

**UNIVERSIDADE DE SÃO PAULO**  
**INSTITUTO DE MATEMÁTICA E ESTATÍSTICA**  
**BACHARELADO EM MATEMÁTICA APLICADA E COMPUTACIONAL**

**LEONARDO KIYOSHI KINOSHITA ASSAHIDE**

Um estudo sobre *Markov Switching Generalized Linear Model* para dados discretos com  
distribuição de Poisson

**Orientador:** Prof<sup>ª</sup>. Dra. Airlane Pereira Alencar

São Paulo - SP

2018

Prof. Dr. Vahan Agopyan  
Reitor da Universidade de São Paulo

Prof. Dr. Junior Barrera  
Diretor do Instituto de Matemática e Estatística

Prof. Dr. Fábio Armando Tal  
Chefe do Departamento de Matemática Aplicada

**LEONARDO KIYOSHI KINOSHITA ASSAHIDE**

Um estudo sobre *Markov Switching Generalized Linear Model* para dados discretos com  
distribuição de Poisson

Trabalho de Conclusão de Curso apresentado ao Bacharelado em Matemática Aplicada e Computacional - Habilitação: Estatística Econômica do Instituto de Matemática e Estatística da Universidade de São Paulo.

**ORIENTADORA: PROF<sup>a</sup>. DRA. AIRLANE PEREIRA ALENCAR**

São Paulo - SP

2018

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

## FICHA CATALOGRÁFICA

Assahide, Leonardo Kiyoshi Kinoshita.

Um estudo sobre *Markov Switching Generalized Linear Model* para dados discretos com distribuição de Poisson. São Paulo , 2018.

84 p. : il. ; 30cm

Trabalho de Conclusão de Curso, apresentado ao Instituto de Matemática e Estatística da Universidade de São Paulo.

Orientadora: Alencar, Airlane Pereira.

1. *Markov Switching*. 2. *Generalized Linear Model (GLM)*. 3. Poisson.

# FOLHA DE APROVAÇÃO

Nome: ASSAHIDE, Leonardo Kiyoshi Kinoshita

Título: Um estudo sobre *Markov Switching Generalized Linear Model* para dados discretos com distribuição de Poisson

Trabalho de Conclusão de Curso apresentado ao Bacharelado em Matemática Aplicada e Computacional - Habilitação: Estatística Econômica do Instituto de Matemática e Estatística da Universidade de São Paulo.

Aprovado em: \_\_\_\_\_

## BANCA EXAMINADORA

Prof<sup>a</sup>. Dra. Airlane Pereira Alencar

Instituição: Departamento de Estatística – IME-USP

Assinatura: \_\_\_\_\_

Prof<sup>a</sup>. Dra. Elisabeti Kira

Instituição: Departamento de Estatística – IME-USP

Assinatura: \_\_\_\_\_

Prof. Dr. Francisco Marcelo Monteiro da Rocha

Instituição: Universidade Federal de São Paulo - UNIFESP

Assinatura: \_\_\_\_\_



## RESUMO

ASSAHIDE, L. K. K. **Um estudo sobre *Markov Switching Generalized Linear Model* para dados discretos com distribuição de Poisson.** 2018. Trabalho de Conclusão de Curso - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2018.

Este trabalho busca estudar o número de internações hospitalares devido a doenças respiratórias de pessoas com mais de 60 anos na cidade de São Paulo, com esse objetivo foi utilizado um modelo *Markov Switching Generalized Linear Model* com distribuição de Poisson para verificar a existência de múltiplos regimes. A suposição de múltiplos regimes está baseada no fator sazonal na cidade de São Paulo, onde o inverno frio e seco amplia os efeitos maléficos da poluição. Dentre as especificações utilizadas, o modelo de *Markov Switching* com efeitos sazonais apresenta melhores resultados, validando a hipótese de sazonalidade, porém este modelo apresenta instabilidade ao estimar as probabilidades de transição entre os regimes.

**Palavras-chave:** 1. *Markov Switching*. 2. *Generalized Linear Model* (GLM). 3. Poisson.

## ABSTRACT

ASSAHIDE, L. K. K. **A study on *Markov Switching Generalized Linear Model for discrete data with the Poisson distribution***. 2018. Course Completion Work - Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2018.

This study aims to study the number of hospitalizations due to respiratory diseases of people over 60 years old in the city of São Paulo, with this objective was used a Markov Switching Generalized Linear Model with Poisson distribution to verify the existence of multiple regimes. The assumption of multiple regimes is based on the seasonal factor in the city of São Paulo, where cold and dry winter magnifies the harmful effects of pollution. Among the specifications used, the Markov Switching model with seasonal effects presents better results, validating the hypothesis of seasonality, but this model presents instability when estimating the transition probabilities between the regimes.

