

Isabel Cristina Gomes

**REGRESSÃO POLINOMIAL LOCAL
BIVARIADA: ESTIMAÇÃO E
APLICAÇÕES**

Universidade Federal de Minas Gerais

Programa de Pós-Graduação em Estatística

Belo Horizonte – MG

2010

Isabel Cristina Gomes

**REGRESSÃO POLINOMIAL LOCAL
BIVARIADA: ESTIMAÇÃO E
APLICAÇÕES**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito para obtenção do título de Mestre em Estatística.

Orientador: Prof. Gregorio Saravia Atuncar

Co-Orientador: Prof. Marcelo Azevedo Costa

Belo Horizonte – MG

2010

Aos meus pais, Abel e Delma,
e ao meu irmão Anderson.

Agradecimentos

Agradeço a Deus por estar me conduzindo tão caprichosamente pelo meu caminho, por me dar força, luz e assim permitir que a cada dia eu possa aumentar mais a minha fé;

Aos meus pais, Abel e Delma, por todo amor, carinho, compreensão e sabedoria;

Ao meu irmão Anderson, meu maior exemplo de vida;

Aos meus orientadores, Gregorio e Marcelo, que com muita dedicação, paciência e sabedoria me conduziram na elaboração dessa dissertação;

Aos amigos Alysson, Fabrícia, Isabella, Deisi e Allan por torcerem por mim e por fazerem parte da minha vida de forma tão especial;

A amiga Elaine por ter dividido comigo um pouco das suas inquietudes, e assim mudar a minha visão sobre o mundo;

Ao Grupo de Pesquisa em Economia da Saúde (*GPES*), por ter me suportado e por ter me iniciado no mundo da pesquisa. Em especial à Mariângela por seu carinho e paciência, a Gisele por sua amizade e espontaneidade, e a Emília por seu imenso carinho e pela indicação;

Aos colegas de mestrado, por dividirem comigo as angústias e dificuldades do curso, e também pelas confraternizações no CEU ou nos bares. Em especial, aos colegas que me acompanharam na graduação e no mestrado, Raphael e Grazi;

A todos os que acreditam em mim e de alguma forma me ajudaram nessa empreitada.

Resumo

A regressão polinomial local é uma abordagem não paramétrica para análise de regressão, aplicável quando a relação entre a variável resposta e as covariáveis não é satisfatoriamente estabelecida por um modelo paramétrico. A estimação da superfície é feita em cada ponto aplicando a função de regressão linear a certa quantidade de observações vizinhas ao ponto. Para tanto, é necessário determinar o tamanho da vizinhança ao redor do ponto no qual a função de regressão é estimada (janela) e a função que pondera essas observações vizinhas (função núcleo). O objetivo desse trabalho é estimar um modelo de regressão não paramétrica para casos em que se tem uma variável resposta e duas preditoras, todas elas contínuas, para pontos no interior do suporte da densidade conjunta das covariáveis. No texto são discutidas formas de se obter a janela global (a mesma para todos os pontos) e local (é diferente para cada ponto), e são apresentadas propostas de estimações para a variância condicional, a matriz Hessiana e o coeficiente de determinação. As simulações mostraram que o ajuste por janela diagonal global produz resultados melhores, com menores erros e maior aproximação à superfície teórica, do que as janelas escalares global e local. O coeficiente de determinação obtido nas aplicações em dados reais pelo ajuste não paramétrico é superior ao do modelo paramétrico, melhorando a explicação da variabilidade da resposta e permitindo identificar os pontos onde o ajuste foi razoável.

Palavras-chave: Regressão polinomial local; Variância condicional; Parâmetro de suavização; Coeficiente de determinação.

Abstract

The local polynomial regression is a nonparametric approach to regression analysis, relevant when the relation among the response and the predictors cannot be well established by a parametric model. The surface estimation is done at each point by application of linear regression function to determined amount of observations in neighborhood of the point. Therefore, it is necessary to determine the size of the neighborhood around the point in which the regression function will be estimated (bandwidth), and the function that sets weights to the neighbors observations (kernel). The purpose of this dissertation is to estimate a nonparametric regression model for cases which we have one response and two predictors, all continuous, to points at interior of support of the joint density function of predictors. In the text will be discussed ways to obtain the global bandwidth (the same to all points) and local (is different to each point), and will be presented purposes of estimation to conditional variance, Hessian matrix and determination coefficient. The simulation results shows that the fit by global diagonal bandwidth produces better results, with lower errors and better approximation to theoretical surface, when compared to constants bandwidths global and local. The determination coefficient obtained in applications to real data in nonparametric fit is upper than the parametric model, making better the explanation of the variability of response and allowing indentify the points where the adjust was reasonable.

Keywords: Local Polynomial Regression; Conditional Variance; Smoothing Parameter; Determination coefficient.

Conteúdo

1	Introdução	1
1.1	Revisão Bibliográfica	3
1.2	Objetivos	5
1.3	Organização da dissertação	6
2	Métodos	7
2.1	O problema	7
2.2	A matriz Hessiana	11
2.3	A variância condicional	13
2.4	Propriedades do estimador m	14
2.5	O parâmetro de suavização	16
2.5.1	Janela Escalar Global	17
2.5.2	Janela Diagonal Global	18
2.5.3	Janela Escalar Local	19
2.6	A densidade conjunta das covariáveis	20
2.6.1	A distribuição <i>a priori</i>	22
2.6.2	Estimadores da matriz de janelas	23
2.7	Coefficiente de Determinação	24
3	Simulações	26
3.1	Comparação dos métodos de escolha do parâmetro de suavização	26
3.2	Estimação da variância condicional	35
4	Aplicações	40
4.1	Aplicação 1	40
4.2	Aplicação 2	45
5	Conclusões	49
	Referências Bibliográficas	52

A	Demonstração do Teorema da Seção 2.4	55
B	Funções de regressão teóricas e estimadas, por modelos e situações	61
C	Funções de variância teóricas e estimadas, por modelos e situações	74

Lista de Figuras

3.1	Estimação do modelo 3.1 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.	31
3.2	Estimação do modelo 3.1 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.	32
3.3	Estimação do modelo 3.1 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.	33
3.4	Estimação do modelo 3.1 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.	34
3.5	Diagrama de dispersão R^2 estimado <i>vs</i> variância condicional estimada para o Modelo 3.1 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.	36
3.6	Diagrama de dispersão R^2 estimado <i>vs</i> variância condicional estimada para o Modelo 3.2 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.	37
3.7	Diagrama de dispersão R^2 estimado <i>vs</i> variância condicional estimada para o Modelo 3.3 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.	38
3.8	Diagrama de dispersão R^2 estimado <i>vs</i> variância condicional estimada para o Modelo 3.4 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.	39

4.1	Diagramas de dispersão para as variáveis: (a) IMRS <i>vs</i> Tx de mortalidade; (b) Tx frequencia ao ensino médio <i>vs</i> Tx de mortalidade e (c) IMRS <i>vs</i> Tx frequencia ao ensino médio.	42
4.2	Gráficos de análise de resíduos do modelo paramétrico: (a) Histograma dos resíduos, (b) Dispersão entre os valores ajustados e os resíduos padronizados e (c) QQ-plot dos resíduos.	43
4.3	Gráficos de ajuste do modelo não paramétrico por janela diagonal global (a) Diagrama de dispersão R^2 <i>vs</i> variância condicional estimada; (b) Gráfico da superfície do R^2 local; (c) Gráfico da superfície ajustada e (d) Gráfico de contorno da superfície ajustada.	44
4.4	Diagramas de dispersão para as variáveis: (a) Temperatura <i>vs</i> Ozônio; (b) Velocidade do vento <i>vs</i> Ozônio e (c) Temperatura <i>vs</i> Velocidade do vento	46
4.5	Gráficos de análise de resíduos do modelo paramétrico: (a) Histograma dos resíduos, (b) Dispersão entre os valores ajustados e os resíduos padronizados e (c) QQ-plot dos resíduos.	47
4.6	Gráficos de ajuste do modelo não paramétrico por janela diagonal global (a) Diagrama de dispersão R^2 <i>vs</i> variância condicional estimada; (b) Gráfico de superfície do R^2 local; (c) Gráfico da superfície ajustada e (d) Gráfico de contorno da superfície ajustada.	48
B.1	Estimação do modelo 3.2 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	62
B.2	Estimação do modelo 3.2 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	63
B.3	Estimação do modelo 3.2 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	64
B.4	Estimação do modelo 3.2 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	65

B.5	Estimação do modelo 3.3 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	66
B.6	Estimação do modelo 3.3 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	67
B.7	Estimação do modelo 3.3 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	68
B.8	Estimação do modelo 3.3 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	69
B.9	Estimação do modelo 3.4 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	70
B.10	Estimação do modelo 3.4 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	71
B.11	Estimação do modelo 3.4 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	72
B.12	Estimação do modelo 3.4 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local	73
C.1	Estimação da variância do modelo 3.1 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global.	74
C.2	Estimação da variância do modelo 3.1 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global.	75
C.3	Estimação da variância do modelo 3.1 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global.	75
C.4	Estimação da variância do modelo 3.1 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global.	75
C.5	Estimação da variância do modelo 3.2 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global.	76
C.6	Estimação da variância do modelo 3.2 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global.	76
C.7	Estimação da variância do modelo 3.2 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global.	76

C.8	Estimação da variância do modelo 3.2 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global. . . .	77
C.9	Estimação da variância do modelo 3.3 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global. . . .	77
C.10	Estimação da variância do modelo 3.3 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global. . . .	77
C.11	Estimação da variância do modelo 3.3 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global. . . .	78
C.12	Estimação da variância do modelo 3.3 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global. . . .	78
C.13	Estimação da variância do modelo 3.4 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global. . . .	78
C.14	Estimação da variância do modelo 3.4 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global. . . .	79
C.15	Estimação da variância do modelo 3.4 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global. . . .	79
C.16	Estimação da variância do modelo 3.4 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global. . . .	79

Lista de Tabelas

3.1	Valores médios dos EQMI's e do R^2 global para os 4 modelos nas 4 situações	29
3.2	Valores médios dos EQMI's da estimação da variância condicional para os 4 modelos nas 4 situações	35
4.1	Estatísticas descritivas - Aplicação 1	41
4.2	Modelo paramétrico para a Aplicação 1	42
4.3	Estatísticas descritivas - Aplicação 2	45
4.4	Modelo paramétrico para a Aplicação 2	46

Capítulo 1

Introdução

A análise de regressão é uma metodologia estatística que permite modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. Na regressão paramétrica a forma do relacionamento funcional entre essas variáveis é conhecida, e é possível interpretar os parâmetros estimados. Já na regressão não paramétrica não há a necessidade de um conhecimento *a priori* da forma a ser assumida pela função. A consequência da flexibilidade do modelo não paramétrico é a estimação de um número elevado de parâmetros, os quais, por isso, não permitem interpretação.

Uma vez apropriado o ajuste do modelo paramétrico, este é preferível por permitir interpretação dos parâmetros e por estar disponível na maioria dos *softwares* estatísticos. A modelagem não paramétrica se apresenta então como uma alternativa para os casos em que o modelo paramétrico não é razoável.

A análise descritiva dos dados pode fornecer pistas a favor da inadequação do modelo paramétrico, como uma relação não linear apresentada na dispersão entre a variável resposta e uma das preditoras. Na análise de resíduos pode ser verificado o atendimento às suposições de normalidade e homocedasticidade dos mesmos. O não atendimento às suposições descritas sobre os resíduos fornece evidências contra o uso do modelo paramétrico.

Na literatura estatística há diversos métodos para estimação de modelos de regressão não paramétricos como regressão local, *splines*, modelos aditivos generalizados e *projection pursuit*, en-

tre outros (Fan e Gijbels, 1996). Nessa dissertação foi utilizado o modelo de regressão local.

A modelagem local consiste em aplicar um modelo de regressão polinomial localmente, isto é, para cada ponto x , modelar a função de regressão ao redor do ponto aplicando a técnica de regressão polinomial a uma determinada fração de dados em volta de x (Fan e Gijbels, 1995). A estimação da superfície de regressão é feita através de suavização. Para tanto, há dois pontos fundamentais: a determinação do tamanho da vizinhança ao redor do ponto x no qual a função de regressão é estimada e a função que pondera os pontos utilizados na estimação local (Marquetti e Viali, 2004). Geralmente essa função atribui maior peso aos pontos mais próximos de x e menor peso aos mais distantes.

1.1 Revisão Bibliográfica

Em 1979, Cleveland propôs o método LOWESS (*Locally Weighted Scatterplot Smoothing*), cuja motivação foi a procura de um método robusto não influenciado pela presença de *outliers* na variável resposta. O LOWESS está restrito ao caso univariado (somente uma covariável). Opsomer e Ruppert (1997) trataram o problema da regressão polinomial local bivariada através da ótica dos modelos aditivos para o caso homoscedástico. Yang e Tschernig (1999) consideraram o caso heteroscedástico e propuseram formas de encontrar janela escalar global e matriz diagonal para o caso multivariado. Ye, Hyndman e Li (2006) trabalharam o caso heteroscedástico bivariado na estimação da função de regressão, porém, utilizavam a mesma janela para as duas covariáveis. Ruppert e Wand (1994) estudaram os estimadores de regressão local multivariada e conseguiram obter expressões para o vício e variância do estimador para todos os pontos de estimação, inclusive os de fronteira.

A respeito da estimação não-paramétrica da densidade bivariada, Wand e Jones (1994) propuseram a regra de referência normal para o caso univariado, e Zhang, King e Hyndman (2006) utilizaram um algoritmo *Markov chain Monte Carlo* (MCMC) para obter uma estimativa para a matriz de janelas no caso multivariado. Lima e Atuncar (2009) utilizaram a método bayesiano através da minimização da função perda para obter a matriz de janelas.

Para a estimação da variância condicional da função de regressão, Ruppert et al. (1997), e Fan e Yao (1998), estudaram o caso univariado e propuseram um estimador baseado na aplicação da regressão linear local dos resíduos quadráticos. Cai, Levine e Wang (2009) abordaram o caso multivariado com efeitos fixos usando uma generalização do estimador baseado em diferenças.

Sobre qualidade do ajuste não paramétrico, Huang e Chen (2008) propuseram uma forma de se obter a tabela ANOVA e o coeficiente de determinação, localmente e globalmente, nos casos univariado e multivariado.

No presente texto, o problema da estimação do parâmetro de suavização foi abordado de três formas: estimação da janela escalar global (método do vizinho mais próximo), janela diagonal global (método do vizinho mais próximo) e janela escalar local (minimização do erro quadrático médio). O estimador local de janela depende de quantidades desconhecidas, como a densidade conjunta das covariáveis, a variância condicional e a matriz Hessiana (matriz de derivadas de segunda ordem da função de regressão). Para a estimação da densidade bivariada foi utilizada uma abordagem Bayesiana baseada na minimização da função perda (Lima e Atuncar, 2009). Para a estimação da variância condicional foi reproduzida a ideia de Fan e Yao (1998) para o caso bivariado, ou seja, aplicação da regressão não paramétrica aos resíduos quadráticos. A estimação da matriz Hessiana foi feita através de um modelo de regressão cúbica local que, por expansão em série de Taylor, fornece estimadores naturais para as entradas dessa matriz. Para avaliar a qualidade do ajuste, foi reproduzida a metodologia proposta por Huang e Chen (2008) para a estimação do coeficiente de determinação no caso univariado.

1.2 Objetivos

O objetivo desse trabalho é estimar e implementar computacionalmente um modelo de regressão não paramétrica:

$$m(x_1, x_2) = E(Y | (X_1, X_2) = (x_1, x_2))$$

para o caso em que se tem uma variável resposta e duas preditoras, todas elas contínuas, através de regressão linear local, avaliar a qualidade do ajuste através do coeficiente de determinação e da variância local, testar os modelos estimados através de simulação e aplicar a problemas reais. Esse texto será direcionado à estimação da função de regressão em pontos no interior do suporte da função f , a função de densidade conjunta de (X_1, X_2) .

Para atingir o objetivo dessa dissertação, serão apresentados o estudo e a aplicação de alguns estimadores: da matriz de derivada segunda da função m (matriz Hessiana), da variância condicional de Y e da densidade conjunta das covariáveis (X_1, X_2) .

1.3 Organização da dissertação

O conteúdo desse texto está organizado em cinco capítulos e três apêndices. O capítulo 1 compreende a introdução, revisão de literatura e objetivos. O capítulo 2 aborda os aspectos metodológicos, começando com uma apresentação do problema de estimação da função de regressão de forma não paramétrica, discutindo a escolha da função núcleo, a obtenção da matriz Hessiana, a estimação do parâmetro de suavização, da variância condicional, da densidade conjunta das covariáveis e do coeficiente de determinação. O capítulo 3 apresenta os resultados das simulações. O capítulo 4 as aplicações a problemas reais. O capítulo 5 contém as conclusões e propostas de trabalhos futuros. No Apêndice A está a demonstração do teorema estabelecido por Ruppert e Wand (1994) para obtenção de expressões para o vício e a variância do estimador de regressão polinomial local. O Apêndice B e o Apêndice C contém, respectivamente, as superfícies estimadas das funções de regressão e das variâncias condicionais, obtidas nas simulações.

Capítulo 2

Métodos

2.1 O problema

O modelo de regressão não paramétrica com duas variáveis preditoras é definido por:

$$y_i = m(x_{1i}, x_{2i}) + v^{1/2}(x_{1i}, x_{2i})\epsilon_i, \quad i = 1, \dots, n. \quad (2.1)$$

onde $v(\cdot) = \text{Var}(y|x_1, x_2)$ é finita e conhecida como função de variância condicional. Como $v(\cdot)$ pode variar com os valores de x_1 e x_2 , o modelo é dito ser heteroscedástico. Os ϵ_i 's são variáveis aleatórias com média zero e variância unitária e independentes dos (x_1, x_2) 's e m é a função de regressão.

Seja uma amostra aleatória $(X_{11}, X_{12}, Y_1), (X_{21}, X_{22}, Y_2), \dots, (X_{1n}, X_{2n}, Y_n)$ proveniente de uma população (X_1, X_2, Y) , onde $Y_i, i = 1, \dots, n$ é o vetor de respostas e os X'_{ij} s, $i = 1, \dots, n$ e $j = 1, 2$ são vetores de variáveis preditoras em \mathbb{R}^2 com densidade comum f tendo suporte $\text{supp}(f) \subseteq \mathbb{R}^2$. O modelo (2.1) pode então ser reescrito por:

$$Y_i = m(X_{1i}, X_{2i}) + v^{1/2}(X_{1i}, X_{2i})\epsilon_i, \quad i = 1, \dots, n. \quad (2.2)$$

onde, no contexto de efeitos aleatórios,

$$m(x_1, x_2) = E(Y|(X_1, X_2) = (x_1, x_2)) \quad (2.3)$$

$$v(x_1, x_2) = Var(Y|(X_1, X_2) = (x_1, x_2)) \quad (2.4)$$

são, respectivamente, a média e a variância condicionais de Y dado $(X_1, X_2) = (x_1, x_2)$.

A regressão local, dessa forma, não estima a função de regressão globalmente, e sim na vizinhança de cada ponto $x = (x_1, x_2)$. Para isso, utiliza um parâmetro para determinar o tamanho da vizinhança do ponto, a chamada matriz de suavização B , e uma função K , conhecida como função núcleo, que pondera o conjunto de pontos ao redor do ponto x . O problema de estimação da função de regressão m pelo método de mínimos quadrados ponderados consiste em minimizar:

$$\sum_{i=1}^n \{Y_i - \beta_0 - \beta_1(X_{i1} - x_1) - \beta_2(X_{i2} - x_2)\}^2 K_B(X_{i1} - x_1, X_{i2} - x_2) \quad (2.5)$$

onde B é uma matriz não-singular, simétrica, positiva definida e de dimensão 2×2 e K é um núcleo bivariado tal que $\int_{\mathbb{R}^2} K(u) du = 1$ e

$$K_B(u) = \frac{1}{|B|^{1/2}} K(B^{-1/2}u) \quad (2.6)$$

onde $|B|$ é o determinante da matriz B .

