Chapter 18: Linear Model

18.1 INTRODUCTION

The **linear model** is fundamental to econometrics. These notes discuss the probabilistic assumptions underlying the linear model and how they allow us to understand the properties of the Ordinary Least Squares (OLS) estimator.

18.2 THE MODEL

We specify a model describing how the data

$$y, X \in \mathbb{R}^{n\times(K+1)}$$

were generated. We assume there is an underlying random mechanism. The observed data represents one realization from an infinite set of potential outcomes. The model focuses on how y responds to X, with minimal assumptions about X itself. We assume throughout that

$$\text{rank}(X) = K < n.$$

This assumption is immediately verifiable from the data.

Assumption A

1. **Linearity in Parameters:** There exists a vector $\beta \in \mathbb{R}^K$ such that with probability one,

   $$E(y|X) = X\beta.$$

   Intuition: The expected value of the dependent variable y, given the independent variables X, is a linear function of X. The coefficients of this linear function are represented by the vector β. *Real world example:* Suppose y is wage and X includes education and experience. β would contain coefficients for education and experience's respective impacts on wages.

2. **Conditional Variance:** There exists a positive definite finite $n \times n$ matrix $\Sigma(X)$ such that with probability one,

   $$\text{var}(y|X) = \Sigma(X).$$

   Intuition: The variance of y given X is not necessarily constant. It is captured by the matrix $\Sigma(X)$, which can depend on X. *Real world example:* The variability of wages might be higher for individuals with higher education levels, meaning that the variance of wage is not constant and depends on X.

3. **Conditional Normality:** Conditional on X, we have with probability one,

   $$y \sim N(X\beta, \Sigma(X)).$$

   Intuition: Given X, the dependent variable y follows a normal distribution with mean $X\beta$ and variance $\Sigma(X)$. *Real world example:* If we look at individuals with the same levels of education and experience, their wages will be normally distributed around a mean wage determined by their specific education and experience.

4. **Conditional Normality with Homoskedasticity and no Autocorrelation:** Conditional on X, we have with probability one

$$y \sim N(X\beta, \sigma^2 I)$$

Intuition: This is a simplification of assumption 3. Now the variance is a constant (σ^2) multiplied by the identity matrix. This is a special case, important for OLS.

Regression Model in Familiar Form

We can rewrite the regression model using an error term. Define

$$\varepsilon = y - X\beta = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Then,

$$y = X\beta + \varepsilon.$$

Assumptions A1 and A2 are equivalent to

$$E(\varepsilon|X) = 0,$$

$$E(\varepsilon\varepsilon^T|X) = \Sigma(X).$$

with probability one.

The model is often stated in terms of the unobservable error ε rather than directly on the observable y. The assumptions about ε are weak (and incomplete, as we only specify some moments).

Assumption A1 Explained:

$E(\varepsilon|X) = 0$ can be expressed as (local notation change)

$$E\left(\varepsilon_i|X_1, \ldots, X_n\right) = 0, \text{ where } X_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{pmatrix} \in \mathbb{R}^K,$$

with probability one.

Intuition: This condition means that the expected value of each error term ε_i is zero, regardless of the values of the regressors X. The error term is uncorrelated with both past, present, and future values of the regressors. This holds if we have fixed design (e.g. $x_i = i$) or with i.i.d. data (x_i, ε_i). In a time series context, this condition is called **strong exogeneity**, and implies, among other conditions, no lagged dependent variable included in X.

Condition 18.3 Explained:

The condition $E(\varepsilon\varepsilon^T|X) = \Sigma(X)$ implies that the conditional variance is finite with probability one. The $n \times n$ covariance matrix $\Sigma(X)$ can depend on the covariates X and allows for correlation between error terms.

Special Cases for $\Sigma(X)$:

1. **Homoskedastic and Uncorrelated Case:**

$$\Sigma(X) = \sigma^2 I_n,$$

where I_n is the $n \times n$ identity matrix. This means the errors have constant variance (σ^2) and are uncorrelated with each other.

2. **Heteroskedastic Case:**

$$\Sigma(X) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix},$$

where the σ_i^2 may all be distinct. This allows for different variances for each observation.

3. **Clustered or Block-Diagonal Case:**

$$\Sigma(X) = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_p \end{pmatrix},$$

where B_j are non-zero blocks with dimensions $n_j \times n_j$, $j = 1, \ldots, p$, such that $\sum_{j=1}^{p} n_j = n$, and the off-diagonal elements are zero. This represents situations where groups of observations have correlated errors, but errors are uncorrelated across groups.

4. **Time Series Case (Autocorrelation):** $\Sigma(X)$ may have all entries non-zero, but the entries get smaller as they move further away from the diagonal. This represents situations where closer observations in time are more correlated than those farther apart.

Exercises

Exercise 1

Solution 1

Explain the meaning of the assumption $\text{rank}(X) = K < n$ in the context of the linear model. What would be the implications if $\text{rank}(X) < K$?

Exercise 2

Solution 2

Describe the difference between assumptions A1 and A3 regarding the distribution of y conditional on X.

Exercise 3

Solution 3

Given the model $y = X\beta + \varepsilon$, derive the expression for $E(\varepsilon\varepsilon^T|X)$ under the assumption of homoskedasticity and no autocorrelation.

Exercise 4

Solution 4

Provide an example of a real-world scenario where the assumption $E(\varepsilon|X) = 0$ is likely to be violated. Explain why.

Exercise 5

Solution 5

Explain the concept of **strong exogeneity** in a time series context and how it relates to the assumption $E(\varepsilon|X) = 0$.

Exercise 6

Solution 6

Consider a dataset of student test scores (y) and hours studied (X). Formulate a linear model, and discuss potential sources of heteroskedasticity in the error term.

Exercise 7

Solution 7

What does it mean for the matrix $\Sigma(X)$ to be positive definite in Assumption A2? What are the implications if $\Sigma(X)$ is not positive definite?

Exercise 8

Solution 8

Explain the difference between the clustered and heteroskedastic cases for the covariance matrix $\Sigma(X)$.

Exercise 9

Solution 9

Suppose you are analyzing panel data on firm profits over several years. How might you expect the error terms to be correlated, and which form of $\Sigma(X)$ would be appropriate?

Exercise 10

Solution 10

Explain how Assumption A1, $E(y|X) = X\beta$, implies that the relationship between y and X is linear in parameters. Provide an example of a model that is linear in parameters and one that is not.

