

Heterocedasticidade

Econometria

Alexandre Gori Maia

Ementa:

- Definição;
- Identificação: Análise Gráfica, Goldfeld-Quandt, Breusch-Pagan, White;
- Correção: MQP e MQGF;
- Estimadores Robustos para a Variância;

Bibliografia Básica:

- Maia, Alexandre Gori (2017). *Econometria: conceitos e aplicações*. Cap. 12.

Teorema de Gauss-Markov

1) Relação linear entre os regressores x e Y:

O modelo só é válido para relações lineares.

2) Os valores de x são fixos em repetidas amostras, não aleatórios:

Quem varia é o regressando, o regressor é fixo, qualquer que seja a amostra. Em outras palavras, dados os valores controlados de x, Y variará aleatoriamente segundo uma distribuição de probabilidade, com valor esperado dado por $E(Y|x)$.

3) Esperança condicional dos erros igual a zero, ou seja, $E(e|x)=0$:

É a mesma coisa afirmar que $E(Y|x)=x_i \beta$.

4) A variabilidade dos erros é constante, ou seja, $E(e_i^2)=\sigma^2$

Os erros são homocedásticos, ou seja, sua variância é uma constante.

5) Os erros são não autocorrelacionados, ou seja, $E(e_i e_j)=0$:

Não há relação entre valores dos erros.

Heterocedasticidade - Definição

Homocedasticidade:

A variância dos erros e , condicionada aos valores das variáveis explanatórias, será constante.

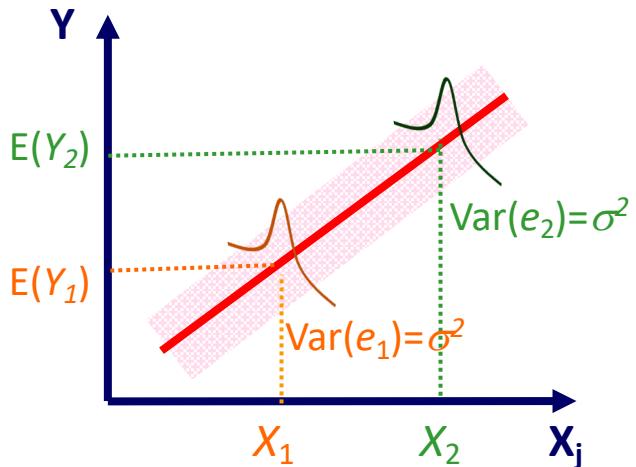
$$\text{Var}(e_i / X_{1_i}, X_{2_i}, \dots, X_{k_i}) = \sigma^2$$

Heterocedasticidade:

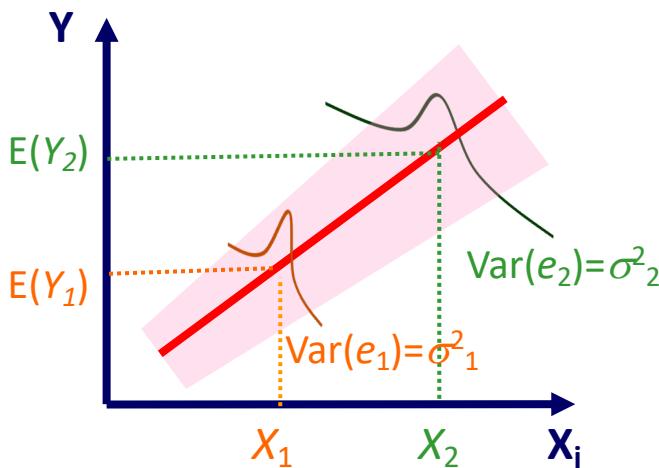
A variância dos erros será diferente para cada valor condicional de X_j .

$$\text{Var}(e_i / X_{1_i}, X_{2_i}, \dots, X_{k_i}) = \sigma_i^2$$

Homocedasticidade



Heterocedasticidade



Heterocedasticidade - Causas

Principais causas da heterocedasticidade:

- **Natureza das variáveis:** alguns relacionamentos apresentam tipicamente tendência à heterocedasticidade. Por exemplo, renda e poupança.

$$Y_i = \alpha + \beta X_i + e_i \quad \Rightarrow \quad \text{Var}(e_i) = \sigma_i^2$$

- **Valores extremos:** a ocorrência de um valor extremo na amostra pode inflacionar a variabilidade em um determinado ponto do ajuste.

- **Falhas na especificação do modelo:** a heterocedasticidade pode também ser devida à omissão de importantes variáveis no modelo.

$$Y_i = \alpha + \beta X_i + e_i \quad \Rightarrow \quad Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

- **Transformação dos dados:** a transformação das variáveis (por exemplo, proporção ao invés de valores absolutos) ou da forma funcional (modelo log-duplo ao invés de linear) pode eliminar a heterocedasticidade.

$$Y_i = \alpha + \beta X_i + e_i \quad \Rightarrow \quad \ln(Y_i) = \alpha + \beta \ln(X_i) + e_i$$

