VALOREM
Revista da Sociedade Brasileira de Engenharia de Avaliações

MÉTODO DOS MÍNIMOS QUADRADOS PONDERADOS: VANTAGENS E APLICAÇÃO NA ENGENHARIA DE AVALIAÇÕES

Weighted least squares method: advantages and application in valuation engineering

Luiz Fernando Palin Droubi
http://orcid.org/0000-0002-5971-7220
Droubi Engenharia e Avaliações LTDA, Florianópolis, Brasil.
droubiaval@gmail.com

Lutemberg de Araújo Florencio
http://orcid.org/0009-0006-3478-8272
Banco do Nordeste do Brasil (BNB), Recife, Brasil.
lutembergflorencio@yahoo.com.br

RESUMO

O método dos mínimos quadrados ponderados (MQP) é uma extensão do método dos mínimos quadrados ordinários (MQO), empregado para ajuste de modelos de regressão linear. De outra forma, pode-se afirmar que o estimador MQO constitui um caso particular do estimador MQP, em que os pesos das observações são todos idênticos. O estimador MQP é especialmente eficaz quando a variância dos erros não é constante ao longo da escala do modelo, sendo considerado um estimador com propriedades ideais ou ótimas, haja vista as suas características de não tendenciosidade e variância mínima, mesmo na presença de erros não-constantes, desde que adotada uma ponderação adequada das observações. A suposição de homoscedasticidade, que postula que os erros têm variância constante ao longo do modelo, muitas vezes não é válida, evidenciando a relevância do estimador MQP. Neste contexto, o presente artigo objetiva demonstrar, a partir de uma aplicação com 190 dados de apartamentos situados na região central de Florianópolis, disponibilizados em Zilli (2020), a superioridade do ajuste de modelos de regressão linear via estimador MQP, em detrimento do estimador MQO. Este artigo argumenta que a aplicação do estimador MQP é preferível a outras abordagens para contornar a heterocedasticidade na Engenharia de Avaliações, como a transformação da variável dependente (com os desafios de retransformação) ou a utilização de erros heteroscedástico-consistentes.

Palavras-Chave: Regressão linear ponderada; Heteroscedasticidade; Método dos mínimos quadrados ordinários; Transformação de variáveis; Critério de Informação de Akaike (AIC).

ABSTRACT

The weighted least squares method (MQP) is an extension of the ordinary least squares method (OLS), used to adjust linear regression models. In another way, it can be stated that the OLS estimator constitutes a particular case of the MQP estimator, in which the weights of the observations are all identical. The MQP estimator is especially effective when the error variance is not constant throughout the model scale, being considered an estimator with ideal or optimal properties, given its characteristics of non-bias and minimum variance, even in the presence of non-biased errors. constant, as long as an appropriate weighting of observations is adopted. The assumption of homoscedasticity, which postulates that errors have constant variance throughout the model, is often not valid, highlighting the relevance of the MQP estimator. In this context, this article aims to demonstrate, based on an application with 190 data from apartments located in the central region of Florianópolis, available in Zilli (2020), the superiority of adjusting linear regression models via the MQP estimator, to the detriment of the MQO. This article argues that the application of the MQP estimator is preferable to other approaches to overcome heteroscedasticity in Valuation Engineering, such as transforming the dependent variable (with the challenges of retransformation) or using heteroscedastic-consistent errors.

Keywords: Weighted linear regression; Heteroscedasticity; Ordinary least squares method; Transformation of variables; Akaike Information Criterion (AIC).

Preenchimento dos Editores

INFORMAÇÕES SOBRE O ARTIGO
Submetido em 08/04/2024
Publicado em 12/05/2024

Comitê Científico Interinstitucional
Editor-Responsável: Carlos Augusto Zilli
Avaliado pelo Sistema Double Blind Review
(SEER/OJS – Versão 3)

1. INTRODUÇÃO

A avaliação de imóveis a partir do método comparativo direto de dados de mercado, com tratamento científico, é um processo complexo que envolve a análise de regressão em busca de modelos que expliquem de maneira satisfatória a variabilidade observada nos preços dos imóveis, com base na variação dos regressores, no mercado que se estuda.

Tradicionalmente, modelos de regressão linear com o uso do estimador de mínimos quadrados ordinários (MQO) têm sido amplamente utilizados para estimar os valores de imóveis com base em variáveis explicativas, conforme pode ser observado em Ozgur *et al.* (2016).

Todavia, conforme alerta Dantas (2003), em que pese a intensa utilização do modelo de regressão linear na Engenharia de Avaliações, há grande probabilidade de os resultados dos estudos realizados serem tendenciosos, ineficientes ou inconsistentes, por negligenciarem ou conflitarem com os pressupostos básicos do modelo de regressão.

Neste contexto, a suposição da premissa da homoscedasticidade inerente ao estimador MQO em modelos de regressão linear pode não ser adequada para dados imobiliários, onde a variabilidade dos preços tende a não ser constante. Para lidar com essa heterogeneidade, o uso do método dos mínimos quadrados ponderados (MQP) se destaca como uma abordagem mais apropriada. O MQP ajusta o modelo de regressão atribuindo pesos diferentes às observações, de acordo com a variabilidade de cada uma. Essa técnica permite que as estimativas dos coeficientes de regressão sejam mais robustas e menos suscetíveis a vieses causados por variâncias não constantes.

