Heterocedasticidade
Definição, Conseqüências, Identificação e Correção

Profa. Rosângela Ballini

Bibliografia:

1. WOLDRIDGE, J.M. (2006). Introductory Econometrics: a Modern Approach. Cap. 8.

Teorema de Gauss-Markov

Para que os estimadores de MQO sejam os Melhores Estimadores Lineares Não Viesados (MELNV, ou BLUE):

1) O modelo de regessão é linear nos parâmetros;
2) Os valores de X são fixos em repetidas amostras;
3) Ausência de colinearidade perfeita;
4) $E(u_j|X_1,\ldots,X_k) = 0$;
5) $Var(u_j|X_1,\ldots,X_k) = \sigma^2$ constante;
6) $corr(u_i,u_j) = 0,\ i \neq j$;
7) O termo estocástico u_j se distribui normalmente.

Definição

A suposição de homocedasticidade implica que, condicional às variáveis explicativas, a variância do erro é constante. Ou seja,

$$Var\left(u_j | X_{1j}, X_{2j}, \ldots, X_{kj}\right) =$$

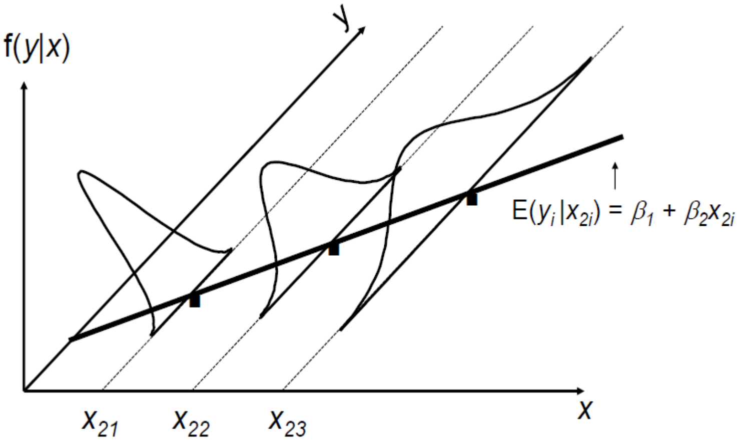
$$= \sigma^2 \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma^2 & 0 & \ldots & 0 \\ 0 & 0 & \sigma^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \sigma^2 \end{pmatrix}$$

Definição

Considerando que as hipóteses de 1, 2, 3, 4 e 6 sejam válidas (ou seja, apenas a suposição 5 está sendo violada), heterocedasticidade ocorre quando a dispersão em torno da regressão se altera em função dos valores da(s) variável(eis) explanatória(s), ou seja,

$$Var\left(u_j | X_{1j}, X_{2j}, \ldots, X_{kj}\right) = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ 0 & 0 & \sigma_3^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \sigma_n^2 \end{pmatrix}$$

Exemplo



Causas da heterocedasticidade

1. Natureza das variáveis: algumas relações apresentam tipicamente tendência à heterocedasticia.
2. Valores extremos: ocorrência de um valor extremo pode inflacionar a variabilidade em um determinado ponto do ajuste.
3. Falhas na especificação do modelo: heterocedasticidade pode ser devido à omissão de variáveis importantes do modelo ou forma funcional incorreta.

Conseqüências da Heterocedasticidade

Seja o modelo de regressão linear simples, com as variáveis dependente e independente centradas, ou seja, $y_j = (Y_j - \bar{Y})$ e $x_j = (X_j - \bar{X})$:

$$y_j = \beta_1 x_j + u_j \tag{1}$$

Estimador de β_1:

$$b_1 = \frac{\sum x_j y_j}{\sum x_j^2} = \frac{\sum x_j(\beta_1 x_j + u_j)}{\sum x_j^2} = \beta_1 + \frac{\sum x_j u_j}{\sum x_j^2} \tag{2}$$