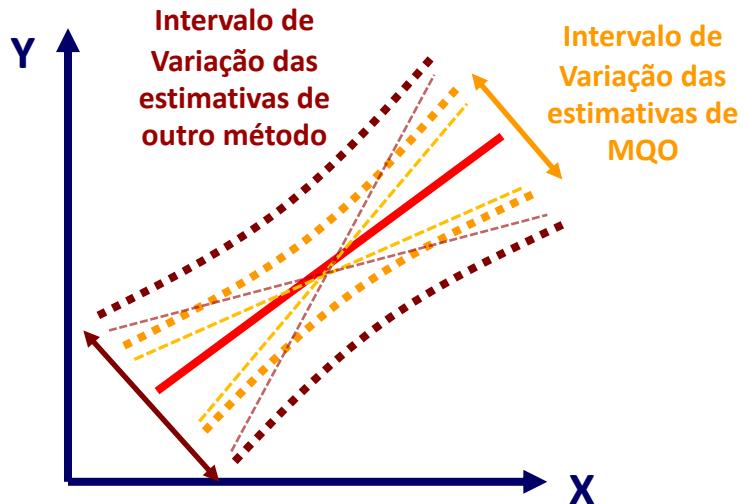
Consequências - Ineficiência

Ineficiência dos Estimadores de MQO

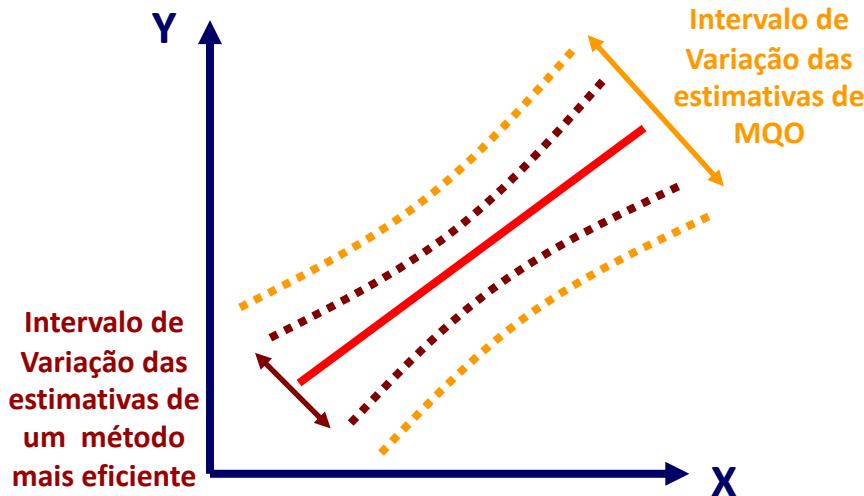
Na presença de heterocedasticidade nos erros, os estimadores de MQO continuam sendo não viesados e consistentes, mas deixam de ser eficientes (ou seja, não possuem mais variância mínima). Em outras palavras, seja $\hat{\beta}$ o estimador de MQO, então existe outro estimador $\hat{\beta}^*$ tal que:

$$\text{Var}(\hat{\beta}^*) < \text{Var}(\hat{\beta})$$

Homocedasticia



Heterocedasticia



Consequências – Tendenciosidade

Tendenciosidade da Variância dos Estimadores:

Outra importante consequência da heterocedasticidade é o viés do estimador da variância de $\hat{\beta}$, mesmo para amostras grandes (inconsistência). Como resultado, as estatísticas de teste t e F deixam de ser válidas, pois dependem da variância do estimador. Em outras palavras, na presença de heterocedasticidade teremos:

$$E(S_{\hat{\beta}}^2) \neq Var(\hat{\beta})$$

Seja o modelo:

$$Y_i = \alpha + \beta X_i + e_i$$

Pelo MQO:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \quad \text{e} \quad S_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2}$$

não viesado na ausência de heterocedasticidade e viesado na sua presença

No caso de homocedasticia:

$$Var(e_i) = \sigma^2$$



$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

No caso de heterocedasticia:

$$Var(e_i) = \sigma_i^2$$



$$Var(\hat{\beta}) = \frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{(\sum_{i=1}^n x_i^2)^2}$$

Identificação

Principais testes para se detectar a heterocedasticidade:

