
Modelo e Estimadores

2.1 Introdução

Modelos lineares heteroscedásticos apresentam uma grande aplicabilidade em diversos problemas práticos. A não constância das variâncias dos erros impede que os modelos lineares heteroscedásticos herdem certas propriedades matemáticas dos modelos lineares homoscedásticos. A estrutura usual de covariâncias dos estimadores dos parâmetros que indexam o modelo de regressão linear homoscedástico não é mais válida em casos de heteroscedasticidade. [White \(1980\)](#) propôs um estimador da matriz de covariâncias do estimador de mínimos quadrados ordinários do vetor de parâmetros de regressão que é consistente tanto sob homoscedasticidade quanto sob heteroscedasticidade de forma desconhecida. Notou-se na literatura que o estimador de White é viesado em amostras pequenas e moderadas, sobretudo quando os dados contêm pontos de alavanca. Variantes do estimador de White foram propostos por [MacKinnon e White \(1985\)](#), [Davidson e MacKinnon \(1993\)](#), [Cribari-Neto \(2004\)](#) e [Cribari-Neto, Souza e Vasconcellos \(2007, Errata: v. 37, n. 20, p. 3329–3330, 2008\)](#). Este capítulo apresentará tais estimadores.

2.2 Modelagem de Regressão

Seja $L(\cdot)$ uma função qualquer, tal que, $\forall 0 < u < v$, temos:

- i) $0 = L(0) \leq L(u) \leq L(v)$;
- ii) $0 = L(0) \leq L(-u) \leq L(-v)$.

Qualquer função $L(\cdot)$ que satisfaz as propriedades acima é chamada de função perda. Alguns exemplos de função perda são:

- Função perda quadrática, em que $L(u) = u^2$;
- Função perda absoluta, em que $L(u) = |u|$;
- Perda degrau, em que para algum $\delta > 0$, $L(u) = 0$ se $|u| < \delta$ e 1 caso contrário.

No caso mais simples, queremos prever o comportamento de uma variável de interesse y condicional a uma variável explicativa x . O melhor preditor de y condicional em x é aquele que minimiza a função perda esperada, ou seja, é aquele que resolve $\min \mathbb{E}(L(y - f)|x)$, em que f é um preditor e $\mathbb{E}(L(y - f)|x)$ é a função perda esperada. A função f é função de x e de quantidades fixas e desconhecidas. Em alguns casos, tais quantidades são estimáveis a partir dos dados. Elas são chamadas de parâmetros. Aqui serão agrupadas no vetor β .

Para o caso da função perda quadrática, o melhor preditor de y condicional a x é a média condicional de y dado x . No caso em que L é a função perda absoluta temos que a mediana condicional de y dado x é o melhor preditor para y . Os modelos de regressão, em geral, fazem uso da função perda quadrática. Regressão é qualquer aspecto da distribuição condicional de y em x tomado como função de x .

Modelos lineares de regressão foram uma das primeiras formas de análise regressiva estudadas com rigor matemático. Esses modelos são amplamente utilizados porque são mais fáceis de serem ajustados que os modelos não-lineares. Um outro motivo de sua vasta utilização diz respeito à facilidade de se obter as propriedades estatísticas dos estimadores resultantes.

Em termos mais gerais, modelagem de regressão refere-se à modelagem estocástica que relaciona matematicamente uma variável de interesse y_i a um conjunto de variáveis explicativas, $x_{1i}, x_{2i}, \dots, x_{pi}$, em que, $i = 1, 2, \dots, n$, sendo n a quantidade de dados. A variável y_i também pode ser referenciada como variável resposta, regressando ou variável dependente, enquanto as variáveis $x_{1i}, x_{2i}, \dots, x_{pi}$ podem ser chamadas de variáveis independentes, regressores ou covariáveis. Em forma matricial, y é um vetor $n \times 1$ e X é uma matriz de variáveis explicativas de dimensão $n \times p$, em que p é o número de parâmetros lineares de regressão. Dessa forma, temos que o comportamento de y é afetado por variações na matriz de regressores X e por quantidades desconhecidas presentes em qualquer modelo estocástico denominadas de erros aleatórios, que serão agrupados no vetor ε de dimensão $n \times 1$.