Valor médio de b_1:

$$E(b_1) = \beta_1 + \frac{E(\sum x_j u_j)}{\sum x_j^2} = \beta_1 \tag{3}$$

Consequências para os Estimadores de MQO

Supondo o modelo (1) homocedástico, o cálculo da variância do estimador b_1 é dado por:

$$Var(b_1) = \frac{\sigma^2}{\sum x_j^2} \tag{4}$$

Para o caso de presença de heterocedasticidade, a variância do estimador b_1 é dada por:

$$Var(b_1) = \frac{\sum x_j^2 \sigma_j^2}{\left(\sum x_j^2\right)^2} \tag{5}$$

Consequências para os Estimadores de MQO

Na forma matricial, temos que os estimadores são dados por:

$$\hat{\beta}^{(MQO)} \overset{Hip.3}{=} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \overset{Hip.1}{=} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{X}\beta + \mathbf{u}\right) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'u$$

Com valor médio dado por:

$$E[\hat{\beta}^{(MQO)}] = E\left[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'u\right] \overset{Hip.2}{=} \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\left[u\right] \overset{Hip.4}{=} \beta$$

A hipótese de homocedasticidade (suposição 5) NÃO desempenha papel algum na demonstração de que os estimadores de MQO do vetor de parâmetros do modelo de regressão linear é não viesado e consistente.

Consequências para os Estimadores de MQO

Na forma matricial, para a variância:

$$Var\left(\hat{\beta}^{(MQO)}\right) = E\left[\left(\hat{\beta}^{(MQO)} - \beta\right)\left(\hat{\beta}^{(MQO)} - \beta\right)'\right] =$$

$$= E\left\{\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right]\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right]'\right\} =$$

$$= \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\left\{\mathbf{u}'\mathbf{u}\right\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right\}$$

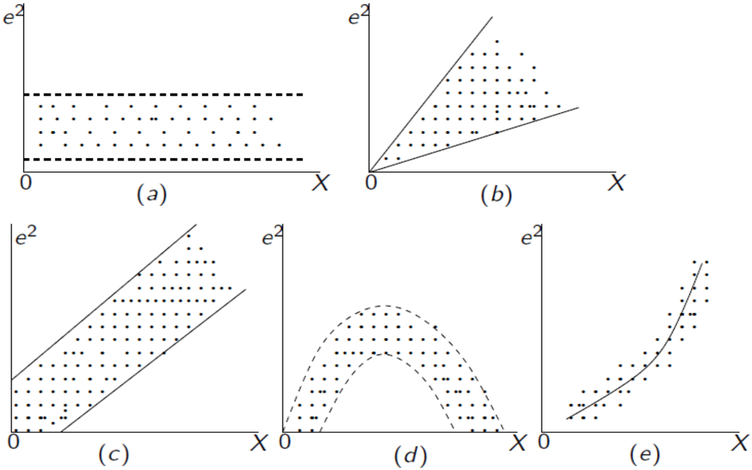
$$Var\left(\hat{\beta}^{(MQO)}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Consequências para os Estimadores de MQO

- O **Teorema de Gauss-Markov**, que afirma que os estimadores de MQO são os melhores estimadores lineares não viesados (**BLUE**), depende de forma crucial da suposição de homocedasticidade. Na presença da heterocedasticidade, os estimadores de MQO não são mais BLUE e nem assintoticamente eficientes.

Identificação de Heterocedasticidade

1) Análise Gráfica: dispersão entre e_j^2 (ou e_j) e X ou $\hat{Y}$.



Aplicação

A partir dos dados hprice.xlsx, estime o modelo de regressão linear múltipla:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

Faça o gráfico de dispersão dos resíduos em função das variáveis independentes e em relação aos valores estimados de *price*.

Identificação de Heterocedasticidade - Testes Paramétricos

2) Métodos Formais (Testes Paramétricos):

a) Teste de Breuch-Pagan;

b) Teste de White.

