Heteroscedasticity

1. Nature of Heteroscedasticity
2. Consequences of Using OLS in Presence of Heteroscedasticity
3. Detection of Heteroscedasticity
4. Remedial Measures

Nature of Heteroscedasticity

Assumption 4 of OLS:

Homoscedasticity $E\left(u_i^2\right) = \sigma^2$

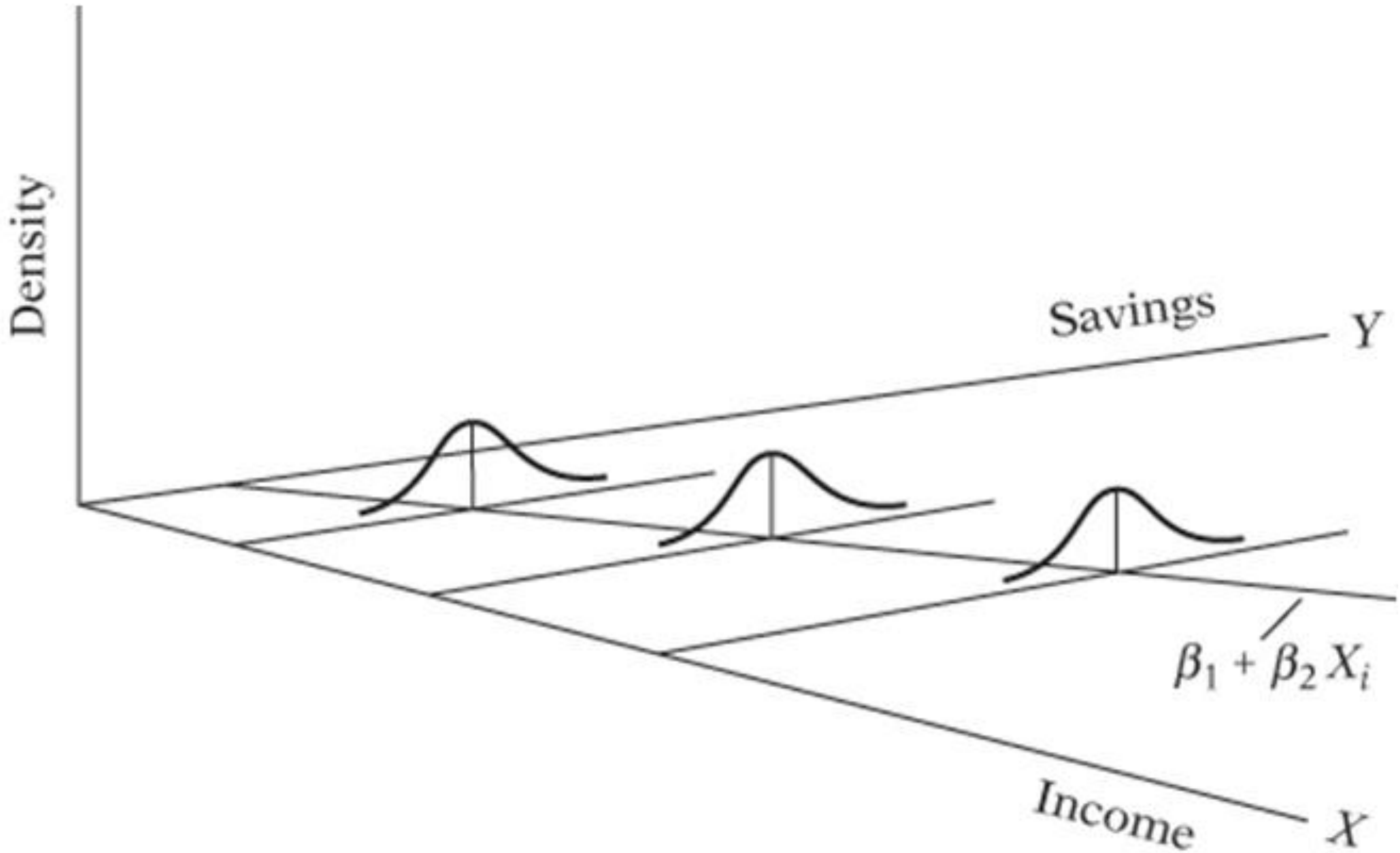
- Conditional variance of Y_i are constant.