Exercise 11

[Solution 11](#)

If y represents household consumption and X includes household income and household size, explain how you would interpret the elements of the β vector.

Exercise 12

[Solution 12](#)

Explain why we model the unobservable ϵ instead of directly modeling the conditional distribution of y given X.

Exercise 13

[Solution 13](#)

Describe a situation where you would expect the errors in a linear regression model to be both heteroskedastic and autocorrelated.

Exercise 14

[Solution 14](#)

Why is Assumption A, specifically condition 3, considered a strong assumption? What are its advantages?

Exercise 15

[Solution 15](#)

Consider the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. If $E(\varepsilon_i | x_i) = x_i$, does this model satisfy Assumption A1? Explain.

Exercise 16

[Solution 16](#)

Explain the implications of using Assumption A4 instead of Assumption A3 for the properties of the OLS estimator.

Exercise 17

[Solution 17](#)

Explain the intuitive meaning of $E(\varepsilon | X) = 0$.

Exercise 18

[Solution 18](#)

Suppose y_i represents the sales of a store and X_i is the advertising spending on day i. Provide a plausible explanation for correlation between the errors of adjacent days.

Exercise 19

[Solution 19](#)

Give an example of a dataset where the clustered error structure would be a reasonable assumption.

Exercise 20

[Solution 20](#)

How does the assumption of the linear model relate to the existence of an infinity of potential outcomes?

Solutions

Solution 1

[Exercise 1](#)

The assumption $\text{rank}(X) = K < n$ means that the matrix X has full column rank, and the number of columns (K, representing the number of regressors plus the constant term) is less than the number of rows (n, representing the number of observations).

- **Full column rank:** This means that no column in X can be written as a linear combination of the other columns. In other words, there is no perfect multicollinearity among the regressors. Each regressor provides unique information.
- $K < n$**:** This condition ensures that we have more observations than parameters to estimate. This is necessary for degrees of freedom in the estimation.

If $\text{rank}(X) < K$, it means that there is perfect multicollinearity. At least one regressor is a perfect linear combination of the others. This makes it impossible to uniquely identify the individual effects of the regressors (β) because multiple combinations of coefficients would produce the same fitted values. The OLS estimator would not be unique.

Solution 2

[Exercise 2](#)

- **Assumption A1:** $E(y|X) = X\beta$. This assumption only specifies the *conditional mean* of y given X. It states that the expected value of y is a linear function of X. It says nothing about the distribution of y other than its mean.
- **Assumption A3:** $y \sim N(X\beta, \Sigma(X))$. This assumption specifies the *entire conditional distribution* of y given X. It states that y is normally distributed with a mean of $X\beta$ and a variance-covariance matrix of $\Sigma(X)$. A3 implies A1 because if y is conditionally normal, its conditional expectation is the mean of that normal distribution, which is $X\beta$.

In short, A3 is a much stronger assumption than A1. A3 implies A1, but A1 does not imply A3.

Solution 3

[Exercise 3](#)

Under homoskedasticity and no autocorrelation, the variance-covariance matrix of the error term is $\Sigma(X) = \sigma^2 I_n$. Therefore:

$$E(\varepsilon\varepsilon^T|X) = \Sigma(X) = \sigma^2 I_n = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}.$$

This means:

- $E(\varepsilon_i^2|X) = \sigma^2$ for all i (homoskedasticity: constant variance).
- $E(\varepsilon_i\varepsilon_j|X) = 0$ for all $i \neq j$ (no autocorrelation: errors are uncorrelated).

Solution 4

Exercise 4

Consider a model where y is income and X includes years of education. It's possible that omitted variables, such as innate ability, are correlated with both education and income. People with higher innate ability might choose to pursue more education *and* also earn higher incomes, even holding education constant.

If innate ability is omitted, it becomes part of the error term ε. Since innate ability is positively correlated with education (X), we have $E(\varepsilon|X) \neq 0$. Specifically, we would expect $E(\varepsilon|X)$ to be an increasing function of X (higher education is associated with a more positive error term due to higher omitted ability).

Solution 5

Exercise 5

In a time series setting, **strong exogeneity** means that the error term at any time period t is uncorrelated with the regressors in *all* time periods (past, present, and future). Formally:

$$E(\varepsilon_t|X_1, \ldots, X_n) = 0,$$

where X_i is the vector of regressors at time i.

This is a stronger condition than simply requiring contemporaneous exogeneity, which would only require $E(\varepsilon_t|X_t) = 0$. Strong exogeneity rules out feedback from y to future values of X. This relates to the text as mentioned that $E(\epsilon|X) = 0$ can be expressed as $E\left(\varepsilon_i|X_1, \ldots, X_n\right) = 0$.

Solution 6

Exercise 6

Linear model:

$$\text{Test Score}_i = \beta_0 + \beta_1 \text{Hours Studied}_i + \varepsilon_i.$$

Heteroskedasticity: The variance of the error term might not be constant. For example:

- Students who study very little might have test scores clustered close to zero (low variance).
- Students who study a moderate amount might have a wide range of scores (high variance) due to differences in learning ability, test-taking skills, etc.
- Students who study a lot might have scores clustered near the top (low variance).

Thus, $\mathrm{Var}(\varepsilon_i | \text{Hours Studied}_i)$ would not be constant, violating the homoskedasticity assumption.

Solution 7

Exercise 7

A matrix $\Sigma(X)$ is **positive definite** if for any non-zero vector z, the quadratic form $z^T \Sigma(X) z$ is strictly positive:

$$z^T \Sigma(X) z > 0 \text{ for all } z \neq 0.$$

Implications:

1. **Variances are positive:** All diagonal elements of $\Sigma(X)$ (which represent variances) must be positive.
2. **Invertibility:** A positive definite matrix is always invertible (non-singular). This is important for many statistical procedures, including OLS estimation.

If $\Sigma(X)$ is *not* positive definite, several problems arise:

- **Zero or negative variances:** It could imply zero or negative variances for some linear combinations of the error terms, which is nonsensical.
- **Singularity:** $\Sigma(X)$ might be singular (non-invertible), leading to problems in estimation.