Testes de Hipótese para Detecção de Heterocedasticidade

Considere a Hipótese de Homocedasticidade:

$$H_0 : Var(u_i | x_{1i}, x_{2i}, \ldots, x_{ki}) =$$

$$= E(u_i^2 | x_{1i}, x_{2i}, \ldots, x_{ki}) = \sigma^2 \qquad (6)$$

Para H_0 ser rejeitada, precisamos encontrar relação entre o u^2 e variáveis explicativas.

Testes de Homocedasticidade - Teste Breuch-Pagan

Da eq (6), temos:

$$u_i^2 = \delta_0 + \delta_1 x_{1i} + \ldots + \delta_k x_{ki} + \nu_i \qquad (7)$$

em que ν é o termo de erro da regressão auxiliar.

Hipótese de Homocedasticidade:

$$H_0 : \delta_1 = \delta_2 = \ldots = \delta_k = 0$$

a qual pode ser verificada utilizando teste LM (Multiplicador de Lagrange) ou teste F.

Como a variável resposta, no modelo (7), não é diretamente observada, teremos que estimá-la.

Ou seja, estime o modelo:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj} + u_j$$

por MQO e obtenha os resíduos e_j e eleve ao quadrado e_j^2.

A equação (7) será dada por:

$$e_j^2 = \delta_0 + \delta_1 X_{1j} + \ldots + \delta_k X_{kj} + \nu_j \tag{8}$$

e obtenha o R-quadrado desta regressão R^2_{aux}.

Testes de Homocedasticidade - Teste Breuch-Pagan

Construa a estatística F:

$$F = \frac{R^2_{aux}/(k)}{(1 - R^2_{aux})/(n - k - 1)} \sim F_{(k,n-k-1)}$$

em que k é o número de regressores.

Ou calcule a estatística LM (Multiplicador de Lagrange):

$$LM = nR^2_{aux} \sim \chi^2_k$$

- Rejeitamos H_0 quando o valor observado da estatística de teste for superior ao crítico.

Teste Breuch-Pagan – Exemplo

Considerando os dados do arquivo hprice.xlsx, estime o modelo:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

e aplique o teste de Breush-Pagan.

No R use a função `bptest()`.

Teste de Homocedasticidade – Teste de White

White (1980)[1] motivado pelo fato de que a suposição

$$Var(u_i|x_{1i}, x_{2i}, \ldots, x_{ki}) = \sigma^2,\ i = 1, 2, \ldots, n$$

poderia ser substituída pela suposição de que bastaria o erro ao quadrado, u^2, ser não correlacionado com todas as variáveis explicativas, com os quadrados das variáveis explicativas e com todos os produtos cruzados entre as variáveis explicativas, escreveu seu clássico artigo.

[1] White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, vol. 48, p. 817–838, 1980

Teste de Homocedasticidade – Teste de White

Por exemplo, quando o modelo de interesse apresentar k = 2 variáveis explicativas:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + u_j \tag{9}$$

O teste de White ficará baseado nos resultados da estimação do seguinte modelo de regressão auxiliar

$$e_j^2 = \alpha_0 + \alpha_1 X_{1j} + \alpha_2 X_{2j} + \alpha_3 X_{1j}^2 + \alpha_4 X_{2j}^2 + \alpha_5 X_{1j} X_{2j} + \nu_j \tag{10}$$

em que e_j é o resíduo da regressão (9).

Teste de Homocedasticidade – Teste de White

Neste caso, a hipótese nula de interesse seria

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$$

a qual, segundo White (1980), pode ser testada usando:

$$LM = nR^2_{aux} \sim \chi^2_5$$

em que R^2_{aux} é o coeficiente de determinação da regressão auxiliar (10).