Heteroscedasticity occurs when this assumption is violated.

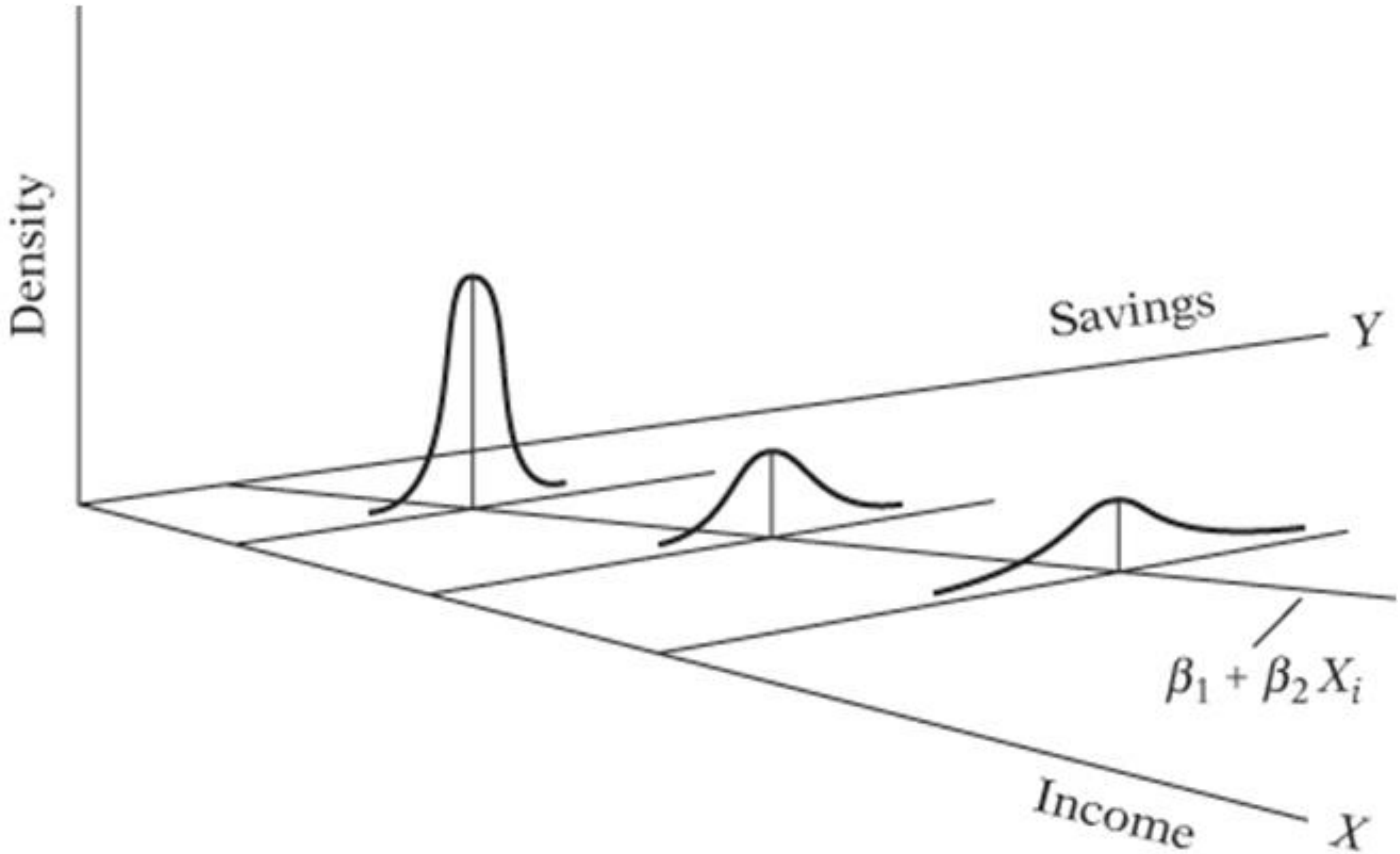
Heteroscedasticity $E\left(u_i^2\right) = \sigma_i^2$

- Conditional variance of Y_i are not constant.