Identificação

Análise gráfica: $\hat{e}^2 \times X_j$

Testes
Estatísticos

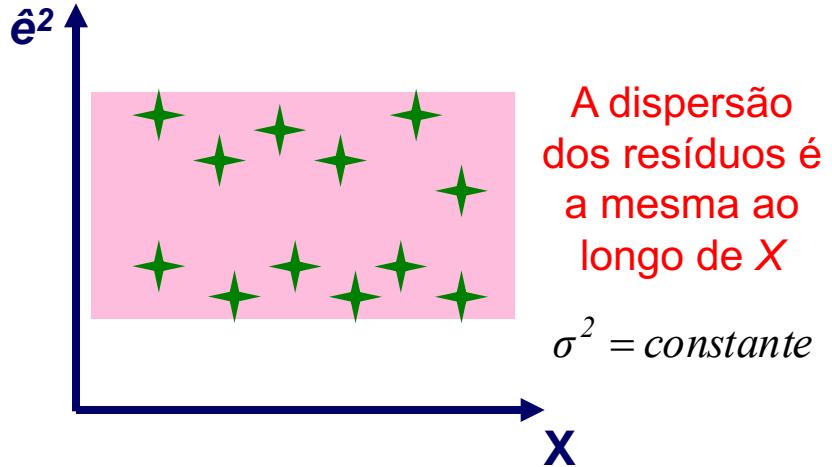
Teste Goldfeld-Quandt

Teste de Breusch-Pagan

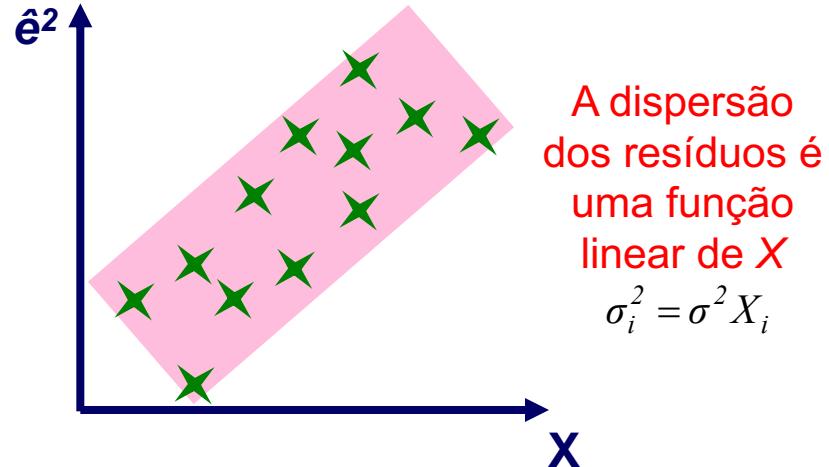
Teste de White

Análise Gráfica

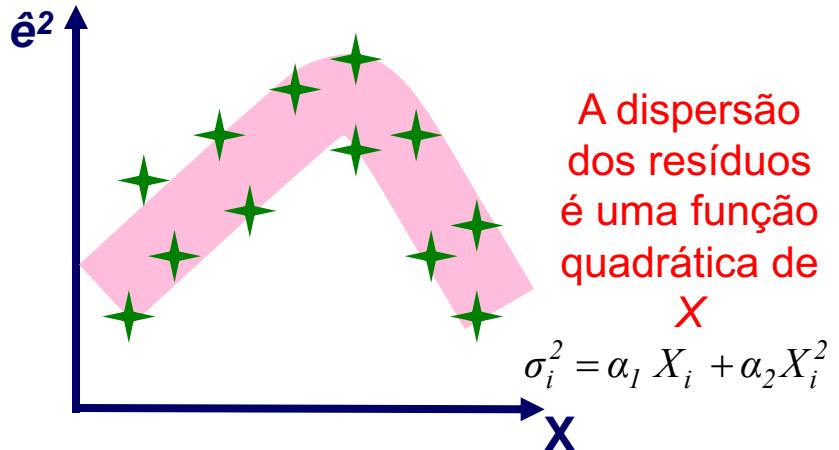
Homocedasticidade



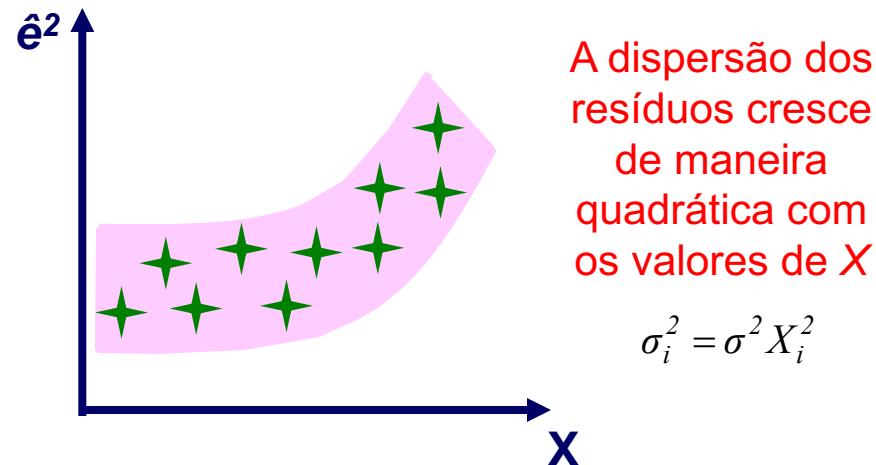
Heterocedasticidade



Heterocedasticidade

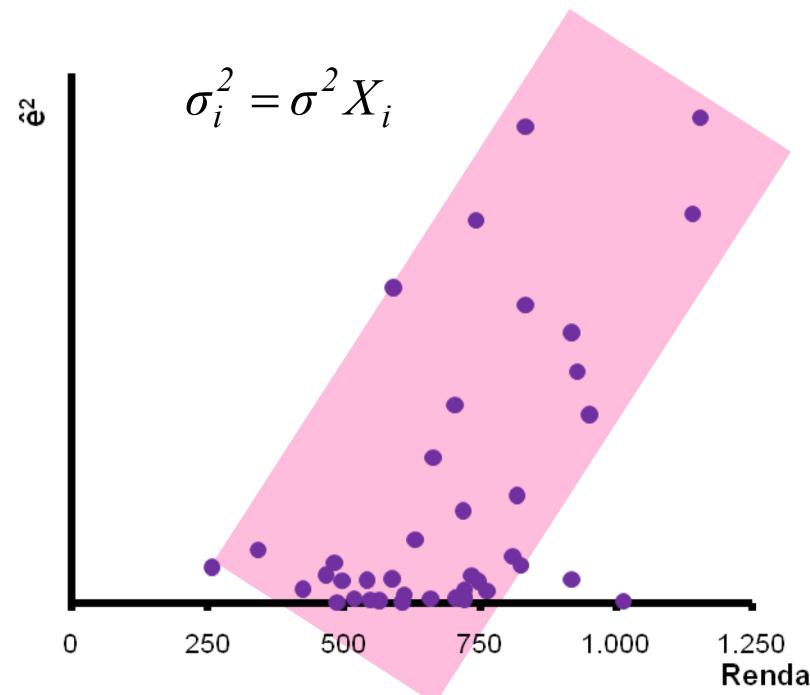
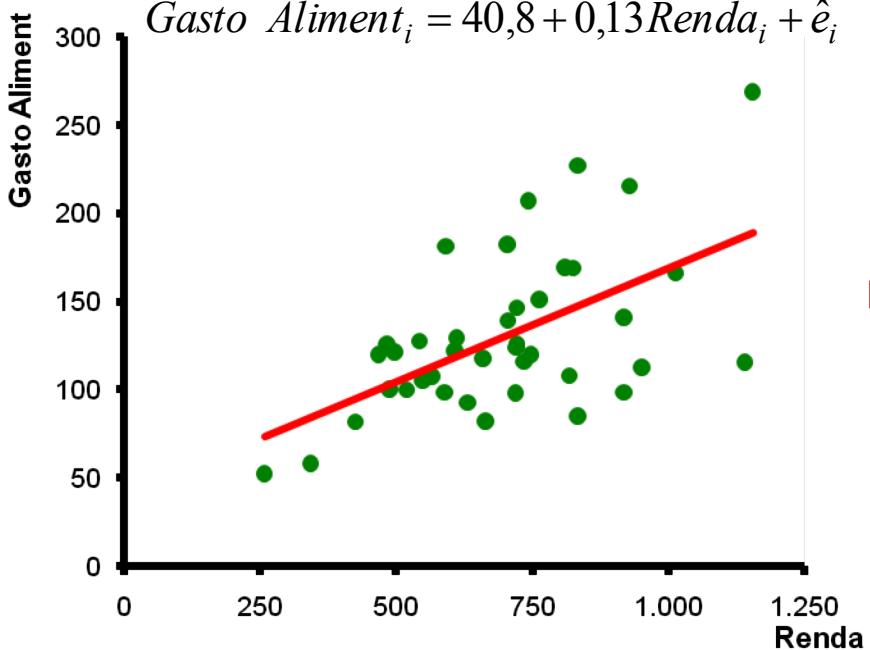