Solution 8

Exercise 8

- **Heteroskedastic Case:** The variances of the error terms can be different for *each* observation. The covariance matrix is diagonal, meaning there is no correlation between errors of different observations.

$$\Sigma(X) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

- **Clustered Case:** Observations are grouped into clusters. Within each cluster, the errors can be correlated, and the variances can differ across clusters. However, errors from different clusters are uncorrelated. The covariance matrix is block-diagonal.

$$\Sigma(X) = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_p \end{pmatrix}$$

  Each B_j is a submatrix representing the covariance structure within cluster j.

The key difference is that heteroskedasticity allows for individual variation in variances, while the clustered case allows for group-wise correlation and variance differences.

Solution 9

Exercise 9

In panel data on firm profits, we might expect errors to be correlated *within* each firm over time (autocorrelation). This is because unobserved factors affecting a firm's profitability (e.g., management quality, firm-specific shocks) are likely to persist over time. However, we might assume that errors are uncorrelated *across* different firms.

The appropriate form of $\Sigma(X)$ would be the **clustered** case, where each firm represents a cluster. Each block B_j on the diagonal of $\Sigma(X)$ would represent the covariance matrix of the error terms for firm j. These blocks could potentially have an autocorrelated structure.

Solution 10

Exercise 10

$E(y|X) = X\beta$ implies linearity in parameters because the expected value of y is a weighted sum of the elements of X, where the weights are the *parameters* in β. The parameters (β) enter the equation linearly; they are not raised to any powers, multiplied together, or transformed in any non-linear way.

- **Linear in parameters:** $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}^2 + \varepsilon_i$. Even though x_{i2} is squared, the *parameter* β_2 is not.
- **Not linear in parameters:** $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2^{\beta_3} x_{i2} + \varepsilon_i$. Here the parameter β_2 is raised to the power of parameter β_3 which is a non-linear transformation of parameters.

Solution 11

Exercise 11

The model would be:

$$\text{Consumption}_i = \beta_0 + \beta_1 \text{Income}_i + \beta_2 \text{Household Size}_i + \varepsilon_i.$$

- β_0: The intercept, representing autonomous consumption (consumption when income and household size are zero).
- β_1: The marginal propensity to consume out of income. It represents the expected increase in consumption for a one-unit increase in income, holding household size constant.
- β_2: The expected change in consumption for a one-unit increase in household size, holding income constant.

Solution 12

Exercise 12

We model the unobservable ϵ because it represents the combined effects of all factors influencing y that are *not* included in X. It's a convenient way to capture the inherent randomness and unexplained variation in y. Directly modeling the *conditional* distribution $P(y|X)$ would be very complex and might require strong distributional assumptions that we don't want to, or can't, make. By modeling the error we can focus on $E(y|X)$ instead of the complete distribution, as a first step.

Solution 13

Exercise 13

Consider daily stock prices (y) regressed on some economic indicators (X).

- **Heteroskedasticity:** Volatility in stock markets tends to cluster. Days with large price swings (either positive or negative) are often followed by other days with large swings. Thus, the variance of the error term might be higher on days following large market movements.
- **Autocorrelation:** Unobserved factors influencing stock prices (e.g., news, investor sentiment) can persist for several days. Therefore, the error term on one day is likely to be correlated with the error term on the following day.

Solution 14

Exercise 14

Assumption A, condition 3 ($y \sim N(X\beta, \Sigma(X))$) is strong because it assumes that the conditional distribution of y is *normal*. This is a specific distributional assumption that may not hold in all real-world situations. Many variables are not normally distributed.

Advantages:

- **Simplifies inference:** Normality makes statistical inference (hypothesis testing, confidence intervals) much easier. Many statistical results are derived under the assumption of normality.
- **Maximum Likelihood:** With normality the OLS estimator coincides with the Maximum Likelihood Estimator (MLE).

Solution 15

Exercise 15

No, this model does *not* satisfy Assumption A1. Assumption A1 requires $E(\varepsilon|X) = 0$. In this case, $E(\varepsilon_i|x_i) = x_i$, which is not zero unless $x_i = 0$. The expected value of the error term depends on the value of the regressor, violating the assumption.

Solution 16

Exercise 16

Assumption A4 ($y \sim N(X\beta, \sigma^2 I)$) assumes both normality and homoskedasticity (and no autocorrelation) of the errors. Assumption A3 ($y \sim N(X\beta, \Sigma(X))$) only assumes normality, allowing for a more general covariance structure $\Sigma(X)$.

Implications for OLS:

- **Under A4:** The OLS estimator is the Best Linear Unbiased Estimator (BLUE) and also the Maximum Likelihood Estimator (MLE). Its standard errors are straightforward to calculate.
- **Under A3:** The OLS estimator is *still* unbiased and consistent, but it is no longer necessarily the most efficient (it's not BLUE). We need to use different formulas for standard errors (e.g., robust standard errors) to account for heteroskedasticity or autocorrelation.

Solution 17

Exercise 17

$E(\varepsilon|X) = 0$ means that, on average, the errors are zero for *any* given value of the regressors X. The unobserved factors captured by ε are not systematically related to the regressors. There's no systematic over- or under-prediction for any particular values of X.

Solution 18

Exercise 18

Suppose there's a major, unexpected event (e.g., a product recall or a positive news story) that affects store sales on day i. This event is not included in X_i (advertising spending), so it will be part of the error term ε_i. If the effects of this event linger for several days (e.g., customers continue to avoid the store after the recall or continue to be drawn in by the positive news), then ε_{i+1} will also be affected. This creates a positive correlation between ε_i and ε_{i+1}.

Solution 19

Exercise 19

A dataset of student test scores within classrooms across multiple schools. Students within the same classroom (cluster) are likely to have correlated errors due to shared factors like teacher quality, classroom environment, and peer effects. However, students in different classrooms (different clusters) are less likely to have correlated errors.

Solution 20

Exercise 20

The linear model assumes that the observed data (y, X) is just *one* possible realization from a random process. There's an underlying data-generating process, and we only see one draw from it. There's an infinity of *potential* outcomes that we *could* have observed, but we only observe one. The error term ε captures the difference between the observed outcome and the expected outcome given X, representing the influence of unobserved factors and randomness in this data-generating process.