**Keywords:** 1. *Markov Switching*. 2. *Generalized Linear Model (GLM)*. 3. Poisson.

## LISTA DE FIGURAS

	<b>Página</b>
Figura 1 - Distribuição da população, por grupos de idade . . . . .	13
Figura 2 - Número mensal de internações de pessoas com mais de 60 anos em São Paulo . . . . .	25
Figura 3 - Número mensal de internações observado e predito pelo pelo modelo GLM com efeitos sazonais . . . . .	30
Figura 4 - Quantidade de internações em relação ao Regime 0 (estimado pelo modelo <i>Markov Switching</i> com efeitos sazonais) . . . . .	30
Figura 5 - Probabilidades Suavizadas pelo modelo <i>Markov Switching</i> com efeitos sazonais . . . . .	31
Figura 6 - Número mensal de internações observado e predito pelo pelo modelo GLM com variável defasada . . . . .	32
Figura 7 - Quantidade de internações em relação ao Regime 0 (estimado pelo modelo <i>Markov Switching</i> com variável defasada . . . . .	33
Figura 8 - Probabilidades Suavizadas pelo modelo <i>Markov Switching</i> com variável defasada . . . . .	34
Figura 9 - Resíduos do Modelo <i>Markov Switching</i> GLM Poisson com efeitos sazonais	39
Figura 10- Normal Q-Q Plot do Modelo <i>Markov Switching</i> GLM Poisson com efeitos sazonais . . . . .	39
Figura 11- ACF/PACF dos Resíduos e dos Quadrados do Resíduo do Modelo <i>Markov Switching</i> GLM Poisson com efeitos sazonais . . . . .	40
Figura 12- Resíduos do Modelo <i>Markov Switching</i> GLM Poisson com variável defasada	43
Figura 13- Normal Q-Q Plot do Modelo <i>Markov Switching</i> GLM Poisson com variável defasada . . . . .	43
Figura 14- ACF/PACF dos Resíduos e dos Quadrados do Resíduo do Modelo <i>Markov Switching</i> GLM Poisson com variável defasada . . . . .	44

## LISTA DE TABELAS

### Página

Tabela 1 - Estatísticas Descritivas do Número de internações mensais de pessoas com mais de 60 anos devido a doenças respiratórias em São Paulo . . . . .	25
Tabela 2 - Estimativas e erros padrão para os modelos GLM e <i>Markov Switching</i> GLM	29
Tabela 3 - Matriz de Probabilidades de Transição no modelo <i>Markov Switching</i> com efeitos sazonais . . . . .	31
Tabela 4 - Estimativas e erros padrão para os modelos GLM e <i>Markov Switching</i> GLM	32
Tabela 5 - Matriz de Probabilidades de Transição no modelo <i>Markov Switching</i> com variável defasada . . . . .	33
Tabela 6 - Modelo GLM Poisson com efeitos sazonais . . . . .	37
Tabela 7 - Modelo <i>Markov Switching</i> GLM Poisson com efeitos sazonais . . . . .	38
Tabela 8 - Modelo GLM Poisson com variável defasada . . . . .	41
Tabela 9 - Modelo <i>Markov Switching</i> GLM Poisson com variável defasada . . . . .	42

# SUMÁRIO

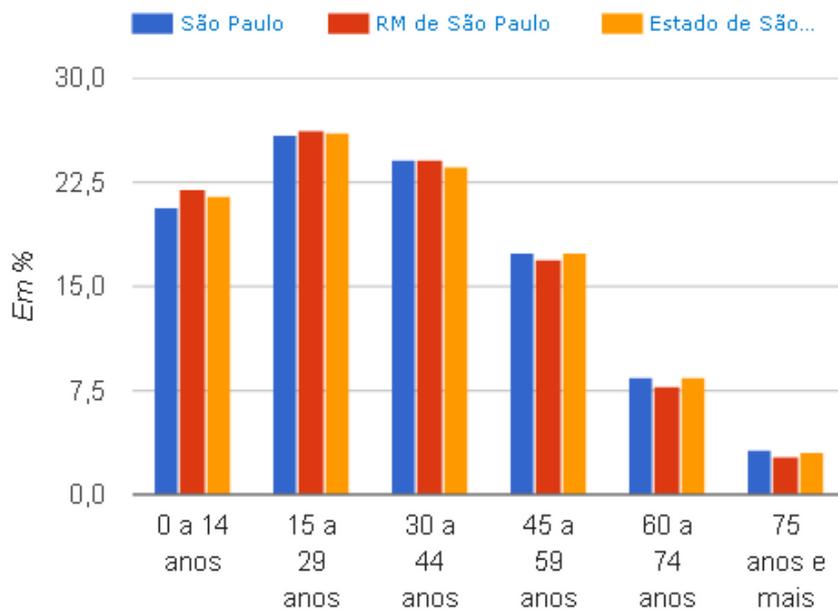
	Página
<b>RESUMO</b> . . . . .	<b>7</b>
<b>ABSTRACT</b> . . . . .	<b>8</b>
<b>LISTA DE FIGURAS</b> . . . . .	<b>9</b>
<b>LISTA DE TABELAS</b> . . . . .	<b>10</b>
<b>1 INTRODUÇÃO</b> . . . . .	<b>13</b>
<b>2 REVISÃO BIBLIOGRÁFICA</b> . . . . .	<b>14</b>
2.1 Introdução ao modelo . . . . .	14
2.2 Modelo com mudança de regime . . . . .	16
2.2.1 Caso com Independência na troca de regime . . . . .	17
2.2.2 Caso de <i>Markov Switching</i> . . . . .	18
2.3 Série Cronológica e <i>Markov Switching</i> . . . . .	19
2.3.1 Resolvendo o problema do regime não observado . . . . .	20
2.3.2 Filtro de Probabilidades . . . . .	21
2.4 Questões relacionadas ao <i>Markov Switching</i> . . . . .	22
2.4.1 Algoritmo de Suavização de Kim . . . . .	22
2.4.2 Probabilidades no Estado de Equilíbrio para começar o Filtro . . . . .	23
2.4.3 Expectativa de duração dos Regimes no modelo <i>Markov Switching</i> . . . . .	24
<b>3 BASE DE DADOS</b> . . . . .	<b>25</b>
<b>4 METODOLOGIA APLICADA</b> . . . . .	<b>26</b>
4.1 Modelo com efeitos sazonais mensais . . . . .	26
4.2 Modelo com variável defasada . . . . .	27
<b>5 RESULTADOS</b> . . . . .	<b>28</b>
5.1 Modelo com efeitos sazonais mensais . . . . .	28
5.2 Modelo com variável defasada . . . . .	31