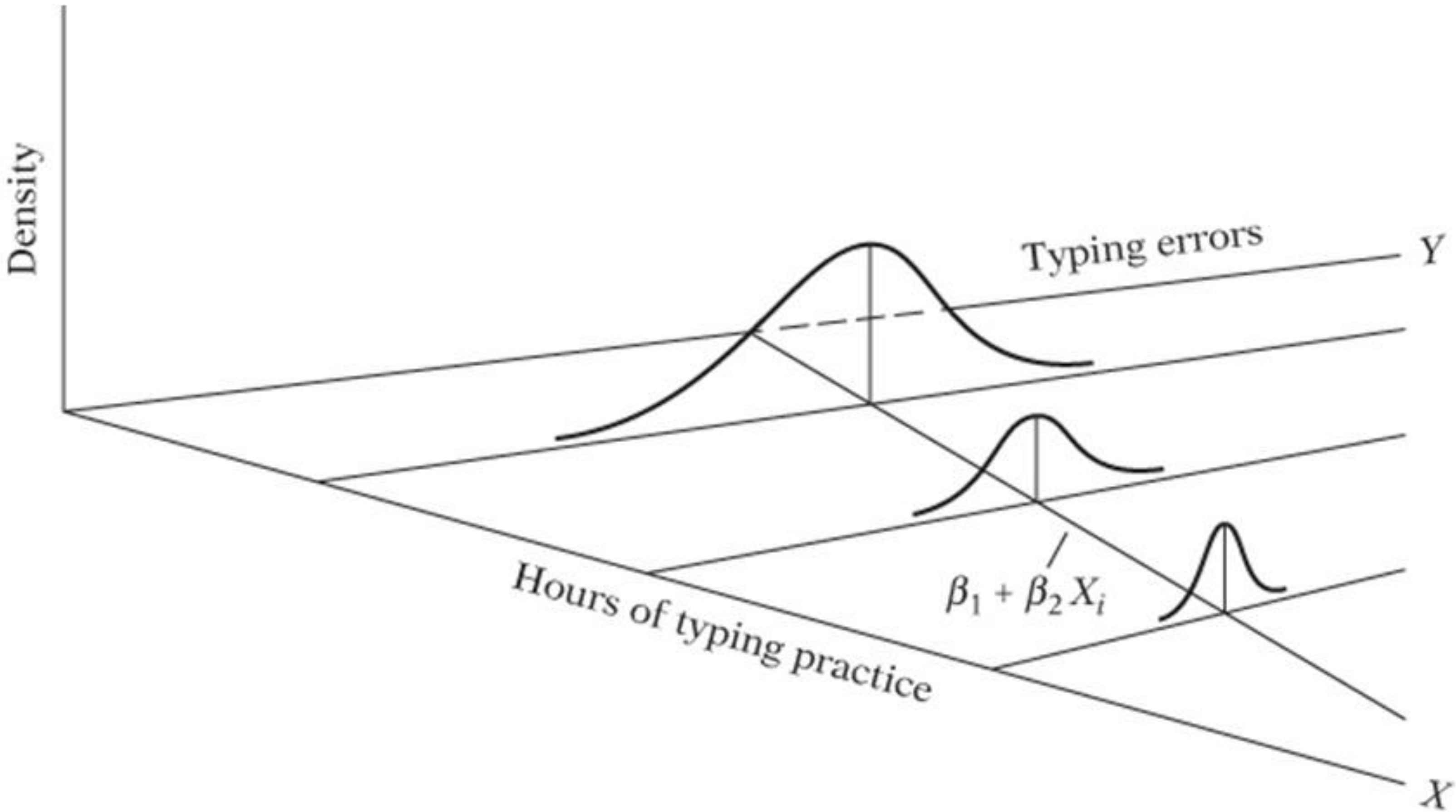
Nature of Heteroscedasticity



Nature of Heteroscedasticity



Nature of Heteroscedasticity

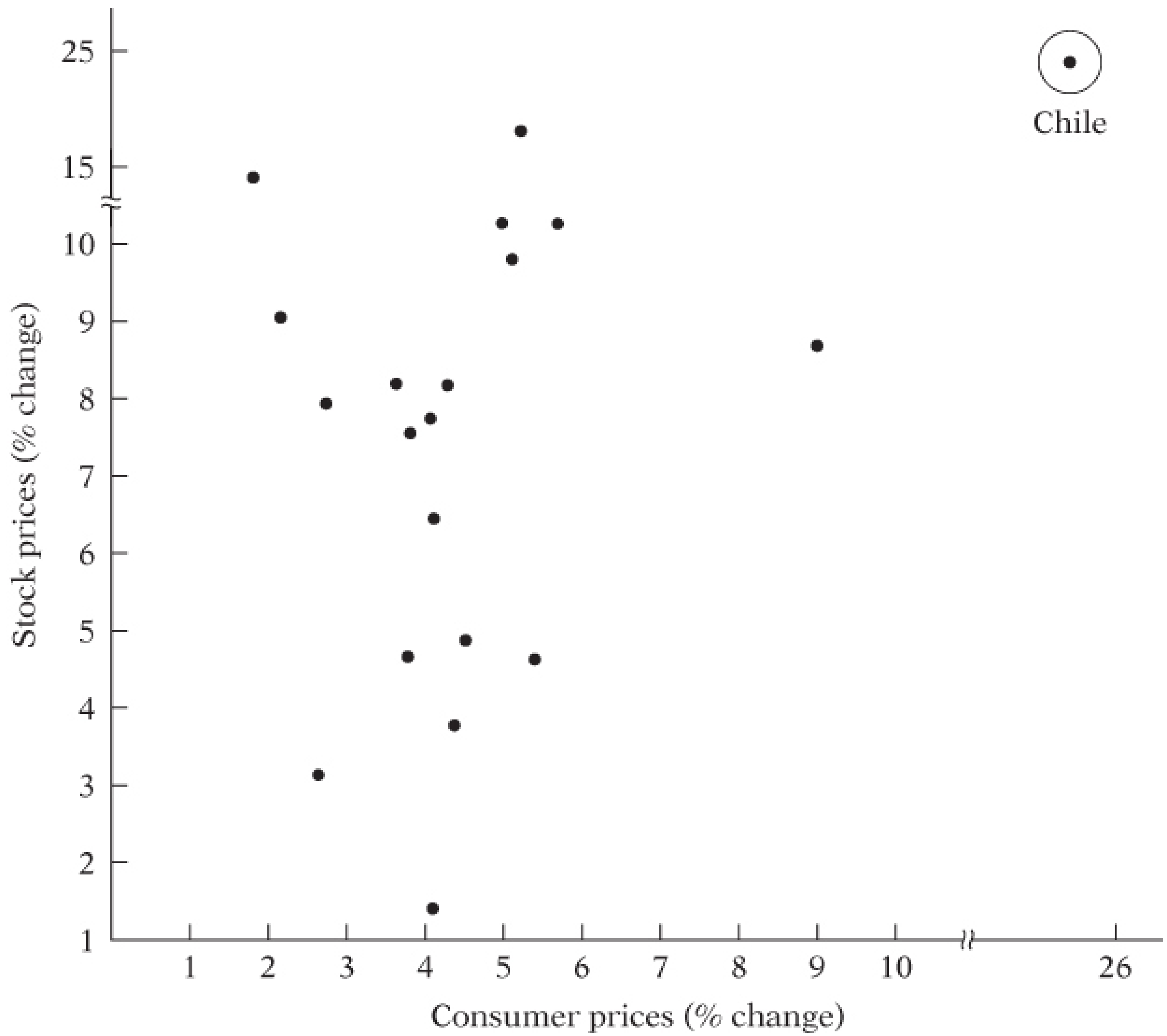


Nature of Heteroscedasticity

Possible Reasons of Heteroscedasticity

1. Error Learning Models
2. Discretionary Income
3. Different Technology
4. Outliers
5. Specification Errors--Omitted Variable

Nature of Heteroscedasticity



OLS Estimation in Presence of Heteroscedasticity

Let $E\left(u_i^2\right) = \sigma_i^2$

Two-variable Model

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Variance of the OLS estimator of β_2 will be:

<u>Homoscedasticity</u>

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2}$$

<u>Heteroscedasticity</u>

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

OLS Estimation in Presence of Heteroscedasticity

With heteroscedasticity, will OLS estimators still be BLUE?