Teste de Homocedasticidade – Teste de White

- Comparado ao teste proposto por Breusch-Pagan, a equação auxiliar (10) envolve 3 regressores a mais.
- Dependendo do número de variáveis explicativas constantes do modelo original de interesse, a perda de graus de liberdade será bastante grande.
- É importante observar se o tamanho da amostra que estamos usando para a estimação é suficientemente grande para a aplicação de tal metodologia.

Teste de Homocedasticidade – Teste de White

Observações:

1. A equação auxiliar envolvida no teste pode ser útil na identificação da forma funcional da heterocedasticidade ou de erro de especificação, ou de ambos.
2. Para minimizar a perda de graus de liberdade, o teste pode ser conduzido utilizando a seguinte regressão auxiliar:

$$e_j^2 = \gamma_0 + \gamma_1 \hat{y}_j + \gamma_2 \hat{y}_j^2 + \nu$$

Exemplo – Teste de White

Considerando os dados do arquivo hprice.xlsx, estime o modelo:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

e aplique o teste de White.

Heterocedasticidade

Verificado heterocedasticidade, deve-se:

1. Ou usar estimadores robustos

2. Ou aplicar Mínimos Quadrados Generalizados (MQG):

   2.1) Mínimos Quadrados Ponderados – σ_i^2 conhecida

   2.2) Mínimos Quadrados Generalizados Factíveis – σ_i^2 desconhecida

Estimadores Robustos

Recentemente, muito se tem desenvolvido com relação ao ajuste de erros padrões e estatísticas de testes para que os mesmos se tornem válidos na presença de heterocedasticidade.

Estes procedimentos são conhecidos como ROBUSTOS pois são válidos, pelo menos para grandes amostras, sendo ou não a variância do erro constante.

Estimadores Robustos

Suponha:

$$Var(\mathbf{u}|\mathbf{X}) = E(\mathbf{u}'\mathbf{u}|\mathbf{X}) = \sigma^2\Omega =$$

$$= \sigma^2 \begin{pmatrix} \omega_1 & & & & \\ & \omega_2 & & & \\ & & \omega_3 & & \\ \dots & \dots & \dots & \dots & \dots \\ & & & & \omega_n \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ \dots & \dots & \dots & \dots & \dots \\ & & & & \sigma_n^2 \end{pmatrix}$$

em que $\sigma_i^2, i = 1, 2, \dots, n$ são parâmetros desconhecidos.

Estimadores Robustos

Supondo válidas as suposições do modelo de regressão, com exceção da suposição da variância constante, temos:

$$Var\left(\hat{\beta}^{(MQO)}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Ou,

$$Var\left(\hat{\beta}^{(MQO)}\right) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum \sigma_i^2 \mathbf{x_i}\mathbf{x_i}'\right)(\mathbf{X}'\mathbf{X})^{-1}$$

Estimadores Robustos

White (1980) demonstrou que uma estimativa simples das quantidades desconhecidas pode ser obtida a partir do cálculo de e_i^2 (quadrado do resíduo de MQO), ou seja,

$$\widehat{Var\left(\hat{\beta}^{(MQO)}\right)} = (\mathbf{X'X})^{-1}\left(\sum e_i^2 \mathbf{x_i x_i'}\right)(\mathbf{X'X})^{-1} \qquad (11)$$

A raiz quadrada dos elementos da diagonal principal desta matriz teremos o que usualmente costuma se chamar de erro-padrão devido a White (ou erro padrão robusto em relação à heterocedasticidade).

Estimadores Robustos

Observações:

1. O erro-padrão robusto pode ser maior ou menor do que o erro-padrão não robusto;
2. Usando um método de estimação robusto, a estatística de teste t também será robusta.
3. Os erros padrão robustos e as estatísticas t robustas são justificadas somente quando o tamanho da amostra se torna grande.

Aplicação

A partir dos dados hprice.xlsx, estime o modelo de regressão linear múltipla:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

e estime os erros padrão robusto e análise o modelo.