Na prática, o uso de modelos de regressão linear ponderado em avaliações de imóveis oferece diversas vantagens. Por exemplo, imóveis localizados em áreas com maior variabilidade de preços podem ser ajustados com pesos menores, enquanto imóveis em regiões mais estáveis recebem pesos maiores, refletindo assim a variação real dos dados. Essa abordagem resulta em um modelo que se adapta melhor às características específicas do mercado imobiliário, proporcionando estimativas mais precisas e confiáveis.

Neste sentido, o objetivo deste artigo é explorar em profundidade o uso de modelos de regressão linear ponderado na avaliação de imóveis. Discutiremos as bases teóricas do MQP, a metodologia para determinar os pesos das observações, e apresentaremos um estudo de caso que ilustra a aplicação prática dessa técnica. Ao final, esperamos demonstrar como o MQP pode melhorar significativamente a precisão das avaliações imobiliárias, contribuindo para tomadas de decisão fundamentadas.

2. REFERENCIAL TEÓRICO

O modelo de regressão linear clássico (MRLC), expresso na equação (1), é a pedra angular de boa parte da teoria econométrica e se baseia em um conjunto de premissas, a saber: (i) linearidade nos parâmetros, (ii) ausência de autocorrelação entre os termos do erro (ϵ_i), (iii) homoscedasticidade ou variância (σ^2) constante de ϵ_i, (iv) ausência de multicolinearidade perfeita e (v) correta especificação do modelo.

$$\boldsymbol{y_i} = \sum \boldsymbol{\beta_j x_{ij}} + \boldsymbol{\epsilon_i} \tag{1}$$

Onde y_i é a variável dependente, ou regressando, que se pretende explicar, a partir de uma ou mais variáveis explicativas, ou regressores, x_j. Os coeficientes β_j da equação são desconhecidos e, portanto, devem ser estimados, o que pode ser feito a partir de diversos métodos, sendo o mais comum e amplamente difundido, o método dos mínimos quadrados ordinários (MQO). O termo ϵ_i é conhecido como distúrbio ou, simplesmente, termo de erro, sendo uma variável aleatória (estocástica) que tem propriedades probabilísticas conhecidas.

Uma vez atendidas as premissas do MRLC, o estimador MQO (Equação 2) assume determinadas propriedades resumidas no conhecido Teorema de Gauss Markov, que diz que, na classe dos estimadores lineares não tendenciosos, os de mínimo quadrados ordinários têm a variância mínima. Em resumo, o estimador MQO é o melhor estimador linear não tendencioso.

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{MQO}} = argmin_{\boldsymbol{\beta}} \ \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y_i} - \boldsymbol{X_i\beta})^2 \tag{2}$$

A solução do problema de minimização acima é determinada (única) e pode ser obtida matricialmente a partir da equação (3) a seguir:

$$\widehat{\boldsymbol{\beta}}_{MQO} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y} \tag{3}$$

No caso da verificação da homoscedasticidade ($Var(y|X = t) = \sigma^2,\ \forall\, t$), então (Matloff, 2009, p. 401):

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{MQO}) = \sigma^2(\boldsymbol{X^T X})^{-1} \tag{4}$$

Nos casos práticos, o valor de σ^2 é desconhecido e, portanto, deve ser estimado. Segundo Matloff (2009, p. 402), o estimador natural para σ^2 é o da equação (5) abaixo:

$$s^2 = \frac{1}{n-p-1}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

Onde *n* é o número de dados amostrais e *p* o número de parâmetros do modelo. Nota-se que o MRLC não faz qualquer hipótese relativa à distribuição do termo de erro ϵ_i, apenas exige que o valor médio de ϵ_i seja igual a zero e sua variância, uma constante finita. Neste contexto, Gujarati (2006, p. 53) destaca que "se o objetivo for apenas estimar os parâmetros do modelo (β_k), o estimador MQO é suficiente".

Contudo, na Engenharia de Avaliações, o objetivo não é apenas o de estimar os parâmetros $\boldsymbol{\beta}_k$ do modelo, mas também testar a significância dos coeficientes da regressão ou ainda, de tecer inferências sobre o quanto o valor de mercado do imóvel $(\widehat{PU})$ se aproxima da verdadeira função geradora do preços $E(PU|X_i)$. Por esta razão, conhecer a natureza da distribuição de probabilidade de ϵ_i assume papel central para a realização dos testes de hipóteses e construção de intervalos de confiança.

Acrescentando, então, a premissa da normalidade dos ϵ_i às premissas do modelo de regressão linear clássico, obtém-se o que se denomina de modelo normal de regressão linear clássico (MNRLC). Consigna-se que dada a premissa de normalidade, a distribuição de probabilidade do estimador MQO pode ser facilmente derivada, porque uma das propriedades da distribuição normal é que qualquer função linear de variáveis com distribuição normal também é normalmente distribuída, ou seja, permite-se deduzir as distribuições de probabilidade de β_k, normal, e de σ^2, relacionada ao qui-quadrado (Gujarati, 2006).

