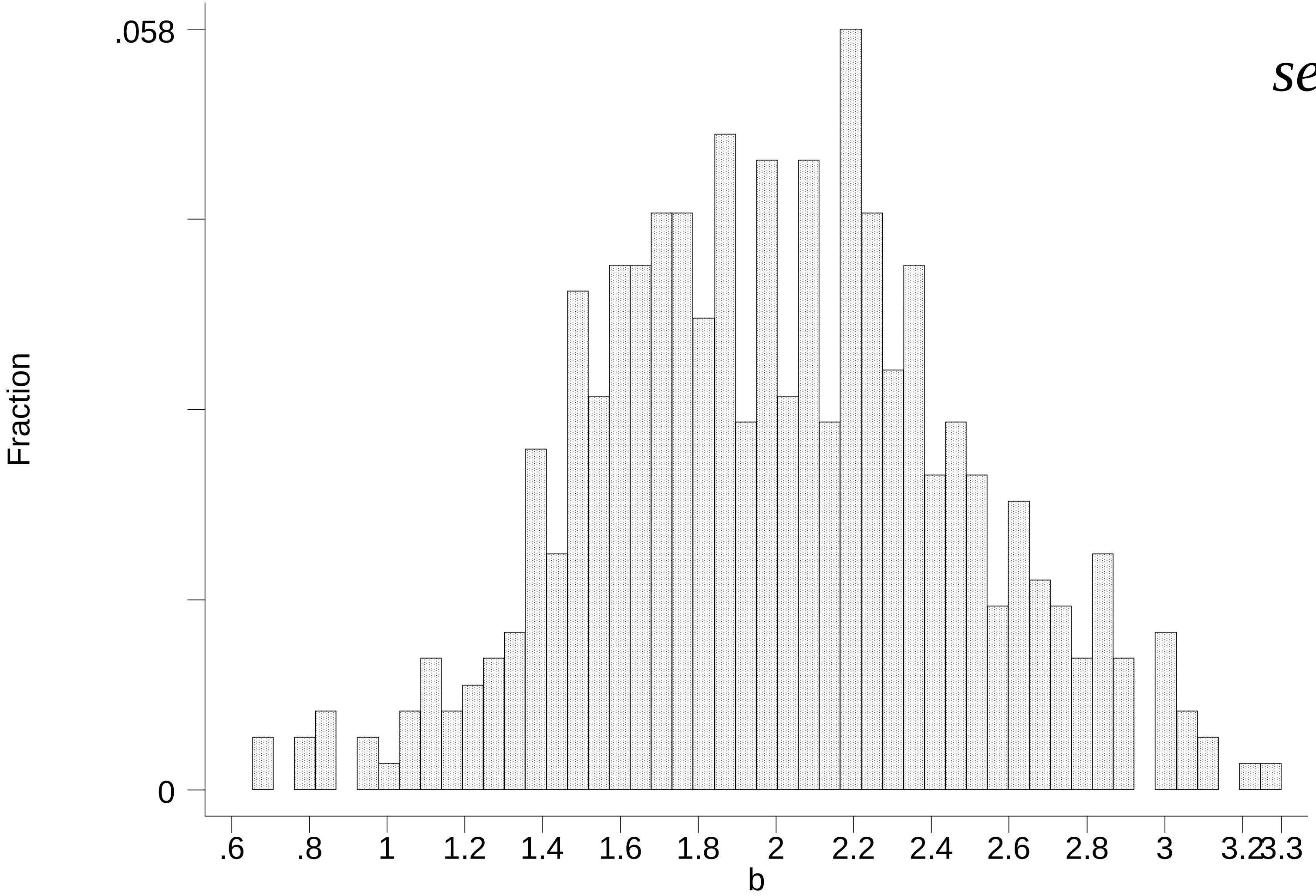