R Script Examples

R Script 1: Simulating a Linear Model with Homoskedastic Errors

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.2
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(MASS) # For mvrnorm function
```

```

Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

    select
```

```
# Set seed for reproducibility
set.seed(123)

# Number of observations
n <- 100

# Number of regressors (excluding the intercept)
k <- 2

# Generate regressors (X)
X <- matrix(runif(n * k), nrow = n, ncol = k)
X <- cbind(1, X) # Add a column of 1s for the intercept

# True coefficients (beta)
beta <- c(2, -1, 0.5)

# Error variance (sigma^2)
sigma2 <- 4

# Generate errors (epsilon) from a normal distribution with mean 0 and variance sigma2
#   This satisfies the homoskedasticity assumption (Sigma(X) = sigma2 * I)
epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2))

# Generate dependent variable (y)
# y = X %*% beta + epsilon corresponds to the linear model y = Xβ + ε
y <- X %*% beta + epsilon

# Combine into a data frame
data <- data.frame(y = y, X = X[, -1]) # Exclude the intercept column

# Visualize the relationship between y and one of the regressors (X.1)
ggplot(data, aes(x = X.1, y = y)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) + # Add the true regression line
  labs(title = "Simulated Linear Model with Homoskedastic Errors",
       x = "X1",
       y = "y")
```

```
`geom_smooth()` using formula = 'y ~ x'
```

Explanation:

1. **Setup:** We load the necessary libraries (`tidyverse` for data manipulation and visualization, `MASS` for multivariate normal simulation). We set a seed for reproducibility.
2. **Define Parameters:** We define the number of observations (`n`), the number of regressors (`k`), the true coefficient vector (`beta`), and the error variance (`sigma2`).
3. **Generate Regressors (X):** We generate the regressor matrix `X`. `runif(n * k)` creates a vector of `n*k` uniformly distributed random numbers, which are then arranged into an `n x k` matrix. We then `cbind` a column of 1s to `X` to represent the intercept term. This step relates to the text where it defines $X \in \mathbb{R}^{n\times(K+1)}$.
4. **Generate Errors (ε):** We generate the error terms, `epsilon`, from a normal distribution with mean 0 and standard deviation `sqrt(sigma2)`. This ensures $E(\varepsilon|X) = 0$ (Assumption A1) and $\mathrm{Var}(\varepsilon|X) = \sigma^2 I_n$ (homoskedasticity, Assumption A2 specialized to equation 18.4).
5. **Generate Dependent Variable (y):** We generate the dependent variable `y` using the linear model equation: $y = X\beta + \varepsilon$. This corresponds directly to the core equation of the linear model in the text.
6. **Combine into Data Frame:** We combine `y` and `X` into a data frame for easier manipulation.

7. **Visualization:** We create a scatter plot of `y` against one of the regressors (`X.1`) and add a smoothed line using `geom_smooth(method = "lm")`, which fits a linear model. This visualizes the linear relationship between y and X, reflecting $E(y|X) = X\beta$ (Assumption A1).

R Script 2: Simulating a Linear Model with Heteroskedastic Errors

```
library(tidyverse)

# Set seed for reproducibility
set.seed(456)

# Number of observations
n <- 100

# Number of regressors (excluding the intercept)
k <- 2

# Generate regressors (X)
X <- matrix(runif(n * k), nrow = n, ncol = k)
X <- cbind(1, X) # Add a column of 1s for the intercept

# True coefficients (beta)
beta <- c(2, -1, 0.5)

# Generate heteroskedastic errors
#   Here, the variance depends on the first regressor (X[, 2])
sigma2_i <- exp(X[, 2])  # Variance increases exponentially with X[, 2]
epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2_i))

# Generate dependent variable (y)
y <- X %*% beta + epsilon

# Combine into a data frame
data <- data.frame(y = y, X = X[, -1])

# Visualize the heteroskedasticity
ggplot(data, aes(x = X.1, y = y)) +
  geom_point() +
  labs(title = "Simulated Linear Model with Heteroskedastic Errors",
       x = "X1",
       y = "y") +
    geom_smooth(method = "lm", se=FALSE)
```

```
`geom_smooth()` using formula = 'y ~ x'
```

```
ggplot(data, aes(x = X.1, y = epsilon^2)) +
    geom_point() +
  labs(title = "Squared Residuals vs. X1 (Illustrating Heteroskedasticity)",
         x = "X1",
       y = "Squared Residuals")
```

Explanation:

1. **Setup:** Similar to the previous example, we load `tidyverse`, set a seed, and define `n`, `k`, `X`, and `beta`.

2. **Generate Heteroskedastic Errors:** This is the key difference. Instead of a constant `sigma2`, we create `sigma2_i`, which varies for each observation. Here, we make the variance a function of the first regressor (`X[, 2]`): `sigma2_i <- exp(X[, 2])`. This means the variance of the error term increases exponentially with the value of `X[, 2]`. We then generate `epsilon` from a normal distribution with mean 0 and standard deviation `sqrt(sigma2_i)`. This satisfies Assumption A2, where $\text{Var}(y|X) = \Sigma(X)$, but *not* the homoskedastic special case. $\Sigma(X)$ is now a diagonal matrix with the `sigma2_i` values on the diagonal.
3. **Generate Dependent Variable (y):** Same as before: $y = X\beta + \varepsilon$.
4. **Combine into Data Frame:** We combine y and X into a data frame.
5. **Visualization:**
    - The first plot shows the relationship between y and `X.1`. The heteroskedasticity might not be immediately obvious from this plot alone.
    - The second plot is more revealing. We plot the *squared residuals* (`epsilon^2`) against `X.1`. Since the variance of ε depends on `X.1`, we expect to see a pattern in this plot (in this case, an increasing spread), indicating heteroskedasticity.

R Script 3: Simulating a Linear Model with Autocorrelated Errors (AR(1))

```r
library(tidyverse)

# Set seed for reproducibility
set.seed(789)

# Number of observations
n <- 100

# Number of regressors (excluding the intercept)
k <- 2

# Generate regressors (X)
X <- matrix(runif(n * k), nrow = n, ncol = k)
X <- cbind(1, X) # Add a column of 1s for the intercept

# True coefficients (beta)
beta <- c(2, -1, 0.5)

# Autocorrelation parameter (rho)
rho <- 0.7

# Generate autocorrelated errors
epsilon <- numeric(n)
epsilon[1] <- rnorm(1)  # Initialize the first error term
for (t in 2:n) {
  epsilon[t] <- rho * epsilon[t-1] + rnorm(1) # AR(1) process
}