Vale observar que, dada a premissa de que $\epsilon_i \sim N(0, \sigma^2)$, então y_i, sendo uma função linear de ϵ_i, também está distribuído normalmente com média e variância dadas por $y_i \sim N(\boldsymbol{X\beta}, \sigma^2)$. Neste contexto, conforme Teorema de Rao, o estimador MQO têm a variância mínima dentro de toda a classe dos estimadores não tendenciosos, sejam ou não lineares.

Contudo, observa-se que a verificação dos pressupostos básicos do MNRLC nem sempre é garantida, haja vista que questões como falta de normalidade e heterocedasticidade são bastante comuns em dados imobiliários. Na ausência da verificação dos pressupostos do MNRLC diversos procedimentos podem ser adotados (ver Droubi e Florencio, 2024).

Especificamente sobre a violação do pressuposto da homoscedasticidade, Matloff (2017, p. 124) afirma que o estimador MQO permanece "não-viesado, consistente e assintoticamente normal", porém deixa de ter variância mínima ou ser eficiente. Gujarati (2006, p. 345) ressalta que, havendo heteroscedasticidade, os testes t e F fornecerão resultados inexatos, resultando em conclusões equivocadas.

Na Engenharia de Avaliações, uma providência corretiva à heteroscedasticidade, muito comum na prática, é a tentativa de transformação da variável dependente. Nem sempre o avaliador, contudo, entende todas implicações decorrentes do use deste artifício matemático.

Para poder estudar as melhores soluções é necessário primeiramente definir e compreender bem o problema. Na Engenharia de Avaliações trabalha-se com preços (totais ou unitários), que é uma variável limitada inferiormente pelo valor zero e sem limite superior, ou seja, a distribuição da variável preços (ou preços unitários) é assimétrica à direita, donde é razoável esperar que a hipótese da normalidade dificilmente verificar-se-á, na prática, na escala original da variável preços. Decorre deste fato que uma transformação conveniente pode gerar uma variável com distribuição aproximadamente normal: por exemplo, se verificado que os preços (ou preços unitários) apresentam distribuição lognormal, nada mais natural que aplicar uma transformação logarítmica à variável para obter uma variável de distribuição normal e assim obter um modelo em que a hipótese da normalidade provavelmente será verificada. Ocorre que nem sempre a distribuição dos preços (ou preços unitários) será lognormal e a transformação adequada da variável dependente deve ser estudada. Ocorre, contudo, que a aplicação de transformações da variável dependente, se por um lado podem resolver o problema da falta de normalidade e homoscedasticidade, trazem uma série de outros problemas, sendo o principal deles a dificuldade de interpretação do modelo, além do problema da retransformação, ou seja, o problema de como voltar à escala original (dos preços), uma vez que o modelo foi ajustado com a variável dependente transformada.

Caso o avaliador opte por permanecer na escala original, ainda que seja possível a construção de um modelo com a verificação da hipótese da normalidade, dificilmente o avaliador irá também verificar a hipótese da homoscedasticidade. Não é de se esperar na prática que os preços flutuem ao redor do valor de mercado com variância constante ao longo de toda a escala. Pelo contrário, para os imóveis com valor de mercado relativamente mais baixos (por exemplo, R$ 5.000,00/m²), espera-se uma flutuação de preços relativamente menor ao redor do valor de mercado (por exemplo, para estes imóveis com valor de mercado de R$ 5.000,00/m², uma variação de preços ao redor de ± R$ 1.000,00/m²), do que para os imóveis com maior valor de mercado (por exemplo, para os imóveis com valor de mercado de R$ 10.000,00/m², uma flutuação de preços de ± R$ 2.000,00/m²). Desta forma, o caberá ao avaliador ainda, após o ajuste do modelo com erros não-constantes, dar um tratamento adequado a este problema.

Ocorre que é quase natural prever que os preços sejam proporcionais ao valor de mercado (ou valor esperado). Nenhum avaliador haveria de discordar veemente da hipótese de que os preços flutuem mais ou menos 10 ou 20% ao redor do valor de mercado dos imóveis. Assim, por exemplo, se para os imóveis com valor de mercado de R$ 5.000,00/m² os preços variam entre R$ 4.500,00/m² e R$ 5.500,00/m², com a hipótese da proporcionalidade, os imóveis com valor de mercado de R$ 10.000,00/m² podem ter preços que variam de R$ 9.000,00/m² a R$ 11.000,00/m², ou seja, neste caso seria razoável propor um modelo em que os preços flutuam 10% (ou outro percentual a ser estimado pelo modelo) ao redor do valor de mercado. É natural, portanto, supor que os resíduos (ou a diferença entre os preços observados e o valor de mercado) sejam proporcionais ao valor de mercado, que a transformação logarítmica seja utilizada, dado que em um modelo de regressão ajustado com esta transformação é um modelo que na escala original é multiplicativo (para maiores detalhes sobre a interpretação de modelos log-linear e log-log, ver Droubi e Florencio, 2024), pois se:

$$\ln(PU) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon \qquad (6)$$

Então:

$$PU = e^{\beta_0}.e^{\beta_1 x_1}. \ldots .e^{\beta_k x_k}.e^{\varepsilon} \qquad (7)$$