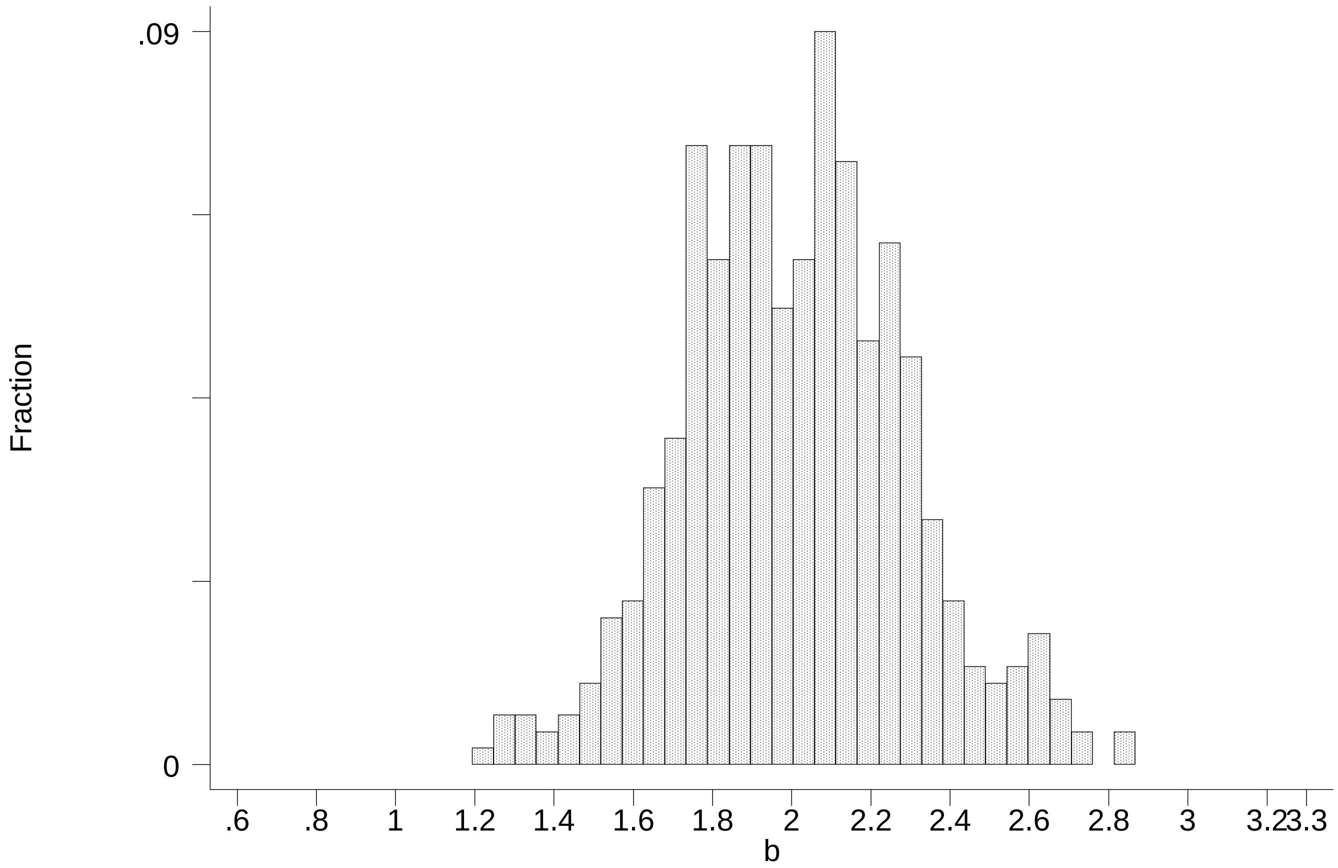