É importante frisar que quando mencionamos modelos lineares de regressão não estamos falando em uma relação linear entre as variáveis envolvidas, mas sim em linearidade nos parâmetros do modelo. Ou seja, a esperança condicional de y , $\mathbb{E}(y|X)$, é uma função linear dos parâmetros do vetor de parâmetros β de dimensão $p \times 1$.

2.3 Modelos Lineares Heteroscedásticos de Regressão

Os parâmetros dos modelos lineares de regressão são tipicamente estimados pelo Método dos Mínimos Quadrados ou Mínimos Quadrados Ordinários (MQO). MQO é o método de estimação mais amplamente utilizado em econometria. Essa metodologia consiste em minimizar a soma das perdas quadráticas e não requer suposição distribucional sobre eles. Observe que em nenhum momento será necessário conhecer a distribuição de y . Consideremos o modelo linear de regressão, dado por

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

em que y é um vetor $n \times 1$ de observações de interesse (regressando), X é uma matriz fixa e conhecida de variáveis independentes de dimensão $n \times p$ com posto coluna completo, ou seja, $\text{posto}(X) = p < n$, $\beta = (\beta_1, \dots, \beta_p)'$ é um vetor de dimensão $p \times 1$ de parâmetros lineares e $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ é um vetor $n \times 1$ de erros aleatórios. A solução de mínimos quadrados é alcançada minimizando $\varepsilon^2 = \varepsilon'\varepsilon = \sum_{i=1}^n \varepsilon_i^2$. Ou seja, $\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \varepsilon_i^2$. Substituindo ε por $y - X\beta$ temos que

$$S(\beta) = (y - X\beta)'(y - X\beta) = y'y - y'X\beta - \beta'X'y + \beta'X'X\beta.$$

Derivando com respeito a β e igualando esta derivada a zero, segue que

$$\frac{\partial S}{\partial \beta} = -2X'y + 2X'X\beta = 0 \Rightarrow X'X\hat{\beta} = X'y.$$

Ou seja,

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (2.1)$$

Em se tratando de modelos lineares de regressão, as suposições que tipicamente são feitas são:

S1: O modelo $y = X\beta + \varepsilon$ é, de fato, o modelo verdadeiro;

S2: $\mathbb{E}(\varepsilon_i) = 0$, $i = 1, \dots, n$;

S3: $\mathbb{E}(\varepsilon_i^2) = \text{var}(\varepsilon_i) = \sigma_i^2$, $i = 1, \dots, n$;

S3': $\text{var}(\varepsilon_i) = \sigma^2$, $i = 1, \dots, n$ ($0 < \sigma^2 < \infty$);

S4: $\mathbb{E}(\varepsilon_i \varepsilon_s) = 0$, $\forall i \neq s$;

S5: $\lim_{n \rightarrow \infty} n^{-1}(X'X) = Q$, em que Q é uma matriz positiva-definida.

Sob [S1] e [S2], temos que o estimador $\hat{\beta} = (X'X)^{-1}X'y$ é não viesado para β , ou seja, em média iguala-se ao parâmetro verdadeiro. Sob essas suposições, temos que $\mathbb{E}(y) = X\beta$. Logo, $\mathbb{E}(\hat{\beta}) = (X'X)^{-1}X'\mathbb{E}(y) = (X'X)^{-1}X'X\beta = I_p\beta = \beta$, em que I_p é uma matriz identidade de dimensão $p \times p$. Tal estimador coincide com o estimador de máxima verossimilhança sob normalidade dos erros.

Quando as suposições [S1], [S2], [S3] e [S4] são válidas, a matriz de covariâncias de ε é dada por

$$\Omega = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}.$$

Essa matriz se reduz a $\Omega = \sigma^2 I_n$ quando $\sigma_i^2 = \sigma^2 > 0$ com $i = 1, \dots, n$ (vale [S3']), sendo I_n matriz identidade de dimensão $n \times n$. Esse resultado pode ser visto abaixo. Note que

$$\mathbb{E}(\varepsilon\varepsilon') = \begin{bmatrix} \mathbb{E}(\varepsilon_1^2) & \mathbb{E}(\varepsilon_1\varepsilon_2) & \cdots & \mathbb{E}(\varepsilon_1\varepsilon_n) \\ \mathbb{E}(\varepsilon_2\varepsilon_1) & \mathbb{E}(\varepsilon_2^2) & \cdots & \mathbb{E}(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}(\varepsilon_n\varepsilon_1) & \mathbb{E}(\varepsilon_n\varepsilon_2) & \cdots & \mathbb{E}(\varepsilon_n^2) \end{bmatrix}.$$