<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>35</b>
<b>A</b>	<b>ANEXO</b>	<b>37</b>
A.1	Resultados Completos dos Modelos com efeitos sazonais	37
A.2	Análises dos Resíduos do Modelo <i>Markov Switching</i> Poisson com efeitos sazonais	39
A.3	Resultados Completos dos Modelos com variável defasada	41
A.4	Análises dos Resíduos do Modelo <i>Markov Switching</i> Poisson com variável defasada	43

# 1 INTRODUÇÃO

Existe uma grande preocupação com as doenças principalmente em recém nascidos e em idosos, dois dos maiores grupos de riscos. A poluição aliada ao clima seco da cidade de São Paulo tem um forte impacto positivo nos casos de doenças respiratórias, levando a um maior volume de internações nestes grupos.

Hoje, as pessoas com mais de 60 anos representam 12% da população total na cidade de São Paulo, um total de 1.7 milhões de pessoas. O gráfico 1<sup>1</sup> apresenta a distribuição da população por faixa etária para a cidade, região metropolitana e Estado de São Paulo:



Fonte: IBGE. Censo Demográfico 2010 (resultados preliminares); Elaboração: Fundação Seade.

Figura 1: Distribuição da população, por grupos de idade

Assim como no restante do Brasil, o IBGE está alertando para o envelhecimento populacional, estudos do instituto indicam que a população de idosos em São Paulo pode crescer até 18% no próximos 5 anos.

Este trabalho busca estudar o número de internações devido a doenças respiratórias de pessoas com mais de 60 anos no município de São Paulo, e assim, contribuir para o entendimento do comportamento das internações. Será realizado uma análise através do modelo *Markov Switching* com distribuição de Poisson.

<sup>1</sup><http://produtos.seade.gov.br/produtos/retratosdesp/view/index.php?locId=3550308&indId=3&temaId=1>

## 2 REVISÃO BIBLIOGRÁFICA

Existe uma grande literatura sobre modelos com quebras estruturais dependendo das amostras ou regimes, indicando mudanças de nível, e podendo não ser lineares. As mudanças de nível em uma série temporal podem ocorrer em datas específicas ou ocorrerem ao longo de um período de tempo. Caso estas mudanças de nível sejam observadas, o modelo pode ser resolvido incluindo variáveis indicadoras dos regimes. Essas quebras estruturais podem ser identificadas por meio de testes de hipóteses (métodos inferenciais), por exemplo através do Teste-F, proposto por Chow (1960).

Inicialmente, os modelos propostos consideravam apenas 1 troca de regime na amostra, como por exemplo em Quandt (1958), Quandt (1960), Farley and Hinich (1970) e Kim and Siegmund (1989). Posteriormente, Quandt (1972), Goldfeld and Quandt (1973), Brown et al. (1975), Ploberger et al. (1989), e Kim and Maddala (1991) consideram modelos com a possibilidade de mais de 1 troca de regime.

Outra questão interessante é a probabilidade de troca de regime; alguns modelos, como por exemplo Quandt (1972), consideram que a probabilidade de troca de regime é independente do regime vigente (e de efeitos passados). Enquanto Goldfeld and Quandt (1973) flexibilizam esta hipótese e permitem que as probabilidades sejam condicionais aos regimes vigentes e de efeitos passados ao realizar o modelo de *Markov Switching*.

