1. According to my output, it indicates that the pandas and scipy packages are not biased, as their results match your custom function calculations. Since my custom function and scipy package provide consistent results, it indicates that scipy and pandas are correctly implementing unbiased estimators. My results show that scipy correctly adjusts for sample size and provides results that align with your custom calculations. Both pandas and scipy use well-established statistical definitions and algorithms for moments. For skewness and kurtosis, scipy adjusts the calculations to be unbiased estimators, aligning with standard definitions.

2.

a. Comparing OLS and MLE with Normality Assumption

	OLS Results	**MLE with Normality Assumption**
Intercept (β_0)	-0.0874	-0.0874
Slope (β_1)	0.7753	0.7753
Sigma (σ)	1.0038	14.1953
R-square (R^2)	0.3456	0.3456

Comparison:

- **Beta Values**: The OLS and MLE with normality assumption yield identical beta values (**Intercept β_0** and **Slope β_1**).
- **Standard Deviation**: The MLE with Normality Assumption: 14.1953 The MLE with normality assumption has a significantly larger σ compared to OLS. This suggests that the MLE with normality assumption has a higher estimated residual variability. In contrast, OLS provides a smaller estimate of the residual variability.
- **Consistency in R^2**: Since R^2 is the same for both OLS and MLE with normality

Conclusion: I think OLS fits better due to the reason below

- **Residual Variability: The OLS model provides a lower estimate for σ, indicating less residual variability compared to MLE with normality assumption.**
- **Robustness to Outliers:** The higher variability in the normality assumption model's sigma indicates issues with robustness or outliers, the OLS model's lower sigma might be more appropriate.
- **Goodness of Fit (R^2):** Since the **R^2** values are nearly identical, this metric alone does not strongly favor one model over the other.

b. Comparing MLE under Normality and T Distribution Assumptions

	MLE with Normality Assumption	MLE under T Distribution Assumption
Intercept (β_0)	-0.0874	-0.0881
Slope (β_1)	0.7753	0.7722
Sigma (σ)	14.1953	1.1898
R-square (R^2)	0.3456	0.3456

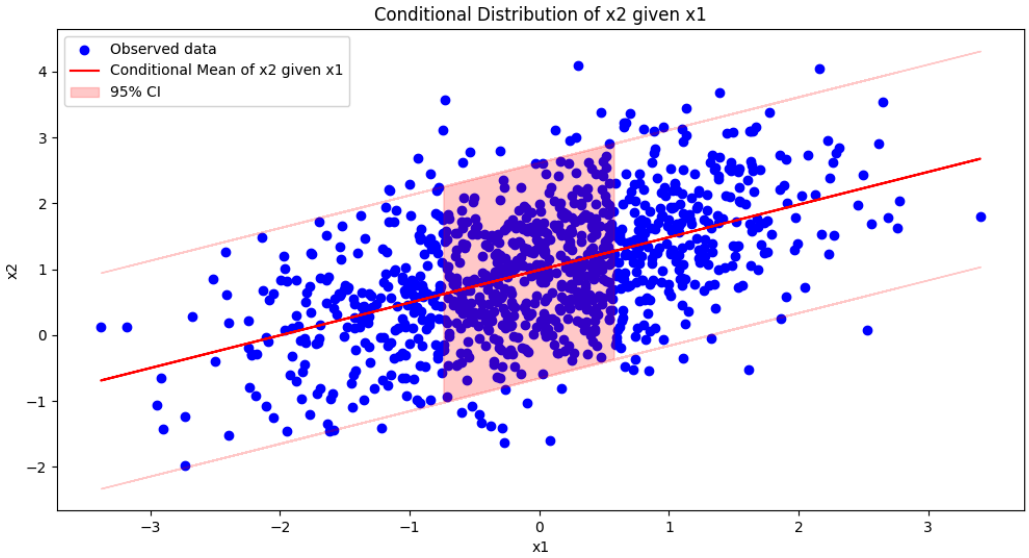
Comparison:

- **Intercept (β_0) and Slope (β_1):** The intercept and slope values are quite similar under both assumptions, with very small differences between them. This suggests that the regression line is quite stable across different assumptions.
- **Sigma (σ):**
    - **The Normality Assumption Model** has a much larger sigma, which might indicate higher variability or less robustness to outliers.
    - **T Distribution Assumption Model** has a much smaller sigma, which suggests that it might be better at handling heavy tails or outliers in the residuals.

Conclusion: I think T-distribution fits better due to the reason below

- **Robustness to Outliers:** The higher variability in the normality assumption model's sigma indicates issues with robustness or outliers, the T-distribution model's lower sigma might be more appropriate.
- **Residual Analysis:** Also, the larger value of sigma σ in the Normality Assumption Model typically indicates that the residuals have a larger spread. So, the T distribution model might be a better choice.
- **Goodness of Fit (R^2):** Since the R^2 values are nearly identical, this metric alone does not strongly favor one model over the other.

c.



d. Extra credit

To derive the maximum likelihood estimators (MLE) for β and σ^2 in the linear regression model

$Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I)$, follow these steps below:

1. Find the MLE for β

Given the assumptions, the residuals ϵ are normally distributed with mean 0 and variance σ^2.