A função núcleo deve ser contínua, simétrica e atribuir maior peso a observações mais próximas e menor peso a observações mais distantes do ponto no qual a função será estimada. O que determina se um ponto é próximo ou distante é a matriz B . Para o caso bivariado, Fan e Gijbels (1996) utilizaram o núcleo esférico de Epanechnikov em suas simulações, o qual tem a seguinte forma:

$$K_E(u_1, u_2) = \frac{4}{S_2} (1 - u_1^2 - u_2^2) I_{(u_1^2 - u_2^2 < 1)} \quad (2.7)$$

onde $S_2 = 2\pi/\Gamma(1)$ denota a área da superfície da esfera bidimensional. Os momentos

$$\mu_2(K_E) = \int_{\mathbb{R}^2} uu^T K_E(u) du$$

$$R(K_E) = \int_{\mathbb{R}^2} K_E(u)^2 du$$

são facilmente obtidos por:

$$\mu_2(K_E) = \frac{1}{6}, \quad R(K_E) = \frac{8}{3S_2} \quad (2.8)$$

Assim como em Lima e Atuncar (2009), nesse trabalho será utilizado o núcleo gaussiano nas simulações e aplicações, cujo formato no caso bivariado é dado por:

$$K_G(u_1, u_2) = (2\pi)^{-1} \exp \left\{ -1/2(u_1, u_2)^T (u_1, u_2) \right\} \quad (2.9)$$

A minimização de (2.5) é realizada com respeito a $\beta = (\beta_0, \beta_1, \beta_2)$, onde então:

$$\beta_0 = m(x_1, x_2) \quad \text{e} \quad \beta_j = \frac{\partial m}{\partial x_j}(x_1, x_2), \quad j = 1, 2. \quad (2.10)$$

A solução do problema através do método de mínimos quadrados ponderados é dada por:

$$\hat{m}((x_1, x_2); B) = e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x Y \quad (2.11)$$

onde e_1^T é um vetor de dimensão 3×1 com a primeira entrada igual a 1 e as demais iguais a zero. Para a expressão acima, assume-se ainda que $X_x^T W_x X_x$ é não singular, e as matrizes são as seguintes:

$$X_x = \begin{bmatrix} 1 & X_{11} - x_1 & X_{21} - x_2 \\ 1 & X_{12} - x_1 & X_{22} - x_2 \\ \vdots & \vdots & \vdots \\ 1 & X_{1n} - x_1 & X_{2n} - x_2 \end{bmatrix}$$

$$W_x = \text{diag} \{K_B(X_{11} - x_1, X_{21} - x_2), \dots, K_B(X_{1n} - x_1, X_{2n} - x_2)\}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Se for denotado por $\hat{\beta}$ o estimador do vetor β , os estimadores da função de regressão bivariada m e de suas 2 primeiras derivadas parciais de primeira ordem são:

$$\hat{m}(x_1, x_2) = \hat{\beta}_0, \text{ e } \frac{\partial \hat{m}}{\partial x_j}(x_1, x_2) = \hat{\beta}_j, \quad j = 1, 2. \quad (2.12)$$

Através da segunda parte da expressão (2.12) podem ser estimadas as entradas do vetor de derivadas parciais de primeira ordem da função m , o vetor $\hat{D}_m(x)$:

$$\hat{D}_m(x) = \begin{bmatrix} \partial \hat{m} / \partial x_1 \\ \partial \hat{m} / \partial x_2 \end{bmatrix}$$

2.2 A matriz Hessiana

Outra quantidade importante para a modelagem local é a matriz de derivadas parciais de segunda ordem da função m , conhecida como matriz Hessiana, dada por:

$$\mathcal{H}_m(x) = \begin{bmatrix} \frac{\partial^2 m}{\partial x_1^2} & \frac{\partial^2 m}{\partial x_1 \partial x_2} \\ \frac{\partial^2 m}{\partial x_2 \partial x_1} & \frac{\partial^2 m}{\partial x_2^2} \end{bmatrix}$$

As entradas da matriz Hessiana são desconhecidas. Para estimá-las, foi utilizada a generalização do método para obter as derivadas da função m apresentadas por Fan e Gijbels (1995) para o caso univariado. Ruppert e Wand (1994) estudaram o estimador de derivadas de ordens superiores do modelo de regressão local univariado. Eles mostraram que, quando a diferença entre o grau do polinômio e a ordem da derivada a ser estimada era ímpar, o vício do estimador era menor do que quando essa diferença era par. Para esta dissertação será estimada a derivada de segunda ordem, então é necessário estimar um polinômio de terceiro grau. Considere uma expansão de ordem 3 da função $m(x_1, x_2)$ em torno do ponto (x_1^0, x_2^0) :

$$\begin{aligned} m(x_1, x_2) &\approx m(x_1^0, x_2^0) + \frac{\partial m}{\partial x_1^0}(x_1 - x_1^0) + \frac{\partial m}{\partial x_2^0}(x_2 - x_2^0) \\ &+ \frac{1}{2} \frac{\partial^2 m}{(\partial x_1^0)^2}(x_1 - x_1^0)^2 + \frac{\partial^2 m}{\partial x_1^0 \partial x_2^0}(x_1 - x_1^0)(x_2 - x_2^0) + \frac{1}{2} \frac{\partial^2 m}{(\partial x_2^0)^2}(x_2 - x_2^0)^2 \\ &+ \frac{1}{6} \frac{\partial^3 m}{(\partial x_1^0)^3}(x_1 - x_1^0)^3 + \frac{1}{2} \frac{\partial^3 m}{(\partial x_1^0)^2 \partial x_2^0}(x_1 - x_1^0)^2(x_2 - x_2^0) \\ &+ \frac{1}{2} \frac{\partial^3 m}{\partial x_1^0 (\partial x_2^0)^2}(x_1 - x_1^0)(x_2 - x_2^0)^2 + \frac{1}{6} \frac{\partial^3 m}{(\partial x_2^0)^3}(x_2 - x_2^0)^3 \end{aligned} \quad (2.13)$$

Comparando a fórmula (2.13) ao problema de mínimos quadrados ponderados com polinômio de grau 3 dado por:

$$\begin{aligned}
& \sum_{i=1}^n (Y_i - \beta_0 - \beta_1(x_{1i} - x_1^0) - \beta_2(x_{2i} - x_2^0) - \beta_3(x_{1i} - x_1^0)^2 & (2.14) \\
& -\beta_4(x_{1i} - x_1^0)(x_{2i} - x_2^0) - \beta_5(x_{2i} - x_2^0)^2 \\
& -\beta_6(x_{1i} - x_1^0)^3 - \beta_7(x_{1i} - x_1^0)^2(x_{2i} - x_2^0) \\
& -\beta_8(x_{1i} - x_1^0)(x_{2i} - x_2^0)^2 - \beta_9(x_{2i} - x_2^0)^3)K_B(x_{1i} - x_1^0, x_{2i} - x_2^0)
\end{aligned}$$

São obtidos os estimadores para as entradas da matriz Hessiana:

$$\frac{\partial^2 \hat{m}}{(\partial x_1^0)^2} = 2\hat{\beta}_3 \quad (2.15)$$

$$\frac{\partial^2 \hat{m}}{\partial x_1^0 \partial x_2^0} = \hat{\beta}_4 \quad (2.16)$$

$$\frac{\partial^2 \hat{m}}{(\partial x_2^0)^2} = 2\hat{\beta}_5 \quad (2.17)$$

2.3 A variância condicional

A literatura de regressão não paramétrica apresenta vários métodos para a estimação da variância condicional. Nesse trabalho, foi usada a generalização do método que Fan e Yao (1998) apresentaram para o caso univariado. Dada a função de regressão $m(\cdot)$, o problema de estimar $v(\cdot)$ torna-se um problema de regressão não paramétrica quando toma-se como ponto de partida a relação:

$$E[r|(X_1, X_2) = (x_1, x_2)] = v^2 \quad (2.18)$$

onde $r = [Y - m(X_1, X_2)]^2$. O estimador linear local de v^2 é \hat{v}^2 , e considerando $\hat{r} = [Y - \hat{m}(X_1, X_2)]^2$, pode ser obtido através da minimização de:

$$\left\{ \sum_{i=1}^n [r_i - \alpha - \beta_1(X_{1i} - x_1) - \beta_2(X_{2i} - x_2)]^2 K_B(X_{1i} - x_1, X_{2i} - x_2) \right\} \quad (2.19)$$

Basicamente, a estimação da variância condicional segue o seguinte algoritmo (Fan e Yao, 1998):

1. Estima-se $\hat{m}(x_1, x_2)$.
2. Calculam-se os resíduos quadráticos $\hat{r}_i = [Y_i - \hat{m}(X_{1i}, X_{2i})]^2$, $i = 1, \dots, n$.
3. Estima-se $\hat{\alpha}$ utilizando a Fórmula 2.19.

2.4 Propriedades do estimador m

Se for denotado por $R(K) = \int_{\mathbb{R}^2} K(u)^2 du$, e atendendo às seguintes suposições (Ruppert e Wand, 1994):

- Suposição A_1 : O núcleo K tem suporte compacto tal que o seu segundo momento é definido por $\mu_2(K)I = \int_{\mathbb{R}^2} uu^T K(u) du$, com $\mu_2(K) \neq 0$. Os momentos de ordem ímpar de K são iguais a zero (condição satisfeita por núcleos esfericamente simétricos).
- Suposição A_2 : O ponto $x = (x_1, x_2)$ está contido em $\text{supp}(f)$. Em x , v é contínua, f é continuamente diferenciável e todas as derivadas de segunda ordem de m são contínuas. Além disso, $f(x) > 0$ e $v(x) > 0$.
- Suposição A_3 : A sequência de matrizes de janelas B_n é tal que $n^{-1}|B_n|$ e cada entrada de B_n tendem a zero quando $n \rightarrow \infty$, com B_n permanecendo simétrica e positiva definida.

Podem ser obtidas expressões para o vício e para a variância de \hat{m} , condicionados à amostra. Antes de enunciar o Teorema, é conveniente definir o que é ordem em probabilidade. Pode-se dizer que uma sequência de variáveis aleatórias Y_n é tal que $Y_n = o_p(k_n)$, se $Y_n/k_n \rightarrow 0$ em probabilidade (Definição 2.4.1) e que $Y_n = O_p(k_n)$, se Y_n/k_n é limitado em probabilidade (Definição 2.4.2) (Lehmann e Casela, 1998).

Definição 2.4.1: Convergência em Probabilidade

Y_n converge para Y em probabilidade se, para todo $\epsilon > 0$, $P(|Y_n - Y| \geq \epsilon) \rightarrow 0$ quando $n \rightarrow \infty$ (Barry James, 2004).

Notação: $Y_n \xrightarrow{P} Y$.

Definição 2.4.2: Sequência limitada em Probabilidade

Uma sequência de variáveis aleatórias Y_n é limitada em probabilidade se, dado um $\epsilon > 0$, há um M e n_0 tal que $P(|Y_n| > M) < \epsilon$ para todo $n > n_0$ (Lehmann e Casela, 1998).

Ou seja, dizer que $Y_n = o_p(k_n)$ significa dizer que, em relação a n , a ordem de k_n é maior que a ordem de Y_n . Similarmente, dizer que $Y_n = O_p(k_n)$ significa que, em relação a n , a ordem de Y_n é similar a ordem de k_n .

Teorema (Ruppert e Wand, 1994): Seja $x = (x_1, x_2)$ um elemento fixo no interior de $\text{supp}(f)$. Assumindo verdadeiras as suposições A1 - A3 e denotando por $\text{tr}(B)$ o traço da matriz B , então:

$$E \{ \hat{m}(x; B) - m(x) | X_1, \dots, X_n \} = 1/2 \mu_2(K) \text{tr}(B \mathcal{H}_m(x)) + o_p(\text{tr}(B)) \quad (2.20)$$

$$\text{Var} \{ \hat{m}(x; B) | X_1, \dots, X_n \} = \{ n^{-1} |B|^{-1/2} R(K) / f(x) \} v(x) (1 + o_p(1)) \quad (2.21)$$

Essas expressões representam o vício e variância condicionais do estimador da função m . Particularmente, sob algumas condições, $\hat{m}(x; B)$ é assintoticamente normal com vício e variância assintótica dadas por essas expressões. A demonstração do Teorema pode ser vista no Apêndice A.

2.5 O parâmetro de suavização

A matriz de janelas representa o parâmetro de suavização responsável por controlar o tamanho da vizinhança do ponto x onde a função núcleo será aplicada. A escolha da matriz B está ligada à variabilidade e ao vício da estimativa da função de regressão: se as entradas de B forem pequenas, o vício da estimativa tende a ser pequeno, porém a variabilidade tende a ser grande; se as entradas de B forem grandes, o vício tenderá a ser grande e a variabilidade a ser pequena. Quando as entradas de B são muito grandes, a superfície estimada aproxima-se de uma regressão linear (Fan e Gijbels, 1995 e Marquetti e Viali, 2004).

A matriz de janelas é dita global quando é a mesma para todos os pontos de interesse $x = (x_1, x_2)$ em que a função será estimada. A escolha da janela global falha quando encontra vizinhanças vazias, o que pode ocorrer nas caudas da distribuição ou de forma frequente quando se estima uma função de mais de uma dimensão. Para resolver o problema das vizinhanças vazias, pode-se obter a matriz de janelas local, a qual pode ser escolhida convenientemente de modo a ter sempre um número pré-determinado de pontos (Marquetti e Viali, 2004).

É comum usar para matriz de janelas uma estrutura da forma $B = bI$. O uso de uma matriz de janelas da forma diagonal significa que as duas covariáveis são independentes e que os elipsóides têm seus eixos na mesma direção dos eixos das variáveis preditoras. Já para uma matriz de janelas completa os eixos dos elipsóides correspondem aos autovetores dessa matriz (Ruppert e Wand, 1994). A seguir serão apresentadas duas formas de se obter a janela global (escalar e diagonal) e uma forma de se obter a janela escalar local.

2.5.1 Janela Escalar Global

O método do vizinho mais próximo

Para a utilização do método do vizinho mais próximo para o caso bivariado, consideremos a distância entre pontos em \mathbb{R}^2 , calculada através da distância euclidiana:

$$\rho(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (2.22)$$

A vizinhança de uma observação particular $x_k = (x_{k1}, x_{k2})$ é determinada pela janela associada b_k dada pelo r -ésimo menor valor obtido de $\rho(x_k, x_i)$, para $i = 1, \dots, n$. O valor de r pode ser obtido como o inteiro mais próximo da quantidade nd , onde n é o tamanho da amostra e d é tal que $0 < d < 1$. Segundo Cleveland (1979), para regressão polinomial de ordem 1 escolher d entre 0,2 e 0,8 produz resultados satisfatórios. Para esse caso, se as covariáveis forem medidas em escalas diferentes recomenda-se usar a forma padronizada delas.

2.5.2 Janela Diagonal Global

O método do vizinho mais próximo

Uma vez que, em certas situações, as unidades de medida das variáveis preditoras podem ser diferentes, e conseqüentemente também a variância delas, a utilização de uma janela diagonal torna-se atraente. O método do vizinho mais próximo nesse caso consiste em aplicar a cada uma das preditoras uma janela diferente. A ideia é basicamente aplicar o método da janela escalar global a cada uma das covariáveis.

A distância euclidiana, neste caso, será calculada para cada uma das preditoras:

$$\rho_j(x_{ij}) = |x_{ij} - x_j|, \quad i = 1, 2, \dots, n \text{ e } j = 1, 2. \quad (2.23)$$

A janela b_j da j -ésima variável, $j = 1, 2$, será dada pelo r -ésimo valor de ρ_j , após ordenação do mesmo. O valor de r pode ser obtido da mesma forma que foi obtido para janela escalar global. A matriz diagonal então será:

$$B = \text{diag} \left\{ \hat{b}_1, \hat{b}_2 \right\} \quad (2.24)$$

2.5.3 Janela Escalar Local

Minimização do erro quadrático médio

Ye, Hyndman e Li (2006) abordaram o problema da minimização das expressões obtidas por Ruppert e Wand (1994) para o vício e a variância do estimador da regressão local (Fórmulas 2.20 e 2.21). Os autores objetivavam encontrar uma expressão para um estimador de janela escalar local, agregando a informação da heteroscedasticidade do modelo para melhorar a estimação do parâmetro de suavização. Nas simulações apresentadas no artigo, mostraram que a janela escalar local apresentava melhores resultados que a global. A expressão por eles obtida para o caso bivariado foi:

$$\hat{b}_{JEL}(x) = \left(\frac{\hat{v}(x)}{n\pi\hat{f}(x)s^2(\hat{\mathcal{H}}_m(x))} \right)^{1/6} \quad (2.25)$$

onde $s(\hat{\mathcal{H}}_m(x))$ é a soma dos termos da matriz Hessiana. Como a expressão depende de quantidades desconhecidas, é preciso estimá-las e substituir dentro da fórmula (método *plug-in*).

2.6 A densidade conjunta das covariáveis

Seja X o vetor (X_1, X_2) com densidade conjunta f . O estimador da densidade f no ponto (x_1, x_2) é:

$$\tilde{f}[(x_1, x_2)|B] = \int_{\mathbb{R}^2} K_B [(X_1, X_2) - (x_1, x_2)] d\hat{F}_n(x_1, x_2) \quad (2.26)$$

Considerando a função núcleo definida pela Fórmula 2.6 e $\hat{F}_n(x_1, x_2)$ a função de distribuição empírica. O núcleo estimador de f será dado por:

$$\hat{f}[(x_1, x_2)|B] = \frac{1}{n} \sum_{i=1}^n |B|^{-1/2} K_B [(X_1, X_2) - (x_1, x_2)] \quad (2.27)$$

Para a estimação da matriz de janelas 2×2 , Lima e Atuncar (2009) utiliza um método Bayesiano que consiste na minimização da função de perda *a posteriori* definida como a distância entre a matriz B real e a matriz B estimada (\hat{B}). Tal proposta é descrita a seguir.

Seja $\pi(B)$ a distribuição *a priori* de B . A distribuição *a posteriori*, dado (x_1, x_2) , é dada por:

$$\pi \{B|(x_1, x_2)\} = \frac{f[(x_1, x_2)|B] \pi(B)}{\int f[(x_1, x_2)|B] \pi(B) dB} \quad (2.28)$$

Nota-se que $\pi(B|(x_1, x_2))$ não pode ser obtida diretamente porque a densidade $f((x_1, x_2)|B)$ não é conhecida. Mas como a densidade é igual a:

$$E_f \{K_B((X_1, X_2) - (x_1, x_2))\}$$

Um estimador natural seria:

$$\hat{f}((x_1, x_2)|B) = n^{-1} \sum_{i=1}^n K_B [(X_1, X_2) - (x_1, x_2)]$$

Pode então ser obtido um estimador para a distribuição *a posteriori* substituindo $f((x_1, x_2)|B)$ por $\hat{f}((x_1, x_2)|B)$ na fórmula 2.28. O núcleo utilizado foi o gaussiano.

Seja \hat{B} um estimador de B e $L = L(\hat{B}, B)$ a função perda. Então, a perda esperada *a posteriori* associada a \hat{B} é dada por:

$$E \left\{ L(\hat{B}, B)|(x_1, x_2) \right\} = \int L(\hat{B}, B)\pi(B|(x_1, x_2))dB \quad (2.29)$$

E assim, seguindo a regra de Bayes, o estimador da matriz B é a matriz solução do seguinte problema:

$$\hat{B}(x_1, x_2) = \text{Min}_{B \in \psi} \left\{ E \left[L(\hat{B}, B|(x_1, x_2)) \right] \right\} \quad (2.30)$$

onde ψ é o espaço das matrizes positivas definidas. Dois tipos de função perda são comumente usadas, a função de perda entrópica (L_1) e a função de perda quadrática (L_2). As fórmulas e os respectivos estimadores de Bayes são dados abaixo.

$$L_1 = \text{tr}(\hat{B}B^{-1}) - \log(\hat{B}B^{-1}) - 2 \quad (2.31)$$

$$\hat{B}_1(x_1, x_2) = [E_\pi(B^{-1})]^{-1} \quad (2.32)$$

$$L_2 = \text{tr}(B - \hat{B})^2 \quad (2.33)$$

$$\hat{B}_2(x_1, x_2) = E_\pi(B) \quad (2.34)$$

2.6.1 A distribuição *a priori*

A distribuição *a priori* utilizada para B foi a Wishart inversa, sob a notação $B \sim WI_2(q, V)$, cuja função densidade é dada por:

$$\pi(B) = \frac{|V|^{q/2} |B|^{-(q+3)/2} \exp[-\text{tr}(VB^{-1})/2]}{2^q \Gamma_2(q/2)} \quad (2.35)$$

onde $q \geq 2$ representa os graus de liberdade, V é uma matriz 2×2 positiva definida, chamada matriz escala e $\Gamma_2(\cdot)$ é função gama generalizada. A distribuição de Wishart é uma generalização da distribuição gama de dimensões múltiplas, e faz parte de uma família de distribuições de probabilidade para modelar matrizes não-negativas-definidas. Esses tipos de distribuições são importantes na estimação de matrizes de covariância em estatística multivariada.

Sobre os parâmetros q e V da distribuição, estudos empíricos mostraram que a utilização de $q = 7$ é razoável no caso bivariado e V pode ser estimado pela matriz de covariâncias de (X_1, X_2) . Se a matriz de covariâncias for desconhecida, pode ser estimada pela matriz de covariâncias amostral.

2.6.2 Estimadores da matriz de janelas

Através dos resultados mostrados anteriormente, é possível obter o estimador de Bayes para a matriz B em cada ponto (x_1, x_2) . Se for utilizada a função de perda L_1 , o estimador é dado por:

$$\hat{B}_1(x_1, x_2) = \frac{[\sum_{i=1}^n \omega_i \Delta_i^{-1}]^{-1}}{q+1} \quad (2.36)$$

E se for utilizada a função de perda L_2 , o estimador é dado por:

$$\hat{B}_2(x_1, x_2) = \frac{\sum_{i=1}^n \omega_i \Delta_i}{q+1} \quad (2.37)$$

Onde:

$$\Delta_i = \left\{ [(X_1, X_2) - (x_1, x_2)] [(X_1, X_2) - (x_1, x_2)]^T + V \right\} \quad (2.38)$$

E:

$$\omega_i = \frac{|\Delta_i|^{-(q+1/2)}}{\sum_{i=1}^n |\Delta_i|^{-(q+1/2)}} \quad (2.39)$$

2.7 Coeficiente de Determinação

Huang e Chen (2008) propuseram formas de calcular o coeficiente de determinação R^2 tanto para casos com apenas uma variável preditora quanto para casos com várias preditoras. Os coeficientes de determinação locais e global foram estimados utilizando a mesma ideia proposta para o caso univariado.