# Generate dependent variable (y)
y <- X %*% beta + epsilon

# Combine into a data frame
data <- data.frame(y = y, X = X[, -1], Time = 1:n)

# Visualize the autocorrelation
ggplot(data, aes(x = Time, y = epsilon)) +
  geom_line() +
  labs(title = "Autocorrelated Errors over Time",
       x = "Time",
       y = "Error Term")
```

```
acf(epsilon, main = "ACF of Error Terms") # Autocorrelation function
```

Explanation:

1. **Setup:** We load `tidyverse`, set the seed, define `n`, `k`, `X`, and `beta` as before.
2. **Autocorrelation Parameter (ρ):** We define `rho`, the autocorrelation coefficient, which determines the strength of the correlation between consecutive error terms.
3. **Generate Autocorrelated Errors:** This is the core part. We simulate an AR(1) process:
   - We initialize the first error term, `epsilon[1]`, with a standard normal random variable.
   - For subsequent time periods (`t in 2:n`), we generate `epsilon[t]` using the AR(1) equation: $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, where u_t is a standard normal random variable. This means the current error term is equal to `rho` times the previous error term plus a new random shock. This corresponds to a special case of Assumption A2 where the off-diagonal entries of $\Sigma(X)$ get smaller the further apart they are.
4. **Generate Dependent Variable (y):** We generate the dependent variable: $y = X\beta + \varepsilon$.
5. **Combine into a Data Frame**: We create a data frame including the y and X variables.
6. **Visualization:**
   - We plot the error term (`epsilon`) over time. The AR(1) process will typically exhibit some “stickiness” or persistence, where positive errors tend to be followed by positive errors, and negative errors by negative errors.
   - We use the `acf()` function (Autocorrelation Function) to display the autocorrelation of the error terms. This will show significant autocorrelation at lag 1 (and possibly other lags, depending on the value of `rho`).

R Script 4: Illustrating the Rank Condition

```
library(tidyverse)

# Set seed
set.seed(101)

# Number of observations
n <- 50

# Create a matrix X with rank deficiency
X <- matrix(runif(n * 3), nrow = n, ncol = 3)
X[, 3] <- 2 * X[, 1] + 3 * X[, 2]  # Make the third column a linear combination of the first two

# Calculate the rank of X
rank_X <- qr(X)$rank

# Print the rank
print(paste("Rank of X:", rank_X))
```

```
[1] "Rank of X: 2"
```

```
# Attempt to calculate (X'X)^-1 (will fail)
# solve(t(X) %*% X) # This line would produce an error

# Create a matrix X with full column rank
X_full <- matrix(runif(n * 3), nrow = n, ncol = 3)

# Calculate the rank of X_full
```

```
rank_X_full <- qr(X_full)$rank
print(paste("Rank of X_full:", rank_X_full))
```

```
[1] "Rank of X_full: 3"
```

```
# Calculate (X_full'X_full)^-1 (will succeed)
solve(t(X_full) %*% X_full)
```

```
            [,1]        [,2]        [,3]
[1,]  0.13624535 -0.04690227 -0.08378334
[2,] -0.04690227  0.20430759 -0.10415225
[3,] -0.08378334 -0.10415225  0.20643596
```

Explanation:

1. **Setup:** Load libraries, and set seed.
2. **Create Rank-Deficient Matrix:** We create a matrix `X` where the third column is a linear combination of the first two columns. This ensures that `X` does *not* have full column rank. This directly demonstrates a violation of the condition $\text{rank}(X) = K < n$ stated in the text.
3. **Calculate Rank:** We use `qr(X)$rank` to calculate the rank of `X`. The `qr()` function performs QR decomposition, and its `$rank` component gives the rank of the matrix.
4. **Illustrate Problem:** We (commented out) attempt to calculate $(X'X)^{-1}$. This will result in an error because $(X'X)$ is singular when `X` does not have full column rank. This is a critical issue in OLS estimation, as $(X'X)^{-1}$ is required to calculate the OLS estimator.
5. **Create Full-Rank Matrix:** We create another matrix, `X_full`, where the columns are linearly independent (generated from independent random uniform variables).
6. **Calculate Rank (Full Rank):** We calculate the rank of `X_full`, which will be equal to the number of columns (3 in this case).
7. **Calculate Inverse:** We successfully calculate $(X_{full}'X_{full})^{-1}$, demonstrating that the inverse exists when the matrix has full column rank.

R Script 5: Illustrating Clustered Errors

```
library(tidyverse)
library(mvtnorm) # For multivariate normal simulation

# Set seed
set.seed(202)

# Number of clusters
n_clusters <- 10

# Observations per cluster
n_per_cluster <- 5

# Total observations
n <- n_clusters * n_per_cluster

# Generate regressors (X) - for simplicity, just one regressor
X <- matrix(runif(n), nrow = n, ncol = 1)
X <- cbind(1, X) # Add intercept

# True coefficients (beta)
beta <- c(1, 2)

# Generate clustered errors
#   Within each cluster, errors are correlated; between clusters, they are uncorrelated.
epsilon <- numeric(n)
```

```
rho <- 0.8 # Within-cluster correlation

for (i in 1:n_clusters) {
  start_index <- (i - 1) * n_per_cluster + 1
  end_index <- i * n_per_cluster

  # Covariance matrix for the cluster
  cluster_cov <- matrix(rho, nrow = n_per_cluster, ncol = n_per_cluster)
  diag(cluster_cov) <- 1  # Variance of 1 for each error

  # Generate errors for the cluster using rmvnorm
  epsilon[start_index:end_index] <- rmvnorm(1, mean = rep(0, n_per_cluster), sigma = cluster_cov)
}

# Generate dependent variable (y)
y <- X %*% beta + epsilon

# Create data frame, including cluster ID
data <- data.frame(y = y, X = X[,-1], Cluster = rep(1:n_clusters, each = n_per_cluster))