Heterocedasticidade



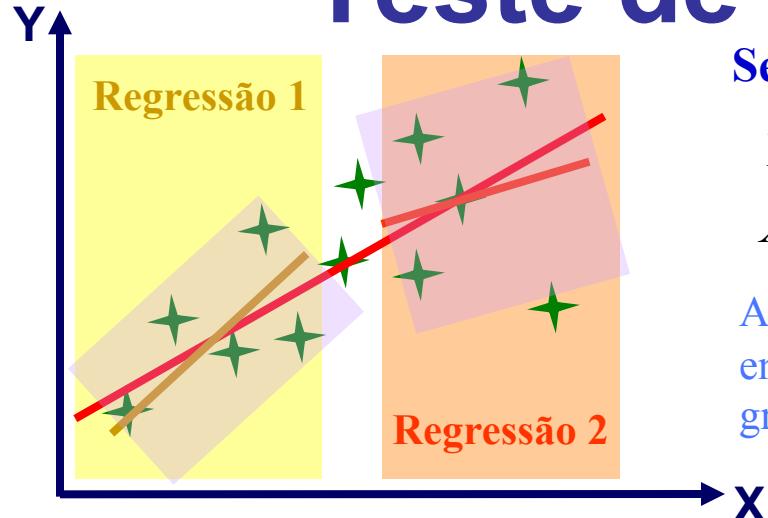
Análise Gráfica

Sejam os dados de 40 famílias para gastos com alimentação (Y) e renda (X):



A dispersão dos resíduos em função da variável X (renda) sugere que, à medida que a renda cresce, a dispersão dos resíduos também aumenta, indicando a presença de heterocedasticidade, em uma relação aparentemente linear.

Teste de Goldfeld-Quandt



Sejam os valores da amostra:

$$Y_1 \quad Y_2 \quad \dots \quad Y_n$$

$$X_1 \quad X_2 \quad \dots \quad X_n$$

Onde: $X_1 < X_2 < \dots < X_n$

A omissão de c observações centrais objetiva acentuar a diferença entre o grupo com variância pequena ($SQReg_1$) e com variância grande ($SQReg_2$).

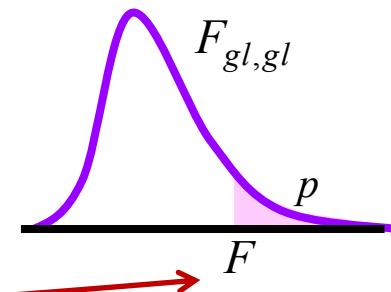
Para testar a hipótese nula da homocedasticidade:

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 < \sigma_2^2 \end{cases}$$



$$F = \frac{SQRes_2 / gl}{SQRes_1 / gl}$$

onde $gl = \frac{n-c}{2} - (k+1)$

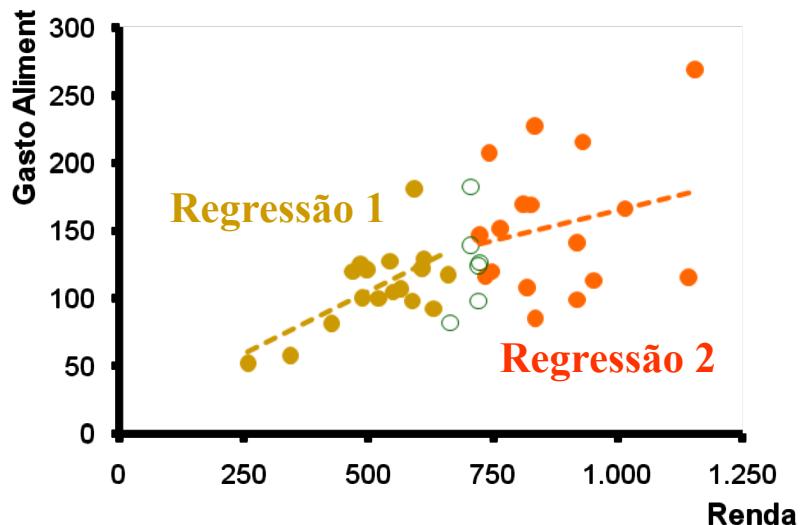


Passos para efetuar o teste de Goldfeld-Quandt

- 1- Ordenar as observações da amostra de acordo com os valores de X ;
- 2- Omitir c observações centrais para dar mais poder ao teste (c costuma ser igual a 4 para $n=30$ e $c=10$ para $n=60$) e separar observações em duas subamostras de $(n-c)/2$ observações;
- 3- Ajustar uma regressão para cada subamostra (cada regressão terá k variáveis independentes);
- 4- Testar hipótese da igualdade dos erros quadráticos médios a partir da estatística F.