$se(\hat{b}) = 0.48$

N=10 Observations. Model $Y_i = 1 + 2X_i + u_i$

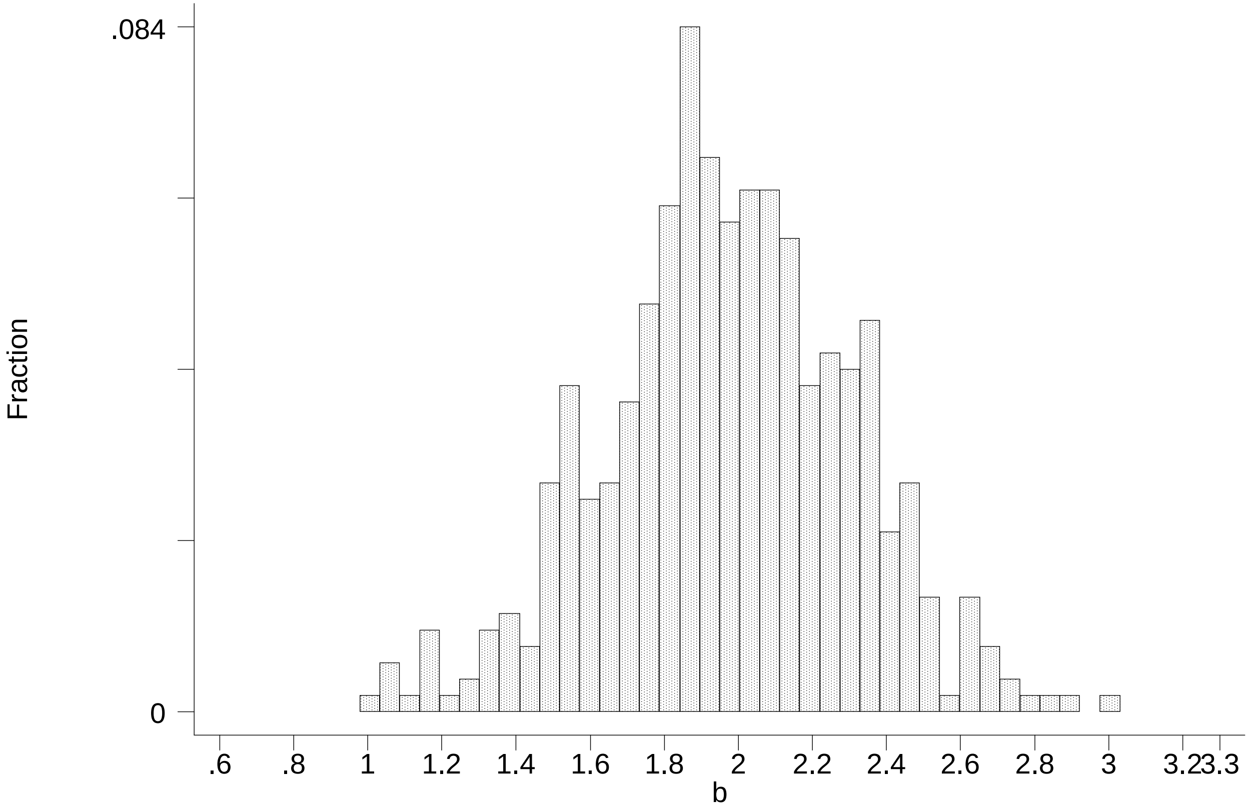
Var(X)= 0.12 Var(u)=0.25



$$se(\hat{b}) = 0.28$$

N=10 Observations. Model $Y_i = 1 + 2X_i + u_i$

Var(X)= 0.12 Var(u)=0.09



$$se(\hat{b}) = 0.33$$

N=20 Observations. Model $Y_i = 1 + 2X_i + u_i$

Var(*X*)= 0.11 Var(*u*)=0.25

Precision and Standard Errors

- We have shown that the OLS estimator (under our assumptions) is unbiased.
- But how sensitive are our results to random changes to our sample? The variance of the estimator is a measure of this.
- Consider first the slope coefficient. As we showed this can be decomposed into two parts: The true value and the estimation error:

$$\hat{b} = b + \frac{\sum_{i=1}^{N} (X_i - \overline{X})(u_i - \overline{u})}{\sum_{i=1}^{N} (X_i - \overline{X})^2}$$

- We also showed that $E(\hat{b} \mid X) = b$

- The definition of the variance is $V(\hat{b} \mid X) = E[(\hat{b}-b)^2 \mid X]$
- Now note that

$$E[(\hat{b}-b)^2 \mid X] = E\left(\left\{ \frac{\sum_{i=1}^{N}(X_i-\bar{X})(u_i-\bar{u})}{\sum_{i=1}^{N}(X_i-\bar{X})^2} \right\}^2 \mid X \right) =$$

$$\frac{1}{\left[\sum_{i=1}^{N}(X_i-\bar{X})^2\right]^2} E\left(\left\{ \sum_{j=1}^{N} \sum_{i=1}^{N}(X_i-\bar{X})(X_j-\bar{X})(u_i-\bar{u})(u_j-\bar{u}) \right\} \mid X \right)$$

- Now note that: $$\left(\sum_{i=1}^{N} z_i\right)^2 = \sum_{j=1}^{N} \sum_{i=1}^{N} z_i z_j$$

- For example $$\left(\sum_{i=1}^{2} z_i\right)^2 = (z_1 + z_2)^2 = z_1^2 + z_2^2 + 2 z_1 z_2 =$$

$$\sum_{j=1}^{2} \sum_{i=1}^{2} z_i z_j$$

- Applying this to the expression above we get that

$$E[(\hat{b}-b)^2 \mid X] = E\left(\left\{ \frac{\sum_{i=1}^{N}(X_i - \overline{X})(u_i - \overline{u})}{\sum_{i=1}^{N}(X_i - \overline{X})^2} \right\}^2 \mid X \right) =$$

$$\frac{1}{\left[\sum_{i=1}^{N}(X_i - \overline{X})^2\right]^2} \left(\left\{ \sum_{j=1}^{N} \sum_{i=1}^{N}(X_i - \overline{X})(X_j - \overline{X}) E[(u_i - \overline{u})(u_j - \overline{u}) \mid X] \right\} \right)$$