Ou seja, se o modelo é ajustado com a transformação logarítmica, pode-se concluir que os erros, isolando-se o termo de erro na equação (4), são:

$$\varepsilon_i = l\,n\left(\frac{PU_i}{\widehat{PU}_i}\right) \qquad (8)$$

Pela análise da equação (8), nota-se que os resíduos do modelo logaritmizado estão no conhecido formato de *log-retornos*, ou seja, o erro é proporcional ao valor de mercado $\widehat{\mathrm{PU}}$. Isto significa que, uma vez obtidos os resíduos do modelo transformado (com a transformação logarítmica), basta exponenciá-los para obter o percentual deste erro em relação ao valor de mercado ajustado pelo modelo. Caso esta hipótese esteja correta, o avaliador irá verificar a hipótese da homoscedasticidade no seu modelo transformado.

O caminho acima, contudo, não é a única possibilidade. Caso a avaliador tenha uma hipótese para os erros do modelo, como a hipótese acima, é possível que, ao invés de utilizar o estimador MQO (Equação 2) o avaliador utilize o estimador de mínimos quadrados ponderados (MQP), da Equação (9):

$$\boldsymbol{\beta_{MQP}} = argmin_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{w_i}(\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta})^2 \qquad (9)$$

Em que w_i são pesos aplicados a cada observação, definidos de maneira inversamente proporcional à variância dos erros, *i.e.*, $w_i = 1/\sigma^2(\boldsymbol{X_i})$ (ver MATLOFF 2017, p. 133).

Em suma, uma vez que se admita que é possível conhecer (ou estimar) a função $\sigma^2(\boldsymbol{X})$, para que se possa então estabelecer uma matriz diagonal **W,** cuja diagonal deve ser composta dos pesos convenientes[1], então é possível aplicar o método dos mínimos quadrados ponderados da equação (9) e evitar, assim, a transformação da variável dependente. Definida a matriz **W**, o estimador MQP pode ser assim obtido (ver Romano e Wolf, 2017):

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{MQP}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{y} \tag{10}$$

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MQP}}) = \sigma^2(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-1} \tag{11}$$

Mais uma vez, nos casos práticos, o valor de σ^2 é desconhecido e, portanto, deve ser estimado, conforme a equação (12) abaixo:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-p-1}\sum_{i=0}^{n} w_i(y_i - \hat{y}_i)^2 \tag{12}$$

Usualmente, os pesos são obtidos a partir de valores ajustados de uma regressão auxiliar executada com os resíduos da aplicação do método MQO (ε) contra os valores ajustados do estimador MQO ($\hat{y}_i$) (caso a magnitude do erro seja proporcional a estes, ou seja, caso $Var(y|X) = \sigma^2(\hat{y}_i)$), ou contra os valores de um dos regressores (caso $Var(y|X) = \sigma^2(X_i)$). É importante, portanto, que os pesos sejam estabelecidos após a devida análise do comportamento dos resíduos de um modelo inicial MQO.

3. METODOLOGIA

Para exemplificar a utilização do método MQP, será utilizada uma amostra com 190 dados de apartamentos situados na região central de Florianópolis, disponibilizados em Zilli (2020). Na Tabela 1 são apresentadas as estatísticas descritivas das variáveis que caracterizam os referidos apartamentos.

Com estes dados será ajustado um modelo de regressão linear múltipla ordinário, para obtenção dos resíduos $\hat{\varepsilon}_i$ e dos valores ajustados $\hat{y}_i$. Com estes dois últimos, será ajustado um modelo de regressão linear simples auxiliar dos resíduos absolutos contra os valores ajustados ($|\hat{\varepsilon}_i| \sim \hat{y}_i$), visando obter a estimativa da função $\sigma(X_i)$. Com os valores ajustados da regressão auxiliar, serão compostos os pesos w_i, que por sua vez serão utilizados para a ponderação das observações e aplicação do método dos mínimos quadrados ponderados.

Tabela 1 - Descrição das variáveis efetivamente utilizadas.