A SSE local foi definida como o quociente entre a expressão 2.5 e a soma dos pesos locais:

$$SSE((x_1, x_2); B) = \frac{n^{-1} \sum_{i=1}^n \{Y_i - \hat{Y}\}^2 K_B(X_{1i} - x_1, X_{2i} - x_2)}{n^{-1} \sum_{i=1}^n K_B(X_{1i} - x_1, X_{2i} - x_2)} \quad (2.40)$$

O denominador da fórmula 2.40 é o núcleo estimador da função de densidade. A SST local é definida como:

$$SST((x_1, x_2); B) = \frac{n^{-1} \sum_{i=1}^n \{Y_i - \bar{Y}\}^2 K_B(X_{1i} - x_1, X_{2i} - x_2)}{\hat{f}((x_1, x_2); B)} \quad (2.41)$$

E a SSR local é definida como:

$$SSR((x_1, x_2); B) = \frac{n^{-1} \sum_{i=1}^n \{\hat{Y} - \bar{Y}\}^2 K_B(X_{1i} - x_1, X_{2i} - x_2)}{\hat{f}((x_1, x_2); B)} \quad (2.42)$$

Assim, pode ser definido o R^2 pontual:

$$R^2((x_1, x_2); B) = 1 - \frac{SSE((x_1, x_2); B)}{SST((x_1, x_2); B)} = \frac{SSR((x_1, x_2); B)}{SST((x_1, x_2); B)} \quad (2.43)$$

Para se obter o coeficiente de determinação global, deve-se integrar as versões locais das somas de quadrados (dos resíduos, da regressão e total) apresentadas anteriormente.

$$SST(B) = \int_{\mathbb{R}^2} SST((x_1, x_2); B) \hat{f}((x_1, x_2); B) dx \quad (2.44)$$

$$SSE(B) = \int_{\mathbb{R}^2} SSE((x_1, x_2); B) \hat{f}((x_1, x_2); B) dx \quad (2.45)$$

$$SSR(B) = \int_{\mathbb{R}^2} SSR((x_1, x_2); B) \hat{f}((x_1, x_2); B) dx \quad (2.46)$$

E o R^2 global torna-se:

$$R^2(B) = 1 - \frac{SSE(B)}{SST(B)} = \frac{SSR(B)}{SST(B)} \quad (2.47)$$

Capítulo 3

Simulações

3.1 Comparação dos métodos de escolha do parâmetro de suavização

Os modelos usados nas simulações são adaptações dos modelos apresentados por Ye, Hyndman e Li (2006) nos estudos numéricos para regressão bivariada.

$$Y = x_1x_2 + \{Var_1\}^{1/2} \epsilon \quad (3.1)$$

$Var_1 = (X_2 - X_1)^2$ para $X_1 \geq 3$ e $X_2 \geq 3$; e 0,5 caso contrário.

$$Y = x_1^2 + e^{-x_2^2} + \{Var_2\}^{1/2} \epsilon \quad (3.2)$$

$Var_2 = (X_2 - X_1)^2$ para $X_1 \geq 3$ e $X_2 \geq 3$; e 0,5 caso contrário.

$$Y = x_1 + 2sen(1,5x_2) + \{Var_3\}^{1/2} \epsilon \quad (3.3)$$

$Var_3 = (X_2 - X_1)^2 + 2$ para $X_1 \geq 3$ e $X_2 \geq 3$; e 0,5 caso contrário.

$$Y = \text{sen}(x_1 + x_2) + 2e^{-2x_2^2} + \{\text{Var}_4\}^{1/2} \epsilon \quad (3.4)$$

$\text{Var}_4 = (X_2 - X_1)^2 + 2$ para $X_1 \geq 3$ e $X_2 \geq 3$; e 0,5 caso contrário.

Os dados foram simulados para 4 modelos (Fórmula 2.2), onde $\epsilon \sim N(0, 1)$, com as covariáveis X_1 e X_2 provenientes de uma distribuição normal bivariada com vetor de médias $\mu = (3, 4)$ e matriz de covariâncias Σ_i , $i = 1, 2, 3, 4$ ilustrando as seguintes situações:

1. Variáveis preditoras independentes (correlação igual a zero) e com variâncias parecidas:

$$\Sigma_1 = \begin{bmatrix} 0,2 & 0 \\ 0 & 0,3 \end{bmatrix}$$

2. Variáveis preditoras independentes (correlação igual a zero), mas com variâncias diferentes:

$$\Sigma_2 = \begin{bmatrix} 0,2 & 0 \\ 0 & 2,0 \end{bmatrix}$$

3. Variáveis preditoras dependentes, com correlação positiva $\rho = 0,46$ e variâncias parecidas:

$$\Sigma_3 = \begin{bmatrix} 1,0 & 0,5 \\ 0,5 & 1,2 \end{bmatrix}$$

4. Variáveis preditoras dependentes, com correlação negativa $\rho = -0,46$ e variâncias parecidas:

$$\Sigma_4 = \begin{bmatrix} 1,0 & -0,5 \\ -0,5 & 1,2 \end{bmatrix}$$

Foram geradas 100 amostras aleatórias de tamanho 200 de cada modelo. Para avaliar a qualidade do ajuste, foi utilizado o erro quadrático médio integrado (EQMI), obtido a partir da integração dos erros quadráticos médios (EQM):

$$EQM = n^{-1} \{\hat{m}(x_{i1}, x_{i2}) - m(x_{i1}, x_{i2})\}^2 \quad (3.5)$$

A integração dos erros quadráticos médios foi feita através de metodologia de integração numérica. Além da comparação do EQMI para os 3 métodos de seleção do parâmetro de suavização, foi possível avaliar a qualidade do ajuste através da comparação das curvas teóricas e estimadas e do valor médio do R^2 global. As simulações foram feitas no *software* livre *R* (*R Development Core Team*) versão 2.7.2. As amostras de distribuição normal bivariada foram simuladas com funções do pacote *mixAK* (Komarek, 2010) e as superfícies apresentadas com funções do pacote *rgl* (Adler e Murdoch, 2009).

A Tabela 3.1 apresenta os valores médios dos EQMI's e dos R^2 globais para os 3 métodos, por modelo e situação. Para todos os modelos, a janela diagonal global apresenta os menores EQMI's médios nas 4 situações. A janela escalar local parece gerar um vício maior, porque na maioria dos casos avaliados apresentou EQMI médio maior que os outros dois métodos. A situação 2 apresentou maiores EQMI's para todos os modelos.

Em relação aos valores médios dos R^2 globais, os maiores valores ocorreram para a janela escalar global nas situações 1 e 2 e para a janela diagonal global para as situações 3 e 4, no caso do modelo 3.1. Para os modelos 3.2 e 3.3, em todas as situações o maior R^2 global médio foi da janela diagonal global. Para o modelo 3.4, na situação 2 o maior R^2 ocorreu para janela diagonal global, enquanto para as outras situações ocorreu na janela escalar global. O potencial de explicação da variabilidade da resposta é pequeno para o modelo 3.4 nas 4 situações. Isso provavelmente ocorre porque o valor médio da variável resposta é zero, de forma que o ruído domina a informação.

Tabela 3.1: Valores médios dos EQMI's e do R^2 global para os 4 modelos nas 4 situações

Modelos	Estatísticas	Situação 1	Situação 2	Situação 3	Situação 4
Modelo 3.1					
Janela escalar global	EQMI	0,0021	0,0190	0,0092	0,0048
	R^2	0,9809	0,9917	0,9987	0,9956
Janela diagonal global	EQMI	0,0017	0,0162	0,0079	0,0038
	R^2	0,9803	0,9916	0,9988	0,9957
Janela escalar local	EQMI	0,0021	0,0210	0,0091	0,0051
	R^2	0,9790	0,9909	0,9985	0,9949
Modelo 3.2					
Janela escalar global	EQMI	0,0022	0,0195	0,0097	0,0068
	R^2	0,9875	0,9580	0,9985	0,9990
Janela diagonal global	EQMI	0,0020	0,0167	0,0074	0,0052
	R^2	0,9862	0,9458	0,9984	0,9990
Janela escalar local	EQMI	0,0023	0,0202	0,0096	0,0085
	R^2	0,9850	0,9414	0,9979	0,9986
Modelo 3.3					
Janela escalar global	EQMI	0,0080	0,0353	0,0247	0,0143
	R^2	0,6488	0,6667	0,8106	0,7272
Janela diagonal global	EQMI	0,0077	0,0303	0,0159	0,0103
	R^2	0,6434	0,6228	0,8074	0,7214
Janela escalar local	EQMI	0,0083	0,0463	0,0250	0,0171
	R^2	0,6266	0,5395	0,7552	0,6485
Modelo 3.4					
Janela escalar global	EQMI	0,0079	0,0371	0,0151	0,0114
	R^2	0,1817	0,2410	0,2840	0,2748
Janela diagonal global	EQMI	0,0073	0,0318	0,0136	0,0102
	R^2	0,1795	0,2446	0,2825	0,2692
Janela escalar local	EQMI	0,0085	0,0407	0,0164	0,0126
	R^2	0,1443	0,1796	0,1770	0,2217

As Figuras 3.1, 3.2, 3.3 e 3.4 mostram as superfícies teóricas e estimadas para o modelo 3.1 nas 4 situações. As aproximações às curvas teóricas mostraram-se razoáveis. As figuras para os demais modelos nas 4 situações podem ser vistas no Apêndice B. Para os modelos 3.2, 3.3 e 3.4 pode-se notar boas aproximações à superfície teórica. Porém, para os modelos 3.3 e 3.4, podem ser observados problemas de fronteira nas estimações.

Tecnicamente, as superfícies estimadas por janela escalar local teriam um potencial de suavização maior que a dos outros métodos. Esse fato pode ser explicado pela incorporação das estimativas da variância condicional local e da matriz Hessiana na estimação da

janela local. Segundo Ruppert e Wand (1994), cada entrada da matriz Hessiana reflete uma medida de curvatura da função m no ponto x em uma direção particular. Assim, quanto maior for o 'alisamento' exposto pela superfície, maior também será o vício da estimativa da função de regressão.

Diante desses resultados, o método escolhido para a estimação da variância condicional foi a janela diagonal global, por apresentar menores erros.

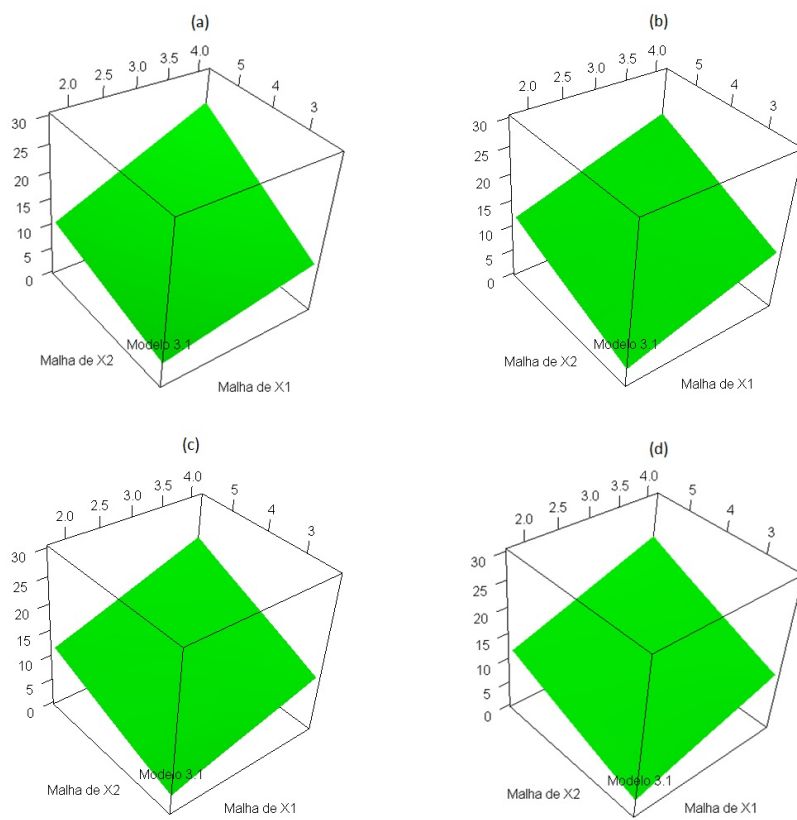


Figura 3.1: Estimação do modelo 3.1 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.

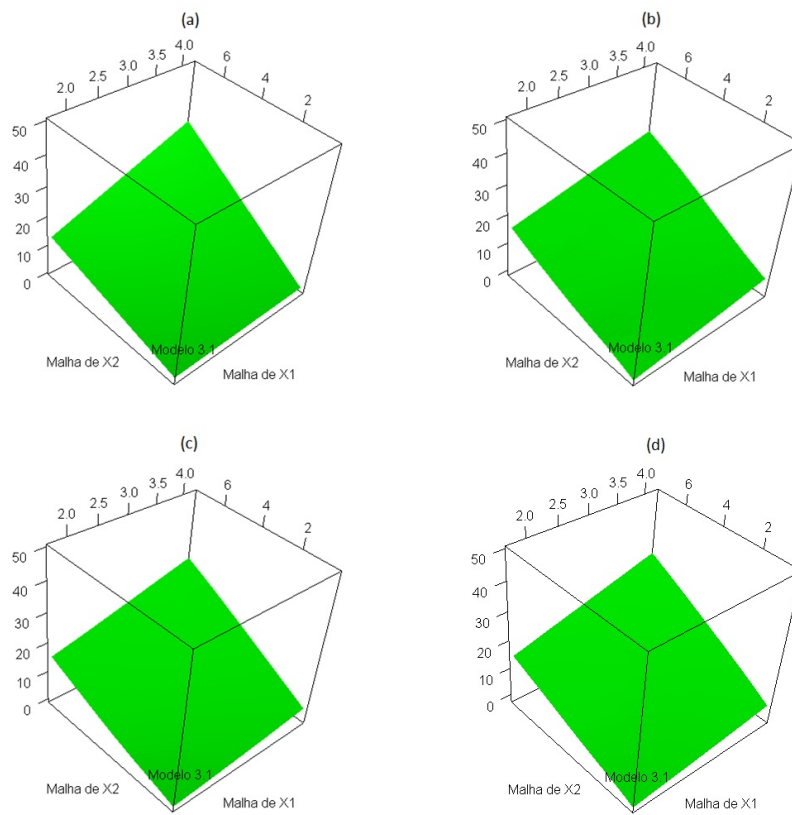


Figura 3.2: Estimação do modelo 3.1 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.

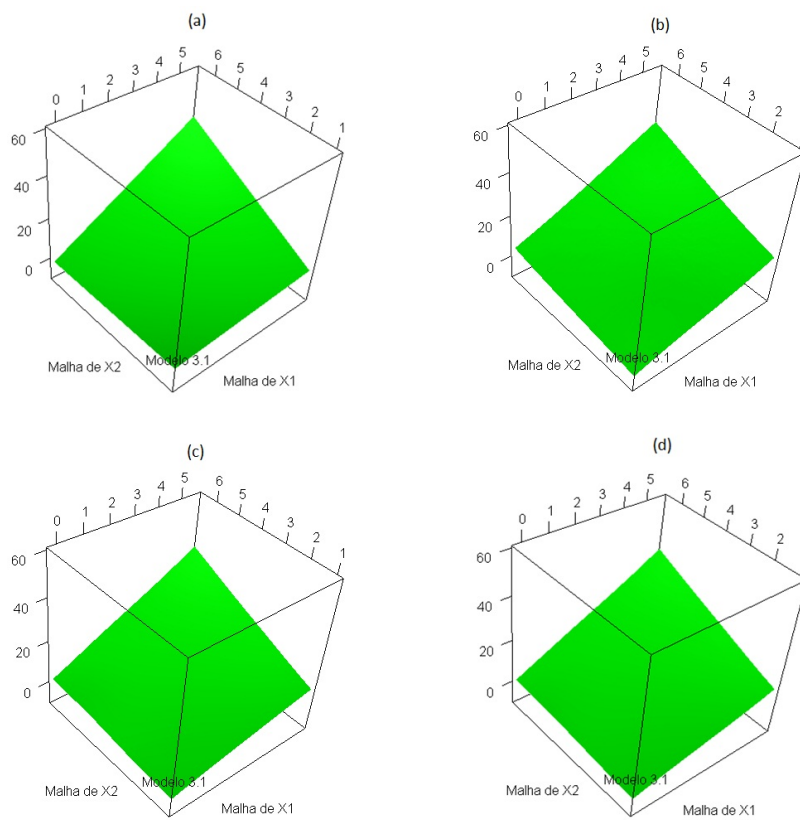


Figura 3.3: Estimação do modelo 3.1 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.

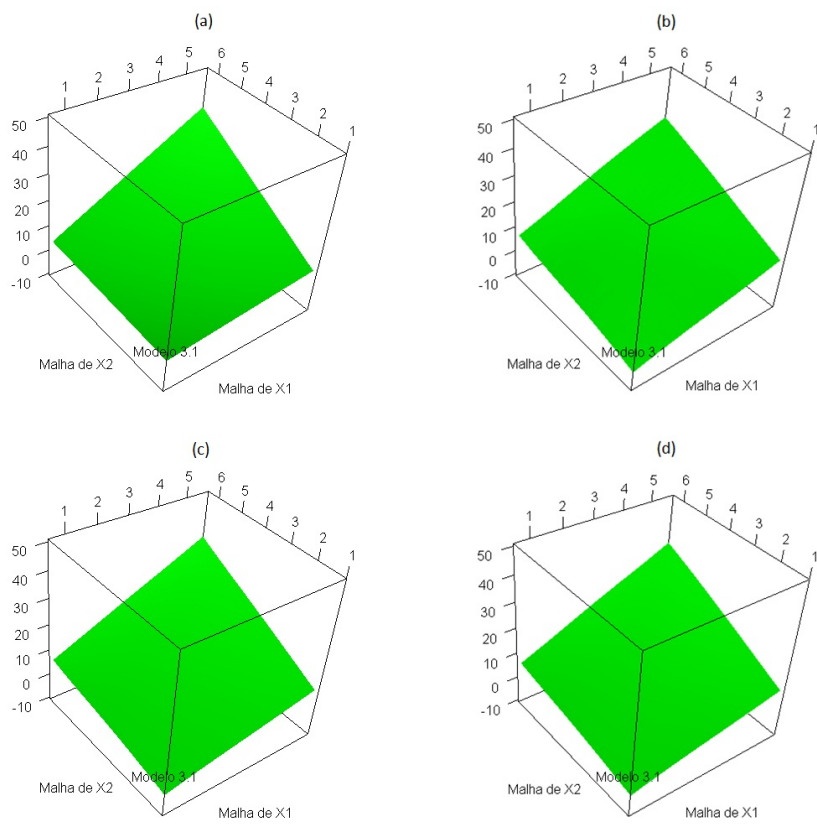


Figura 3.4: Estimação do modelo 3.1 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local.

3.2 Estimação da variância condicional

Para avaliação da estimação da variância condicional também foi utilizado o EQMI. A Tabela 3.2 mostra os EQMI's dos 4 modelos nas 4 situações. Pode ser observado que os menores EQMI's ocorreram na situação 1, e os maiores EQMI's ocorreram na situação 2.

Tabela 3.2: Valores médios dos EQMI's da estimação da variância condicional para os 4 modelos nas 4 situações

Modelos	Situação 1	Situação 2	Situação 3	Situação 4
Modelo 3.1	0,0014	0,7293	0,0773	0,0900
Modelo 3.2	0,0012	1,3507	0,0809	0,0672
Modelo 3.3	0,0048	1,7280	0,0756	0,0696
Modelo 3.4	0,0051	1,2330	0,1223	0,0767

As figuras 3.5, 3.6, 3.7 e 3.8 mostram a dispersão do R^2 estimado *vs* a variância condicional estimada para os quatro modelos nas quatro situações. Nos gráficos estão desenhadas a curva de regressão não paramétrica estimada pelo método LOWESS (Cleveland, 1979) afim de ilustrar como é dada a relação entre esses dois estimadores. Para todos modelos pode-se observar uma tendência de decréscimo do R^2 à medida que aumenta a variância na maioria das situações. Pode ser visto ainda que os modelos 3.1 e 3.2 apresentam R^2 locais maiores, em média, do que os modelos 3.3 e 3.4.

Os gráficos com as superfícies teóricas e estimadas de variância para os 4 modelos nas 4 situações estão apresentados no Apêndice C. As aproximações foram razoáveis à curva teórica, com problemas de fronteira aparecendo somente no modelo 3.4.