The probability density function (pdf) for each observation y_i, given x_i is:

$$f(y_i \mid x_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right)$$

For *n* observations, the joint likelihood function is:

$$L(\beta, \sigma^2 \mid Y, X) = \prod_{i=1}^{n} f(y_i \mid x_i, \beta, \sigma^2)$$

$$= L(\beta, \sigma^2 \mid Y, X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right)$$

take the natural logarithm of the likelihood function

$$\ell(\beta, \sigma^2 \mid Y, X) = \log L(\beta, \sigma^2 \mid Y, X)$$

$$= \ell(\beta, \sigma^2 \mid Y, X) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(y_i - x_i\beta)^2}{2\sigma^2}$$

$$= \ell(\beta, \sigma^2 \mid Y, X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i\beta)^2$$

take the partial derivative of the log-likelihood

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial}{\partial \beta}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i\beta)^2\right]$$

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(y_i - x_i\beta)$$

Setting this equal to zero

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i x_i \beta = 0$$

$$X^T Y = X^T X \beta$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

2. Find the MLE for σ^2

Substitute $\widehat{\beta}$ back into the log-likelihood function:

$$\ell(\beta, \sigma^2 \mid Y, X) = -\tfrac{n}{2} \log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i \hat{\beta})^2$$

Take partial derivative with respect to σ^2**:**

$$\tfrac{\partial \ell}{\partial \sigma^2} = -\tfrac{n}{2\sigma^2} + \tfrac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (y_i - x_i \hat{\beta})^2$$

Set equation above to zero:

$$-\tfrac{n}{2\sigma^2} + \tfrac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (y_i - x_i \hat{\beta})^2 = 0$$

$$\hat{\sigma}^2 = \tfrac{1}{n} \sum_{i=1}^{n} (y_i - x_i \hat{\beta})^2$$

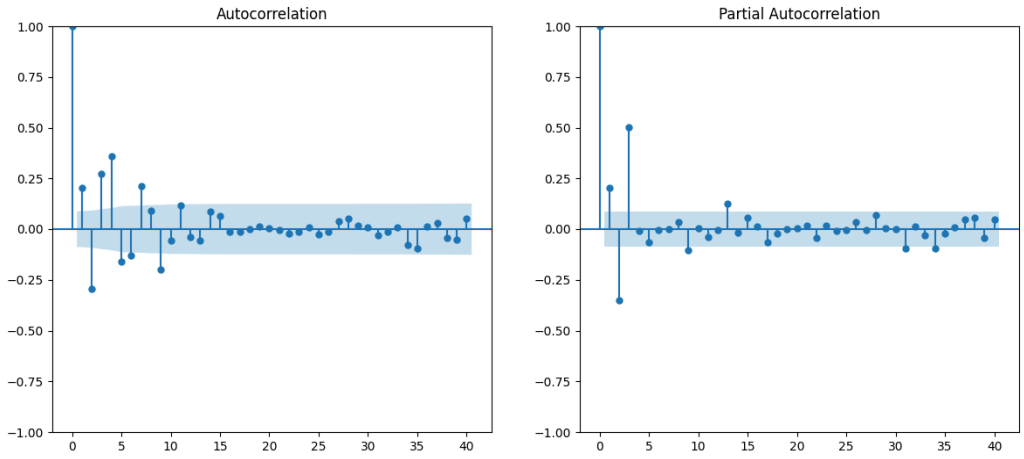
Summary of MLEs:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\hat{\sigma}^2 = \tfrac{1}{n} \sum_{i=1}^{n} (y_i - x_i \hat{\beta})^2 = \frac{1}{n} \sum_{i=1}^{n} r_i^2$$

3. The PACF plot shows a sharp cut-off after a few lags, it suggests that an AR model might be a good fit. Specifically, there's a sharp cutoff at lag 3, it suggests that an AR(n) model will be the best fit.

a.



b. **Best fit**

i. Reason 1: As we can see from the graph above, the PACF has a significant spike at lag 3 and drops off afterward, it suggests an **AR(3)** model is the best fit.

ii. To be more precise, let's use code to calculate and verify our assumption.

	Best AR Model: AR(3)	**Best MA Model: MA(3)**
AIC	**1428.26**	**1536.87**
BIC	**1449.31**	**1557.94**

Comparison:

- AIC: The AR(3) model has a lower AIC (1428.26) compared to the MA(3) model (1536.87). This suggests that AR(3) fits the data better in terms of AIC.
- BIC: The AR(3) model also has a lower BIC (1449.31) compared to the MA(3) model (1557.94). This indicates that AR(3) is preferable in terms of BIC as well.

Conclusion: To choose the best model overall between AR and MA models, I compare the best models from each type using their AIC and BIC values. The AR(3) model is the best overall fit when considering both AIC and BIC values. It has lower values for both criteria compared to the MA(3) model.