Variável	Descrição	Tipo	Média	Desvio-Padrão	Mínimo	Mediana	Máximo
PU	Preço Unitário (R$/m^2)	Var. num. contínua	8.188,04	2.663,76	3.608,00	7.955,00	17.500,00
AP	Área Privativa (m^2)	Var. num. contínua	103,60	42,39	37,00	94,00	295,00
DABM	Distância à Beira-Mar (m)	Var. num. contínua	666,47	534,1	67,00	477,00	2,211,00
NG	Número de Garagens	Var. num. Discreta	1,48	0,70	0	1	4
ND	Número de Dormitórios	Var. num. Discreta	2,7	0,71	1	3	4
NB	Número de Banheiros	Var. num. Discreta	2,4	1,1	1	2	5
PSN	Piscina (Sim ou Não)	Var. dicotômica	-	-	-	-	-
PADRÃO	Padrão Construtivo (Baixo, Médio ou Alto	Var. dicotômica em grupo	-	-	-	-	-
BAIRRO	Bairro do imóvel	Var. dicotômica em grupo	-	-	-	-	-

[1] Os pesos devem ser estabelecidos de forma a anular a variação existente da variância do termo de erro (heteroscedasticidade) ao longo do intervalo de validade do modelo. Para maiores detalhes sobre como obter os pesos, ver Droubi e Florencio (2024).

4. RESULTADOS E DISCUSSÃO

Os resíduos do modelo MQO podem ser vistos na Figura 1. Nesta Figura pode-se notar que os resíduos apresentam uma tendência de alta de sua magnitude (valores absolutos) com o aumento dos valores ajustados.

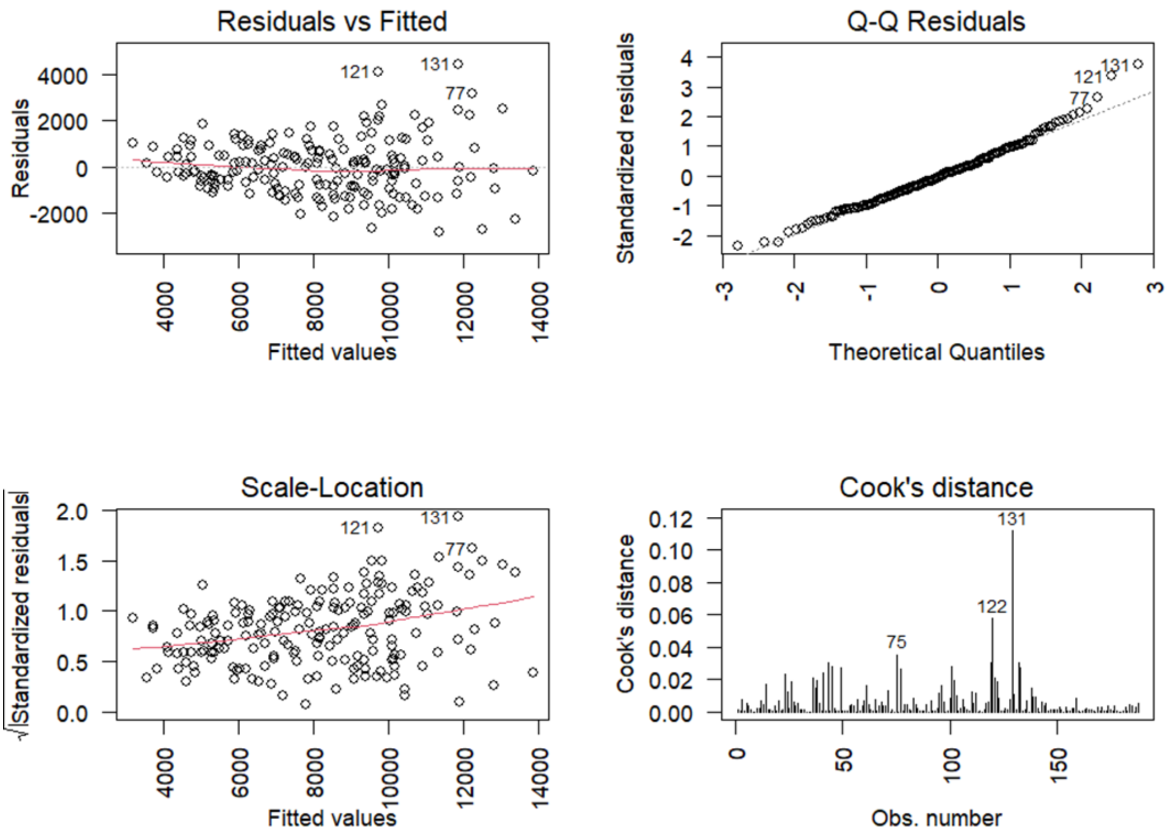

Figura 1 - Resíduos do modelo de regressão linear ordinária.

Os valores absolutos dos resíduos deste modelo foram então regredidos contra os valores ajustados do mesmo modelo. Este modelo pode ser mais bem visualizado pela análise da Figura 2, em que a reta tracejada em vermelho representa o modelo de regressão simples auxiliar.