A partir do modelo de Goldfeld and Quandt (1973), Hamilton (1989) realiza um modelo de *Markov Switching* com mudanças estruturais observando o comportamento da variável dependente. Nas próximas subseções, será apresentado o desenvolvimento do modelo de *Markov Switching* com distribuição de Poisson, este modelo teórico foi baseado em Kim (1999).

### 2.1 Introdução ao modelo

Nelder and Wedderburn (1972) propuseram os Modelos Lineares Generalizados (GLM, em inglês *Generalized Linear Models*) para flexibilizar e generalizar os modelos lineares gaussianos. O modelo GLM permite que a variável explicada tenha outras distribuições além da normal. Além disso, no GLM a resposta marginal de  $\mathbb{E}[y_t]$  está relacionada com as covariáveis por meio da função de ligação  $g(\cdot)$ , que deve ser monótona e diferenciável; e denominada função de relação. Neste caso, suporemos que  $y_t$  possui distribuição de Poisson, e consequentemente a relação funcional entre a média  $\mu_t$  de  $y_t$  e o preditor linear será logarítmica.

O Modelo Linear Generalizado (GLM) com distribuição de Poisson, ainda sem considerar quebras estruturais pode ser definido como  $y_t \sim \text{Poisson}(\mu_t)$  com:

$$\ln(\mu_t) = \mathbf{x}_t^T \boldsymbol{\beta}, \quad t = 1, 2, \dots, T$$

Por definição da Poisson:

$$\begin{aligned} \mathbb{E}[y_t] &= \mu_t \\ g(\mu_t) &= \ln(\mu_t) = \mathbf{x}_t^T \boldsymbol{\beta} \\ \mu_t &= e^{\mathbf{x}_t^T \boldsymbol{\beta}} \end{aligned}$$

Desta forma, a probabilidade será:

$$\begin{aligned} \mathbb{P}(\mathbf{y} \mid \mathbf{x}) &= \frac{e^{-\mu_t} \mu_t^{\mathbf{y}}}{\mathbf{y}!} = \frac{e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}}} (e^{\mathbf{x}_t^T \boldsymbol{\beta}})^{\mathbf{y}}}{\mathbf{y}!} \\ \mathbb{P}(y_1, \dots, y_T \mid x_1, \dots, x_T; \boldsymbol{\beta}) &= \prod_{t=1}^T \frac{e^{y_t \mathbf{x}_t^T \boldsymbol{\beta}} e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}}}}{y_t!} \end{aligned}$$

As funções de verossimilhança e log-verossimilhança são dadas, respectivamente, por:

$$\begin{aligned} L(\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{Y}) &= \prod_{t=1}^T \frac{e^{y_t \mathbf{x}_t^T \boldsymbol{\beta}} e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}}}}{y_t!} \\ \ln L(\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{Y}) &= \sum_{t=1}^T (y_t \mathbf{x}_t^T \boldsymbol{\beta} - e^{\mathbf{x}_t^T \boldsymbol{\beta}} - \ln(y_t!)) \end{aligned}$$

Dados os valores das covariáveis e da variável resposta  $y_i$ , o termo  $\ln(y_i!)$  é um constante, desta forma podemos ignorar o último termo e reescrever a log-verossimilhança

$$\ln L(\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{Y}) = \sum_{t=1}^T (y_t \mathbf{x}_t^T \boldsymbol{\beta} - e^{\mathbf{x}_t^T \boldsymbol{\beta}})$$

## 2.2 Modelo com mudança de regime

O modelo com *Markov Switching* é útil para capturar as trocas de regime em uma série temporal. Sendo  $S_t$  a variável aleatória que define o regime, será apresentado o modelo com quebras estruturais com 2 regimes:

$$\ln(\mu_t) = \mathbf{x}_t^T \boldsymbol{\beta}_{S_t},$$

Onde  $t = 1, 2, \dots, T$  e  $S_t = 1$  ou  $2$ . Com  $\boldsymbol{\beta}_{S_t}$ :

$$\begin{aligned} \boldsymbol{\beta}_{S_t} &= \boldsymbol{\beta}_0(1 - S_t) + \boldsymbol{\beta}_1 S_t \\ \mathbf{x}_t^T \boldsymbol{\beta}_{S_t} &= \mathbf{x}_t^T \boldsymbol{\beta}_0(1 - S_t) + \mathbf{x}_t^T \boldsymbol{\beta}_1 S_t \end{aligned}$$

Portanto, sob o regime 0 o parâmetro será  $\boldsymbol{\beta}_0$ . Analogamente, sob o regime 1 o parâmetro será  $\boldsymbol{\beta}_1$ . Se as quebras estruturais  $S_t$  fossem conhecidas, seria apenas um modelo com variáveis. Entretanto, neste caso será preciso prever  $S_t$ , e a função de log-verossimilhança será dada por:

$$\begin{aligned} \ln L &= \sum_{t=1}^T \ln(f(y_t | S_t)) \\ f(y_t | S_t) &= \frac{e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}_{S_t}}} (e^{\mathbf{x}_t^T \boldsymbol{\beta}_{S_t}})^{y_t}}{y_t!} \end{aligned}$$