- From **Assumption 2**

$$Var(u_i \mid X) \equiv E((u_i - \overline{u})^2 \mid X) = \sigma^2$$

- From **Assumption 3**

$$E[(u_i - \overline{u})(u_j - \overline{u}) \mid X] = 0$$

- Hence

• Hence we obtain the final formula for the variance of the slope coefficient

$$E[(\hat{b}-b)^2 \mid X] =$$

$$\frac{1}{\left[\sum_{i=1}^{N}(X_i-\overline{X})^2\right]^2}\left(\left\{\sum_{j=1}^{N}\sum_{i=1}^{N}(X_i-\overline{X})(X_j-\overline{X})E[(u_i-\overline{u})(u_j-\overline{u}) \mid X]\right\}\right) =$$

$$\frac{1}{\left[\sum_{i=1}^{N}(X_i-\overline{X})^2\right]^2}\left[\sum_{i=1}^{N}(X_i-\overline{X})^2\right]\sigma^2 = \frac{\sigma^2}{\left[\sum_{i=1}^{N}(X_i-\overline{X})^2\right]} = \frac{1}{N}\frac{\sigma^2}{Var(X)}$$

Properties of the variance

- The Variance reflects the precision of the estimation or the sensitivity of our estimates to different samples.
- The higher the variance - the lower the precision.
- The variance increases with the variance of the error term (*noise)*
- The variance decreases with the variance of *X*
- The variance decreases with the sample size.
- The ***standard error*** is the *square root* of the variance:

$$se(\hat{b}) = \frac{1}{\sqrt{N}} \frac{\sigma}{\sqrt{Var(X)}} = \frac{\sigma}{\sqrt{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2}}$$

Efficiency

- An estimator is efficient if within the set of assumptions that we make it provides the most precise estimates in the sense that the variance is the lowest possible *in the class of estimators we are considering.*
- In the exercise sheet I proposed an alternative method of fitting a line: The Wald estimator. You will show that this method also leads to unbiased estimates.
- How do we choose between the OLS estimator and any other unbiased estimator.
- Our criterion is ***efficiency.***

The Gauss Markov theorem

- Given Assumptions 1-4 (see Lecture 2) the Ordinary Least Squares Estimator is a **Best Linear Unbiased Estimator (BLUE)**

- This means that the OLS estimator is the most efficient (least variance) estimator in the class of **linear unbiased** estimators.

Linear Estimators

- An estimator is said to be linear if it can be written as a simple weighted sum of the dependent variable (*Y)*, where the weights do not depend on *Y*.
- Consider the slope coefficient

$$\hat{b} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N} (X_i - \overline{X})^2} = \frac{\sum_{i=1}^{N} (X_i - \overline{X}) Y_i}{\sum_{i=1}^{N} (X_i - \overline{X})^2} = \sum_{i=1}^{N} w_i Y_i$$

where

$$w_i = \frac{(X_i - \overline{X})}{\sum_{i=1}^{N} (X_i - \overline{X})^2}$$

Proof of the Gauss Markov Theorem

- Outline of proof: A. Construct an alternative unbiased and linear estimator. B. Show that the new estimator can never have a smaller variance than the OLS estimator
- Step 1: An alternative unbiased estimator for the slope coefficient will have the form

$$\tilde{b} = \sum_{i=1}^{N} k_i Y_i$$

- Unbiasedness requires $\sum_{i=1}^{N} k_i = 0$ and $\sum_{i=1}^{N} k_i X_i = 1$

- To see why consider the following

$$\hat{b} = \sum k_i Y_i = \sum k_i (a + bX_i + u_i)$$

- For this to equal *b* plus estimation error we must have the conditions overleaf.
- To prove the GAUSS-MARKOV theorem first construct the variance of this arbitrary unbiased linear estimator:

$$Var(\tilde{b}) = E(\sum_{i=1}^{N} k_i u_i - 0)^2 = \quad (by\ assumption\ \ 3)$$

$$\sum_{i=1}^{N} k_i^2 E u_i^2 = \quad (by\ assumption\ \ 2)$$

$$\sigma^2 \sum_{i=1}^{N} k_i^2$$

- Now we need to show that this variance can never be smaller than the OLS variance.
- Consider the expression for the difference in the two variances:

$$\sigma^2 \sum_{i=1}^{N} k_i^2 - \frac{\sigma^2}{\left[\sum_{i=1}^{N} (X_i - \overline{X})^2\right]} = \sigma^2 \sum_{i=1}^{N} k_i^2 \left[1 - \frac{1}{\left[\sum_{i=1}^{N} (X_i - \overline{X})^2\right] \sum_{i=1}^{N} k_i^2} \right] =$$

$$\sigma^2 \left(\sum_{i=1}^{N} k_i^2 \right) \left[1 - \frac{\left(\sum_{i=1}^{N} k_i \, X_i \right)\left(\sum_{i=1}^{N} k_i \, X_i \right)}{\left[\sum_{i=1}^{N} (X_i - \overline{X})^2\right] \sum_{i=1}^{N} k_i^2} \right] = \sigma^2 \left(\sum_{i=1}^{N} k_i^2 \right) \left[1 - \left(correlation(X_i, k_i) \right)^2 \right] \geq 0$$