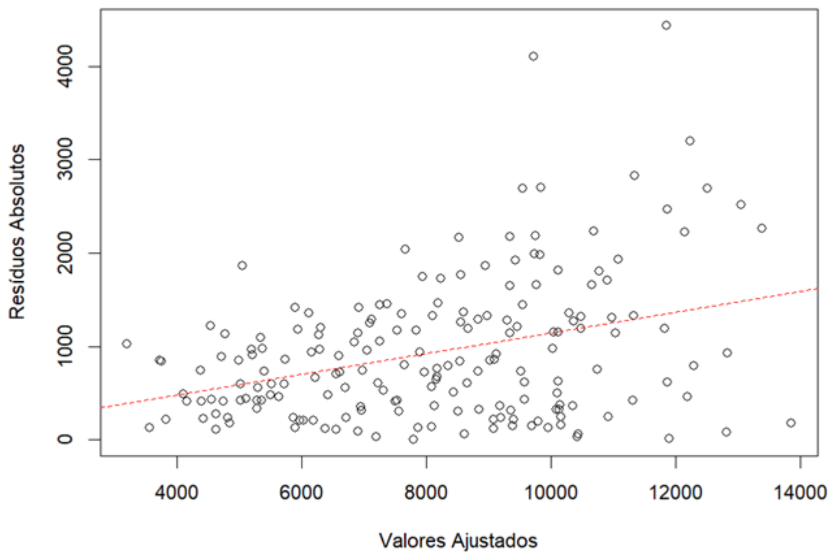

Figura 2 - Modelo de regressão linear simples auxiliar.

Por fim, os resíduos do modelo de regressão linear ponderada, utilizando-se como pesos o inverso do quadrado dos valores ajustados do modelo de regressão auxiliar, podem ser visualizados na Figura 3.

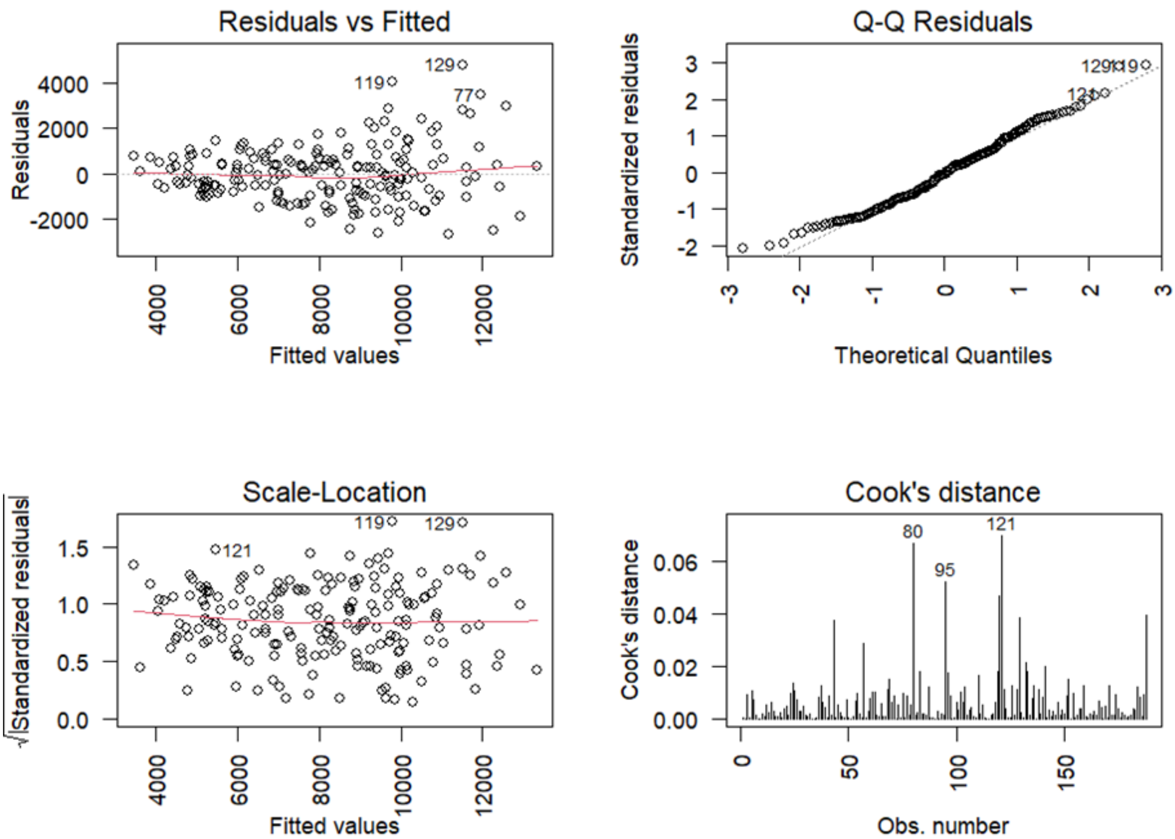

Figura 3 - Resíduos do modelo de regressão linear ponderada.

É possível perceber, pela análise dos gráficos *Scale-Location* das Figuras 1 e 3, que a tendência de aumento da magnitude dos resíduos foi eliminada, o que significa que o objetivo de obtenção de um modelo homoscedástico foi atingido com sucesso. A homoscedasticidade também foi verificada com o teste de Breusch-Pagan studentizado, tendo-se obtido para este teste, para o modelo MQP, uma estatística BP = $5{,}27.10^{-5}$. No modelo MQO o valor desta mesma estatística foi BP = 38,36, falhando na verificação da homoscedasticidade.