Definindo  $\boldsymbol{\psi}_t$  como toda informação disponível até o período  $t$ :

$$\boldsymbol{\psi}_t = \{x_t, \dots, x_1; y_t, \dots, y_1; \mu_{t-1}, \dots, \mu_1\}$$

Como  $S_t$  não é observado, podemos encontrar as probabilidades utilizando os seguintes passos:

1. Considerar a densidade conjunta de  $y_t$  e da variável não observada  $S_t$ , sendo o produto das densidades condicionais e marginais individuais:

$$f(y_t, S_t | \boldsymbol{\psi}_{t-1}) = f(y_t | \boldsymbol{\psi}_{t-1})f(S_t | \boldsymbol{\psi}_{t-1})$$

2. Para obter a densidade marginal de  $y_t$ , considere a  $S_t$  fora da densidade conjunta ao somar

os possíveis valores ponderados pelas probabilidade:

$$\begin{aligned}
f(y_t | \boldsymbol{\psi}_{t-1}) &= \sum_{S_t=0}^1 f(y_t, S_t | \boldsymbol{\psi}_{t-1}) \\
&= \sum_{S_t=0}^1 f(y_t | \boldsymbol{\psi}_{t-1}, S_t) f(S_t | \boldsymbol{\psi}_{t-1}) \\
&= \sum_{S_t=0}^1 \frac{e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}_{S_t}}} (e^{\mathbf{x}_t^T \boldsymbol{\beta}_{S_t}})^y}{y!} \mathbb{P}[S_t | \boldsymbol{\psi}_{t-1}] \\
&= \frac{e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}_0}} (e^{\mathbf{x}_t^T \boldsymbol{\beta}_0})^y}{y!} \mathbb{P}[S_t = 0 | \boldsymbol{\psi}_{t-1}] + \frac{e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}_1}} (e^{\mathbf{x}_t^T \boldsymbol{\beta}_1})^y}{y!} \mathbb{P}[S_t = 1 | \boldsymbol{\psi}_{t-1}]
\end{aligned}$$

Então a função de log-verossimilhança:

$$\ln L = \sum_{t=1}^T \ln \left\{ \sum_{S_t=0}^1 f(y_t | S_t, \boldsymbol{\psi}_{t-1}) \mathbb{P}[S_t | \boldsymbol{\psi}_{t-1}] \right\}$$

Então a função de densidade marginal anterior pode ser interpretada com a média ponderada pelas densidades condicionais dadas por  $S_t$ . Será preciso estimar antes as probabilidades  $\mathbb{P}[S_t = 0 | \boldsymbol{\psi}_{t-1}]$  e  $\mathbb{P}[S_t = 1 | \boldsymbol{\psi}_{t-1}]$  antes de calcular a função de log-verossimilhança. Para este cálculo, assume-se que a variável  $S_t$  tem distribuição Bernoulli apresentando probabilidade  $\mathbb{P}[S_t = 1 | \boldsymbol{\psi}_{t-1}]$ , por exemplo. Além disso, pode-se assumir a troca de regimes seja independente ou não do regime anterior.

### 2.2.1 Caso com Independência na troca de regime

Caso a variável  $S_t$  seja independente dos regimes anteriores:

$$\begin{aligned}
\mathbb{P}[S_t = 1] &= \mathbb{P}[S_t = 1 | \boldsymbol{\psi}_{t-1}] \\
\mathbb{P}[S_t = 1] &= p = \frac{e^{\gamma_0}}{1 + e^{\gamma_0}} \\
\mathbb{P}[S_t = 0] &= 1 - p = \frac{1}{1 + e^{\gamma_0}}
\end{aligned}$$

No caso anterior, as probabilidades não são dependentes de variáveis exógenas, i.e., não será dependente das informações passadas ( $\boldsymbol{\psi}_{t-1}$ ), apenas de uma constante  $\gamma_0$ . Genericamente teremos que  $\mathbb{P}[S_t = j] = \mathbb{P}[S_t = j | \boldsymbol{\psi}_{t-1}]$ .