Vamos considerar o modelo de regressão linear simples:

$$Y_j = \beta_0 + \beta_1 X_j + u_j \tag{12}$$

Haverá heterocedasticidade quando

$$Var(u_j) = E(u_j^2) = \sigma_j^2$$

a qual pode ser representada por:

$$Var(u_j) = E(u_j^2) = \sigma^2 h_j$$

em que h_j indica como o erro u varia para cada observação.

Se conhecemos h_j podemos corrigir o modelo pela equação:

$$\frac{Y_j}{\sqrt{h_j}} = \beta_0 \left(\frac{1}{\sqrt{h_j}} \right) + \beta_1 \left(\frac{X_j}{\sqrt{h_j}} \right) + \left(\frac{u_j}{\sqrt{h_j}} \right) \tag{13}$$

Quando fazemos essa transformação temos as seguintes características do termo de erro transformado $u_j^* = \dfrac{u_j}{\sqrt{h_j}}$:

$$Var(u_j^*) = E(u_j^*)^2 = E\left(\frac{u_j}{\sqrt{h_j}} \right)^2 = \frac{1}{h_j} E(u_j^2) = \sigma^2$$

pois $E(u_j^2) = \sigma^2 h_j$.

A variância do termo erro transformado u_j^* é homocedástica.

Exemplo - MQP

A partir dos dados hprice.xlsx, estime o modelo de regressão linear múltipla:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

Assumindo que a heterocedasticidade é da forma $\sigma_i^2 = \sigma^2 h_i$, em que $h_i = lotsize_i$. Faça a correção do modelo usando está forma funcional para a heterocedasticidade.

Procedimento MQG Factível para Corrigir a Heterocedasticidade

Caso os erros do modelo de regressão sejam heterocedásticos, o caso mais comum é aquele em que a forma da heterocedasticidade é desconhecida.

Sendo assim, precisamos encontrar uma estimativa para a função $h(\mathbf{x})$, que seja gerada a partir de um estimador com boas propriedades.

MQG Factível para Corrigir a Heterocedasticidade

Suponha que:

$$Var(u|X) = \sigma^2 \exp(\delta_0 + \delta_1 X_1 + \ldots + \delta_k X_k) \qquad (14)$$

Note que $h(\mathbf{x}) = \exp(\delta_0 + \delta_1 X_1 + \ldots + \delta_k X_k) > 0$

Para a estimação dos δ_j reescreveremos (14) como:

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 X_1 + \ldots + \delta_k X_k)\nu$$

Tomando o logaritmo, temos:

$$ln(u^2) = \alpha_0 + \delta_1 X_1 + \ldots + \delta_k X_k + erro \qquad (15)$$

em que *erro* tem média zero e é independente de **x**.

MQG Factível para Corrigir a Heterocedasticidade

Ao estimarmos os parâmetros de (15), conseguiremos uma estimativa para a função $h(\mathbf{x})$, que é obtida de

$$\hat{h}(\mathbf{x}) = \exp(\widehat{ln(u^2)}) \tag{16}$$

Usamos (16) para encontrar as estimativas de MQP e os respectivos erros padrões do modelo de interesse.

Exemplo - MQP

A partir dos dados hprice.xlsx, estime o modelo de regressão linear múltipla:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

Use MQG para fazer a correção da heterocedasticidade.

Medidas Corretivas: Transformação dos Dados

Uma transformação em log do tipo:

$$ln\, Y_j = \beta_0 + \beta_1 ln\, X_j + u_j \qquad (17)$$

muitas vezes reduz a heterocedasticidade quando comparada com a regressão

$$Y_j = \beta_0 + \beta_1 X_j + u_j$$

Exemplo - MQP

A partir dos dados hprice.xlsx, estime o modelo de regressão linear múltipla:

$$log(price) = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 bdrms + u$$

Faça o teste de Breush-Pagan.

Também estime o modelo usando estimador robusto.