AULAS 25 E 26
Heteroscedasticidade

Ernesto F. L. Amaral

10 e 15 de junho de 2010
Métodos Quantitativos de Avaliação de Políticas Públicas (DCP 030D)

Fonte:
Wooldridge, Jeffrey M. “Introdução à econometria: uma abordagem moderna”. São Paulo: Cengage Learning, 2008. Capítulo 8 (pp.243-271).

HOMOSCEDASTICIDADE

– A hipótese de homoscedasticidade para a regressão múltipla significa que a variância do erro não observável (*u*), condicional nas variáveis explicativas, é constante.

– A homoscedasticidade não se mantém quando a variância dos fatores não-observáveis muda ao longo de diferentes segmentos da população.

– Por exemplo, a heteroscedasticidade está presente se a variância dos fatores não-observados (*u*) que afetam a renda (*y*) aumenta com a idade (*x*).

– A homoscedasticidade é necessária para estimar os testes de t e F, além dos intervalos de confiança.

– A intenção aqui é de: (1) discorrer sobre as consequências da heteroscedasticidade para estimação de MQO; (2) verificar a presença da heteroscedasticidade; (3) discutir soluções para a ocorrência deste problema.

β_j E R^2 NA HETEROSCEDASTICIDADE

– A heteroscedasticidade não provoca viés ou inconsistência nos estimadores MQO de β_j, enquanto a omissão de uma variável importante teria esse efeito.

– O R^2 da população é:

$$1 - (\text{variância do erro} / \text{variância de y})$$

– Como ambas variâncias no R^2 da população são incondicionais, o R^2 da população não é afetado pela presença de heteroscedasticidade em $Var(u|x_1,..., x_k)$.

– SQR/n estima consistentemente a variância do erro, e SQT/n estima consistentemente a variância de y, seja $Var(u|x_1,..., x_k)$ constante ou não.

– Portanto R^2 e R^2 ajustados são estimadores consistentes do R^2 da população, mantendo ou não a hipótese de homoscedasticidade.

ERROS-PADRÃO NA HETEROSCEDASTICIDADE

– Os estimadores de variâncias [Var(β_j)] são viesados sem a hipótese de homoscedasticidade.

– Como os erros-padrão dos estimadores MQO são baseados diretamente nessas variâncias, eles não mais são válidos para construirmos intervalos de confiança e estatísticas *t*.

– Na presença de heteroscedasticidade, as estatísticas *t* não têm distribuições *t*, as estatísticas *F* não têm distribuição *F*, e a estatística *LM* não tem distribuição qui-quadrada.

– Portanto, as estatísticas que usamos para testar hipóteses não são válidas na presença de heteroscedasticidade.

– Os estimadores MQO são os melhores estimadores lineares não-viesados na hipótese de homoscedasticidade: isso ocorre quando Var(u|x) for constante.

INFERÊNCIA ROBUSTA

– É possível ajustar erros-padrão, estatísticas *t*, *F* e *LM* de forma a torná-las válidas na presença de heteroscedasticidade de forma desconhecida.

– Isso significa que é possível descrever novas estatísticas que funcionam independentemente do tipo de heteroscedasticidade presente na população.

– Esses métodos são os procedimentos robustos em relação à heteroscedasticidade, já que são válidos mesmo que a variância dos erros não seja constante.

– É possível então estimar variâncias consistentes na presença de heteroscedasticidade.

– A aplicação de métodos robustos em relação à heteroscedasticidade é bastante fácil, pois muitos programas estatísticos e econométricos calculam essas estatísticas como uma opção.

ESTIMANDO VARIÂNCIA COM HETEROSCEDASTICIDADE

– No caso da regressão simples e sem a hipótese de homoscedasticidade, a variância do estimador é:

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{SQT_x^2}$$

– Quando $\sigma_i^2 = \sigma^2$ para todo *i*, a fórmula se reduz a: σ^2/SQT_x.