Usando [S1], [S2], [S3'] e [S4], obtemos

$$\Omega = \mathbb{E}(\varepsilon\varepsilon') = \mathbb{E}(\varepsilon^2) = \sigma^2 I_n.$$

O vetor de parâmetros β pode ser estimado usando o método de mínimos quadrados ordinários; equação (2.1). Sem que se faça suposição sobre a distribuição das respostas, minimiza-se a soma dos quadrados dos erros. É importante destacar também que mesmo quando a suposição [S3] é satisfeita o estimador de mínimos quadrados ordinários continua não-viesado. O estimador $\hat{\beta}$ possui as seguintes propriedades:

- (1) Quando as suposições [S1] e [S2] são válidas, $\hat{\beta}$ é um estimador não-viesado para β . Ou seja, em média o estimador iguala-se ao parâmetro em estimação;
- (2) A estrutura de covariância de $\hat{\beta}$ é $\text{cov}(\hat{\beta}) = \Psi_{\hat{\beta}} = (X'X)^{-1}X'\Omega X(X'X)^{-1}$;
- (3) Quando as suposições [S1], [S2] e [S5] valem, o estimador $\hat{\beta}$ é consistente para β , ou seja, converge em probabilidade para o vetor de parâmetros;
- (4) Sob as suposições [S1], [S2], [S3'] e [S4], vale o Teorema de Gauss-Markov, ou seja, $\hat{\beta}$ possui variância mínima na classe dos estimadores lineares e não-viesados de β ;
- (5) Sob [S1], [S2] e [S3], o estimador $\hat{\beta}$ é assintoticamente gaussiano.

É importante destacar que quando a suposição de homoscedasticidade é violada (i.e., os erros seguem padrão heteroscedástico) $\hat{\beta}$ deixa de ser eficiente, ou seja, não vale

mais o Teorema de Gauss Markov. Contudo, o estimador continua não-viesado, consistente e assintoticamente gaussiano. Dessa forma, $\hat{\beta}$ mantém muitas das propriedades desejáveis de um bom estimador.

Sob homoscedasticidade, ou seja, quando as variâncias dos erros são idênticas (suposição [S3']), temos que a estrutura de covariância de $\hat{\beta}$ é dada por $\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$. Essa matriz pode ser facilmente estimada substituindo σ^2 pelo estimador $\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n-p)$, em que $\hat{\varepsilon}$ é o vetor de dimensão $n \times 1$ contendo os resíduos de mínimos quadrados. Esses resíduos são dados por $\hat{\varepsilon} = \{I_n - X(X'X)^{-1}X'\}y = (I_n - H)y$. Os erros-padrão fornecidos pela maioria dos softwares estatísticos são as raízes quadradas dos elementos diagonais de $\hat{\sigma}^2(X'X)^{-1}$.

A matriz $H = X(X'X)^{-1}X'$ é conhecida como “matriz chapéu”, pois $\hat{y} = Hy$. Os elementos da diagonal principal da matriz H assumem valores no intervalo $(0, 1)$ e somam p , i.e., $\text{tr}(H) = p$. Dessa forma, a média dos elementos diagonais de H é igual a p/n .

Os elementos diagonais da matriz H são denotados por h_1, \dots, h_n e são utilizados como medidas dos graus de alavancagem, i.e., o i -ésimo elemento diagonal de H (h_i) mede o grau de influência da i -ésima observação sobre o correspondente valor predito (\hat{y}_i). Uma regra muito usada é concluir que a i -ésima observação é ponto de alavanca se $h_i > 2p/n$ ou $h_i > 3p/n$. Maiores detalhes poderão se encontrados em [Judge, Hill e Griffiths \(1988, p. 893\)](#).