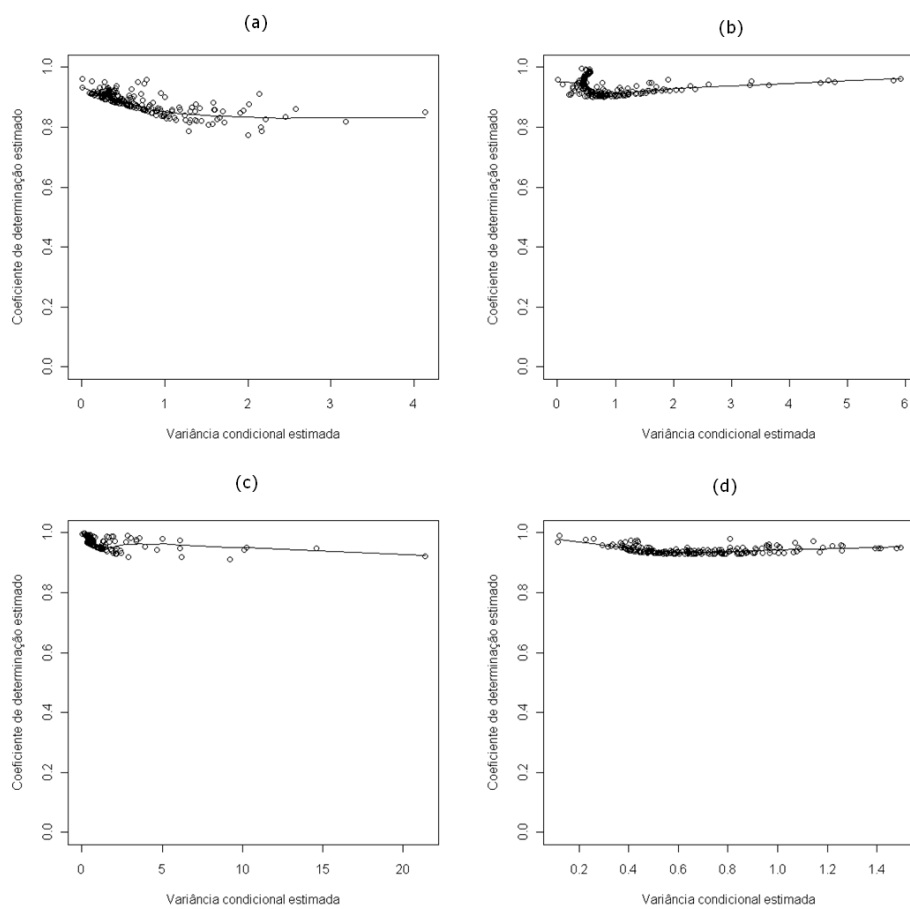


Figura 3.5: Diagrama de dispersão R^2 estimado *vs* variância condicional estimada para o Modelo 3.1 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.

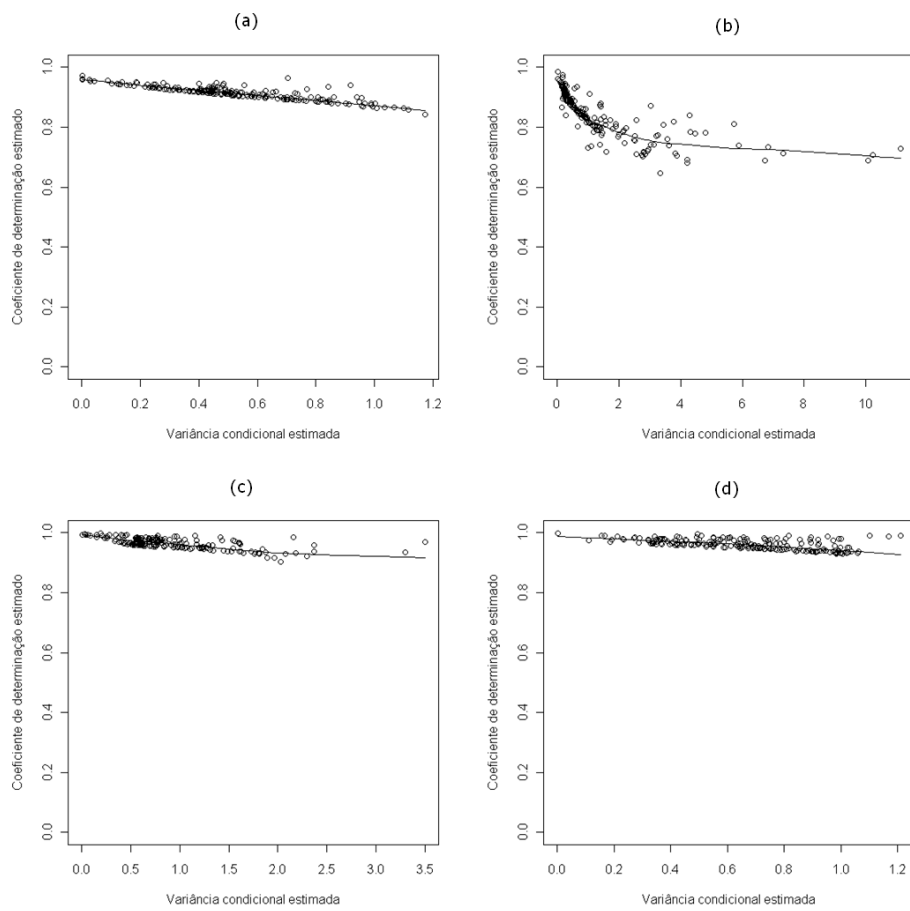


Figura 3.6: Diagrama de dispersão R^2 estimado *vs* variância condicional estimada para o Modelo 3.2 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.

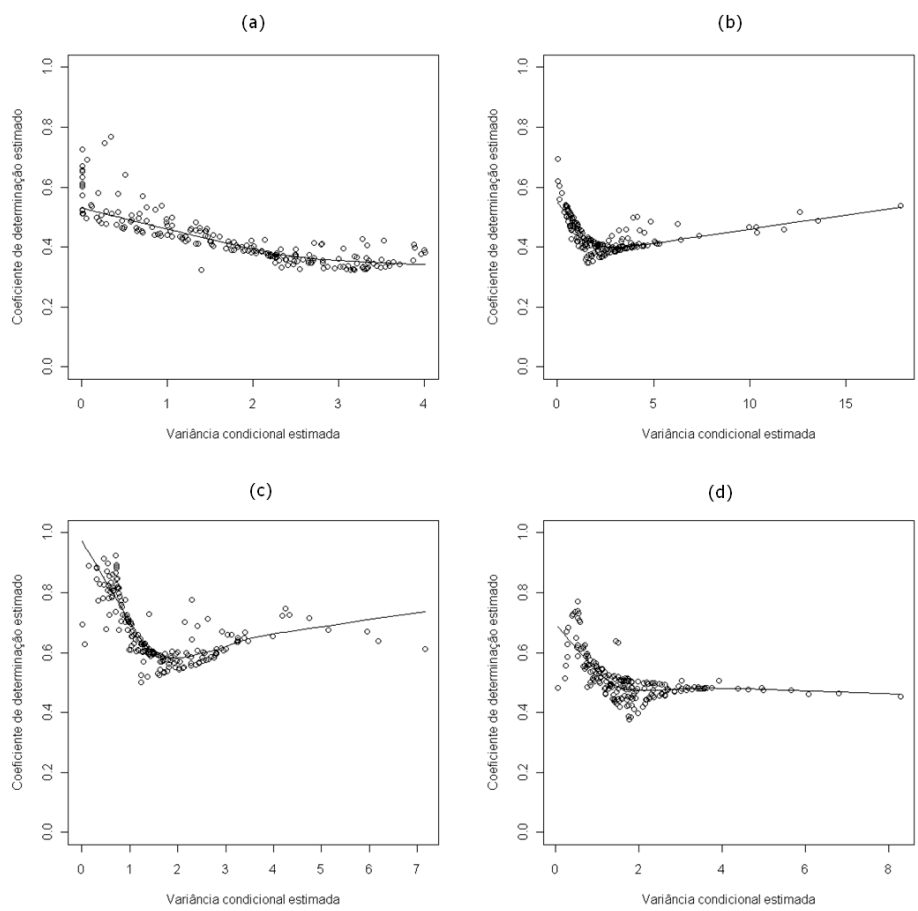


Figura 3.7: Diagrama de dispersão R^2 estimado *vs* variância condicional estimada para o Modelo 3.3 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.

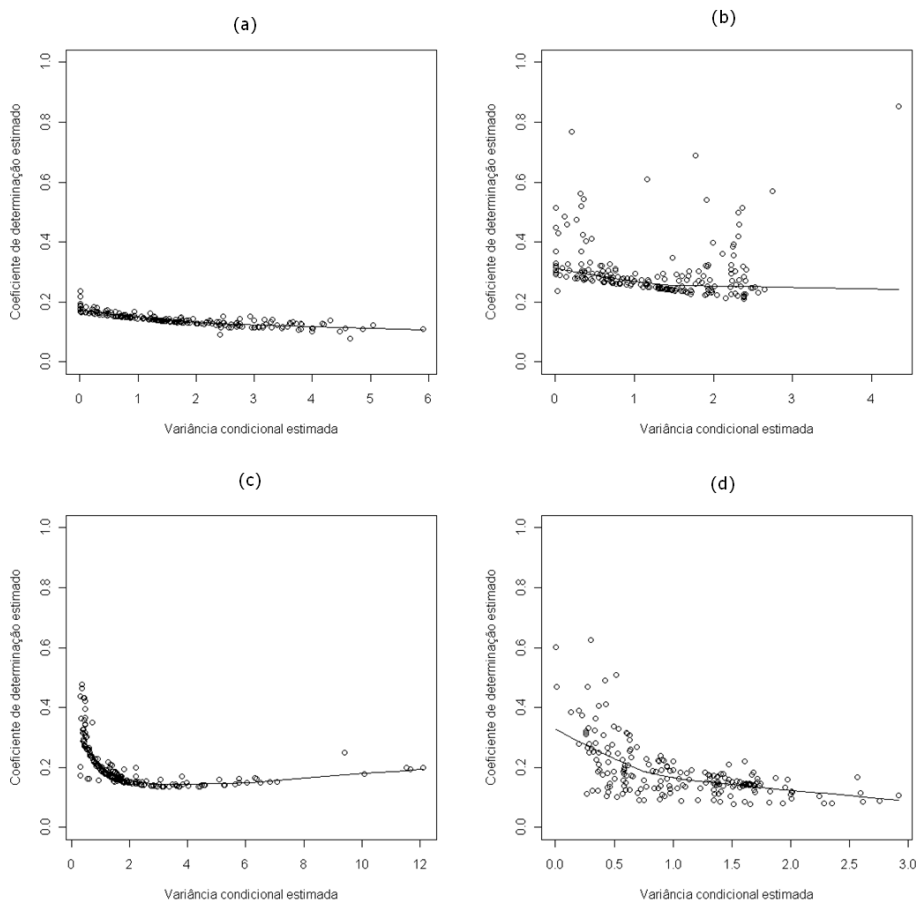


Figura 3.8: Diagrama de dispersão R^2 estimado *vs* variância condicional estimada para o Modelo 3.4 segundo as 4 situações: (a) Situação 1; (b) Situação 2; (c) Situação 3 e (d) Situação 4.

Capítulo 4

Aplicações

4.1 Aplicação 1

A mortalidade por homicídios é crescente, um problema de saúde pública em países em desenvolvimento, afetando principalmente jovens do sexo masculino de 15 a 29 anos. Fato este observado nas principais metrópoles brasileiras. Como objetivo de investigar a relação entre o nível de desenvolvimento do município e a escolaridade com a taxa de mortalidade por homicídios, foi realizado um estudo ecológico com dados de municípios do Estado de Minas Gerais - Brasil com mais de 50.000 habitantes, do ano de 2006.

As variáveis utilizadas foram a taxa de mortalidade por homicídio da população de 15 a 29 anos (por cem mil habitantes), percentual de frequência ao ensino médio de jovens de 15 a 17 anos e o Índice Mineiro de Responsabilidade Social (IMRS). O IMRS é um indicador composto que expressa o nível de desenvolvimento dos municípios mineiros, avaliando a situação de cada um em nove dimensões: renda, saúde, educação, demografia, segurança pública, gestão, habitação e meio ambiente, cultura e desporto e lazer. Os dados foram obtidos do *software* Índice Mineiro de Responsabilidade Social - versão 2009 (Fundação João Pinheiro, 2009).

As estatísticas descritivas das variáveis mostram taxa média de mortalidade por homicídio na população de 15 a 29 anos de 35,95

(desvio-padrão 40,09) por cem mil habitantes. A taxa média de frequência (em percentual) ao ensino médio para jovens de 15 a 17 anos foi de 46,33 (desvio-padrão 7,76) e o IMRS médio foi de 0,65 (desvio-padrão 0,04) para os 67 municípios (Tabela 4.1). A taxa de mortalidade por homicídio apresenta grande dispersão, o que pode ser verificado pela diferença entre o máximo e o mínimo e pelo desvio-padrão. A taxa média de frequência ao ensino médio é inferior a 50% para jovens entre 15 e 17 anos, e a diferença entre os municípios parece não ser tão grande em relação a essa variável e em relação ao IMRS, que apresentou um desvio-padrão pequeno.

Tabela 4.1: Estatísticas descritivas - Aplicação 1

Variáveis	N	Mínimo	Mediana	Média	Desvio-padrão	Máximo
Tx de mortalidade	67	0,00	20,42	35,95	40,09	170,70
IMRS	67	0,55	0,66	0,65	0,04	0,74
Tx freq. ensino médio	67	31,24	47,27	46,33	7,76	62,30

Nos diagramas de dispersão das variáveis duas a duas (Figura 4.1) foi ajustada a curva de regressão não paramétrica do método LOWESS (Cleveland, 1979), afim de fornecer uma indicação do sentido das relações na análise exploratória. É possível verificar nos diagramas que o aumento do IMRS e da frequência ao ensino médio dos jovens de 15 a 17 anos diminuem a taxa de mortalidade por homicídio. A dispersão entre as duas variáveis preditoras evidencia a existência de correlação entre elas. O coeficiente de correlação de Pearson para as duas variáveis foi de 0,40 e a normalidade bivariada das duas não foi rejeitada (p-valor 0,1156) pelo teste feito através do pacote *energy* (Rizzo, 2008) do *software R* versão 2.7.2.

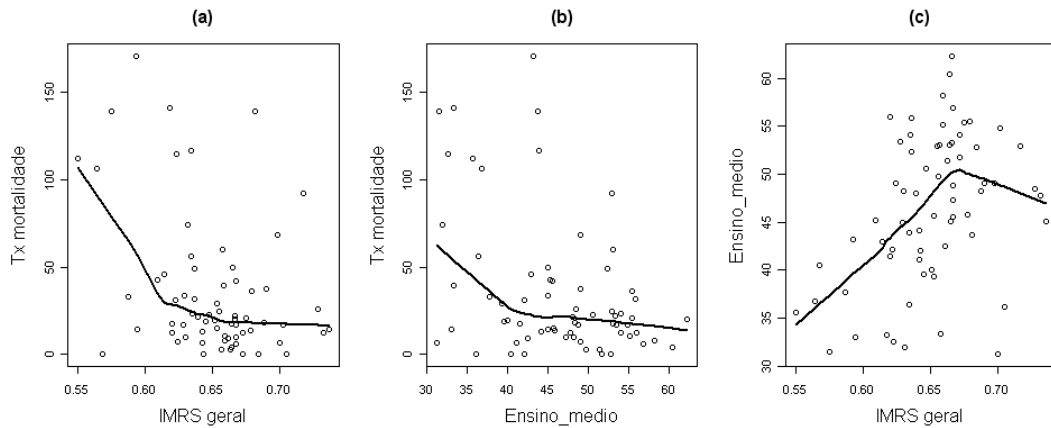


Figura 4.1: Diagramas de dispersão para as variáveis: (a) IMRS *vs* Tx de mortalidade; (b) Tx frequência ao ensino médio *vs* Tx de mortalidade e (c) IMRS *vs* Tx frequência ao ensino médio.

Diante da evidência de relação entre as variáveis predictoras com a variável resposta, foi ajustado o modelo paramétrico de regressão linear múltipla aos dados (Montgomery, 1992). No ajuste paramétrico (Tabela 4.2), as duas variáveis foram significativas (p -valor do teste *t-Student* $< 0,05$), apesar de o erro padrão dos coeficientes ter sido alto. O R^2 obtido no modelo paramétrico foi de 0,1849. A análise dos resíduos (Figura 4.2) revelou violação das suposições de normalidade e homoscedasticidade dos mesmos.

Tabela 4.2: Modelo paramétrico para a Aplicação 1

Parâmetros	Estimativas	Erro Padrão	P-valor
Intercepto	273,89	75,54	0,001
IMRS	-254,89	126,32	0,047
Tx freq. ensino médio	-1,56	0,63	0,016
$R^2 = 0,1849$			
Estatística F: 8,484 ($p = 0,001$)			

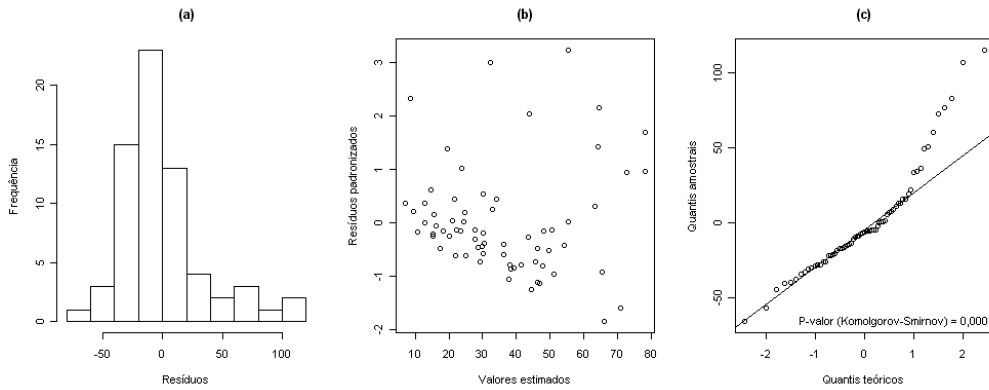


Figura 4.2: Gráficos de análise de resíduos do modelo paramétrico: (a) Histograma dos resíduos, (b) Dispersão entre os valores ajustados e os resíduos padronizados e (c) QQ-plot dos resíduos.

A seguir foi ajustado o modelo não paramétrico de regressão linear local com janela diagonal global aos dados. O R^2 obtido foi de 0,4205. O Gráfico (a) da Figura 4.3 mostra que, à medida que a variância aumenta, o R^2 diminui, sendo uma evidência de bom ajuste. O Gráfico (b) mostra a superfície do R^2 , onde pode ser visto que a região que abrange a taxa do ensino médio de 40 a 55 teve R^2 menor. Na Figura 4.1 (b), pode ser visto que a relação entre o percentual de frequência ao ensino médio e a taxa de mortalidade não é tão marcante nessa região. O Gráfico (c) mostra a superfície estimada, onde pode-se verificar pequena elevação na taxa de mortalidade na região os pontos não foram bem ajustados, segundo a análise do contorno da R^2 . O Gráfico da superfície do R^2 também oferece evidência de que a variância local não é constante nos pontos amostrados. E o Gráfico (d) mostra o contorno da superfície estimada.

A relação encontrada sugere que a mortalidade por homicídio em jovens está relacionada a menores níveis de educação e de desenvolvimento de um local. Dessa forma, as estratégias de enfrentamento da violência devem considerar investimentos no desenvolvimento da população, principalmente ações voltadas para a educação de seus cidadãos.

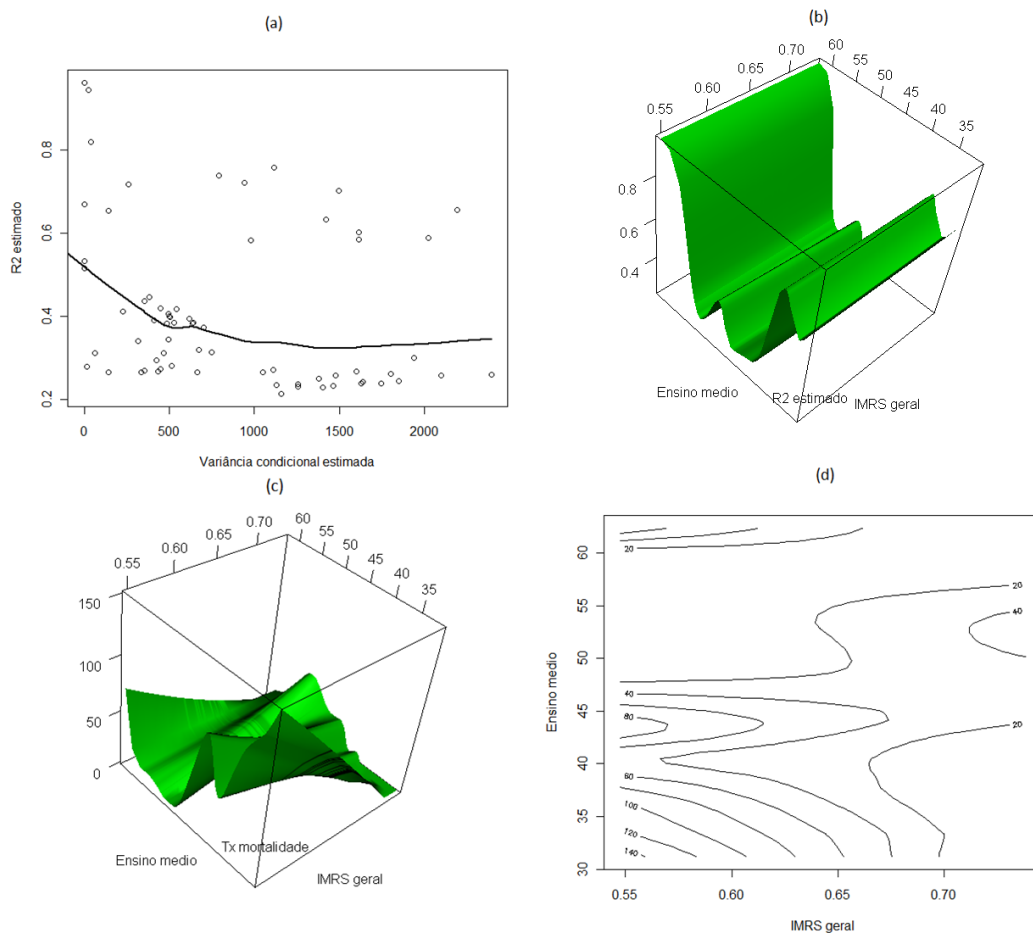


Figura 4.3: Gráficos de ajuste do modelo não paramétrico por janela diagonal global (a) Diagrama de dispersão R^2 vs variância condicional estimada; (b) Gráfico da superfície do R^2 local; (c) Gráfico da superfície ajustada e (d) Gráfico de contorno da superfície ajustada.

