependent variables

Proposition 22. *An estimator of the unconditional probability is given by $\hat{p} = \frac{N}{T}$. It is obtained by optimizing $\hat{p} = argmax_p \log f(N, T, p)$ in which $f(N, T, p)$ denotes the density function of the binomial distribution with N successes for T trials.*

Proof. The binomial log-density function is given by

$$\log f(N,T,p) \;=\; \log \begin{pmatrix} N \\ T \end{pmatrix} + N\log(p) + (T-N)\log(1-p).$$

To maximize this function, we take the derivative and we equal it to zero. It leads to

$$\begin{aligned}
\frac{d\log f(N,T,p)}{dp} &= \frac{N}{p} - \frac{(T-N)}{1-p}(\quad = 0),\\
\frac{N}{\hat{p}} &= \frac{(T-N)}{1-\hat{p}},\\
N - N\hat{p} &= (T-N)\hat{p},\\
\hat{p} &= \frac{N}{T}
\end{aligned}$$

□