Em modelos de regressão heteroscedásticos quando se conhece a matriz Ω de covariâncias dos erros (o que quase nunca ocorre na prática), é possível transformar o modelo $y = \beta X + \varepsilon$ de forma que as suposições necessárias para a validade do Teorema de Gauss-Markov sejam satisfeitas. No contexto em que se conhece a matriz de covariâncias dos erros, o estimador de mínimos quadrados generalizados (EMQG) é dado por $\hat{\beta}_G = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$. Pode-se mostrar que $\mathbb{E}(\hat{\beta}_G) = \beta$ e

$$\text{cov}(\hat{\beta}_G) = \Psi_{\hat{\beta}_G} = (X'\Omega^{-1}X)^{-1}.$$

Quando a suposição [S3'] é válida, ou seja, sob homoscedasticidade, tem-se que $\hat{\beta}_G = \hat{\beta}$ e $\text{cov}(\hat{\beta}_G) = \text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$.

Estimação do vetor β pelo método de mínimos quadrados generalizados é inviável, uma vez que os elementos de Ω (i.e., as variâncias dos n erros) são desconhecidos. Uma alternativa é postular um modelo para as variâncias dos erros e usar esse modelo para obter uma estimativa de Ω . Essa estimativa seria então utilizada no lugar da matriz verdadeira no estimador de mínimos quadrados generalizados. O estimador resultante é conhecido como estimador de mínimos quadrados generalizados viável (EMQGV) e é dado por

$$\hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y.$$

O Teorema de Gauss Markov, contudo, não vale para o EMQGV. De fato, esse estimador não é linear e não é possível mostrar, em geral, que ele é não-viesado. Adicionalmente, para se usar esse estimador é necessário postular um modelo para as n variâncias. Tipicamente, contudo, é mais difícil modelar variâncias do que efeitos médios.

Deve ser notado que mesmo não valendo o Teorema de Gauss Markov, ou seja, quando a suposição [S3'] não é válida (i.e., os erros seguem um padrão heteroscedástico), $\hat{\beta}$ permanece não-viesado, consistente e assintoticamente gaussiano. Dessa forma, podemos basear as estimativas pontuais nesse estimador. Para a obtenção de intervalos de confiança e realização de testes de hipóteses é necessário um estimador consistente da matriz de covariância de $\hat{\beta}$. Para tanto, deve-se usar um estimador $\hat{\Omega}$ de Ω tal que $X'\hat{\Omega}X$ seja consistente para $X'\Omega X$, ou seja, $\text{plim}((X'\hat{\Omega}X)^{-1}(X'\hat{\Omega}X)) = I_p$, onde plim denota limite em probabilidade.

White (1980) propôs um estimador da matriz de covariâncias $\Psi_{\hat{\beta}}$ que é consistente tanto sob homoscedasticidade quanto sob heteroscedasticidade de forma desconhecida. Ele percebeu que não seria preciso estimar a matriz de covariâncias dos erros de forma consistente; note que essa matriz possui n quantidades desconhecidas. Ele notou que basta estimar consistentemente $(X'\Omega X)$, que possui $p(p+1)/2$ elementos desconhecidos independentemente do tamanho da amostra. O estimador de Halbert White é obtido substituindo-se o i -ésimo elemento da diagonal da matriz de covariâncias dos erros (σ_i^2) pelo i -ésimo resíduo MQO ao quadrado, ou seja,

$$\text{HC0} = (X'X)^{-1}X'\hat{\Omega}_0X(X'X)^{-1},$$

em que $\hat{\Omega}_0 = \text{diag}\{\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2\}$.

O estimador de White pode ser muito viesado em amostras de tamanho pequeno a moderado. O viés desse estimador tende a ser negativo, ou seja, ele é um estimador “otimista”, pois tipicamente subestima as variâncias verdadeiras; ver Cribari-Neto e Zarkos (1999), Cribari-Neto e Zarkos (2001) e MacKinnon e White (1985). O viés do estimador é mais acentuado em situações em que os dados incluem pontos de alavanca; ver Chesher e Jewitt (1987). Testes sobre os parâmetros que fazem uso desse estimador são tipicamente liberais, i.e., anticonservativos. Dessa forma, a hipótese nula tende a ser rejeitada acima do esperado.

Em busca de corrigir o viés do estimador HC0, uma sequência de estimadores HC0 corrigidos por viés foi obtida por Cribari-Neto, Ferrari e Cordeiro (2000). Esses resultados foram posteriormente estendidos por Cribari-Neto e Galvão (2003).