4.2 Aplicação 2

O ozônio funciona como um indicador da presença de oxidantes fotoquímicos na atmosfera, por ser o principal produto da reação ocorrida na mistura dos mesmos na presença de luz solar. Estudos recentes mostram que a poluição por ozônio tem forte relação com fatores climáticos (Confalonieri, 2008). Com o objetivo de investigar a relação entre a concentração de ozônio com a temperatura e a velocidade do vento, foi realizado um estudo que envolveu três variáveis: concentração de ozônio (em miligramas por metro cúbico), temperatura (em graus Celsius) e velocidade do vento (em metros por segundo). As unidades amostrais correspondem a valores médios diários das medidas. Os dados utilizados foram obtidos da Companhia Ambiental do Estado de São Paulo-Brasil e referem-se à região de Pinheiros, na Capital do Estado, no ano de 2008.

As estatísticas descritivas das variáveis (Tabela 4.3) mostram que, para as 366 observações, a concentração média diária de ozônio foi de 17,65 (desvio-padrão 12,04). A temperatura média foi de 20,00 (desvio-padrão 2,91), com pouca variação e a velocidade média do vento foi 1,26 (desvio-padrão 0,56).

Tabela 4.3: Estatísticas descritivas - Aplicação 2

Variáveis	N	Mínimo	Mediana	Média	Desvio-padrão	Máximo
Ozônio	366	0,00	15,11	17,65	12,04	62,26
Temperatura	366	12,25	20,06	20,00	2,91	26,19
Velocidade do vento	366	0,00	1,26	1,26	0,56	2,65

O diagrama de dispersão das variáveis duas a duas (Figura 4.4) mostra relação positiva entre a temperatura e a concentração de ozônio, assim como para velocidade do vento e ozônio. O aumento da temperatura leva a diminuição da velocidade do vento. O coeficiente de correlação linear de Pearson entre as duas covariáveis é $-0,1447$ e a normalidade bivariada das mesmas foi rejeitada.

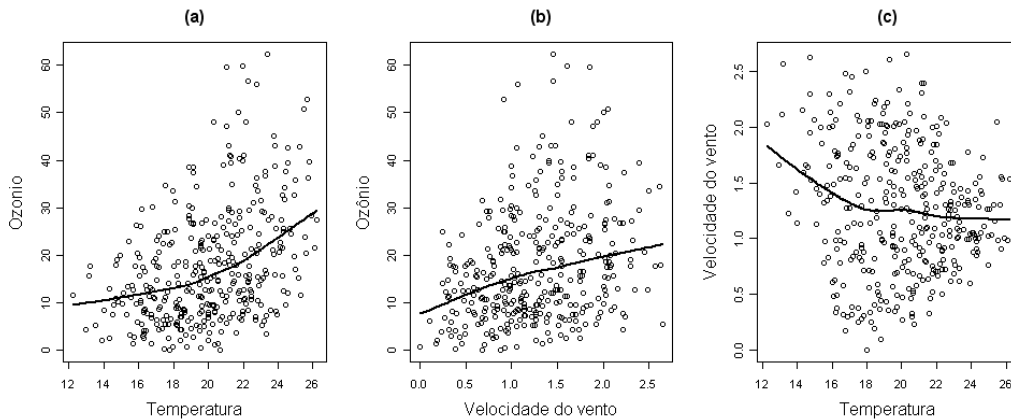


Figura 4.4: Diagramas de dispersão para as variáveis: (a) Temperatura *vs* Ozônio; (b) Velocidade do vento *vs* Ozônio e (c) Temperatura *vs* Velocidade do vento

Em seguida foi ajustado o modelo paramétrico de regressão linear múltipla aos dados. Através desse ajuste (Tabela 4.4), pode-se ver que as duas variáveis são significativas (p -valor do teste *t-Student* $< 0,05$), e o R^2 obtido no modelo paramétrico foi de $0,2853$. A análise dos resíduos (Figura 4.5) revela violação das suposições de normalidade e homoscedasticidade dos resíduos do modelo paramétrico.

Tabela 4.4: Modelo paramétrico para a Aplicação 2

Parâmetros	Estimativas	Erro Padrão	P-valor
Intercepto	-29,89	4,09	0,000
Temperatura	1,90	0,18	0,000
Velocidade do vento	7,57	0,96	0,000
$R^2 = 0,2853$			
Estatística F: 73,86 ($p = 0,000$)			

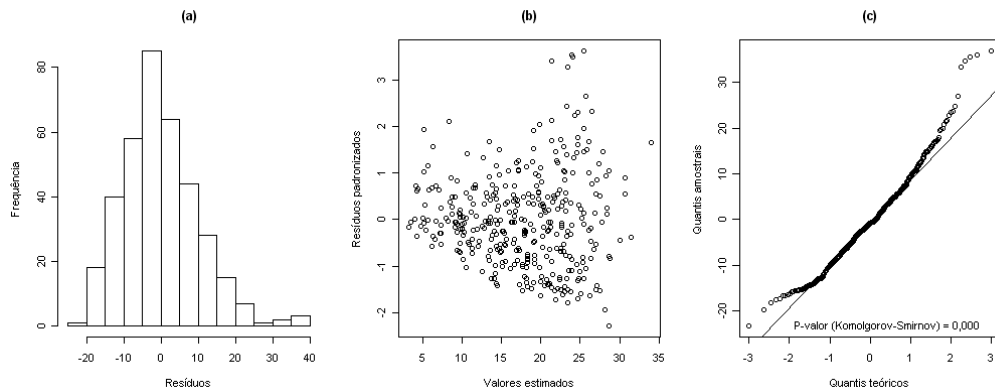


Figura 4.5: Gráficos de análise de resíduos do modelo paramétrico: (a) Histograma dos resíduos, (b) Dispersão entre os valores ajustados e os resíduos padronizados e (c) QQ-plot dos resíduos.

O ajuste não paramétrico de regressão linear local com janela diagonal global alcançou um R^2 global de 0,3772. O Gráfico (a) da Figura 4.6 mostra o decréscimo do R^2 com o aumento da variância, evidência de um bom ajuste. O Gráfico (b) mostra a superfície do R^2 , ilustrando que a variância local não é constante e os menores R^2 ocorreram na região onde a temperatura varia de 18 a 22 °C. Os Gráficos (c) e (d) mostram a superfície estimada e o contorno da mesma, respectivamente, onde verifica-se pequena elevação da concentração de ozônio com o aumento da temperatura e da velocidade do vento, sendo um aumento mais expressivo com o aumento da temperatura. Esse aspecto foi observado também nos dados utilizados por Bowman (2006).

Esses resultados sugerem que o aumento da temperatura média diária favorece a poluição por ozônio, gás extremamente tóxico quando presente na faixa de ar próxima do solo, e que provoca piora de doenças respiratórias, além de ser corrosivo e diminuir a vida útil dos materiais. A poluição por ozônio é, portanto, mais uma conseqüência a ser enfrentada pelo mundo com o aquecimento global.

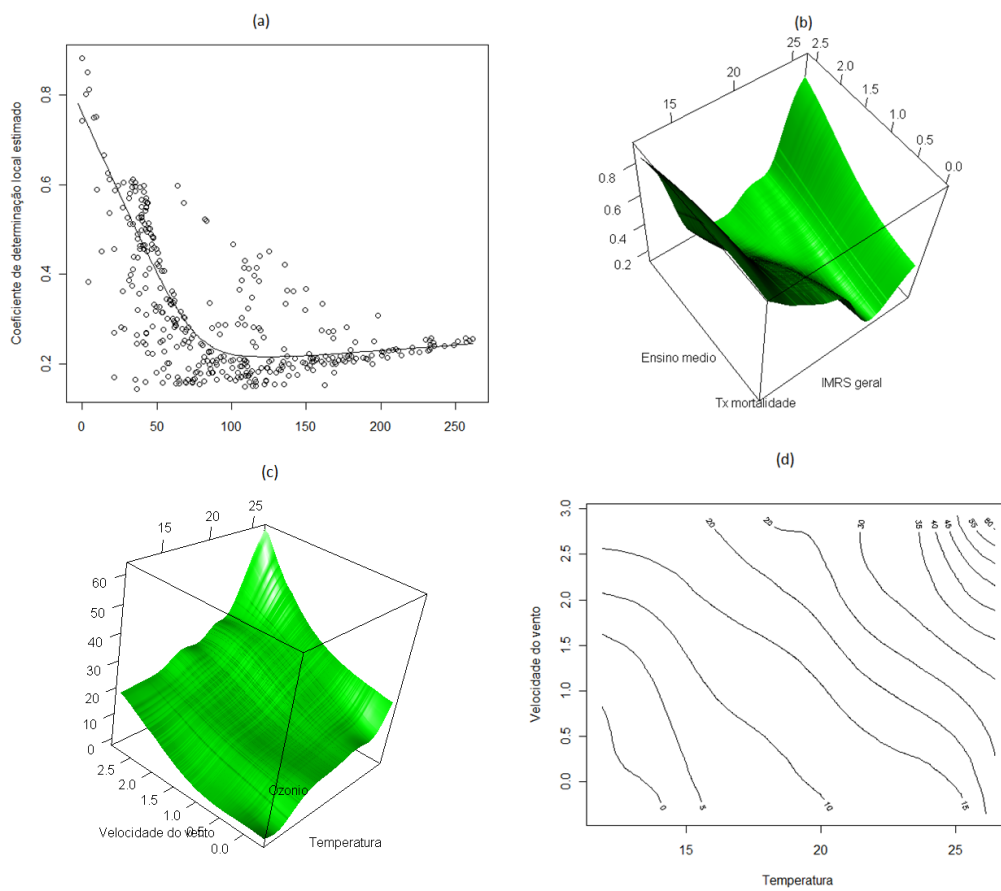


Figura 4.6: Gráficos de ajuste do modelo não paramétrico por janela diagonal global (a) Diagrama de dispersão R^2 vs variância condicional estimada; (b) Gráfico de superfície do R^2 local; (c) Gráfico da superfície ajustada e (d) Gráfico de contorno da superfície ajustada.

Capítulo 5

Conclusões

O aumento do número de variáveis (dimensões) no modelo de regressão não paramétrica dificulta o processo de estimação local em termos de dimensões de matrizes, demonstrações matemáticas, vizinhanças vazias e etc. Diante do desafio que é estudar o caso multivariado, optamos por, nesse trabalho, tratar alguns aspectos da estimação no caso bivariado, onde é possível visualizar curvas e as fórmulas são mais tratáveis.

Para a estimação linear local, foi utilizada o método dos mínimos quadrados ponderados (Ruppert e Wand, 1994), onde os pesos seriam dados pela função núcleo e seriam dependentes da janela. A escolha da janela é um dos pontos mais importantes da estimação. Para o caso multivariado não há tantas opções de métodos na literatura como tem para o caso univariado, porque nesse caso é necessário estimar uma matriz de janelas, e não mais apenas um número real. Neste texto foram apresentados os métodos de estimação por janela escalar global, janela diagonal global e janela escalar local.

Apesar de o terceiro método permitir o ajuste local, assume a mesma janela para as duas covariáveis, o que pode não ser adequado para situações em que as variâncias das mesmas sejam diferentes. Os resultados das simulações mostraram que, nas situações em que as covariáveis eram independentes, a janela diagonal global obteve menor EQMI. A estimação via janela diagonal global para situações em que há correlação entre as preditoras não é a mais adequada,

porém, não teve um resultado ruim.

O ajuste da janela escalar local depende da estimação de algumas quantidades, como a variância condicional, a matriz Hessiana e a densidade bivariada das covariáveis. A estimação da variância condicional local permite verificar se a variabilidade local é a mesma em todos os pontos (homocedasticidade). A proposta de estimação para a matriz Hessiana baseou-se na intuição para o caso univariado proposto por Fan e Gijbels (1995) e considerou a informação de que a diferença entre a ordem da derivada e o grau do polinômio deve ser ímpar (Ruppert e Wand, 1994). Assim, foi proposto um estimador baseado na expansão em série de Taylor de terceira ordem, porém, os resultados obtidos nas simulações considerando o EQMI dos determinantes das Hessianas teóricas e estimadas não foi bom.

Há na literatura um método proposto por Fan e Gijbels (1996) para estimar matriz de janelas locais, o qual depende das estimações das mesmas quantidades desconhecidas de que necessita o método da janela escalar local. O resultado da estimação da Hessiana comprometeu a aplicação desse método, de forma que não o apresentamos nesse texto.

Nas aplicações em dados reais, para o primeiro exemplo envolvendo taxa de mortalidade por homicídio em jovens de 15 a 29 anos, apesar de haver correlação entre as variáveis preditoras e ter sido aplicado o método de janela diagonal global, os resultados foram interessantes e as relações encontradas são parecidas com o que foi observado em estudos anteriores (Macedo et. al. 2001 e Soares Filho et. al. 2007). Houve aumento significativo no R^2 do modelo paramétrico para o não paramétrico, e a análise da variância local dá evidências contra a suposição da homocedasticidade necessária para a aplicação do modelo paramétrico.

O segundo exemplo de aplicação, que envolveu a concentração média diária de ozônio, apresentou também resultados parecidos com os apresentados no estudo de Bowman (2006). Para este exemplo, como a correlação entre as variáveis era baixa, provavelmente a estimação foi mais satisfatória do que para o exemplo anterior. Houve também aumento do R^2 em relação ao modelo paramétrico e a estimação da variância condicional forneceu evidências contra a homocedasticidade.

A utilização de modelos de regressão não paramétrica representam uma alternativa importante para situações em que a aplicação de modelos paramétricos não são adequadas. Em trabalhos posteriores, pretendemos estudar outros métodos de estimação da matriz Hessiana e da variância condicional, estender as análises para situações com mais de 2 covariáveis, trabalhar a comparação de superfícies e estudar técnicas de avaliação da qualidade do ajuste.

Bibliografia

- [1] Adler, D.; Murdoch, D. rgl: 3D visualization device system (OpenGL). R package version 0.81 [Internet]. 2008 - [citado em 10/01/2010]. Disponível em: <http://rgl.neoscientists.org>.
- [2] Cai, T.T.; Levine, M.; Wang, L. Variance function estimation in multivariate nonparametric regression with fixed design. *J Multivar Anal.* 2009; 100 (1): 126-136.
- [3] Campos Filho, F.F. Algoritmos Numéricos. 2^a Ed. Rio de Janeiro: LTC; 2007. 428 p.
- [4] Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc.* 1979; 74(368): 829-836.
- [5] Confalonieri, U.E.C. Mudança climática global e saúde humana no Brasil. *Parc Estrat.* 2008; 27: 323-349.
- [6] Fan, J.; Gijbels, I. Data-Driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J R Stat Soc Series B Stat Methodol.* 1995; 57(2): 371-394.
- [7] Fan, J.; Gijbels, I. Local Polynomial Modelling and its Applications. Londres: CHAPMAN & HALL; 1996. 341 p.
- [8] Fan, J.; Yao, Q. Efficient estimation of conditional variance function in stochastic regression. *Biometrika.* 1998; 85(3): 645-660.
- [9] Gentle, J.E. Matrix Algebra. Nova Yorque: Springer; 2007. 528 p.

- [10] Índice Mineiro de Responsabilidade Social 2009 [Internet]. Belo Horizonte (MG): Fundação João Pinheiro; 2009. Disponível em: [http : //www.datagerais.mg.gov.br/site/int_imrs.php](http://www.datagerais.mg.gov.br/site/int_imrs.php).
- [11] Huang, L.S.; Chen, J. Analysis of variance, coefficient of determination and F-test for local polynomial regression. *Ann Stat.* 2008; 36(5): 2085-2109.
- [12] James, B.R. Probabilidade: Um curso em nível intermediário. Rio de Janeiro: IMPA; 2002. 316 p.
- [13] Komarek, A. A new R package for Bayesian estimation of multivariate normal mixtures allowing for selection of number of components and interval-censored data. *Comput Stat Data Anal.* 2009; 53: 3932-3947.
- [14] Lehmann, E.L.; Casella, G. *Theory of Point Estimation*. Nova Yorque: Springer; 1998. 589 p.
- [15] Lima, M.S.; Atuncar, G.S. A Bayesian method to estimate the optimal bandwidth for multivariate kernel estimator. *J Nonparametr Stat.* Submetido. 2009.
- [16] Macedo, A.C.; Paim, J.S.; Silva, L.M.V.; Costa, M.C.N. Violência e desigualdade social: mortalidade por homicídios e condições de vida em Salvador, Brasil. *Rev Saude Publica.* 2001; 35(6): 515-522.
- [17] Marquetti, A.; Viali, L. Princípios e aplicações de regressão local. *Rev Anal Econ.* 2004; 42: 1-26.
- [18] Montgomery, D.C.; Peck, E.A. *Introduction to Linear Regression Analysis*. Nova Yorque: Wiley; 1992. 527 p.
- [19] Opsomer, J.D.; Ruppert, D. Fitting a bivariate additive model by local polynomial regression. *Ann Stat.* 1997; 25(1): 186-211.
- [20] R Development Core Team. *R: A language and environment for statistical computing* [Internet]. Viena; R Foundation for

- Statistical Computing: 2008. Disponível em: <http://www.R-project.org>.
- [21] Rizzo, M.L.; Szekely, G.J. energy: E-statistics (energy statistics) [Internet]. 2008. R package version 1.1-0.
- [22] Ruppert, D.; Wand, M.P. Multivariate locally weighted least squares regression. *Ann Stat.* 1994; 22(3): 1346-1370.
- [23] Ruppert, D.; Wand, M.P.; Holst, U.; Hossjer, O. Local Polynomial Variance-Function Estimation. *Technometrics.* 1997; 39(3): 262-273.
- [24] Soares Filho, A.M.; Alencar, A.P.; Souza, M.F.M.; Silva, M.M.A.; Gazal-Carvalho, C.; Morais Neto, O.L.; Malta, D.C. Análise da Mortalidade por homicídios no Brasil. *Epidemiol Serv Saude.* 2007; 16(1): 7-18.
- [25] Wand, M.P.; Jones, M.C. Kernel Smoothing. Londres: CHAPMAN & HALL; 1994. 212 p.
- [26] Yang, L.; Tschernig, R. Multivariate bandwidth selection for local linear regression. *J R Stat Soc Series B Stat Methodol.* 1999; 61(4): 793-815.
- [27] Ye, A.; Hyndman, R.J.; Li, Z. Local linear multivariate regression with variable bandwidth in the presence of heteroscedasticity. Forthcoming. [citado em 20/02/2009]:[25 p.]. Manuscrito dos autores disponível em <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2006/wp8-06.pdf>.
- [28] Zhang X, King M, Hyndman R. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Comput Stat Data Anal.* 2006; 50(11): 3009-3031.

Apêndice A

Demonstração do Teorema da Seção 2.4

Algumas lacunas na demonstração foram preenchidas durante o trabalho. Para facilitar a notação, consideremos

$(X_i - x)^T = (X_{i1} - x_1, X_{i2} - x_2)$, para $i = 1, \dots, n$

e ainda $x = (x_1, x_2)$. A esperança da função \hat{m} pode ser vista como:

$$E \{ \hat{m}(x; B) \} = e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x M \quad (\text{A.1})$$

onde $M = [m(X_1), \dots, m(X_n)]^T$. Seja $Q_m(x)$ um vetor de dimensão $n \times 1$ dado por:

$$Q_m(x) = [(X_1 - x)^T \mathcal{H}_m(x)(X_1, -x), \dots, (X_n - x)^T \mathcal{H}_m(x)(X_n, -x)]^T \quad (\text{A.2})$$

O Teorema de Taylor implica que:

$$M = X_x \begin{bmatrix} m(x) \\ D_m(x) \end{bmatrix} + \frac{1}{2} Q_m(x) + R_m(x) \quad (\text{A.3})$$

onde $R_m(x)$ é um vetor de termos remanescentes da série de Taylor. Quando $R_m(x)$ é multiplicado pela esquerda por $e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x$, o resultado é um escalar de ordem desprezível comparado ao termo decorrente de $Q_m(x)$. Desde que $Q_m(x)$ seja diferente de zero, em todos os casos esse escalar é um termo de ordem menor, em probabilidade, que $\text{tr}(B)$. Então, das expressões (A.1) e (A.2),

$$E \{ \hat{m}(x; B) - m(x) | X_1, \dots, X_n \} = \quad (\text{A.4})$$

$$\frac{1}{2} e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x \{ Q_m(x) + R_m(x) \}$$

Observe que o termo $D_m(x)$ proveniente na expressão (A.3) desaparece na expressão acima, em consequência da multiplicação pelo vetor e_1^T .