Teste de Goldfeld-Quandt

Sejam os gastos de 40 famílias com alimentação em função da renda:

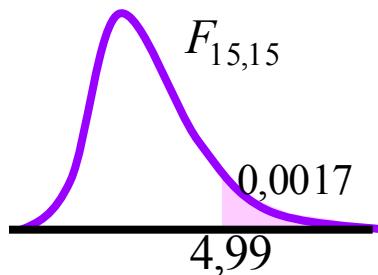


Das 40 observações originais, foram eliminadas 6 observações centrais para dar mais poder ao teste. Restaram dois subconjuntos com 17 observações cada.

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 < \sigma_2^2 \end{cases}$$

estatística de teste: $F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{2629,9}{526,7} = 4,99$

E para calcular a probabilidade de erro do tipo I:



Rejeita-se H_0 , ou seja, pode-se afirmar que há diferença entre as variâncias (heterocedasticidade) com uma probabilidade de erro de apenas 0,17%

Amostra 1

$$Gasto\ Aliment_i = 12,6 + 0,18 Renda_i + \hat{e}_i$$

Fonte	gl	SQ	QM	F	p
Ressagem	1	5967,2	5967,2	11,33	0,0042
Resíduos	15	7900,0	526,7		
Total	16	13867,2			

Amostra 2

$$Gasto\ Aliment_i = 75,1 + 0,09 Renda_i + \hat{e}_i$$

Fonte	gl	SQ	QM	F	p
Ressagem	1	2308,6	2308,6	0,88	0,3636
Resíduos	15	39449,1	2629,9		
Total	16	41757,7			

Teste de Breusch-Pagan

Seja o modelo de RLM com $k=2$:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

Para verificarmos se os resíduos quadráticos têm relação com os regressores:

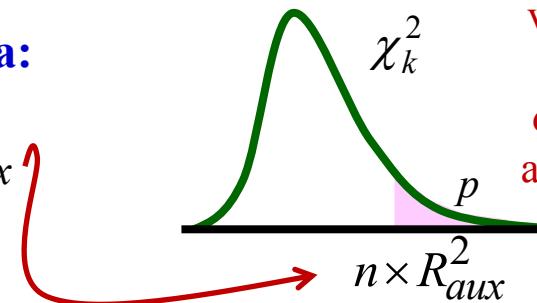
$$\hat{e}_i^2 = \delta_0 + \delta_1 X_{1i} + \delta_2 X_{2i} + u_i$$

Para testar a hipótese nula da homocedasticidade:

$$\begin{cases} H_0 : \delta_1 = \delta_2 = 0 \\ H_1 : \delta_j \neq 0 \end{cases}$$



$$LM = n \times R_{aux}^2$$



Onde k é o número de variáveis explanatórias, R_{aux}^2 o coeficiente de determinação do ajuste auxiliar e n o número de observações

Passos para efetuar o teste de White

- 1- Estimar os resíduos do ajuste de MQO para o modelo original de RLM;
- 2- Ajustar um modelo auxiliar relacionando o quadrado dos resíduos às variáveis independentes do modelo original;
- 3- Calcular a estatística LM pelo produto do número de observações e o R^2 do ajuste auxiliar;
- 4- Calcular o valor p associado à estatística em uma distribuição χ^2 com gl dado pelo número de variáveis explanatórias;

Teste de Breusch-Pagan - Exemplo

Sejam os gastos de 40 famílias com alimentação em função da renda:

Do ajuste original por MQO obtivemos:

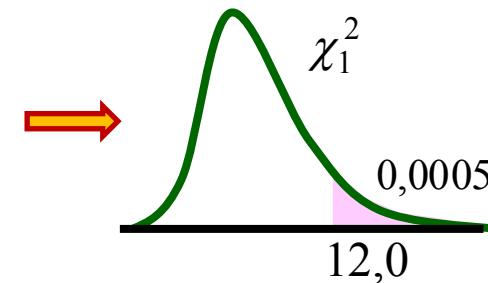
$$\text{Gasto Aliment}_i = 40,8 + 0,13\text{Renda}_i + \hat{e}_i$$

A partir dos resíduos de MQO, ajustamos o seguinte modelo auxiliar:

$$\hat{e}_i^2 = \delta_0 + \delta_1 \text{Renda}_i + u_i \quad \rightarrow \quad \hat{e}_i^2 = -2,279,5 + 5,21 \text{Renda}_i + \hat{u}_i \quad \rightarrow \quad R_{aux}^2 = 0,301$$

O teste de hipóteses será dado por:

$$\begin{cases} H_0 : \delta_1 = 0 \\ H_1 : \delta_1 \neq 0 \end{cases} \quad \rightarrow \quad n \times R_{aux}^2 = 40 \times 0,301 = 12,0$$



A probabilidade de erro ao rejeitar H_0 é de apenas 0,05%. Em outras palavras, há fortíssimas evidências para afirmarmos que os erros são heterocedásticos pois ao fazermos tal afirmação estariámos sujeitos a uma chance de erro de apenas 0,05%.

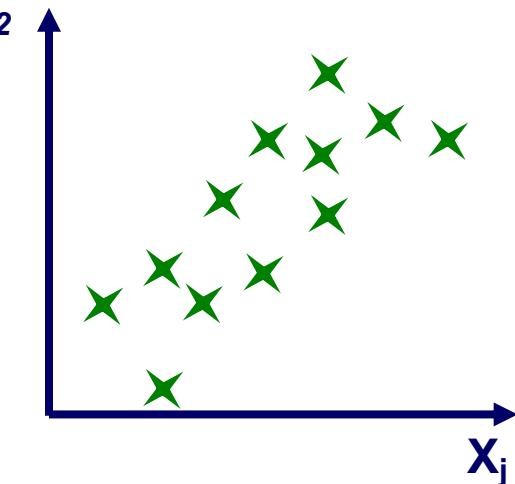
Teste de White

Seja o modelo de RLM com $k=2$:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