A partir de resultados obtidos por Horn, Horn e Duncan (1975), MacKinnon e White (1985) construíram um novo estimador denominado HC2, que usa

$$\hat{\Omega}_2 = \text{diag}\{\hat{\varepsilon}_1^2/(1-h_1), \dots, \hat{\varepsilon}_n^2/(1-h_n)\},$$

em que h_i denota o i -ésimo elemento diagonal da “matriz chapéu” $H = X(X'X)^{-1}X'$, $i = 1, \dots, n$. O estimador HC2 é não-viesado sob homoscedasticidade.

O estimador HC3 foi proposto por [Davidson e MacKinnon \(1993\)](#). Ele é uma aproximação ao estimador jackknife, que é obtido a partir da remoção sequencial de cada observação da amostra. O estimador jackknife é apresentado em [Davidson e MacKinnon \(1993, p. 308 a 309\)](#). O estimador da matriz de covariâncias dos erros utilizado no estimador HC3 é

$$\hat{\Omega}_3 = \text{diag}\{\hat{\varepsilon}_1^2/(1 - h_1)^2, \dots, \hat{\varepsilon}_n^2/(1 - h_n)^2\}.$$

Um estimador alternativo que também faz uso das medidas de alavancagem e inclui termos de ajustes para amostras finitas foi proposto por [Cribari-Neto \(2004\)](#). Tal estimador, denominado HC4, utiliza

$$\hat{\Omega}_4 = \text{diag}\{\hat{\varepsilon}_1^2/(1 - h_1)^{\delta_1}, \dots, \hat{\varepsilon}_n^2/(1 - h_n)^{\delta_n}\},$$

em que $\delta_i = \min\{4, h_i/\bar{h}\} = \min\{4, nh_i/p\}$. Note que $\bar{h} = n^{-1} \sum_{i=1}^n h_i = \text{tr}(H)/n = p/n$. O expoente $\delta_i > 0$ controla o nível de desconto para a i -ésima observação. Como $\delta_i > 0$ e $0 < (1 - h_i) < 1$, tem-se então que $(1 - h_i)^{\delta_i}$ também pertence a $(0,1)$. Dessa forma, o quadrado do i -ésimo resíduo será tanto mais inflacionado quanto maior for h_i relativamente ao seu valor médio. O desconto linear δ_i é truncado em 4, que corresponde ao dobro do desconto utilizado no estimador HC3, de modo que $\delta_i = 4$ quando $h_i > 4\bar{h} = 4p/n$.

Um outro estimador foi proposto por [Cribari-Neto, Souza e Vasconcelos \(2007, Errata: v. 37, n. 20, p. 3329–3330, 2008\)](#). Trata-se do estimador HC5, que usa

$$\hat{\Omega}_5 = \text{diag}\{\hat{\varepsilon}_1^2/\sqrt{(1 - h_1)^{\alpha_1}}, \dots, \hat{\varepsilon}_n^2/\sqrt{(1 - h_n)^{\alpha_n}}\},$$

em que

$$\alpha_i = \min \left\{ \frac{h_i}{\bar{h}}, \max \left\{ \frac{h_i}{\bar{h}}, \frac{kh_{\max}}{\bar{h}} \right\} \right\}.$$

Aqui, k é uma constante cujo valor foi escolhido numericamente a partir de simulações-piloto. Os autores sugerem utilizar $k = 0.7$. Esse estimador leva em conta a alavancagem maximal em todos os n termos de descontos utilizados.

Utilizando um dos estimadores consistentes para a matriz $X'\Omega X$, de dimensão $p \times p$, é possível realizar testes sobre os elementos do vetor β sem especificar a forma da heteroscedasticidade existente, ou seja, não é preciso modelar o comportamento do segundo momento da variável de interesse. O comportamento de alguns desses testes em amostras finitas foram analisados por [Cribari-Neto, Ferrari e Oliveira \(2005\)](#) através do uso de integração numérica. Há ainda resultados de avaliações utilizando Monte Carlo

sobre o comportamento em amostras finitas de testes cujas estatísticas utilizam os estimadores descritos acima. Resultados sobre estimação intervalar robusta à presença de heteroscedasticidade podem ser encontrados em [Cribari-Neto e Lima \(2009\)](#).