Porém pode-se considerar que  $S_t$  não seja independente de seus valores anteriores ou de um conjunto de variáveis  $\mathbf{Z}_{t-1}$ . A matriz  $\mathbf{Z}_{t-1}$  poderá ser definida com um conjunto de va-

lores anteriores, ou variáveis exógenas ao modelo ou as mesmas variáveis, ou uma combinação dos casos. Então as probabilidades serão definidas como:

$$\begin{aligned}\mathbb{P}[S_t = 1 \mid \boldsymbol{\psi}_{t-1}] &= p_t = \frac{e^{\gamma_0 + \mathbf{Z}_{t-1}^T \boldsymbol{\gamma}}}{1 + e^{\gamma_0 + \mathbf{Z}_{t-1}^T \boldsymbol{\gamma}}} \\ \mathbb{P}[S_t = 0 \mid \boldsymbol{\psi}_{t-1}] &= 1 - p_t = \frac{1}{1 + e^{\gamma_0 + \mathbf{Z}_{t-1}^T \boldsymbol{\gamma}}}\end{aligned}$$

### 2.2.2 Caso de *Markov Switching*

No caso da variável  $S_t$  ser dependente de  $S_{t-1}, S_{t-2}, \dots, S_{t-n}$ , então  $S_t$  é chamado como um processo de  $n$ -enésima ordem do processo de *Markov Switching*. Será exemplificado com um processo de 1ª ordem:

$$\mathbb{P}[S_t = 1 \mid S_{t-1} = 1] = p = \frac{e^{p_0}}{1 + e^{p_0}} \quad (1)$$

$$\mathbb{P}[S_t = 0 \mid S_{t-1} = 0] = q = \frac{e^{q_0}}{1 + e^{q_0}} \quad (2)$$

Consequentemente, as probabilidades complementares serão:

$$\mathbb{P}[S_t = 1 \mid S_{t-1} = 0] = 1 - p = \frac{1}{1 + e^{p_0}}$$

$$\mathbb{P}[S_t = 0 \mid S_{t-1} = 1] = 1 - q = \frac{1}{1 + e^{q_0}}$$

Para resolver o problema de ter  $S_t$  não observado, podemos utilizar a mesma estratégia utilizada anteriormente:

1. Dada  $\mathbb{P}[S_t = i \mid \boldsymbol{\psi}_{t-1}]$ ,  $i = 0, 1$  no começo do período  $t$  (ou da iteração), a probabilidade de troca será  $\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}]$ ,  $j = 0, 1$  calculada

$$\begin{aligned}\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}] &= \sum_{i=0}^1 \mathbb{P}[S_t = j, S_{t-1} = i \mid \boldsymbol{\psi}_{t-1}] \\ \mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}] &= \sum_{i=0}^1 \mathbb{P}[S_t = j \mid S_{t-1} = i, \boldsymbol{\psi}_{t-1}] \mathbb{P}[S_{t-1} = i \mid \boldsymbol{\psi}_{t-1}] \\ \mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}] &= \mathbb{P}[S_t = j \mid S_{t-1} = 0, \boldsymbol{\psi}_{t-1}] \mathbb{P}[S_{t-1} = 0 \mid \boldsymbol{\psi}_{t-1}] \\ &\quad + \mathbb{P}[S_t = j \mid S_{t-1} = 1, \boldsymbol{\psi}_{t-1}] \mathbb{P}[S_{t-1} = 1 \mid \boldsymbol{\psi}_{t-1}]\end{aligned}$$

Então  $\mathbb{P}[S_t = j \mid S_{t-1} = i]$ ,  $i = 0, 1; j = 0, 1$  são as probabilidades de transição de estado.

2. Após observar  $y_t$  no fim do período  $t$  (ou da iteração), podemos atualizar a probabilidade:

$$\begin{aligned}\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}, y_t] &= \frac{f(S_t = j, y_t \mid \boldsymbol{\psi}_t)}{f(y_t \mid \boldsymbol{\psi}_{t-1})} \\ &= \frac{f(y_t \mid S_t = j, \boldsymbol{\psi}_t) \mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}]}{\sum_{j=0}^1 f(y_t \mid S_t = j, \boldsymbol{\psi}_{t-1}) \mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}]}\end{aligned}$$

Por iteração, os passos anteriores serão repetidos até encontrar  $\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}]$ , para  $t = 1, 2, \dots, T$ . Entretanto, para começar o processo anterior, é preciso encontrar os valores iniciais para  $t = 1$ , i.e.,  $\mathbb{P}[S_0 \mid \boldsymbol{\psi}_0]$ . Assim, pode-se utilizar as probabilidades de  $S_t$  encontradas nas Equações (1) e (2):

$$\begin{aligned}\pi_0 &= \mathbb{P}[S_0 = 0 \mid \boldsymbol{\psi}_0] = \frac{1 - p}{2 - p - q} \\ \pi_1 &= \mathbb{P}[S_0 = 1 \mid \boldsymbol{\psi}_0] = \frac{1 - q}{2 - p - q}\end{aligned}$$