Para verificar se os resíduos quadráticos têm relação com os regressores, seus quadrados e seus produtos cruzados:

$$\hat{e}_i^2 = \delta_0 + \delta_1 X_{1i} + \delta_2 X_{2i} + \delta_3 X_{1i} X_{2i} + \delta_4 X_{1i}^2 + \delta_5 X_{2i}^2 + u_i$$

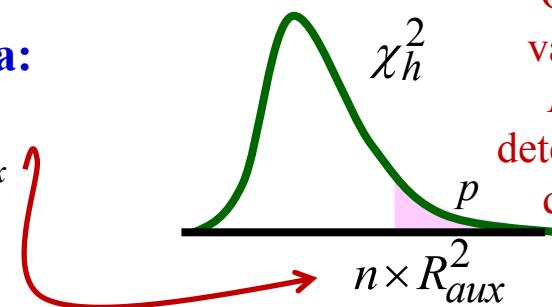


Para testar a hipótese nula da homocedasticidade:

$$\begin{cases} H_0 : \delta_1 = \dots = \delta_5 = 0 \\ H_1 : \delta_j \neq 0 \end{cases}$$



$$LM = n \times R_{aux}^2$$



Onde h é o número de variáveis explanatórias, R_{aux}^2 o coeficiente de determinação e n o número de observações, todos referentes ao ajuste auxiliar

Passos para efetuar o teste de White

- 1- Estimar os resíduos do ajuste de MQO para o modelo original de RLM;
- 2- Ajustar um modelo auxiliar relacionando o quadrado dos resíduos às variáveis independentes do modelo original, seus quadrados e produtos cruzados;
- 3- Calcular a estatística LM pelo produto do número de observações e o R^2 do ajuste auxiliar;
- 4- Calcular o valor p associado à estatística em uma distribuição χ^2 com gl dado pelo número de variáveis explanatórias do ajuste auxiliar;

Teste de White - Exemplo

Sejam os gastos de 40 famílias com alimentação em função da renda:

Do ajuste original por MQO obtivemos:

$$\text{Gasto Aliment}_i = 40,8 + 0,13\text{Renda}_i + \hat{e}_i$$

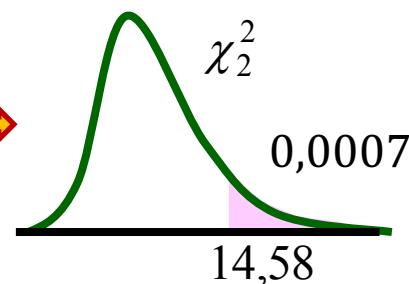
A partir dos resíduos de MQO, ajustamos o seguinte modelo auxiliar:

$$\hat{e}_i^2 = \delta_0 + \delta_1 \text{Renda}_i + \delta_2 \text{Renda}_i^2 + u_i$$

$$\hat{e}_i^2 = 1921 - 7,413\text{Renda}_i + 0,009\text{Renda}_i^2 + \hat{u}_i \quad \longrightarrow \quad R_{aux}^2 = 0,365$$

O teste de hipóteses será dado por:

$$\begin{cases} H_0 : \delta_1 = \delta_2 = 0 \\ H_1 : \delta_j \neq 0 \end{cases} \quad \longrightarrow \quad n \times R_{aux}^2 = 40 \times 0,365 = 14,6 \quad \longrightarrow$$



A probabilidade de erro ao rejeitar H_0 é de apenas 0,07%. Em outras palavras, há fortíssimas evidências para afirmar que os erros são heterocedásticos pois, ao fazermos tal afirmação, estariámos sujeitos a uma chance de erro de apenas 0,07%.

Correção Heterocedasticidade

Dada a equação de RLM:

$$Y_i = \alpha + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + e_i$$

Haverá heterocedasticidade quando:

$$Var(e_i) = E(e_i^2) = \sigma^2$$

Que pode ainda ser representado por:

$$E(e_i^2) = \sigma^2 v_i$$

onde o fator v_i indica como varia a variância de e para cada observação i

A equação de RLM equivale à transformação:

$$\frac{Y_i}{\sqrt{v_i}} = \alpha \frac{1}{\sqrt{v_i}} + \beta_1 \frac{X_{1i}}{\sqrt{v_i}} + \cdots + \beta_k \frac{X_{ki}}{\sqrt{v_i}} + \frac{e_i}{\sqrt{v_i}}$$

A nova equação será homocedástica, pois:

$$E \left[\left(\frac{e_i}{\sqrt{v_i}} \right)^2 \right] = E \left[\frac{e_i^2}{v_i} \right] = \frac{1}{v_i} E[e_i^2] = \sigma^2$$

A equivalente matricial:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Haverá heterocedasticidade quando:

$$Var(\mathbf{e}) = E(\mathbf{e}\mathbf{e}^T) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} v_1 & 0 & 0 & 0 \\ 0 & v_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & v_n \end{bmatrix} \sigma^2 = \mathbf{V}\sigma^2$$

A equação matricial equivale à transformação :

$$\Lambda \mathbf{y} = \Lambda \mathbf{X}\boldsymbol{\beta} + \Lambda \mathbf{e} \quad \text{onde} \quad \Lambda = \begin{bmatrix} 1/\sqrt{v_1} & 0 & 0 & 0 \\ 0 & 1/\sqrt{v_2} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1/\sqrt{v_n} \end{bmatrix}$$

A nova equação será homocedástica pois:

$$E(\Lambda \mathbf{e} \mathbf{e}^T \Lambda) = \Lambda \mathbf{V} \Lambda \sigma^2 = \mathbf{I} \sigma^2$$

Mínimos Quadrados Ponderados

Ponderação da variáveis:

Uma vez conhecida a matriz \mathbf{V} de variâncias e covariâncias do erros de um modelo de RLM, podemos obter os **MELNV** transformando as variáveis originais. Em outras palavras, seja o modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad \text{em que} \quad \text{Var}(\mathbf{e}) = \begin{bmatrix} v_1 & 0 & 0 & 0 \\ 0 & v_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & v_n \end{bmatrix} \sigma^2 = \mathbf{V}\sigma^2$$