- Best — No
- Linear — Yes
- Unbiased — Yes

Method of Generalized Least Squares

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\frac{Y_i}{\sigma_i} = \beta_1\left(\frac{1}{\sigma_i}\right) + \beta_2\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right)$$

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^*$$

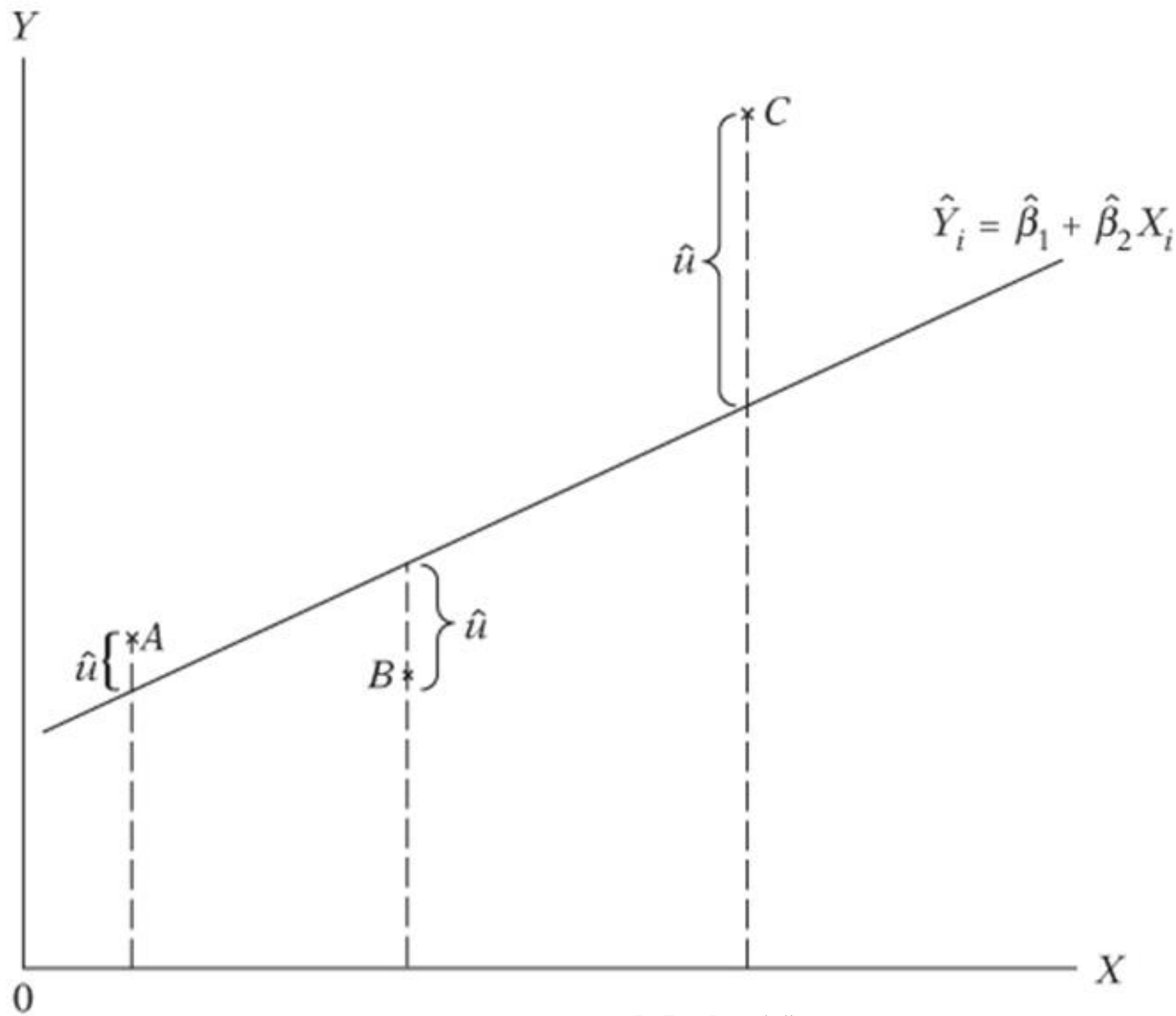
$$\operatorname{var}\left(u_i^*\right) = E\left(u_i^*\right)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 = \frac{1}{\sigma_i^2} E\left(u_i^2\right) = \frac{1}{\sigma_i^2}\left(\sigma_i^2\right) = 1$$