### 2.3 Série Cronológica e *Markov Switching*

Em geral, um modelo autoregressivo de primeira ordem, *Markov Switching* com  $M$  regimes, com média e variância pode ser escrito como:

$$\begin{aligned}\phi(L)(y_t - \mu S_t) &= e_t, e_t \sim N(0, \sigma_{S_t}^2) \\ \mathbb{P}[S_t = j \mid S_{t-1} = i] &= p_{ij}, i, j = 1, 2, \dots, M \\ \sum_{j=1}^M p_{ij} &= 1 \\ \mu S_t &= \mu_1 S_{1t} + \mu_2 S_{2t} + \dots + \mu_M S_{Mt} \\ \sigma_{S_t}^2 &= \sigma_{S_{1t}}^2 S_{1t} + \sigma_{S_{2t}}^2 S_{2t} + \dots + \sigma_{S_{Mt}}^2 S_{Mt} \\ &\begin{cases} S_{mt} = 1, \text{ se } S_t = m \\ S_{mt} = 0, \text{ caso contrário} \end{cases}\end{aligned}$$

O modelo GLM com distribuição de Poisson com  $k$  defasagens, pode ser definido

como:

$$\begin{aligned}\mathbb{E}[y_t | \boldsymbol{\psi}_t] &= \mu_t \\ g(\mu_t) &= \ln(\mu_t) = \mathbf{x}_t^T \boldsymbol{\beta}_{S_t} \\ \mathbb{E}[y_t | \boldsymbol{\psi}_t] &= \mathbf{x}_t^T \boldsymbol{\beta}_{S_t} + \sum_{j=1}^k \phi_{j,S_t} \ln y_{t-j}\end{aligned}$$

E a função de probabilidade como:

$$\begin{aligned}f(y_t | S_t) &= \frac{e^{-e^{\mathbf{x}_t^T \boldsymbol{\beta}_{S_t} + \sum_{j=1}^k \phi_{j,S_t} \ln y_{t-j}}} \left( e^{\mathbf{x}_t^T \boldsymbol{\beta}_{S_t} + \sum_{j=1}^k \phi_{j,S_t} \ln y_{t-j}} \right)^y}{y!} \\ \ln L &= \sum_{t=1}^T \ln(f(y_t | \boldsymbol{\psi}_{t-1}, S_t, S_{t-1}))\end{aligned}$$

Genericamente, podemos escrever o modelo com  $k$  defasagens, i.e. um AR(k) com as probabilidades:

$$\mathbb{P}[S_{t-k+1}, \dots, S_t, S_{t+1} | S_T] = \frac{\mathbb{P}[S_{t-k+2}, \dots, S_t, S_{t+1} | S_T] \mathbb{P}[S_{t-k+1}, \dots, S_t, S_{t+1} | S_T]}{\mathbb{P}[S_{t-k+2}, \dots, S_t, S_{t+1} | S_T]}$$

### 2.3.1 Resolvendo o problema do regime não observado

Para escrever a função de densidade de  $y_t$  dada as informações anteriores contidas em  $\boldsymbol{\psi}_{t-1}$ , será preciso calcular os valores preditos das variáveis não observadas  $S_t$  e  $S_{t-1}$ . Para resolver este problema, vamos usar uma estratégia semelhante a anterior:

1. Encontrar a função de densidade conjunta de  $y_t$ ,  $S_t$  e  $S_{t-1}$  em relação a informação passada  $\boldsymbol{\psi}_{t-1}$ :

$$f(y_t, S_t, S_{t-1} | \boldsymbol{\psi}_{t-1}) = f(y_t | S_t, S_{t-1}, \boldsymbol{\psi}_{t-1}) \mathbb{P}[S_t, S_{t-1} | \boldsymbol{\psi}_{t-1}]$$

2. Para obter a função densidade marginal de  $f(y_t | \boldsymbol{\psi}_{t-1})$  da densidade conjunta  $f(y_t; S_t; S_{t-1})$ ,

somamos a densidade conjunta por todos os valores de  $S_t$  e  $S_{t-1}$ :

$$f(y_t | \boldsymbol{\psi}_{t-1}) = \sum_{S_t=1}^M \sum_{S_{t-1}=1}^M f(y_t, S_t, S_{t-1} | \boldsymbol{\psi}_{t-1})$$

$$f(y_t | \boldsymbol{\psi}_{t-1}) = \sum_{S_t=1}^M \sum_{S_{t-1}=1}^M f(y_t | S_t, S_{t-1}, \boldsymbol{\psi}_{t-1}) \mathbb{P}[S_t, S_{t-1} | \boldsymbol{\psi}_{t-1}]$$

Desta forma a densidade marginal  $f(y_t | \boldsymbol{\psi}_{t-1})$  é ponderada pela média de  $M^2$  das densidades condicionais, com os pesos sendo  $\mathbb{P}[S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1}]$ , sendo  $i = 1, 2, \dots, M$  e  $j = 1, 2, \dots, M$ .

Então a função de log-verossimilhança será dada por:

$$\ln L = \sum_{t=1}^T \left\{ \sum_{S_t=1}^M \sum_{S_{t-1}=1}^M f(y_t | S_t, S_{t-1}, \boldsymbol{\psi}_{t-1}) \mathbb{P}[S_t, S_{t-1} | \boldsymbol{\psi}_{t-1}] \right\}$$

Assim, ainda precisaremos lidar com o problema de calcular  $\mathbb{P}[S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1}]$ .