– Quando $\sigma_i^2 \neq \sigma^2$ (heteroscedasticidade), a variância derivada sob homoscedasticidade não é mais válida.

– Como o erro-padrão é baseado diretamente na estimativa da variância, é preciso estimar a equação acima quando a heteroscedasticidade está presente.

– Sendo u_i os resíduos da regressão simples de y sobre x, um estimador válido da variância para a heteroscedasticidade é:

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\hat{u}_i^2}{SQT_x^2}$$

EM REGRESSÃO MÚLTIPLA

– No caso de: (1) regressão múltipla; (2) r_{ij} ser o *i*-ésimo resíduo da regressão de x_j sobre todas as outras variáveis independentes; e (3) SQR_j ser a soma dos resíduos quadrados da regressão, temos:

$$Var(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SQR_j^2}$$

– A raiz quadrada desta fórmula é o erro-padrão robusto em relação à heteroscedasticidade de beta estimado.

– Os erros-padrão robustos são atribuídos a White (1980).

– A estatística *t* robusta em relação à heteroscedasticidade é calculada após obter os erros-padrão robustos:

$$t = \frac{estimativa - valor\ hipotético}{erro\ padrão}$$

ERROS-PADRÃO USUAIS E ROBUSTOS

– Geralmente, os erros-padrão robustos são frequentemente maiores do que os erros-padrão usuais.

– Os erros-padrão robustos podem ser estimados mesmo sem que se saiba se a heteroscedasticidade está presente.

– Os novos erros-padrão são válidos (assimptoticamente) haja ou não presença de heteroscedasticidade.

– Com frequência, as diferenças entre os erros-padrão usuais e os robustos são pequenas.

– Erros-padrão usuais podem ser usados se a hipótese de homoscedasticidade se mantiver e erros forem normalmente distribuiídos, já que estatísticas t usuais terão distribuições t.

– Em amostras pequenas, as estatísticas t robustas podem ter distribuições que não sejam próximas da distribuição t.

– Em amostras grandes, sempre podemos levar em conta somente os erros-padrão robustos.

ESTATÍSTICAS *F* E *LM*

– É possível obter estatísticas *F* e *LM* robustas em relação à heteroscedasticidade de forma desconhecida.

– A estatística *F* robusta em relação à heteroscedasticidade é chamada de estatística de Wald robusta em relação à heteroscedasticidade.

– O cálculo do teste *F* robusto não tem uma forma simples, mas pode ser computado por alguns programas estatísticos.

MULTIPLICADOR DE LAGRANGE (*LM*) ROBUSTO

– Nem todos programas econométricos calculam estatísticas *F* que sejam robustas em relação à heteroscedasticidade.

– Uma estatística *LM* robusta pode ser obtida manualmente em qualquer programa econométrico:

1. Obtenha os resíduos *u* do modelo restrito.
2. Faça a regressão de cada uma das variáveis independentes excluídas, conforme a hipótese nula, sobre todas as variáveis independentes incluídas, e salve os resíduos (r_1, r_2, ..., r_q).
3. Encontre os produtos de cada r_j por *u* (para todas as observações).
4. Faça a regressão de 1 sobre r_1u, r_2u, ..., r_qu, sem um intercepto.
5. Use a soma dos resíduos quadrados da última regressão para calcular a estatística *LM* robusta (n - SQR), a qual terá distribuição de qui-quadrado.

TESTE DE EXISTÊNCIA DE HETEROSCEDASTICIDADE

– Os erros-padrão robustos em relação à heteroscedasticidade oferecem um método simples para calcular estatísticas *t* que sejam assimptoticamente distribuídas como *t*, haja ou não a presença de heteroscedasticidade.

– Porém, há razões para saber se realmente há presença de heteroscedasticidade, antes de estimar erros-padrão robustos:

  – As estatísticas *t* usuais são preferíveis se não há heteroscedasticidade.
  – É possível obter um estimador melhor que o MQO quando a forma da heteroscedasticidade é conhecida.