Method of Generalized Least Squares

OLS: $$\sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \right)^2$$

GLS: $$\sum w_i \hat{u}_i^2 = \sum w_i \left(Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i \right)^2$$

Method of Generalized Least Squares



Consequences of Using OLS in Presence of Heteroscedasticity

OLS Estimation allowing Heteroscedasticity

Assume σ_i^2 are known.

However, $\operatorname{var}\left(\hat{\beta}_2^*\right) \leq \operatorname{var}\left(\hat{\beta}_2\right)$

- Variances of estimated coefficients are more likely to be larger than they should be.
- t-test and F-test will be smaller than what is appropriate.
- Thus, t-test and F-test are more likely to give inaccurate results.

Consequences of Using OLS in Presence of Heteroscedasticity

OLS Estimation disregarding Heteroscedasticity

- Variances of estimated coefficients are biased estimators.
- Conventional t-test and F-test will be inappropriated.

Detection of Heteroscedasticity

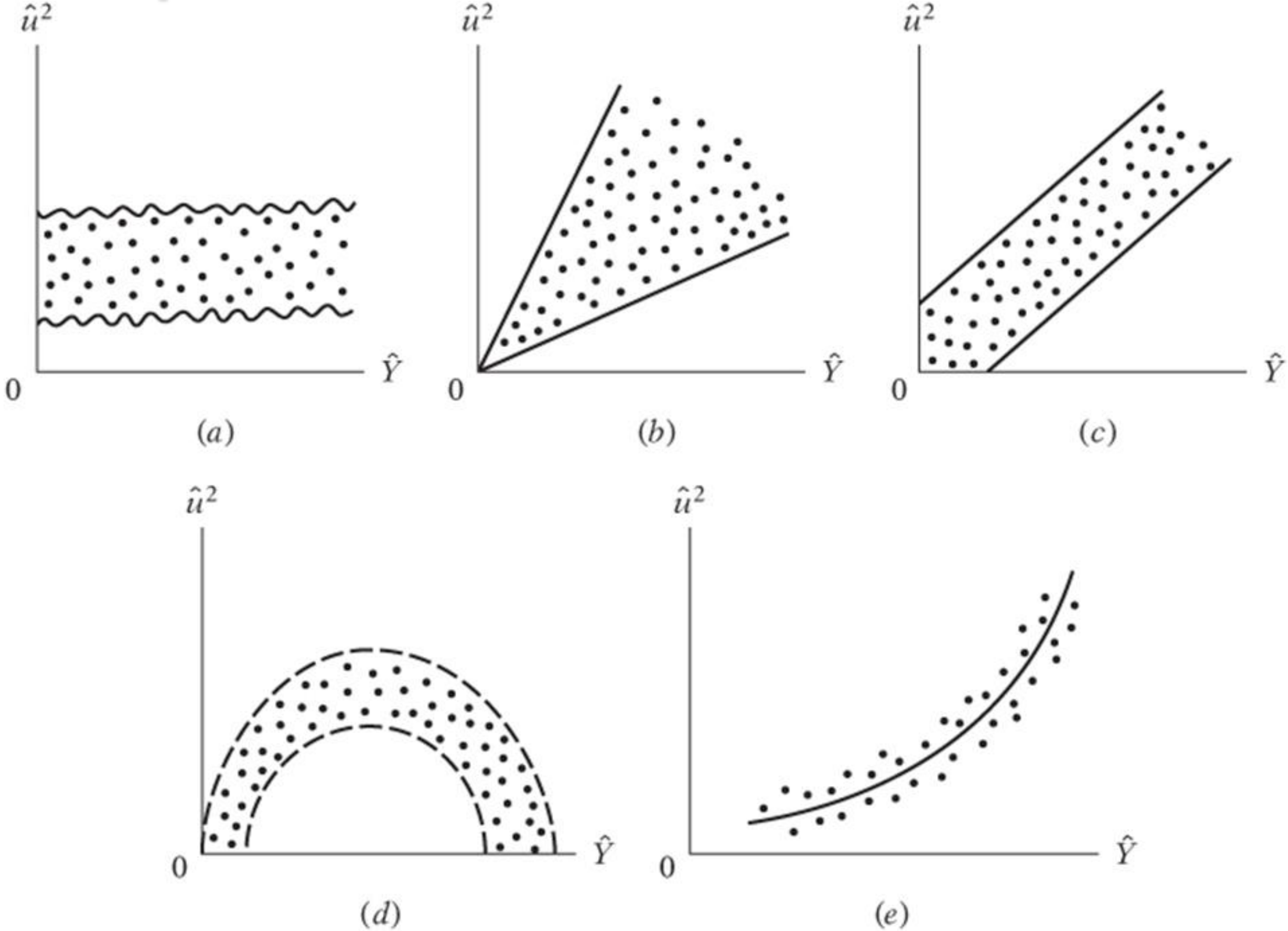
<u>Informal Methods</u>

- Graphical Method

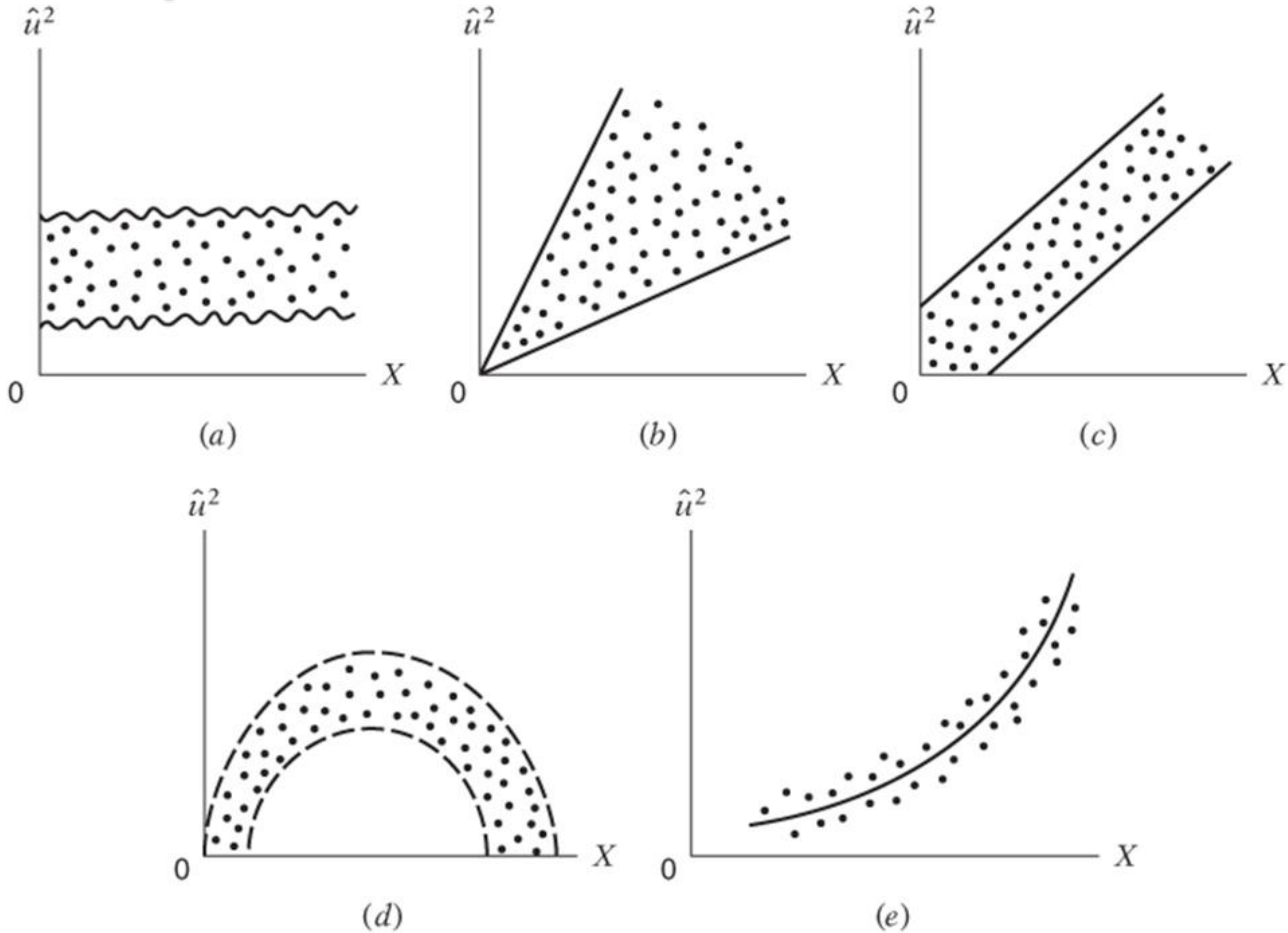
<u>Formal Methods</u>

- Park test
- Glejser test