A seguir, a multiplicação matricial $n^{-1} (X_x^T W_x X_x)$:

$$n^{-1} (X_x^T W_x X_x) = \begin{bmatrix} a_{11}(X_i - x) & a_{12}(X_i - x) \\ a_{21}(X_i - x) & a_{22}(X_i - x) \end{bmatrix}$$

Onde:

$$a_{11}(X_i - x) = n^{-1} \sum_{i=1}^n K_B(X_i - x) \quad (\text{A.5})$$

$$a_{12}(X_i - x) = n^{-1} \sum_{i=1}^n K_B(X_i - x) (X_i - x)^T \quad (\text{A.6})$$

$$a_{21}(X_i - x) = n^{-1} \sum_{i=1}^n K_B(X_i - x) (X_i - x) \quad (\text{A.7})$$

$$a_{22}(X_i - x) = n^{-1} \sum_{i=1}^n K_B(X_i - x) (X_i - x) (X_i - x)^T \quad (\text{A.8})$$

Ao analisar o termo $a_{11}(X_i - x)$ e usando alguns resultados padrões da estimação de densidade (Wand e Jones, 1994), pode-se provar que:

$$n^{-1} \sum_{i=1}^n K_B(X_i - x) = f(x) + o_p(1) \quad (\text{A.9})$$

O resultado acima é proveniente da própria definição de função densidade estimada através da função núcleo. Em relação ao termo $a_{12}(X_i - x) = a_{21}(X_i - x)$, prova-se que:

$$n^{-1} \sum_{i=1}^n K_B(X_i - x)(X_i - x) = \mu_2(K)BD_f(x) + o_p(B1) \quad (\text{A.10})$$

onde 1 é um vetor 2×1 com as entradas iguais a 1 . Para demonstrar esse resultado, utiliza-se o Teorema Multivariado de Taylor:

Teorema Multivariado de Taylor: Seja g uma função d -variada e α_n uma seqüência de vetores $d \times 1$ cujos componentes tendem a zero. Além disso, seja $D_g(x)$ o vetor de derivadas parciais de primeira ordem de g e $\mathcal{H}_g(x)$ a matriz Hessiana de g , uma matriz $d \times d$ com as entradas (i, j) dadas por $\frac{\partial^2 g(x)}{\partial x_i \partial x_j}$ (ver em Wand e Jones, 1994). Então, assumindo que todas as entradas de $\mathcal{H}_g(x)$ são contínuas em uma vizinhança de x , tem-se que :

$$g(x + \alpha_n) = g(x) + \alpha_n^T D_g(x) + 1/2 \alpha_n^T \mathcal{H}_m(x) \alpha_n + o(\alpha_n^T \alpha_n) \quad (\text{A.11})$$

Da expressão (A.10), pode-se ver que:

$$n^{-1} \sum_{i=1}^n K_B(X_i - x)(X_i - x) = \quad (\text{A.12})$$

$$n^{-1} \sum_{i=1}^n |B|^{-1/2} K(B^{-1/2}(X_i - x))(X_i - x)$$

Substituindo $(X_i - x) = B^{1/2}u$, a Fórmula (A.12) pode ser vista como uma esperança, e assim, $\int_{\mathbb{R}^2} K(u)B^{1/2}uf(x + B^{1/2}u)du$. Então, aplica-se o Teorema Multivariado de Taylor em $f(x + B^{1/2}u)$ e, considerando

que a soma de duas funções de mesma ordem gera uma função que conserva a ordem das funções somadas, o resultado da Fórmula (A.10) é demonstrado.

Ao analisar o termo $a_{22}(X_i - x)$, prova-se que:

$$n^{-1} \sum_{i=1}^n K_B(X_i - x)(X_i - x)(X_i - x)^T = \mu_2(K)f(x)B + o_p(B) \quad (\text{A.13})$$

A demonstração desse resultado é similar à demonstração do resultado anterior. Basta aplicar a definição do núcleo, fazer a substituição, montar a fórmula da esperança, aplicar o Teorema de Taylor e usar a propriedade das funções de ordem.

Então, utilizando a teoria de inversão de matriz por blocos (Gentle, 2007), segue que $(n^{-1}(X_x^T W_x X_x))^{-1}$ é igual a:

$$\begin{bmatrix} f(x)^{-1} + o_p(1) & -D_f(x)^T f(x)^{-2} + o_p(1) \\ -D_f(x)^T f(x)^{-2} + o_p(1) & \mu_2(K)f(x)B^{-1} + o_p(B) \end{bmatrix}$$

E fazendo as multiplicações, mostra-se que $n^{-1}X_x^T W_x Q_m(x)$ é igual a:

$$\begin{bmatrix} n^{-1} \sum_{i=1}^n K_B(X_i - x)(X_i - x)^T \mathcal{H}_m(x)(X_i - x) \\ n^{-1} \sum_{i=1}^n \{K_B(X_i - x)(X_i - x)^T \mathcal{H}_m(x)(X_i - x)\} (X_i - x) \end{bmatrix}$$

Assim, da fórmula (A.4), podemos ver que a estimativa do vício esperado do estimador \hat{m} é dada por:

$$\frac{1}{2}f(x)^{-1}E \left[n^{-1} \sum_{i=1}^n K_B(X_i - x)(X_i - x)^T \mathcal{H}_m(x)(X_i - x) \right] \quad (\text{A.14})$$

$$-\frac{1}{2}D_f(x)^T f(x)^{-2}E \left[n^{-1} \sum_{i=1}^n K_B(X_i - x)(X_i - x)^T \mathcal{H}_m(x)(X_i - x)(X_i - x) \right]$$

Porém, pode-se mostrar que a segunda parte da soma é um $o_p\{tr(B)\}$ usando a mesma estratégia de demonstração da Fórmula (A.10) Assim,

$$E \{ \hat{m}(x; B) - m(x) | X_1, \dots, X_n \} = \quad (\text{A.15})$$

$$\frac{1}{2} f(x)^{-1} E \left\{ n^{-1} \sum_{i=1}^n K_B(X_i - x) (X_i - x)^T \mathcal{H}_m(x) (X_i - x) \right\} + o_p(\text{tr}(B))$$

Fazendo a substituição $(X_i - x) = B^{1/2}u$:

$$= \frac{1}{2} f(x)^{-1} \int K(u) (B^{1/2}u)^T \mathcal{H}_m(x) (B^{1/2}u) f(x + B^{1/2}u) du + o_p(\text{tr}(B)) \quad (\text{A.16})$$

E usando que o traço de uma constante é igual a própria constante, tem-se que:

$$= \frac{1}{2} \text{tr} \left\{ B^{1/2} \mathcal{H}_m(x) B^{1/2} \int K(u) u u^T du \right\} + o_p(\text{tr}(B)) \quad (\text{A.17})$$

$$= \frac{1}{2} \mu_2(K) \text{tr}(B \mathcal{H}_m(x)) + o_p(\text{tr}(B)) \quad (\text{A.18})$$

A qual corresponde à Fórmula 2.20 do Teorema da Seção 2.4. Para a demonstração da expressão da variância do estimador da função m , seja $V = \text{diag} \{ v(X_1), \dots, v(X_n) \}$. Então:

$$\text{Var} \{ \hat{m}(x; B) | X_1, \dots, X_n \} = \quad (\text{A.19})$$

$$e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x V W_x X_x (X_x^T W_x X_x)^{-1} e_1$$

O resultado de $X_x^T W_x V W_x X_x$ será uma matriz de dimensão 2×2 , cuja entrada superior esquerda é dada por:

$$n^{-1} \sum_{i=1}^n K_B(X_i - x)^2 v(X_i) \quad (\text{A.20})$$

Usando o núcleo definido na Fórmula 2.6, fazendo a substituição $(X_i - x) = B^{1/2}u$ e tratando a fórmula acima como uma esperança, a expressão torna-se:

$$\begin{aligned} &= |B|^{1/2} \int K^2(u)v(x + B^{1/2}u)f(x + B^{1/2}u)du \{1 + o_p(1)\} \quad (\text{A.21}) \\ &= |B|^{1/2} R(K)v(x)f(x) \{1 + o_p(1)\} \end{aligned}$$

A entrada superior direita é dada por:

$$n^{-1} \sum_{i=1}^n K_B(X_i - x)^2 (X_i - x)^T v(X_i) \quad (\text{A.22})$$

E usando a mesma estratégia utilizada na expressão (A.20) para obter a expressão (A.21), é obtida uma expressão para (A.22) que é um $O_p(|B|^{1/2})$.

O termo inferior direito da matriz é:

$$n^{-1} \sum_{i=1}^n K_B(X_i - x)^2 (X_i - x)(X_i - x)^T v(X_i) \quad (\text{A.23})$$

O qual, através da mesma estratégia utilizada nas expressões anteriores, torna-se:

$$= |B|^{-1/2} B^{1/2} \left\{ \int K^2(u)uu^T du \right\} B^{1/2}v(x)f(x) + o_p(|B|^{-1/2}B) \quad (\text{A.24})$$

Então, fazendo as multiplicações e considerando as expressões (A.21) e de (A.22) a (A.24), tem-se que:

$$Var \{\hat{m}(x; B) | X_1, \dots, X_n\} = n^{-1} |B|^{-1/2} \{R(K)v(x)/f(x)\} \{1 + o_p(1)\} \quad (\text{A.25})$$

A qual corresponde a expressão 2.21 do Teorema.

Apêndice B

Funções de regressão teóricas e
estimadas, por modelos e
situações

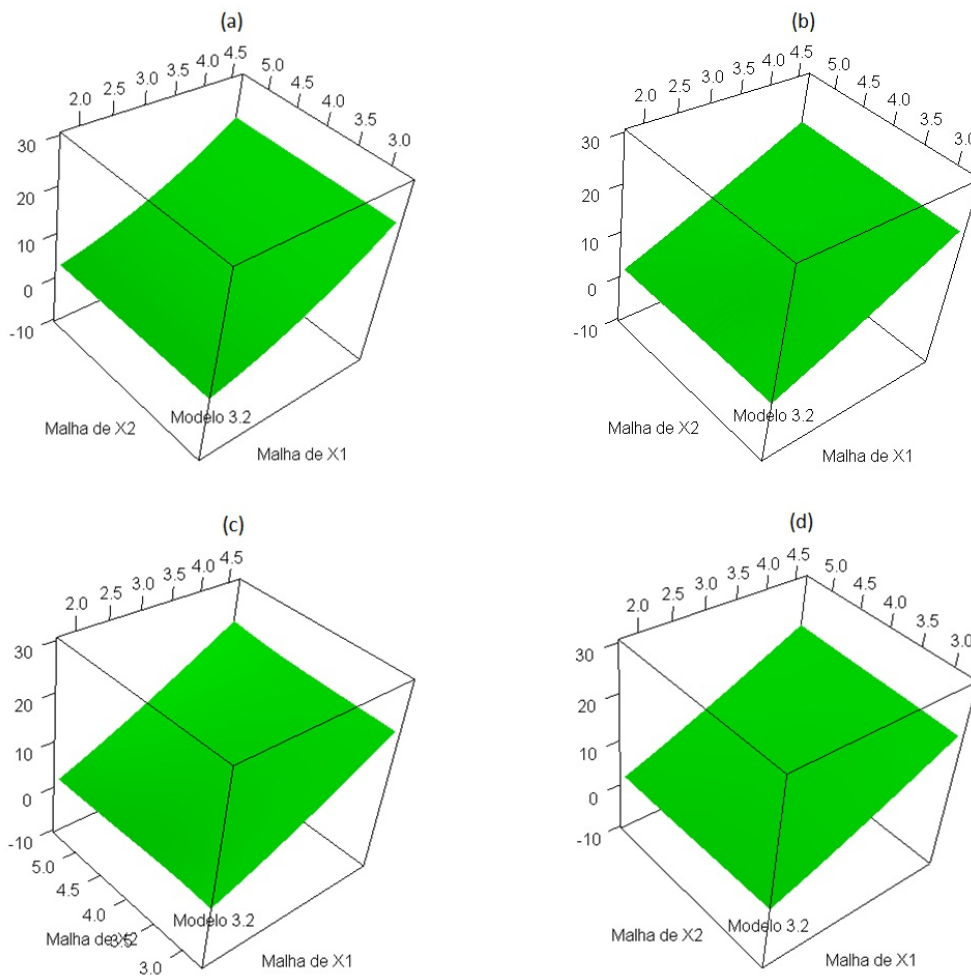


Figura B.1: Estimação do modelo 3.2 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

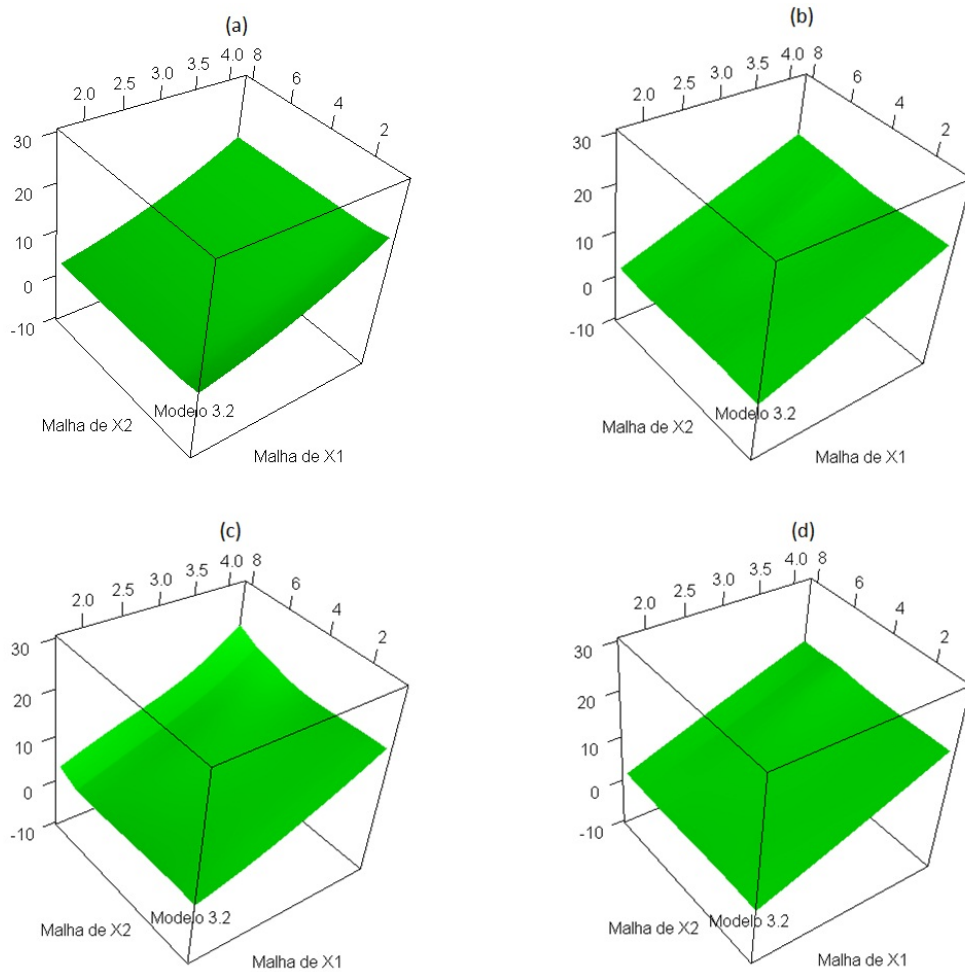


Figura B.2: Estimação do modelo 3.2 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

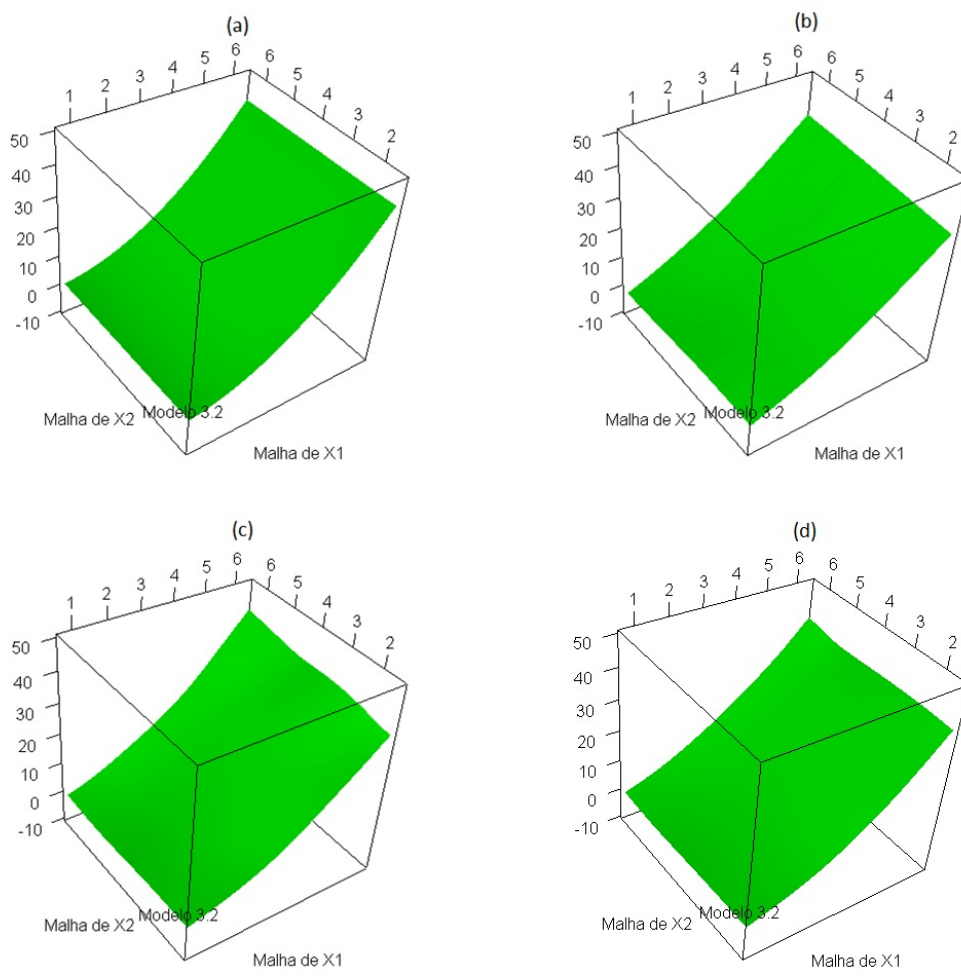


Figura B.3: Estimação do modelo 3.2 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

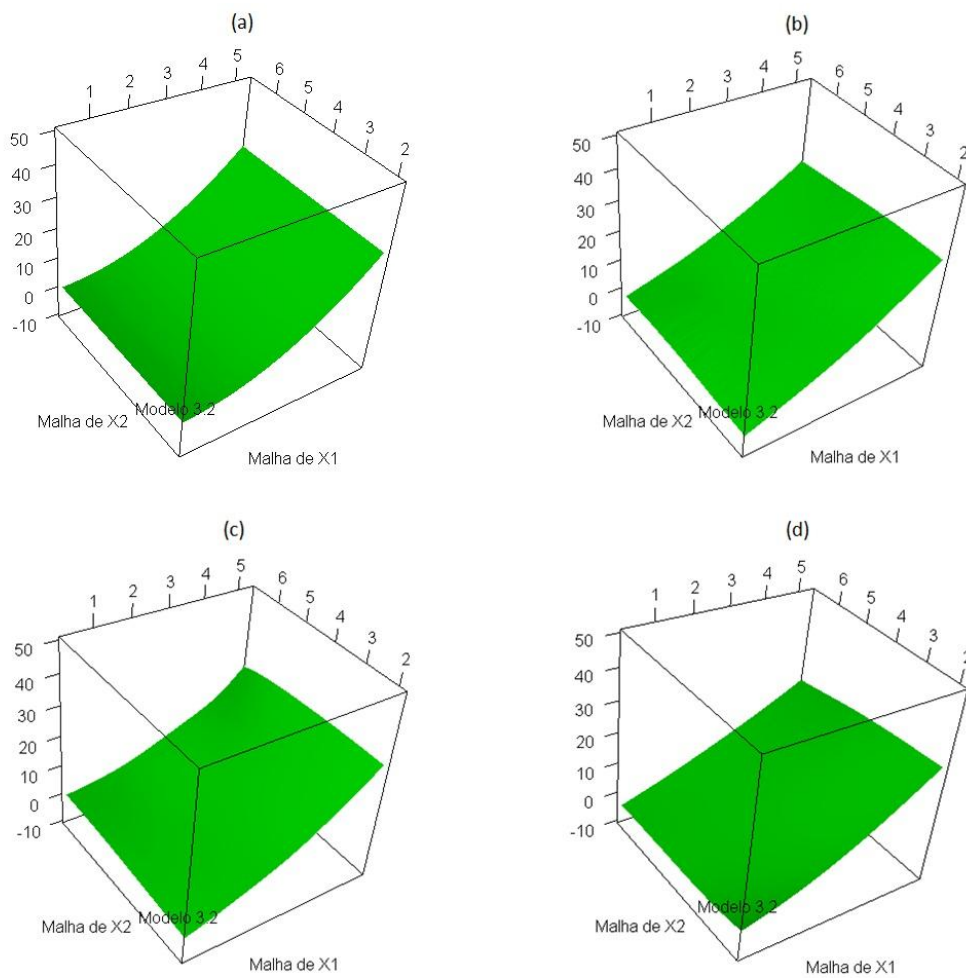


Figura B.4: Estimação do modelo 3.2 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