Por fim, na Tabela 2 podem ser vistos os valores dos coeficientes dos modelos de regressão linear ordinária e ponderada. Entre parentes, após o valor do coeficiente, são mostrados os intervalos de confiança (ao nível de 95%).

Tabela 2 - Coeficientes dos modelos de Regressão Linear Ordinária (MQO, coluna 2) e Ponderada (MQP, coluna 3)

Regressor	Modelo MQO	Modelo MQP
ln(AP/100)	-3.331,34 (-4.285,61 – -2377,07)	-2.960,51 (-3.735,55 – -2.185,47)
ln(DABM/100)	-1.989,25 (-2.813,15 – -1.165,34)	-1.618,72 (-2.402,08 – -835,36)
ln(DABM/100):BAIRRO CENTRO	1.289,45 (413,74 – 2.165,15)	936,73 (106,00 – 1.767,45)
ln(DABM/100):BAIRRO TRINDADE	1.693,67 (673,44 – 2.713,90)	1.374,36 (450,78 – 2.297,93)
NG - 1	1.510,74 (1.143,21 – 1.878,26)	1.402,25 (1.072,02 – 1.732,47)
ND - 3	440,66 (68,90 – 812,42)	347,05 (57,59 – 636,52)
NB - 2	318,49 (27,50 – 609,47)	223,28 (-43,37 – 489,92)
PSN Sim	853,73 (395,98 – 1.311,47)	794,89 (352,97 – 1.236,81)
PADRÃO MÉDIO	1.248,60 (738,28 – 1.758,93)	1.344,02 (985,25 – 1.702,79)
PADRÃO ALTO	3.037,60 (2.460,56 – 3.614,64)	3.036,50 (2.547,08 – 3.525,91)
BAIRRO CENTRO	6.707,95 (5.935,29 – 7.480,62)	6.705,94 (6.017,10 – 7.394,78)
BAIRRO AGRONÔMICA	6.742,63 (5.890,86 – 7.594,39)	6.651,02 (5.830,48 – 7.471,56)
BAIRRO TRINDADE	4.958,97 (3.486,93 – 6.431,01)	4.846,02 (3.628,71 – 6.063,32)
Notas: AIC	3.227,54	3.188,05

A vantagem de facilidade de interpretação do modelo é evidente pela análise da Tabela 2: o modelo foi construído sem intercepto e centralizado nas características de um imóvel paradigma com 100m² de área privativa (AP), 1 vaga de garagem (NG), 3 dormitórios (ND), 2 banheiros (NB), situado a 100m (1 quadra) do mar (DABM) e sem piscina (PSN). Para o padrão baixo (nível de referência), o valor do metro quadrado deste imóvel paradigma é de R$ 6.705,94 para o bairro Centro, R$ 6.651,02 para o bairro Agronômica e R$

4.846,02 para o bairro Trindade. Para imóveis de padrão médio, adicionar R$ 1.344,02 por metro quadrado a estes valores. Para os imóveis de padrão alto, adicionar R$ 3.036,50 por metro quadrado. Caso exista piscina no imóvel, adicionar R$ 794,84 por metro quadrado. Cada garagem adicional, adicionar R$ 1.402,25 por metro quadrado. No bairro Agronômica, cada quadra de distância ao mar reduz o valor do metro quadrado em R$ 1.618,72*ln(x), em que x é o número de quadras de distância ao mar. Por fim, a cada 100 m² de área privativa o valor do metro quadrado cai em R$ 2.960,51ln(y), em que y é igual a AP/100.

Na comparação entre os modelos deve ser notada não apenas a mudança dos coeficientes, mas também a notória melhora nos seus intervalos de confiança. Seja o coeficiente da variável que representa o preço médio dos imóveis no Bairro Centro: no modelo MQO, este valor era de R$ 6.707,95, com intervalo de confiança entre R$ 5.935,29 e R$ 7.480,62. Já no modelo de mínimos quadrados ponderados (MQP), o preço médio para o centro foi estimado em R$ 6.705,94 (praticamente o mesmo valor estimado pelo modelo MQO), porém com intervalo de confiança entre R$ 6.017,10 e R$ 7.394,78, um intervalo significativamente mais preciso (de 23,04% de amplitude no modelo MQO para 20,54% no modelo MQP). No Bairro Trindade, é possível perceber uma melhora ainda maior da estimação do coeficiente. Para o bairro Agronômica a melhora na estimação é mais sutil, porém ela também ocorre. Em todos os outros coeficientes há redução da amplitude dos intervalos de confiança em valores absolutos (em termos percentuais há intervalos com amplitude maiores no modelo MQP, porém isto se deve à substancial redução na magnitude da estimação do valor central do coeficiente). Por fim, a melhora da estimação também é evidenciada pela análise do Critério de Informação de Akaike (AIC_{MQP} = 3.188,05 *versus* AIC_{MQO} = 3.277,54)

Apenas a título de comparação, foi ajustado um modelo logaritmizado equivalente, *i.e.,* um modelo MQO com a mesma especificação dos modelos da Tabela 2, porém com a variável dependente transformada com a função logaritmo natural. Para este modelo foi encontrado um AIC igual a -186,40 na escala logarítimica. Transformado para a escala original (ver AKAIKE, 1978, p. 224), isto representa um AIC de 3.213,30, ou seja, o modelo MQP tem um ajuste melhor do que o modelo logaritmizado.