# Visualize errors within a few clusters
ggplot(data, aes(x = seq_along(y), y = epsilon, color = factor(Cluster))) +
  geom_line() +
  geom_point() +
  labs(title = "Clustered Errors",
       x = "Observation within Cluster",
       y = "Error Term",
       color = "Cluster")
```

Explanation:

1. **Setup:** Load `tidyverse` and `mvtnorm` (for simulating from a multivariate normal distribution). Set seed.
2. **Define Parameters:** Define the number of clusters (`n_clusters`), observations per cluster (`n_per_cluster`), total observations (`n`), regressors (`X`), coefficients (`beta`), and the within-cluster correlation (`rho`).
3. **Generate Clustered Errors:** This is the key part.
    - We loop through each cluster.
    - Inside the loop, we define the start and end indices for the current cluster within the overall `epsilon` vector.
    - We create a covariance matrix `cluster_cov` for the current cluster. The diagonal elements are 1 (variance of each error term), and the off-diagonal elements are `rho` (the correlation between errors within the cluster).
    - We use `rmvnorm()` to generate `n_per_cluster` error terms from a multivariate normal distribution with mean 0 and the specified `cluster_cov`. This ensures that the errors within the cluster are correlated as desired. This corresponds to a special case of Assumption A2 with a block-diagonal matrix, as shown in the text.
4. **Generate Dependent Variable (y):** We create `y` using the usual linear model equation.
5. **Create Data Frame:** Create a dataframe including cluster ids.
6. **Visualization:** We plot the error terms for the first three clusters. Observations within each cluster are connected by lines, and different clusters have different colors. This visually demonstrates the within-cluster correlation.
7. **Sample Correlation:** We calculate the sample correlation matrix of the errors within the first cluster to show that the generated errors indeed have a correlation close to the specified `rho`.

YouTube Video Suggestions for Linear Model Concepts

Here are some YouTube video suggestions related to the concepts in the provided text, along with explanations of their relevance and verification of their availability (as of October 26, 2023):

Video 1: The Linear Regression Model

- **Title:** "1.1. The linear regression model"
- **Channel:** Ben Lambert
- **Link:** https://www.youtube.com/watch?v=zITIFTsivN8
- **Availability:** Verified - video is currently available.
- **Relevance:** This video provides a good introduction to the **linear regression model**, covering the basic equation $y = X\beta + \epsilon$. It discusses the components of the model, including the dependent variable (y), independent variables (X), coefficients (β), and the error term (ϵ). It lays the groundwork for understanding the model presented in section 18.2 of the text. It visually relates data with the linear model and the error term.

Video 2: Assumptions of Linear Regression

- **Title:** "6. Regression Assumptions"
- **Channel:** zedstatistics
- **Link:** https://www.youtube.com/watch?v=0MFpOQRY-t4
- **Availability:** Verified - video is currently available.
- **Relevance:** This video discusses the key **assumptions of linear regression**, many of which align with Assumptions A1-A4 in the text. It covers topics like linearity ($E(y|X) = X\beta$), homoskedasticity ($\text{Var}(y|X) = \sigma^2 I$), and the lack of autocorrelation. Although it doesn't delve into the matrix notation of $\Sigma(X)$ as deeply as the text, it provides a solid intuitive understanding of the assumptions.

Video 3: Multicollinearity

- **Title:** "Econometrics // Lecture 6: Multicollinearity"
- **Channel:** Wooldridge Lectures
- **Link:** https://www.youtube.com/watch?v=4T1w9NEK54c
- **Availability:** Verified - video is currently available.
- **Relevance:** This video discusses **multicollinearity**, which is directly related to the assumption $\text{rank}(X) = K < n$ (Equation 18.1 in the text). It explains what multicollinearity is (linear dependence among regressors), its consequences (inflated standard errors, unstable coefficients), and how to detect and address it. The video makes clear why full column rank is crucial for OLS estimation.

Video 4: Heteroskedasticity

- **Title:** "Heteroskedasticity"
- **Channel:** Steve Grams
- **Link:** https://www.youtube.com/watch?v=cREHqG274Ag
- **Availability:** Verified - video is currently available.
- **Relevance:** This video explains **heteroskedasticity**, which is a violation of the assumption of constant error variance. The text discusses this in the context of $\Sigma(X)$ not being equal to $\sigma^2 I_n$. The video provides a clear visual explanation of heteroskedasticity (non-constant spread of residuals) and discusses its consequences and potential remedies.

Video 5: Autocorrelation

- **Title:** "Autocorrelation"
- **Channel:** DATAtab

- **Link:** https://www.youtube.com/watch?v=w_w6ZkEeUo8
- **Availability:** Verified - video is currently available.
- **Relevance**: This video provides a detailed explanation of **autocorrelation**. This is the condition where errors from different time periods are correlated with each other. The video gives an intuitive explanation, useful examples, and visual aids of how to detect autocorrelation. The text describes autocorrelation as a case where $\Sigma(X)$ may have all entries non-zero but have the property that entries get smaller and smaller the further they are away from the diagonal.

Video 6: Strong Exogeneity

- **Title**: "What is Exogeneity? - Causal Research"
- **Channel**: John Antonakis
- **Link**: https://www.youtube.com/watch?v=efeDP-FagjA
- **Availability**: Verified - video is currently available
- **Relevance**: This video introduces and provides intuitive examples of **strong exogeneity**. The video explains what it means for an independent variable to be exogenous, as the condition that the independent variable cannot be correlated with the error term. The text describes strong exogeneity as a condition where the error term is uncorrelated with past, present and future values of the independent variable.

These videos provide a good visual and intuitive complement to the more formal mathematical treatment in the text. They cover the core assumptions and potential issues in the linear model, aligning well with the concepts discussed in Chapter 18.

Multiple Choice Exercises

MC Exercise 1

MC Solution 1

The assumption $\text{rank}(X) = K < n$ in the linear model implies:

a. The number of regressors is greater than the number of observations.
b. There is perfect multicollinearity among the regressors.
c. The matrix X has full column rank, and the number of observations is greater than the number of regressors (including the intercept).
d. The error terms are normally distributed.

MC Exercise 2

MC Solution 2

Assumption A1, $E(y|X) = X\beta$, states that:

a. The variance of y given X is constant.