Então: $\Lambda\mathbf{y} = \Lambda\mathbf{X}\boldsymbol{\beta} + \Lambda\mathbf{e}$ em que $\Lambda = \begin{bmatrix} 1/\sqrt{v_1} & 0 & 0 & 0 \\ 0 & 1/\sqrt{v_2} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1/\sqrt{v_n} \end{bmatrix}$ e $\Lambda^T\Lambda = \mathbf{V}^{-1}$

Apresentará erros homocedásticos, pois: $E(\Lambda\mathbf{e}\mathbf{e}^T\Lambda) = \Lambda\mathbf{V}\Lambda^T\sigma^2 = \mathbf{I}\sigma^2$

Mínimos Quadrados Ponderados

Estimadores de MQP:

Seja o modelo de RLM:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad \text{com heterocedasticidade dada por } \text{Var}(\mathbf{e}) = \mathbf{V}\sigma^2$$

Então $\Lambda\mathbf{y} = \Lambda\mathbf{X}\boldsymbol{\beta} + \Lambda\mathbf{e}$ será homocedástico, pois $\text{Var}(\Lambda\mathbf{e}) = \mathbf{I}\sigma^2$

Caso a matriz Λ seja conhecida, eu posso aplicar MQO às variáveis transformadas $\Lambda\mathbf{y}$ e $\Lambda\mathbf{X}$ para estimar os coeficientes $\boldsymbol{\beta}$.

Alternativamente, eu posso obter os estimadores de MQP diretamente por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \Lambda^T \Lambda \mathbf{X})^{-1} \mathbf{X}^T \Lambda^T \Lambda \mathbf{y} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

Raciocínio análogo é válido para obter os estimadores da variância:

$$\text{SQRes} = \mathbf{y}^T \mathbf{V}^{-1} \mathbf{y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} \quad \text{e} \quad S_{\hat{\boldsymbol{\beta}}}^2 = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \hat{\sigma}^2$$

Esse método é denominado de **Mínimos Quadrados Ponderados (MQP)**, um caso específico do método de **Mínimos Quadrados Generalizados (MQG)**.

Os estimadores de MQP são não tendenciosos e os mais eficientes (MELNV) na presença de heterocedasticidade.

MQP com V conhecida - Exemplo

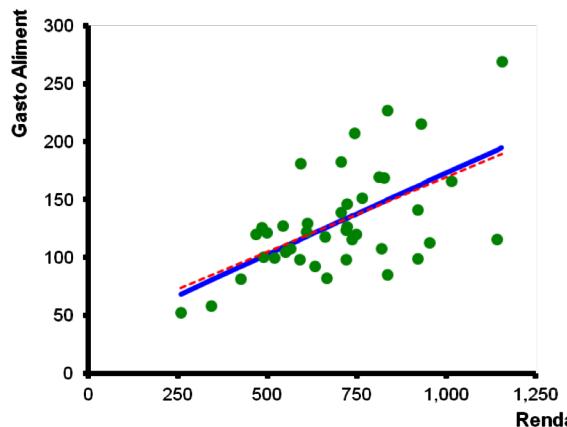
Sejam os gastos de 40 famílias com alimentação em função da renda:

As estimativas de MQO são:

$$\text{Gasto Aliment}_i = 40,8 + 0,13\text{Renda}_i + \hat{e}_i$$

Supondo que:

$$Var(e_i) = X_i \sigma^2 \quad \text{arrow} \quad Var(e) = \begin{bmatrix} X_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & X_{40} \end{bmatrix} \sigma^2 = \mathbf{V} \sigma^2$$



Para obter as estimativas de MQP, aplicamos MQO ao modelo transformado:

$$\frac{\text{Aliment}_i}{\sqrt{\text{Renda}_i}} = \alpha \frac{1}{\sqrt{\text{Renda}_i}} + \beta \frac{\text{Renda}_i}{\sqrt{\text{Renda}_i}} + \frac{e_i}{\sqrt{\text{Renda}_i}} \quad \text{arrow} \quad \Lambda \mathbf{y} = \Lambda \mathbf{X} \boldsymbol{\beta} + \Lambda \mathbf{e}$$

em que $\Lambda^T \Lambda = \mathbf{V}^{-1}$

As estimativas de MQP seriam:

$$\text{Aliment}_i = 32.0 + 0.14\text{Renda}_i + \hat{e}_i^*$$

As estimativas de MQP são não tendenciosas e as mais eficientes. O estimador da variância também é não tendencioso. Mas a validade dessas estimativas depende de um pressuposto forte, que a variância dos erros seja de fato uma função linear da renda.

MQGF - V Desconhecida

Mínimos Quadrados Generalizados Factíveis (MQGF):

Seja o modelo de RLM:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad \text{com heterocedasticidade dada por: } \text{Var}(e_i) = v_i \sigma^2$$

Caso o fator v_i seja desconhecido, podemos estimá-lo a partir da relação da dispersão dos resíduos de MQO com as variáveis \mathbf{X} do modelo. Há várias especificações que podem ser assumidas. Uma estratégia simples é assumir que a dispersão seja uma função logarítmica de \mathbf{X} :

$$\ln(\hat{e}_i^2) = \delta_0^* + \delta_1 X_{1i} + \dots + \delta_k X_{ki} + u_i^* \quad \Rightarrow \quad \hat{v}_i = e^{\hat{\delta}_0^* + \hat{\delta}_1 X_{1i} + \dots + \hat{\delta}_k X_{ki}}$$

A transformação exponencial garante que todas as estimativas de \hat{v} sejam positivas (variância não assume valores negativos!).