TESTE DE EXISTÊNCIA DE HETEROSCEDASTICIDADE

– Considere um modelo linear:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

– A hipótese nula de que a homoscedasticidade se mantém é:

$$H_0: Var(u|x_1,x_2,...,x_k) = \sigma^2$$

– Precisamos analisar os dados para saber se a hipótese nula é adequada ou não.

– Se não rejeitamos H_0, concluímos que a heteroscedasticidade não será um problema.

– Como *u* tem esperança condicional zero, $Var(u|x)=E(u^2|x)$, e a hipótese nula será:

$$H_0: E(u^2|x_1,x_2,...,x_k) = E(u^2) = \sigma^2$$

TESTE *F* DE EXISTÊNCIA DE HETEROSCEDASTICIDADE

– Estimamos então esta equação:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + erro$$

– Utilizando o R^2 da equação acima e o número de regressores (k), estimamos a estatística *F*:

$$F = \frac{R^2_{\hat{u}^2}/k}{\left(1 - R^2_{\hat{u}^2}\right)/(n-k-1)}$$

– A estatística *F* tem uma distribuição $F_{k,n-k-1}$ sob a hipótese nula de homoscedasticidade, permitindo o cálculo de sua significância.

TESTE *LM* DE EXISTÊNCIA DE HETEROSCEDASTICIDADE

– A estatística *LM* para a heteroscedasticidade é o tamanho da amostra multiplicado pelo R^2 da equação com u^2 como variável dependente:

$$LM = n * R^2_{\hat{u}^2}$$

– Essa versão *LM* do teste é geralmente chamada teste de Breusch-Pagan da heteroscedasticidade (teste BP).

RESUMINDO O TESTE BP

– Estime o modelo MQO em que *y* é a variável dependente e obtenha os resíduos quadrados (u^2) para cada observação.

– Estime o modelo em que *u* é a variável dependente para obter o R-quadrado.

– Construa a estatística *F* e calcule o p-valor usando a distribuição $F_{k,n-k-1}$.

– Construa a estatística *LM* e calcule o p-valor usando a distribuição de qui-quadrado.

– Se o p-valor ficar abaixo do nível de significância selecionados, então rejeitamos a hipótese nula de homoscedasticidade.

– Se for constatada que não há homoscedasticidade, os erros-padrão robustos em relação à heteroscedasticidade e suas estatísticas de testes poderão ser utilizadas.

– Sabemos ainda que há menos heteroscedasticidade com a variável dependente em forma logarítmica.

TESTE DE WHITE PARA HETEROSCEDASTICIDADE

– A hipótese de homoscedasticidade [$Var(u|x_1,...,x_k)$] pode ser substituída por outra hipótese:

  – O erro quadrado (u^2) é não-correlacionado com:

    – Todas as variáveis independentes (x_j).

    – Os quadrados das variáveis independentes (x_j^2).

    – Todos os produtos cruzados ($x_j x_h$ para $j \neq h$).

– White sugeriu um testar formas de heteroscedasticidade que invalidem os erros-padrão e as estatísticas de testes.

– Para um modelo com três variáveis independentes, temos:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 +$$
$$\delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 +$$
$$\delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + erro$$

–O teste de White para a heteroscedasticidade é a estatística *LM* para testar se todos δ_j na equação sejam zero, exceto δ_0.

TESTE DE WHITE PARA HETEROSCEDASTICIDADE

– O teste de White usa muitos graus de liberdade para modelos com um número moderado de variáveis independentes.

– É possível obter um teste que seja mais facilmente implementado que o teste de White.

– Uma sugestão é usar os valores estimados MQO para verificar a existência de heteroscedasticidade.

– Os valores estimados são apenas funções lineares das variáveis independentes.