- Spearman's rank correlation test
- Goldfeld-Quandt test
- Breusch-Pagan-Godfrey (BPG) test
- White's general heteroscedasticity test

Graphical Method



Graphical Method



Park test

Park suggested: $$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{v_i}$$

or $$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$

Use estimated u_i $$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$
$$= \alpha + \beta \ln X_i + v_i$$

1. Estimate model disregarding heteroscedasticity--obtaining estimated u_i.
2. Regress the above regression equation.

- If β turns out to be significant, it suggest that heteroscedasticity occurs.

Glejser test

Glejser suggested others alternative functional forms:

$$|\hat{u}_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

White's Heteroscedasticity test

White suggested: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

1. Estimate model by OLS and obtain the estimated residuals.

2. Regress $\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_4 X_{2i} X_{3i} + v_i$

3. Compute $n \cdot R^2 \underset{asy}{\sim} \chi^2_{df=5}$

4. If computed χ^2 exceeds the critical value, there is heteroscedasticity.

Remedial Measures

When σ_i^2 is known

- Method of Generalized Least Squares (GLS) will lead to estimators that are BLUE.

When σ_i^2 is not known

- White's heteroscedasticity-consistent variances and standard errors.
- Method of Weighted Least Squares (WLS) can be employed.

Plausible Assumptions about Heteroscedasticity Pattern

Assumption 1 $E\left(u_i^2\right) = \sigma^2 X_i^2$

Assumption 2 $E\left(u_i^2\right) = \sigma^2 X_i$

Assumption 3 $E\left(u_i^2\right) = \sigma^2 \left[E(Y_i)\right]^2$

Assumption 4 Log transformation

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

Weighted Least Squares

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\text{var}(u_i) = \sigma_i^2 = \alpha_1 + \alpha_2 z_i + \varepsilon_i$$

1. Estimate model using OLS and obtain $\hat{u}_i$
2. Estimate $\hat{u}_i^2 = \alpha_1 + \alpha_2 z_i + e_i$ and obtain $\hat{\hat{u}}_i^2$ as predicted value for $\hat{\sigma}_i^2$
3. Then, $$\frac{Y_i}{\hat{\sigma}_i} = \beta_1\left(\frac{1}{\hat{\sigma}_i}\right) + \beta_2\left(\frac{X_i}{\hat{\sigma}_i}\right) + \left(\frac{u_i}{\hat{\sigma}_i}\right)$$

WLS

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^*$$