As estimativas de v_i podem ser substituídas na matriz de ponderações \mathbf{V} , agora denominada $\hat{\mathbf{V}}$, para obtermos os estimadores **consistentes** (embora possam ser viesados) de *Mínimos Quadrados Generalizados Factíveis*:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{y}$$

MQGF com V Desconhecida - Exemplo

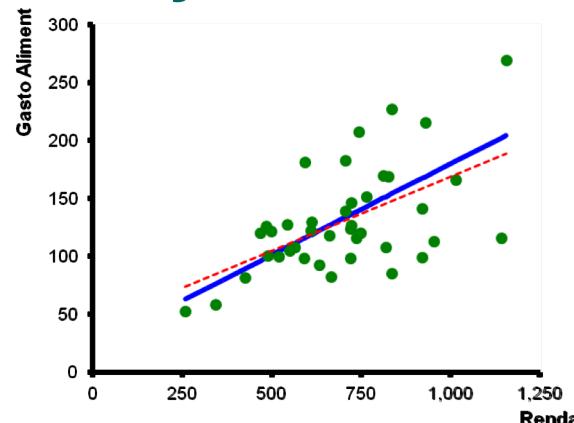
Sejam os gastos de 40 famílias com alimentação em função da renda:

As estimativas de MQO foram:

$$\text{Gasto Aliment}_i = 40,8 + 0,13\text{Renda}_i + \hat{e}_i$$

Supondo agora que:

$$v_i = e_i^2 = e^{\delta_0 + \delta_1 \text{Renda}_i} u_i \quad \Rightarrow \quad \ln(e_i^2) = \delta_0^* + \delta_1 \text{Renda}_i + u_i^*$$



O fator de correção v_i pode então ser estimado por:

$$\ln(\hat{e}_i^2) = 3,363 + 0,004 \text{Renda}_i + \hat{u}_i \quad \Rightarrow \quad \hat{v}_i = e^{3,363 + 0,004 \text{Renda}_i}$$

As variáveis transformadas seriam:

$$\frac{\text{Aliment}_i}{\sqrt{\hat{v}_i}} = \alpha \frac{1}{\sqrt{\hat{v}_i}} + \beta \frac{\text{Renda}_i}{\sqrt{\hat{v}_i}} + \frac{e_i}{\sqrt{\hat{v}_i}} \quad \Rightarrow \quad \widehat{\Lambda}\mathbf{y} = \widehat{\Lambda}\mathbf{y}\boldsymbol{\beta} + \widehat{\Lambda}\mathbf{e} \quad \text{em que} \quad \widehat{\Lambda}^T \widehat{\Lambda} = \widehat{\mathbf{V}}^{-1}$$

E as estimativas de MQGF seriam:

$$\text{Aliment}_i = 22,6 + 0,16 \text{Renda}_i + \hat{e}_i^*$$

As estimativas de MQGF são válidas apenas para amostras grandes (consistentes). Nesse caso, as estimativas de MQGF seriam (assintoticamente) também mais eficientes que as de MQO. Mas, para amostras pequenas, podem ser tendenciosas.

Estimadores Robustos

Estimadores da Variância Robustos à Heterocedasticidade - White:

Seja o modelo de RLS:

$$Y_i = \alpha + \beta X_i + e_i \quad \text{com heterocedasticidade dada por: } \text{Var}(e_i) = v_i \sigma^2$$

Embora os estimadores de MQO para β continuem não viesados, os estimadores de suas variâncias seriam viesados.

Um estimador *robusto* da variância do estimador $\hat{\beta}$, ou seja, igualmente válido na presença ou não de heterocedasticidade, pode ser dado por:

$$\text{Var}(\hat{\beta}) = \frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} \quad \Rightarrow \quad S_{\hat{\beta}}^2 = \frac{\sum_{i=1}^n x_i^2 \hat{e}_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}$$

No caso de uma RLM teríamos:

$$Y_i = \alpha + \sum_{j=1}^k \beta_j X_{j_i} + e_i \quad \Rightarrow \quad S_{\hat{\beta}_j}^2 = \frac{\sum_{i=1}^n \hat{u}_{j_i}^2 \hat{e}_i^2}{\left(\sum_{i=1}^n \hat{u}_{j_i}^2\right)^2}$$

Onde \hat{u}_j é o resíduo do ajuste de X_j em função das demais variáveis independentes

Esse estimador *robusto* à heterocedasticidade, atribuído a White, é válido(não tendencioso) assintoticamente. Para amostras pequenas, as estatísticas t e F baseadas no estimador robusto não serão necessariamente válidas.

Estimador Robusto - Exemplo

Sejam os gastos de 40 famílias com alimentação em função da renda:

As estimativas de MQO foram:

$$\text{Gasto Aliment}_i = \alpha + \beta \text{Renda}_i + e_i \quad \rightarrow \quad \text{Gasto Aliment}_i = 40,8 + 0,13 \text{Renda}_i + \hat{e}_i$$

Caso os erros sejam homocedásticos, a estimativa de MQO para $\text{Var}(\hat{\beta})$ seria:

$$S_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2} = \frac{1.429}{1.532.463} = 0,00093 = 0,031^2$$

Por sua vez, uma estimativa robusta à heterocedasticidade para $\text{Var}(\hat{\beta})$ seria:

$$S_{\hat{\beta}^*}^2 = \frac{\sum_{i=1}^n x_i^2 \hat{e}_i^2}{(\sum_{i=1}^n x_i^2)^2} = \frac{3.421.453.919}{(1.532.463)^2} = 0,0015 = 0,038^2$$

Finalmente, a estatística t para testar a significância de β seria:

$$t = \frac{0,13}{0,038} = 3,36$$

Embora haja evidências de heterocedasticidade no modelo, devemos ter cautela na interpretação desse teste t , já que o estimador robusto pode não ser válido para amostras pequenas ($n=40$).