– Se eles forem elevados ao quadrado, estamos na prática obtendo uma função particular de todos os quadrados e produtos cruzados das variáveis independentes:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + erro$$

– Podemos usar as estatísticas *F* ou *LM* para a hipótese nula:

$$H_0: \delta_1 = 0, \delta_2 = 0$$

RESUMINDO O TESTE DE WHITE

– Estime o modelo MQO em que *y* é a variável dependente e obtenha os resíduos (*u*) e os valores estimados de *y*.

– Calcule os resíduos quadrados (u^2) e os quadrados dos valores estimados.

– Estime o modelo em que *u* é a variável dependente e *y* e y^2 sejam as variáveis independentes para obter o R^2.

– Construa a estatística *F* e calcule o p-valor usando a distribuição $F_{2,n-3}$.

– Construa a estatística *LM* e calcule o p-valor usando a distribuição de qui-quadrado.

– Se o p-valor ficar abaixo do nível de significância selecionados, então rejeitamos a hipótese nula de homoscedasticidade.

CONSIDERAÇÃO IMPORTANTE

- Se omitirmos um ou mais termos quadráticos em um modelo de regressão ou usarmos o modelo em nível ao invés de usar o log, um teste de heteroscedasticidade pode vir a ser significante, rejeitando a hipótese de homoscedasticidade.
- Isso tem levado alguns pesquisadores a verem estes testes como testes de má especificação do modelo:
  - Porém, há outros testes que podem testar melhor a má especificação de formas funcionais das variáveis.
- Ou seja, é mais apropriado:
  - Primeiro, realizar testes específicos de formas funcionais, já que a má especificação da forma funcional é mais importante que a heteroscedasticidade.
  - Depois de satisfeitos com as formas funcionais das variáveis, estimar o teste para verificar a existência de heteroscedasticidade.

ESTIMAÇÃO DE MÍNIMOS QUADRADOS PONDERADOS

– Se for detectada heteroscedasticidade com o uso de testes estatísticos, é possível estimar erros padrão robustos em relação à heteroscedasticidade após a estimação MQO.

– Porém, antes das estatísticas robustas, é possível modelar e estimar a forma específica da heteroscedasticidade, calculando um estimador mais eficiente que o MQO, além de estatísticas *t* e *F* não enviesadas.

– Isso requer mais trabalho, pois é preciso ser específico sobre a natureza de qualquer heteroscedasticidade.

CONSTANTE MULTIPLICATIVA

– Considere que ***x*** representa todas as variáveis explicativas em:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

– Assuma que *h(**x**)* é alguma função das variáveis explicativas que determina a heteroscedasticidade:

$$Var(u|\boldsymbol{x}) = \sigma^2 h(\boldsymbol{x})$$

– Como variâncias devem ser positivas, *h(**x**)*>0 para todos valores possíveis das variáveis independentes.

– Supomos que a função *h(**x**)* é conhecida. Assim, mesmo que o parâmetro populacional σ^2 seja desconhecido, teremos condições de estimá-lo a partir de uma amostra de dados.

EQUAÇÃO TRANSFORMADA

– Com o objetivo de obter estimadores de β_j que tenham propriedades de eficiência melhores que MQO, estimamos esta equação:

$$y_i/\sqrt{h_i} = \beta_0/\sqrt{h_i} + \beta_1\left(x_{i1}/\sqrt{h_i}\right) + \cdots + \beta_k\left(x_{ik}/\sqrt{h_i}\right) + u_i/\sqrt{h_i}$$

– Esta equação é linear em seus parâmetros (RLM.1), a hipótese de amostragem aleatória não se alterou (RLM.2), o termo de erro tem média condicional zero (RLM.3) e não há colinearidade perfeita entre variáveis independentes (RLM.4).

– A equação transformada satisfará as hipóteses do modelo linear clássico, se o modelo original também o fizer, com exceção da hipótese de homoscedasticidade (RLM.5).