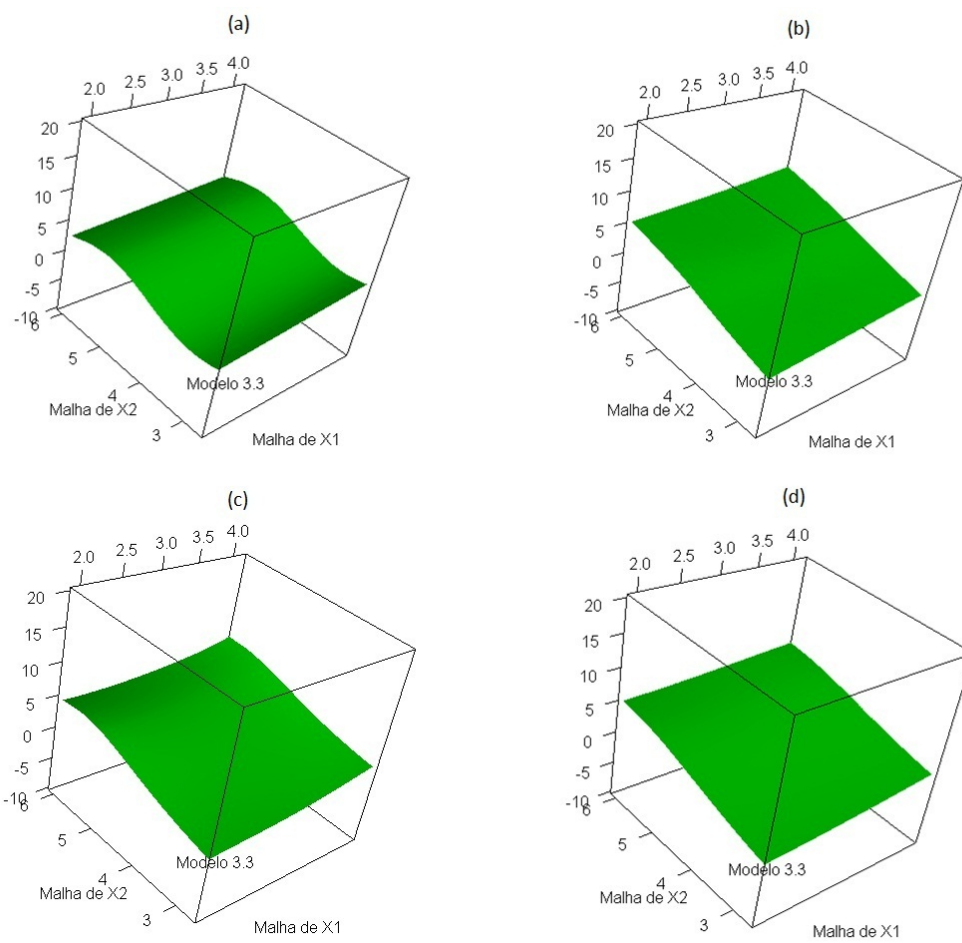


Figura B.5: Estimação do modelo 3.3 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

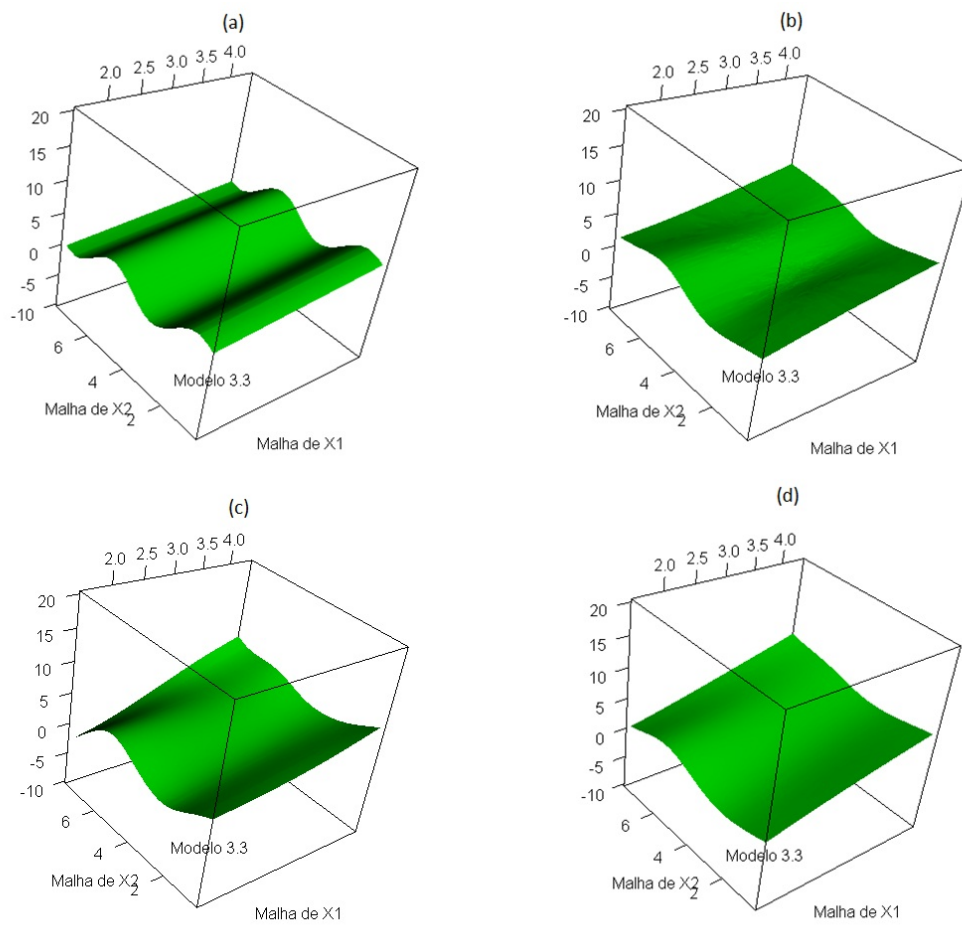


Figura B.6: Estimación do modelo 3.3 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

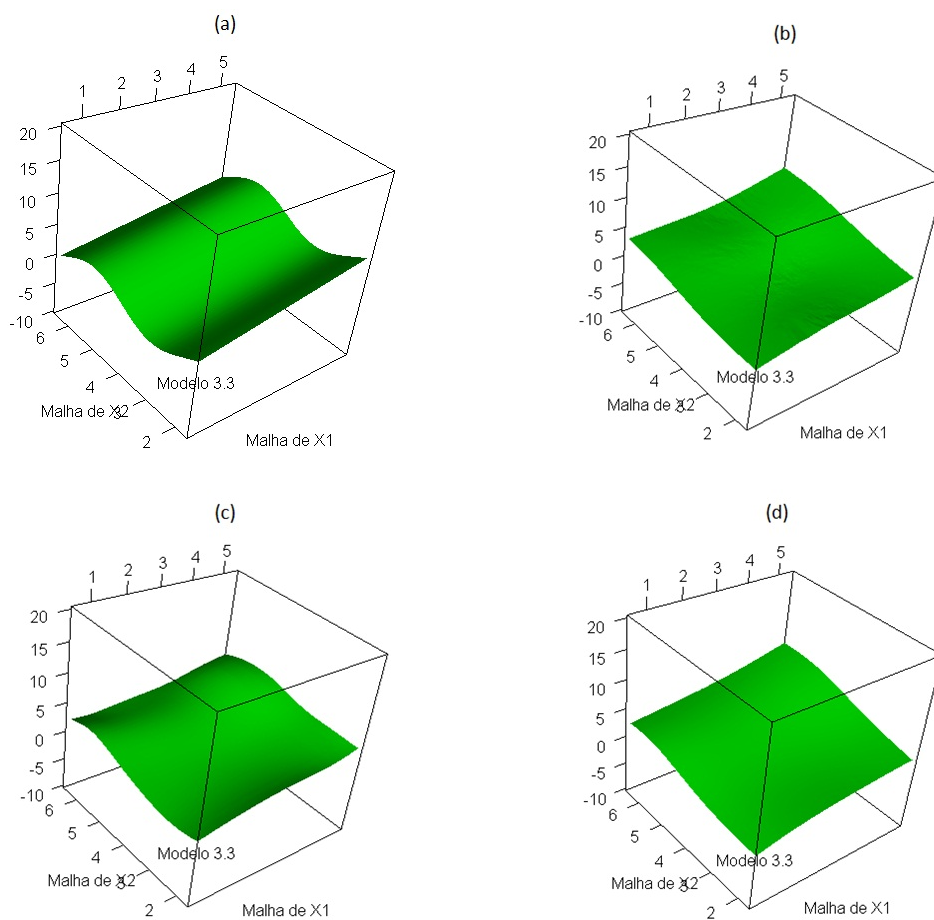


Figura B.7: Estimação do modelo 3.3 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

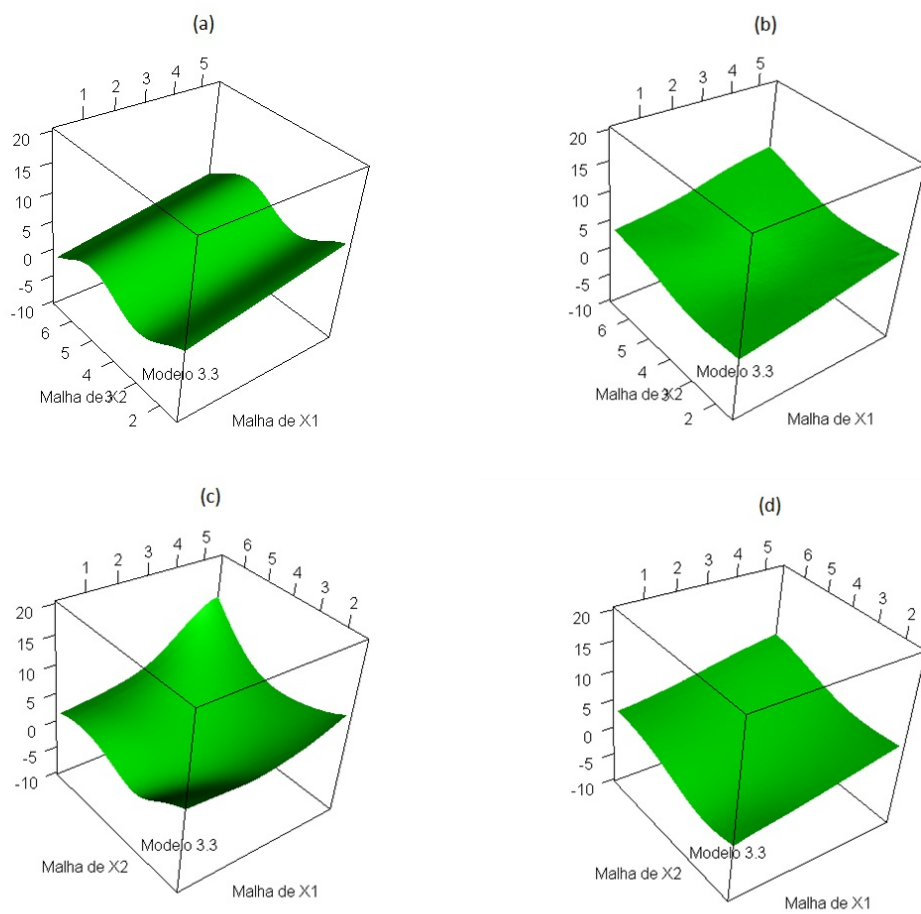


Figura B.8: Estimação do modelo 3.3 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

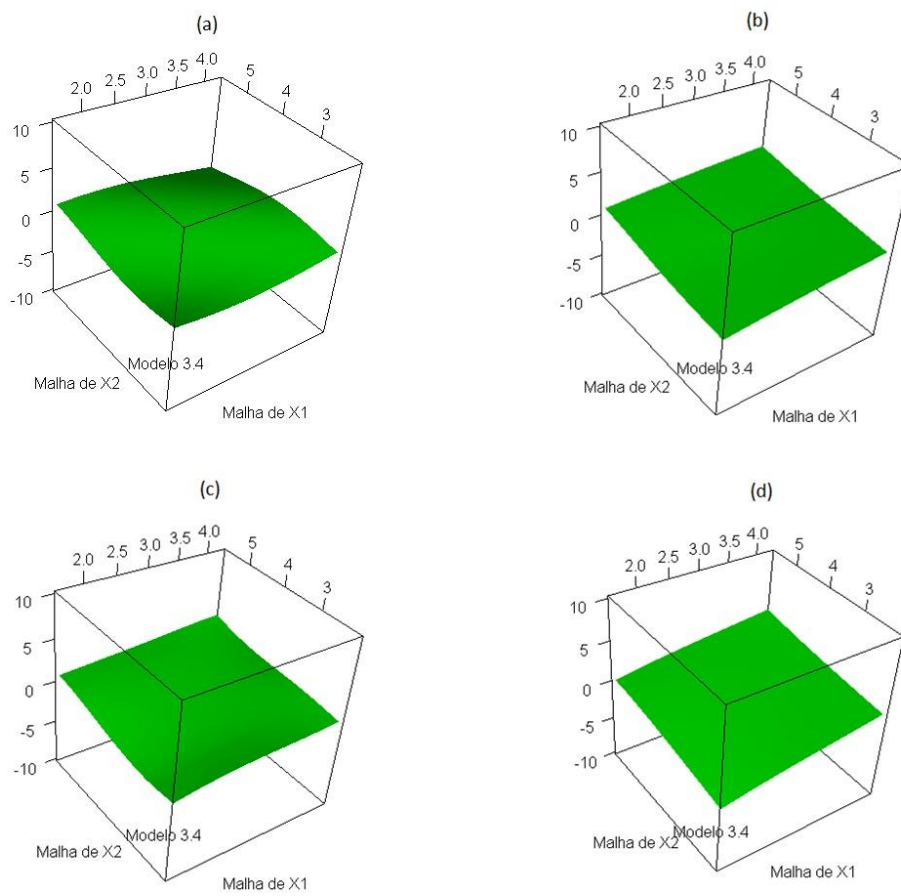


Figura B.9: Estimação do modelo 3.4 na situação 1: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

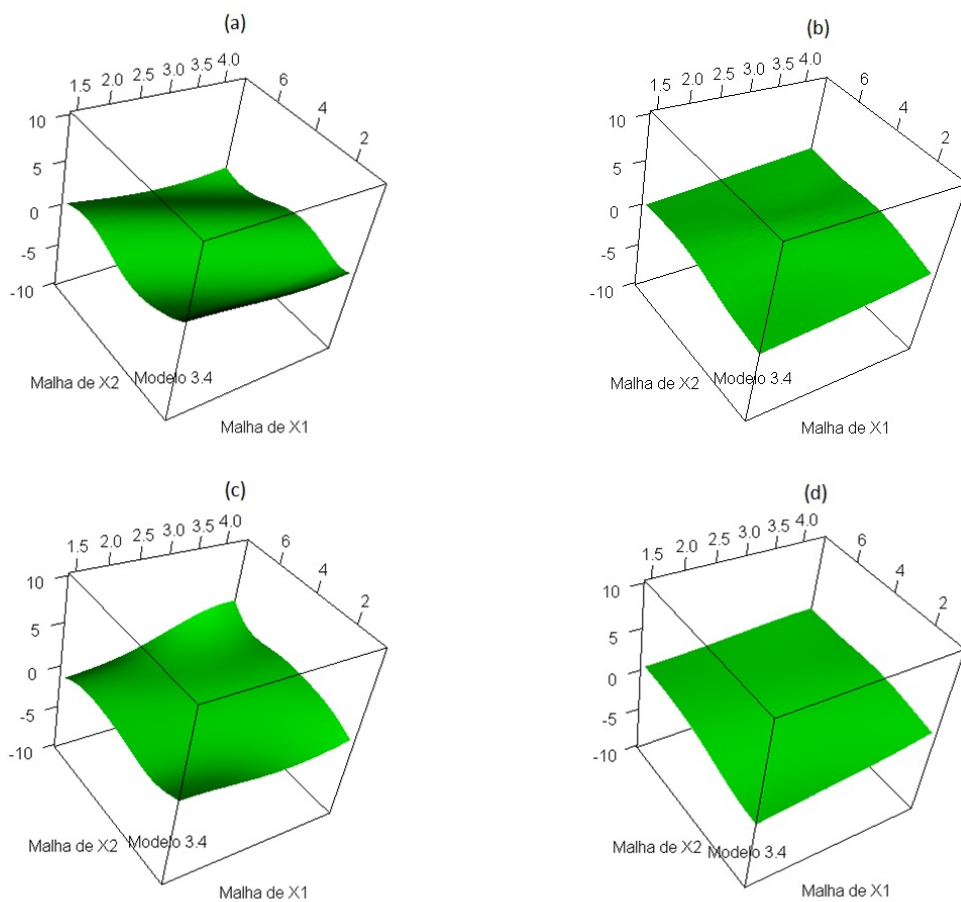


Figura B.10: Estimação do modelo 3.4 na situação 2: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

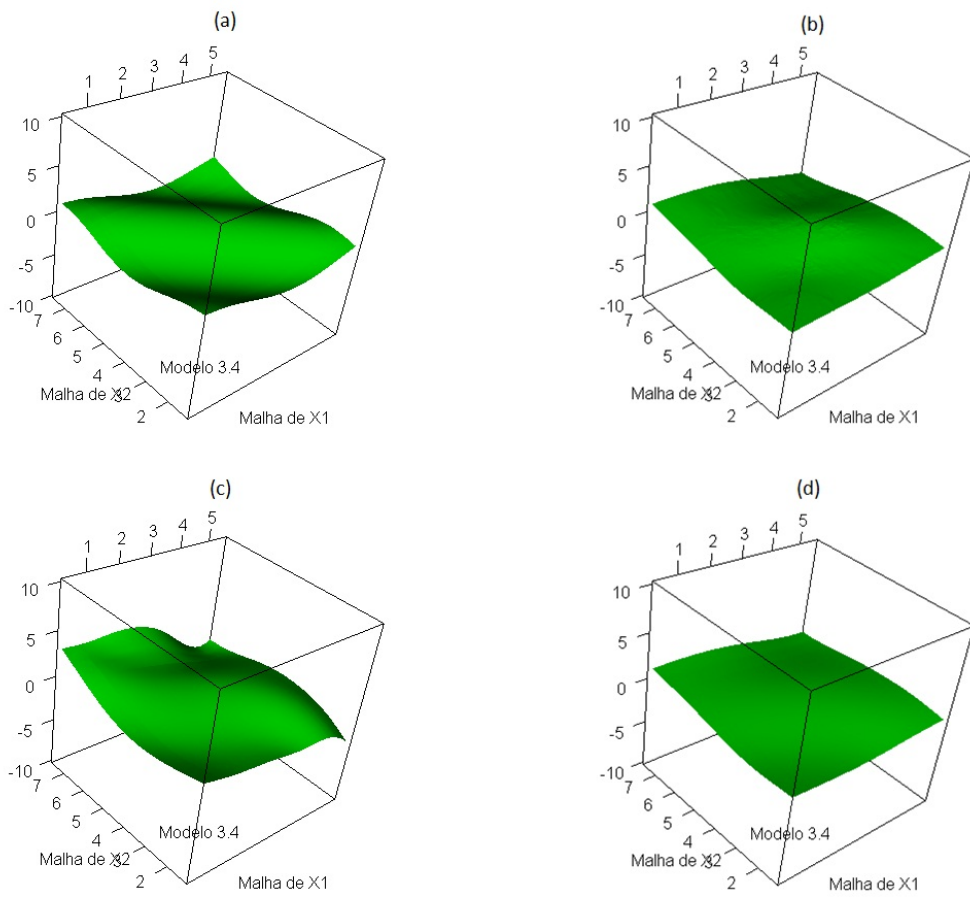


Figura B.11: Estimação do modelo 3.4 na situação 3: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

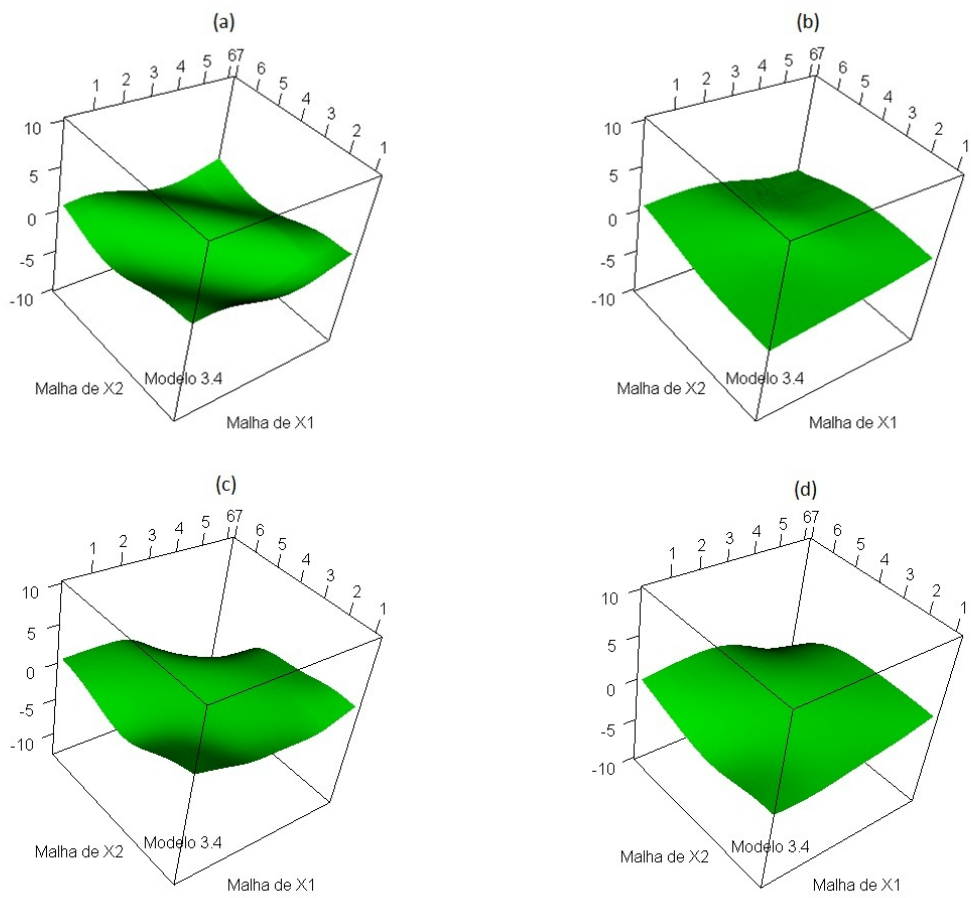


Figura B.12: Estimação do modelo 3.4 na situação 4: (a) Teórico; (b) Janela escalar global; (c) Janela diagonal global e (d) Janela escalar local

Apêndice C

Funções de variância teóricas e estimadas, por modelos e situações

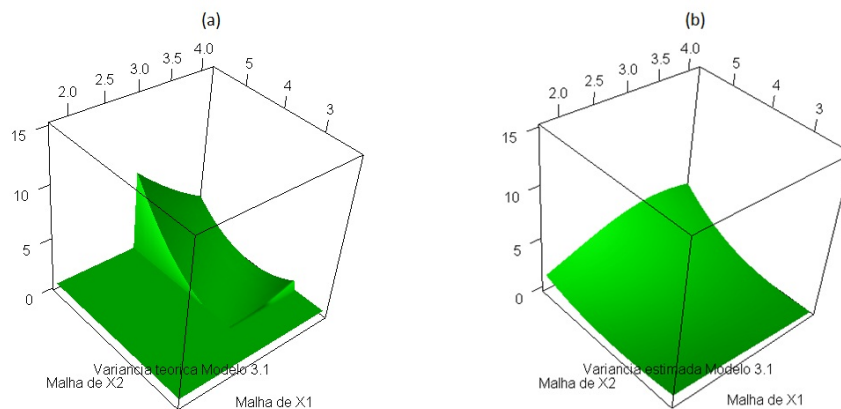


Figura C.1: Estimação da variância do modelo 3.1 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global.