b. The expected value of y given X is a linear function of X.
c. y is normally distributed.
d. The error terms are uncorrelated.

MC Exercise 3

MC Solution 3

Assumption A2, $\text{var}(y|X) = \Sigma(X)$, implies that:

a. The variance of y given X is constant.
b. The variance of y given X can depend on X.
c. The errors are homoskedastic.
d. y is normally distributed.

MC Exercise 4

MC Solution 4

Which assumption implies that y is normally distributed conditional on X?

a. A1
b. A2
c. A3
d. A4 and A1

MC Exercise 5

MC Solution 5

The equation $y = X\beta + \varepsilon$ represents:

a. The conditional expectation of y given X.
b. The variance of y given X.
c. The linear regression model with an error term.
d. The normal distribution of y.

MC Exercise 6

MC Solution 6

$E(\varepsilon|X) = 0$ implies:

a. The errors are homoskedastic.
b. The errors are normally distributed.
c. The expected value of the error term, conditional on X, is zero.
d. The variance of the error term is constant.

MC Exercise 7

MC Solution 7

Strong exogeneity in a time series setting means:

a. The error term is only correlated with the current value of the regressors.
b. The error term is uncorrelated with past, present, and future values of the regressors.
c. The regressors are normally distributed.
d. The error term is homoskedastic.

MC Exercise 8

MC Solution 8

The **homoskedastic** case is characterized by:

a. $\Sigma(X) = \sigma^2 I_n$
b. $\Sigma(X)$ being a diagonal matrix with different values on the diagonal.
c. $\Sigma(X)$ being a block-diagonal matrix.
d. $\Sigma(X)$ having all entries non-zero.

MC Exercise 9

MC Solution 9

The **heteroskedastic** case is characterized by:

a. $\Sigma(X) = \sigma^2 I_n$
b. $\Sigma(X)$ being a diagonal matrix with potentially different values on the diagonal.
c. $\Sigma(X)$ being a block-diagonal matrix.
d. $\Sigma(X)$ having all entries non-zero.

MC Exercise 10

MC Solution 10

The **clustered** error structure is characterized by:

a. $\Sigma(X) = \sigma^2 I_n$
b. $\Sigma(X)$ being a diagonal matrix with different values on the diagonal.
c. $\Sigma(X)$ being a block-diagonal matrix.
d. $\Sigma(X)$ having all entries non-zero and constant.

MC Exercise 11

MC Solution 11

In the time series case, the covariance matrix $\Sigma(X)$ typically has:

a. All entries equal to zero.
b. All entries equal to a constant σ^2.
c. Non-zero entries that decrease in magnitude as they move away from the diagonal.
d. A block-diagonal structure.

MC Exercise 12

MC Solution 12

A positive definite matrix $\Sigma(X)$ implies:

a. All its diagonal elements are negative.
b. It is always singular (non-invertible).
c. For any non-zero vector z, $z^T \Sigma(X) z > 0$.
d. It represents a uniform distribution.

MC Exercise 13

[MC Solution 13]

If $E(\epsilon_i|X_1, \ldots, X_n) = 0$, this implies:

a. The error term ϵ_i is heteroskedastic.
b. Strong exogeneity.
c. ϵ_i follows a normal distribution.
d. Weak exogeneity.

MC Exercise 14

[MC Solution 14]

Assumption A4, $y \sim N(X\beta, \sigma^2 I)$, implies:

a. Heteroskedasticity.
b. Autocorrelation.
c. Homoskedasticity and no autocorrelation.
d. Clustered errors.

MC Exercise 15

[MC Solution 15]

Which of the following statements is true about the linear model?

a. It always assumes that the dependent variable is normally distributed.
b. It always assumes homoskedasticity.
c. It models the conditional expectation of the dependent variable as a linear function of the regressors.
d. It cannot handle time series data.

MC Exercise 16

[MC Solution 16] Consider a model explaining house prices as a function of square footage and the number of bedrooms. Which assumption is most likely to be violated? a) $E(y|X) = X\beta$. b) $\text{rank}(X) = K < n$. c) Homoskedasticity. d) y is normally distributed.

MC Exercise 17

[MC Solution 17]

The notation $\varepsilon = y - X\beta$ defines: a) The predicted value of y. b) The error term of the linear model. c) The variance of y. d) The coefficient vector.

MC Exercise 18

[MC Solution 18]

Which of the following cases of $\Sigma(X)$ would be most applicable for panel data where observations are grouped by individuals over time? a) $\Sigma(X) = \sigma^2 I_n$ b) A diagonal matrix with potentially different values on the diagonal. c) A block-diagonal matrix. d) A matrix with all entries identical and non-zero.

MC Exercise 19

[MC Solution 19]

The linear model assumes there is a random mechanism behind the data generation. This implies: a) The observed data is the only possible outcome. b) The observed data represents one realization of many potential outcomes. c) The data is deterministic. d) The data is normally distributed.

MC Exercise 20

[MC Solution 20]

Assumption A3 is a _____ assumption than Assumption A1: a) Weaker b) Stronger c) Less restrictive d) Simpler

Multiple Choice Solutions

MC Solution 1

[MC Exercise 1]

Correct Answer: c)

- **Explanation:** The assumption $\text{rank}(X) = K < n$ has two parts:
  - $\text{rank}(X) = K$: This means the matrix X has full column rank. No regressor is a perfect linear combination of other regressors.
  - $K < n$: This means the number of regressors (including the intercept, hence K, not $K+1$ as in the text), is less than the number of observations. This is necessary to have degrees of freedom for estimation.
- Options a) and b) are incorrect, and d) is not related to the rank of the matrix.

MC Solution 2

[MC Exercise 2]

Correct Answer: b)

- **Explanation:** $E(y|X) = X\beta$ states that the *expected value* (or mean) of the dependent variable y, given the values of the independent variables X, is a *linear function* of X. The coefficients of this linear function are the elements of the vector β.
- Options a), c), and d) refer to other assumptions or properties, not the definition of the conditional expectation.

MC Solution 3

[MC Exercise 3]

Correct Answer: b)

- **Explanation:** $\text{var}(y|X) = \Sigma(X)$ states that the *variance* of y conditional on X is given by the matrix $\Sigma(X)$. The key here is that $\Sigma(X)$ can *depend on* X. This allows for the possibility of heteroskedasticity (non-constant variance).
- Option a) is incorrect because it would imply homoskedasticity. Options c) and d) are not directly implied by Assumption A2.

MC Solution 4

MC Exercise 4

Correct Answer: c)

- **Explanation:** Assumption A3, $y \sim N(X\beta, \Sigma(X))$, explicitly states that y follows a *normal distribution* with mean $X\beta$ and variance-covariance matrix $\Sigma(X)$.
- Assumptions A1 and A2 only specify the first two moments (mean and variance) and do not assume normality. A4 *also* assumes normality, but with a specific, restrictive form of $\Sigma(X)$, so c) is a more general answer.

MC Solution 5

MC Exercise 5

Correct Answer: c)

- **Explanation:** This is the standard representation of the **linear regression model**, where y is the dependent variable, X is the matrix of regressors, β is the vector of coefficients, and ε is the error term.