MÍNIMOS QUADRADOS GENERALIZADOS (MQG)

– É necessário estimar os parâmetros da nova equação por mínimos quadrados ordinários.

– Os novos betas são estimadores de mínimos quadrados generalizados (MQG).

– Estes estimadores MQG são usados para explicar a heteroscedasticidade nos erros.

– Os erros-padrão, estatísticas *t* e estatísticas *F* podem ser obtidas de regressões que usem as variáveis transformadas.

– Por serem os melhores estimadores lineares não-viesados de beta, os estimadores MQG são mais eficientes que os estimadores MQO.

– A interpretação dos resultados deve ser feita com base na equação original.

– O R^2 indica o quanto da variação do novo y é explicado pelo novo x, o que não é informativo como grau de ajuste.

MÍNIMOS QUADRADOS PONDERADOS (MQP)

– Os estimadores de mínimos quadrados generalizados (MQG) para correção da heteroscedasticidade são chamados de estimadores de mínimos quadrados ponderados (MQP).

– Os novos betas minimizam a soma ponderada dos quadrados dos resíduos.

– A idéia é colocar menos peso nas observações com uma variância de erro mais alta.

– O método MQO atribui pesos iguais a todas as observações, pois isso é melhor quando a variância do erro é idêntica para todas as partições da população.

MÍNIMOS QUADRADOS PONDERADOS (MQP)

– A maioria dos programas econométricos tem um recurso para computar mínimos quadrados ponderados.

– Juntamente com as variáveis dependentes e independentes originais, especificamos a função de ponderação ($1/h_i$).

– Especificamos pesos proporcionais ao inverso da variância.

– Isso nos permite interpretar as estimativas de mínimos quadrados ponderados no modelo original.

– Podemos escrever a equação estimada da maneira habitual.

– As estimativas e os erros-padrão serão diferentes do MQO, mas a maneira como interpretamos essas estimativas, erros-padrão e estatísticas de testes é a mesma.

– Esse procedimento corrige estimativas dos betas (aweight).

– Se considerarmos que a heteroscedasticidade seria um problema para os erros-padrão, deveríamos computar também os erros-padrão robustos (pweight).

MAS NA PRÁTICA...

– Na prática, raramente sabemos como a variância do erro se comporta em relação a uma variável independente.

– Em equações de regressão múltipla, é complicado saber com qual variável independente há heteroscedasticidade nos erros e qual a forma deste problema.

– Existe um caso no qual os pesos necessários para o MQP surgem naturalmente de um modelo econométrico subjacente.

– Isso acontece quando os dados estão em médias de algum grupo ou região, e não em nível individual.

DADOS EM MÉDIAS POR GRUPOS

– Se a equação no nível individual satisfizer a hipótese de homoscedasticidade, então a equação do nível agrupado deverá ter heteroscedasticidade.

– Assim, se para todo grupo *i* e indivíduo *j*:

$$Var(u_{i,j}) = \sigma^2$$

– Então, a variância do termo de erro médio diminui com o tamanho do grupo:

$$Var(\overline{u_i}) = \sigma^2/m_i$$

– Neste caso, $h_i = 1/m_i$.

– Portanto, o procedimento mais eficiente será o dos mínimos quadrados ponderados, com pesos correspondentes ao número de indivíduos nos grupos ($1/h_i = m_i$).

– Isso garante que grupos maiores recebam peso maior, o que oferece método eficiente de estimação dos parâmetros no modelo em nível individual quando temos médias.

HETEROSCEDASTICIDADE NO NÍVEL INDIVIDUAL

– Se no caso anterior existisse heteroscedasticidade no nível individual, então a ponderação adequada dependerá da forma da heteroscedasticidade.

– Por isso, vários pesquisadores simplesmente computam erros-padrão e estatísticas de teste robustos na estimação de modelos que usam dados agrupados.

– Uma alternativa é realizar a ponderação pelo tamanho do grupo (aweight), além de estimar as estatísticas robustas em relação à heteroscedasticidade na estimação MQP (pweight).