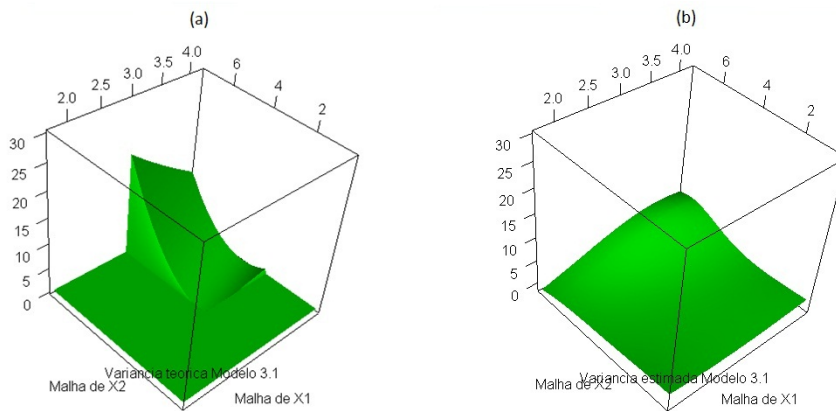


Figura C.2: Estimação da variância do modelo 3.1 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global.

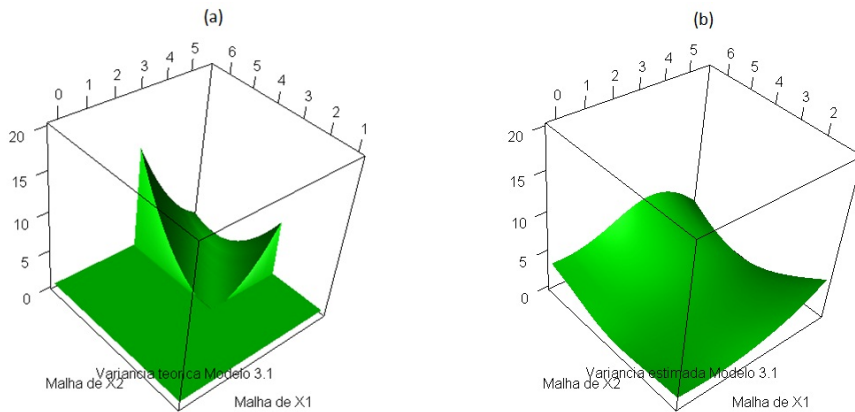


Figura C.3: Estimação da variância do modelo 3.1 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global.

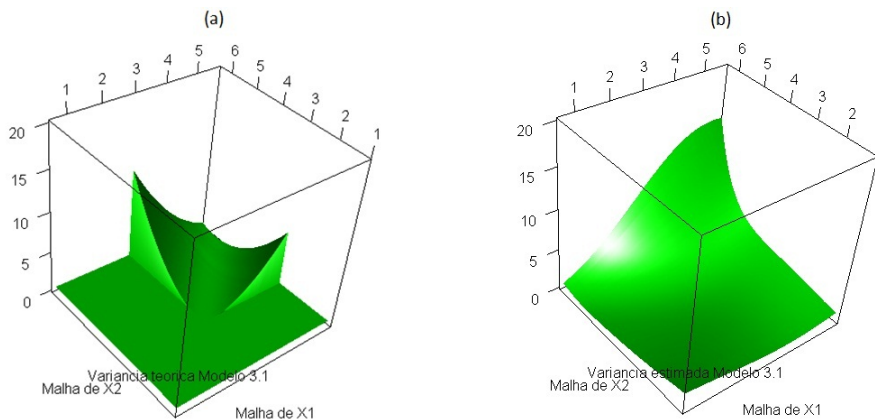


Figura C.4: Estimação da variância do modelo 3.1 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global.

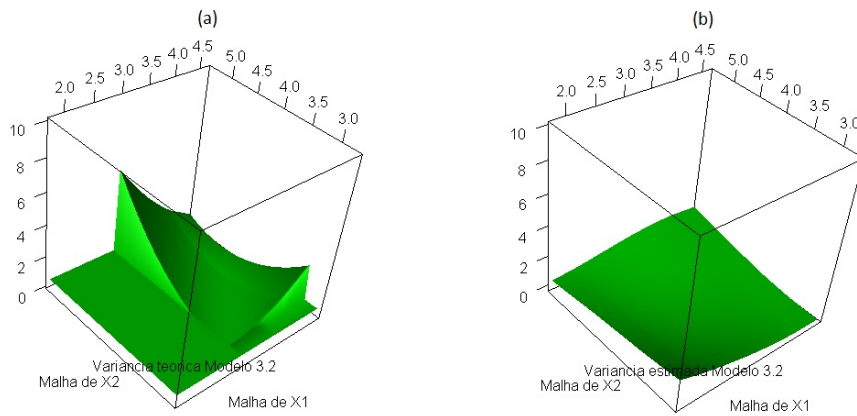


Figura C.5: Estimação da variância do modelo 3.2 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global.

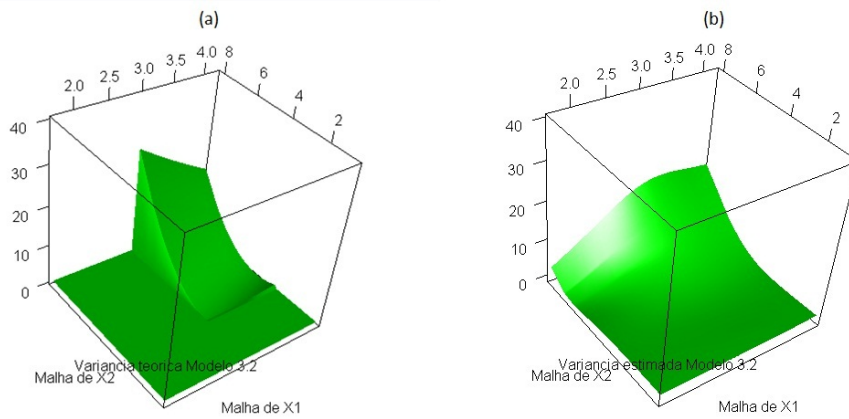


Figura C.6: Estimação da variância do modelo 3.2 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global.

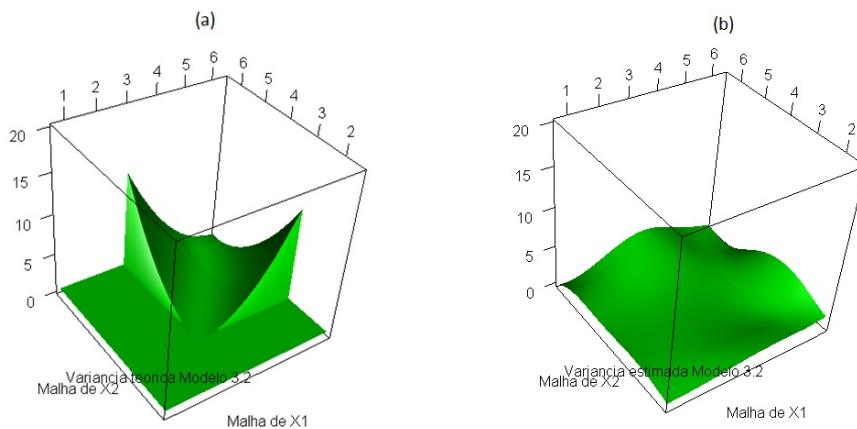


Figura C.7: Estimação da variância do modelo 3.2 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global.

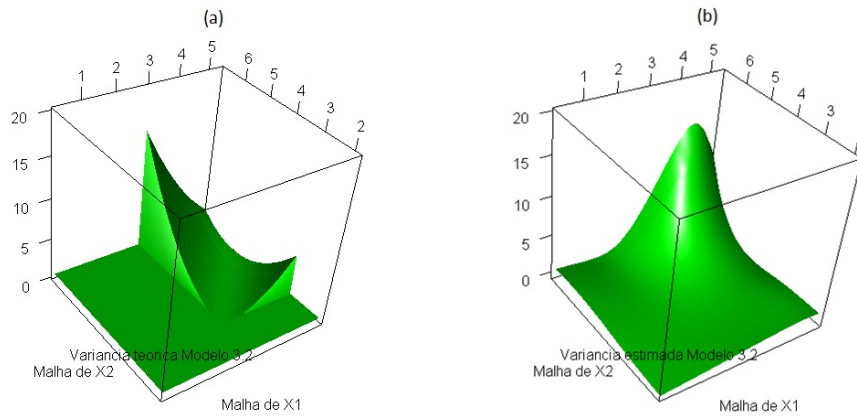


Figura C.8: Estimação da variância do modelo 3.2 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global.

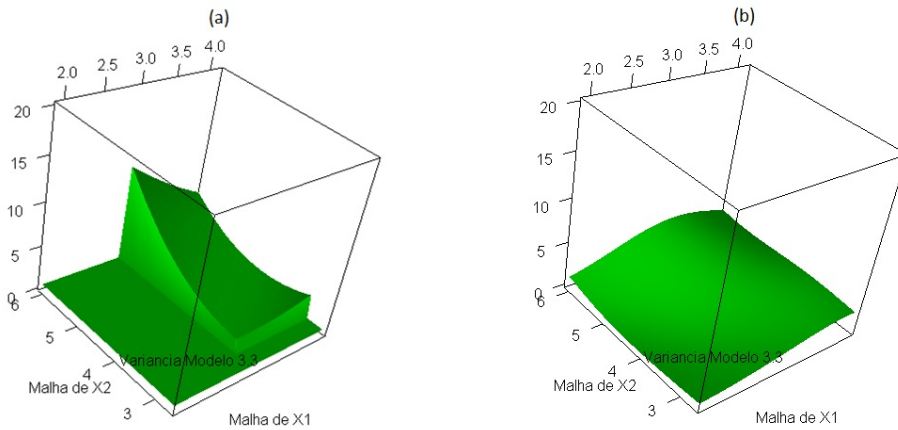


Figura C.9: Estimação da variância do modelo 3.3 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global.

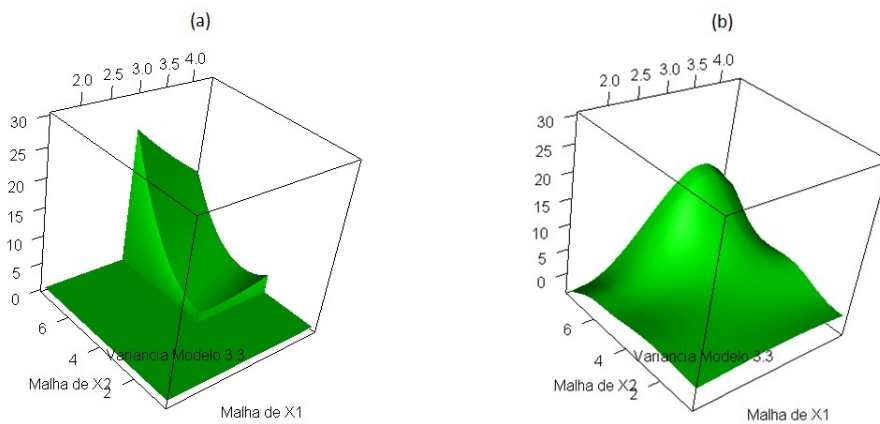


Figura C.10: Estimação da variância do modelo 3.3 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global.

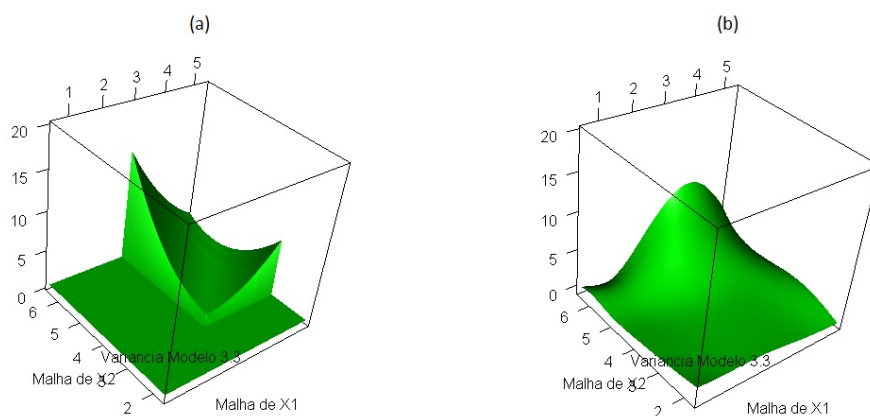


Figura C.11: Estimação da variância do modelo 3.3 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global.

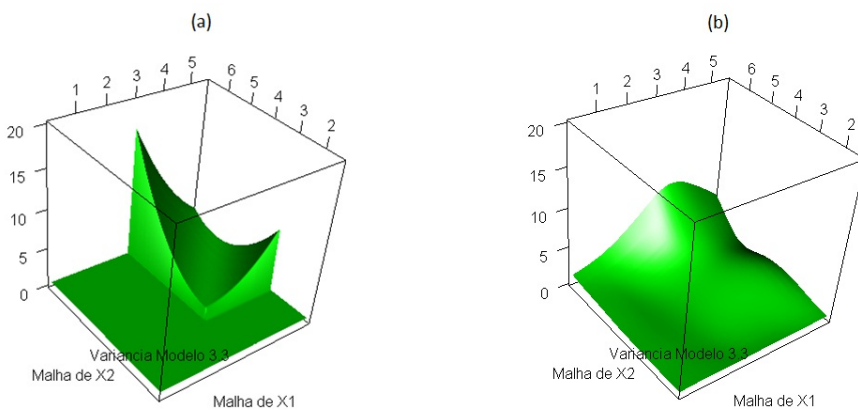


Figura C.12: Estimação da variância do modelo 3.3 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global.

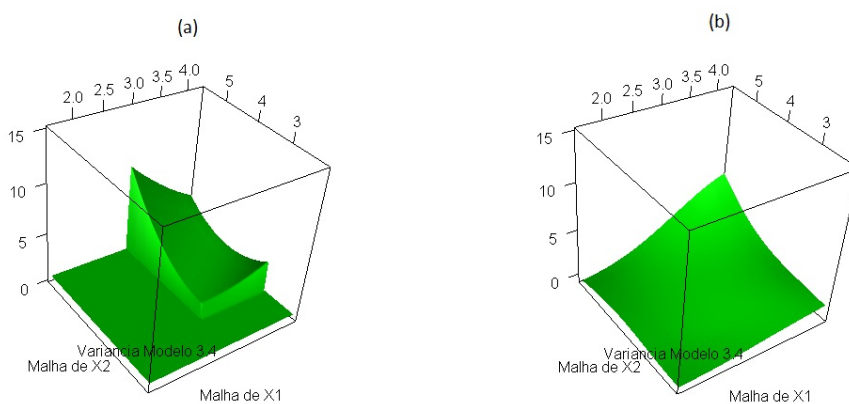


Figura C.13: Estimação da variância do modelo 3.4 na situação 1: (a) Teórica; (b) Estimada por janela diagonal global.

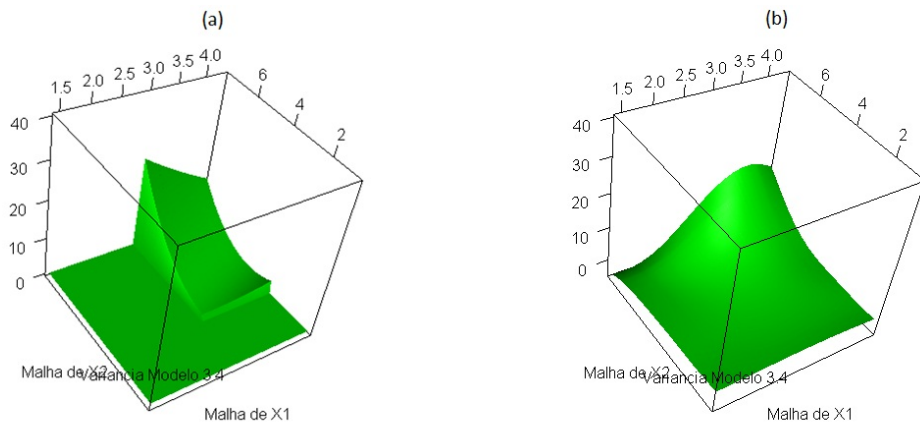


Figura C.14: Estimação da variância do modelo 3.4 na situação 2: (a) Teórica; (b) Estimada por janela diagonal global.

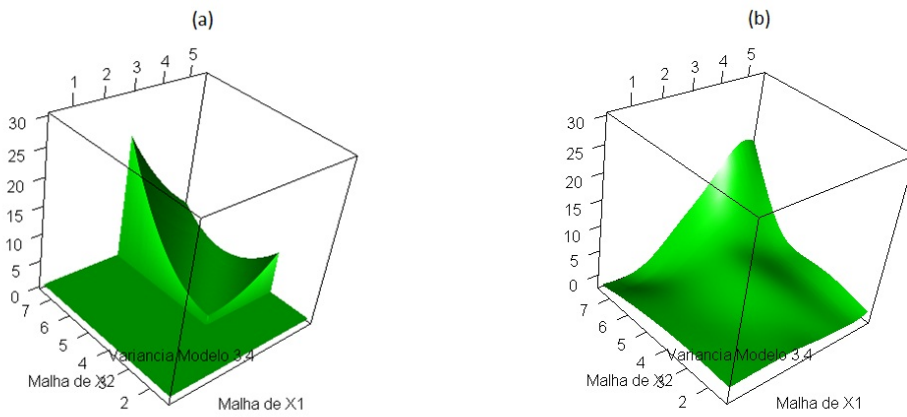


Figura C.15: Estimação da variância do modelo 3.4 na situação 3: (a) Teórica; (b) Estimada por janela diagonal global.

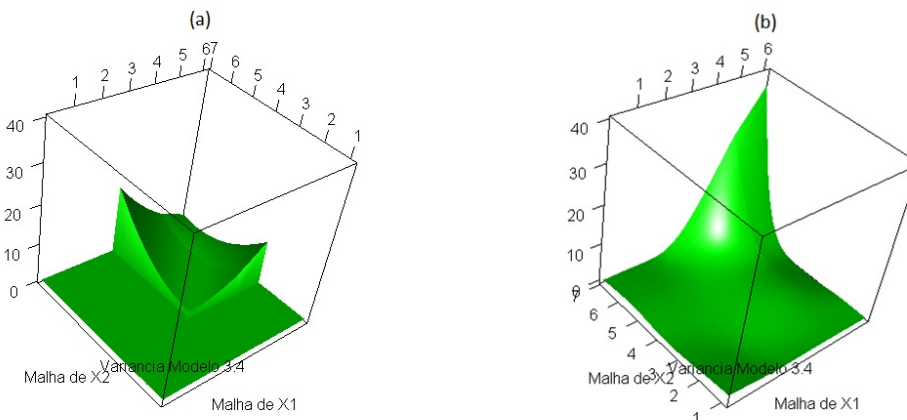


Figura C.16: Estimação da variância do modelo 3.4 na situação 4: (a) Teórica; (b) Estimada por janela diagonal global.