- Options a) and b) refer to specific parts or characteristics of the model, but not the complete equation. Option d) is an assumption, not the definition.

MC Solution 6

MC Exercise 6

Correct Answer: c)

- **Explanation:** $E(\varepsilon|X) = 0$ states that the *expected value* (or mean) of the error term ε, conditional on the values of the regressors X, is *zero*. This means that, on average, the errors are zero for any given values of the regressors.
- Options a), b), and d) refer to other properties of the error term (homoskedasticity, normality, constant variance), which are separate assumptions.

MC Solution 7

MC Exercise 7

Correct Answer: b)

- **Explanation: Strong exogeneity** means that the error term at any time period is uncorrelated with the regressors in *all* time periods (past, present, and future). This is a crucial assumption for causal inference in time series models.
- Option a) describes a weaker form of exogeneity (contemporaneous exogeneity). Options c) and d) are unrelated to exogeneity.

MC Solution 8

MC Exercise 8

Correct Answer: a)

- **Explanation:** The **homoskedastic** case means the variance of the error term is *constant* and the errors are *uncorrelated*. This is represented by $\Sigma(X) = \sigma^2 I_n$, where σ^2 is the constant variance and I_n is the identity matrix, indicating no correlation between different error terms.
- The other options describe heteroskedasticity (b), clustered errors (c), or a general covariance structure (d).

MC Solution 9

MC Exercise 9

Correct Answer: b)

- **Explanation:** The **heteroskedastic** case means the variance of the error term is *not constant*. It can vary across observations. This is represented by $\Sigma(X)$ being a diagonal matrix with potentially *different* values on the diagonal (representing different variances for each observation).
- Option a) represents homoskedasticity. Options c) and d) describe other types of covariance structures.

MC Solution 10

MC Exercise 10

Correct Answer: c)

- **Explanation: Clustered errors** occur when observations are grouped into clusters, and errors within the same cluster are correlated, but errors in different clusters are uncorrelated. This is represented by a block-diagonal matrix $\Sigma(X)$, where each block corresponds to a cluster and captures the within-cluster correlation.
- Option a) represents homoskedasticity. Option b) represents heteroskedasticity. Option d) is too general.

MC Solution 11

MC Exercise 11

Correct Answer: c)

- **Explanation:** In time series data, it's common for observations closer in time to be more correlated than observations farther apart. This means the covariance matrix $\Sigma(X)$ will have non-zero entries, but these entries will *decrease in magnitude* as you move away from the diagonal (representing the correlation between observations further apart in time).
- Options a) and b) imply no autocorrelation. Option d) is more typical of panel data.

MC Solution 12

MC Exercise 12

Correct Answer: c)

- **Explanation:** This is the definition of a **positive definite** matrix. It means that the quadratic form $z^T\Sigma(X)z$ is always positive for any non-zero vector z. This property is important for ensuring that variances are positive and that the matrix is invertible.
- Option a) is incorrect. Positive definite matrices have positive diagonal entries. Option b) is incorrect, positive definite matrices are always invertible. Option d) is unrelated.

MC Solution 13

MC Exercise 13

Correct Answer: b)

- **Explanation:** The condition $E(\epsilon_i|X_1,\ldots,X_n)=0$ means that the expected value of the error term in observation i is zero conditional on *all* values of the regressors in the data set. This encompasses past, present, and future values, indicating strong exogeneity.
- Option a) is unrelated to the conditional expectation. Option c) is not implied by this condition. Option d) is a concept related to, but not the same as $E(\epsilon_i|X_i)=0$.

MC Solution 14

MC Exercise 14

Correct Answer: c)

- **Explanation:** Assumption A4 combines normality with $\Sigma(X)=\sigma^2 I_n$. The $\sigma^2 I_n$ part implies **homoskedasticity** (constant variance) and **no autocorrelation** (errors are uncorrelated).
- The other options represent violations of these assumptions.

MC Solution 15

MC Exercise 15

Correct Answer: c)

- **Explanation:** The core of the linear model is that it models the *conditional expectation* of the dependent variable (y) as a *linear function* of the regressors (X): $E(y|X)=X\beta$.
- Options a) and b) are common assumptions but not always required. Option d) is incorrect; the linear model can be adapted for time series data.

MC Solution 16

MC Exercise 16

Correct Answer: c)

- **Explanation:** It is common for house prices to be **heteroskedastic**. The variance of house prices is very often an increasing function of the size of the house.
- Option a) is likely to be satisfied with the proposed model. Options b) would not likely be violated if we have a sufficiently large number of observations. Finally, while it is not impossible for house prices to be normally distributed, this is not a strict requirement of the linear model (option d).

MC Solution 17

MC Exercise 17

Correct Answer: b)

- **Explanation:** This equation *defines* the **error term** (ε) as the difference between the actual value of the dependent variable (y) and the predicted value based on the linear model ($X\beta$).

MC Solution 18

MC Exercise 18

Correct Answer: c)

- **Explanation:** Panel data (individuals observed over time) often exhibits a **clustered error structure**. Errors are likely to be correlated *within* each individual over time (due to unobserved individual-specific factors), but uncorrelated *across* individuals. This is best represented by a **block-diagonal** matrix, where each block corresponds to an individual and captures the within-individual correlation.
- Option a) represents homoskedasticity and no autocorrelation. Option b) represents heteroskedasticity. Option d) is not a realistic covariance structure.

MC Solution 19

MC Exercise 19

Correct Answer: b)

- **Explanation:** The linear model assumes there's an underlying **random** data-generating process. We only observe *one* possible realization of this process. There's an infinite number of potential outcomes we *could* have observed.
- Option a) is the opposite of what the linear model assumes. Options c) and d) are not inherent to the idea of a random mechanism.

MC Solution 20

MC Exercise 20

Correct Answer: b) * **Explanation:** Assumption A3 ($y \sim N(X\beta, \Sigma(X))$) makes a specific distributional assumption (normality) in addition to specifying the mean and variance. Assumption A1 ($E(y|X) = X\beta$) *only* specifies the conditional mean. A3 therefore implies A1, but not vice versa. Therefore, A3 is the **stronger** assumption.

Author: Peter Fuleky