– Isso assegura que qualquer heteroscedasticidade no nível individual seja representada pela inferência robusta.

MQG FACTÍVEL

– Ao contrário dos exemplos anteriores, a forma exata de heteroscedasticidade não é óbvia na maioria dos casos.

– Em muitos casos podemos modelar a função *h* e utilizar os dados para estimar os parâmetros desconhecidos.

– o uso de h_i-chapéu em lugar de h_i na transformação MQG produz o estimador de mínimos quadrados generalizados factível (MQGF), também chamado de MQG estimado (MQGE).

– Existem várias maneiras de modelar a heteroscedasticidade, mas iremos utilizar um método razoavelmente flexível:

$$Var(u|\boldsymbol{x}) = \sigma^2 exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$$

– É utilizada função exponencial porque modelos lineares não asseguram que os valores previstos sejam positivos, e as variâncias estimadas devem ser positivas para usar o MQP.

ESTIMAÇÃO DO MQG FACTÍVEL

– Para estimar os parâmetros δ_i é preciso transformar a equação anterior em uma forma linear para ser estimada por MQO:

$$log(u^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + e$$

– Na prática (pág. 263):

1. Execute a regressão de y sobre x_1, x_2, ..., x_k e obtenha os resíduos de $\hat{u}$.
2. Crie $log(\hat{u}^2)$ elevando ao quadrado os resíduos MQO e depois calculando seu log natural.
3. Execute a regressão na equação acima dos parâmetros δ_i [ou log(u^2) sobre y, y^2] e obtenha os valores estimados.
4. Calcule o exponencial dos valores estimados, resultando em: $\hat{h}$.
5. Estime a equação $y = \beta_0 + \beta_1 x_1 + \ldots \beta_k x_k + u$, pelo método MQP, usando pesos (aweight) $1/\hat{h}$.

ESTATÍSTICAS *F*

– Ao calcular estatísticas *F*, é importante que os mesmos pesos sejam usados para estimar os modelos com e sem restrições.

– Devemos estimar o modelo sem restrições por MQO com os pesos.

– Usamos os mesmos pesos para estimar o modelo restrito.

– Posteriormente, a estatística *F* pode ser calculada.

– Lembrem-se que o Stata permite utilizar o comando “test” para testar restrições conjuntas após a estimação de um modelo, não sendo necessário calcular manualmente a regressão restrita.

MODELO DE PROBABILIDADE LINEAR REVISITADO

– Quando a variável dependente é binária, o modelo deve conter heteroscedasticidade, a menos que todos parâmetros de inclinação sejam nulos.

– A maneira mais simples de tratar a heteroscedasticidade neste caso é usar a estimação MQO, e calcular os erros-padrão robustos nas estatísticas de testes.

– As estimativas MQO do MPL são simples e geralmente produzem resultados satisfatórios, mas são ineficientes.

– É possível utilizar o MQP para estimar o MPL. No entanto, o método falhará se $\hat{h}$ for negativo (ou zero) em qualquer observação.

ESTIMAÇÃO DO MPL POR MQP

– Estime o modelo por MQO e obtenha os valores estimados de *y*.

– Verifique se todos os valores estimados estão dentro do intervalo unitário:

  – Se assim for, prossiga para o passo seguinte.

  – Caso contrário, alguns ajustes serão necessários para trazer todos os valores estimados para dentro do intervalo unitário:

    – $y_i = 0{,}01$ se $y_i < 0$

    – $y_i = 0{,}99$ se $y_i > 1$

– Construa as variâncias estimadas com esta equação:

$$\hat{h}_i = \hat{y}_i(1 - \hat{y}_i)$$

– Estime a equação $y = \beta_0 + \beta_1 x_1 + \dots \beta_k x_k + u$, pelo método MQP, usando pesos (aweight) $1/\hat{h}$.