Ainda para efeito de comparação, o AIC na escala logarítmica (de -186,40) foi melhor do que o AIC encontrado por Zilli (2020, p. 128) para os modelos de regressão espacial do erro e da defasagem, ou ainda do modelo de regressão linear ordinário de efeitos fixos, que apresentou AIC de -183,31 (Zilli, 2020, p. 139). Acredita-se que este menor AIC se deve à consideração das interações entre as variáveis DABM e BAIRRO, o que não está presente em Zilli (2020).

Apenas os modelos de regressão geograficamente ponderada (GWR) apresentaram melhores valores para o AIC (de -207,57, no melhor caso). É possível que isto se deva, em grande medida, à consideração da variável DABM com efeitos locais (*i.e.*, não-estacionária) na regressão geograficamente ponderada.

Cumpre lembrar que, segundo Zilli (2020, p. 123), o modelo de regressão linear ordinário de efeitos fixos (ou seja, o modelo em que os dados foram agrupados por bairro) não apresentou dependência espacial, motivo pelo qual dispensou-se esta verificação neste trabalho.

5. CONCLUSÕES

O presente artigo evidenciou, por meio de uma aplicação com dados de apartamentos, a superioridade do uso do estimador de mínimos quadrados ponderados em relação ao estimador de mínimos quadrados ordinários em contextos de heterocedasticidade. A análise evidenciou que, quando a variância dos erros não é constante ao longo das observações, o estimador MQP oferece estimativas mais precisas e eficientes dos parâmetros do modelo de regressão, sendo capaz de capturar as nuances dos dados heterogêneos.

Os resultados empíricos mostraram que o modelo MQP apresentou menor variância dos estimadores e maior aderência aos dados observados, corrigindo os problemas de heterocedasticidade que afetam o estimador MQO. A comparação entre os dois métodos revelou que o MQP é não apenas mais robusto, mas também proporciona uma melhor capacidade preditiva, sobretudo quando analisada a amplitude do intervalo de confiança dos coeficientes estimados e a estatística do AIC.

Adicionalmente, para o mesmo conjunto de dados, demonstrou-se ainda, com base no AIC, a preponderância do modelo MQP comparada também com outras modelagens alternativas realizadas, a exemplo do ajuste de um modelo MQO com a variável dependente transformada com a função logaritmo natural; ou um modelo de regressão espacial de erro e de defasagem, encontrado por Zilli (2020, p. 128); ou ainda o modelo de regressão linear ordinário de efeitos fixos, obtido por Zilli (2020, p. 139).

Portanto, este trabalho reafirma a importância de considerar a heterocedasticidade ao realizar análises de regressão em dados reais. A aplicação do MQP deve ser preferida em situações em que a variabilidade dos erros é evidente, assegurando estimativas de parâmetros que são tanto não-viesadas quanto eficientes. Com base nos resultados obtidos, recomenda-se a utilização do método de mínimos quadrados ponderados como alternativa a transformação da variável dependente ou a utilização de erros heteroscedástico-consistentes.

.

REFERÊNCIAS

DANTAS, R.A. (2003). **Modelos espaciais aplicados ao mercado habitacional**: um estudo de caso para a cidade do Recife. Tese (Doutorado em Economia) - Universidade Federal de Pernambuco (UFPE), Recife.

DROUBI, L.F.P; FLORENCIO, L.A. **Engenharia de Avaliações**. O Método Comparativo. São Paulo: No prelo, 2024.

GUJARATI, D.N. **Econometria básica.** Rio de Janeiro: Elsevier, 2006.

MATLOFF, Norman S. **Statistical Regression and Classification**. From Linear Models to Machine Learning. Boca Raton, FL: CRC Press, 2017.

MATLOFF, Norman S. **From Algorithms to Z-Scores:** Probabilistic and Statistical Modeling in Computer Science. Davis, California: 2009. Disponível em: http://heather.cs.ucdavis.edu/~matloff/132/PLN/probstatbook/ProbStatBook.pdf.

OZGUR, C.; HUGHES, Z.; ROGERS, G.; PARVEEN, S. Multiple Linear Regression Applications in Real Estate Pricing. **Business Faculty Publications**, v. 61, 2016. Disponível em: https://scholar.valpo.edu/cba_fac_pub/61

ROMANO, J.P.; WOLF, M. Ressurecting Weighted Least Squares. **Journal of Econometrics**, v. 197, n. 1, p. 1-19, 2017. Disponível em: https://doi.org/10.1016/j.jeconom.2016.10.003.

ZILLI, C. **Regressão Geograficamente Ponderada aplicada na avaliação em massa de imóveis urbanos**. Dissertação de Mestrado. Florianópolis: Universidade Federal de Santa Catarina, 2020. 192 p.