### 2.3.2 Filtro de Probabilidades

O procedimento a seguir permite calcular  $\mathbb{P}[S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1}]$ :

1. Dado  $\mathbb{P}[S_{t-1} = i | \boldsymbol{\psi}_{t-1}]$ ,  $i = 1, 2, \dots, M$ , no começo do tempo  $t$  ou da iteração  $t$ , os pesos  $\mathbb{P}[S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1}]$  com  $i = 1, 2, \dots, M$  e  $j = 1, 2, \dots, M$  são calculados como:

$$\mathbb{P}[S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1}] = \mathbb{P}[S_t = j | S_{t-1} = i] \mathbb{P}[S_{t-1} = i | \boldsymbol{\psi}_{t-1}]$$

Onde  $\mathbb{P}[S_t = j | S_{t-1} = i]$ , com  $i = 1, 2, \dots, M$  e  $j = 1, 2, \dots, M$ , são as probabilidades de transição de regime.

2. Após encontrar  $y_t$  no final de  $t$ , pode-se atualizar as probabilidades dos termos:

$$\begin{aligned} \mathbb{P}[S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1}] &= \mathbb{P}[S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1}, y_t] \\ &= \frac{f(S_t = j, S_{t-1} = i, y_t | \boldsymbol{\psi}_{t-1})}{f(y_t | \boldsymbol{\psi}_{t-1})} \\ &= \frac{f(y_t | S_t = j, S_{t-1} = i, \boldsymbol{\psi}_{t-1}) f(S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1})}{\sum_{S_t=1}^M \sum_{S_{t-1}=1}^M f(y_t | S_t = j, S_{t-1} = i, \boldsymbol{\psi}_{t-1}) f(S_t = j, S_{t-1} = i | \boldsymbol{\psi}_{t-1})} \end{aligned}$$

Com:

$$\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_{t-1}] = \sum_{S_{t-1}=1}^M \mathbb{P}[S_t = j, S_{t-1} = i \mid \boldsymbol{\psi}_{t-1}]$$

Realizando o procedimento anterior para todos os períodos  $t$ , será possível obter os pesos. No modelo de 1ª ordem de *Markov Switching*, utiliza-se probabilidades no *steady-state* ( $\pi_1$ ) para  $t = 1$ :

$$\begin{aligned} \pi_0 &= \mathbb{P}[S_0 = 0 \mid \boldsymbol{\psi}_0] = \frac{1 - p_{11}}{2 - p_{00} - p_{11}} \\ \pi_1 &= \mathbb{P}[S_0 = 1 \mid \boldsymbol{\psi}_0] = \frac{1 - p_{00}}{2 - p_{00} - p_{11}} \end{aligned}$$

Assim, o modelo *Markov Switching* consiste em primeiramente estimar os parâmetros do modelo ao maximizar a função de máxima verossimilhança; e depois fazer inferências sobre os valores preditos de  $S_t$  por período  $t$ .

## 2.4 Questões relacionadas ao *Markov Switching*

### 2.4.1 Algoritmo de Suavização de Kim

A partir dos parâmetros estimados no modelos, é possível fazer inferências de  $S_t$  utilizando as informações da amostra. Assim, calculamos a probabilidade suavizada  $\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_T]$  ao invés de  $\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_T]$ , sendo  $t = 1, 2, \dots, T$ . Para exemplificar, a explicação esta apresentada para o modelo AR(1), i.e., uma defasagem. Considerando que a probabilidade conjunta de que  $S_t = j$  e  $S_{t+1} = k$  são baseadas em todas as informações:

$$\begin{aligned} \mathbb{P}[S_t = j, S_{t+1} = k \mid \boldsymbol{\psi}_T] &= \mathbb{P}[S_{t+1} = k \mid \boldsymbol{\psi}_T] \mathbb{P}[S_t = j \mid S_{t+1} = k, \boldsymbol{\psi}_T] \\ \mathbb{P}[S_t = j, S_{t+1} = k \mid \boldsymbol{\psi}_T] &= \mathbb{P}[S_{t+1} = k \mid \boldsymbol{\psi}_T] \mathbb{P}[S_t = j \mid S_{t+1} = k, \boldsymbol{\psi}_T] \\ \mathbb{P}[S_t = j, S_{t+1} = k \mid \boldsymbol{\psi}_T] &= \frac{\mathbb{P}[S_{t+1} = k \mid \boldsymbol{\psi}_T] \mathbb{P}[S_t = j \mid S_{t+1} = k, \boldsymbol{\psi}_T]}{\mathbb{P}[S_{t+1} = k \mid \boldsymbol{\psi}_T]} \\ \mathbb{P}[S_t = j, S_{t+1} = k \mid \boldsymbol{\psi}_T] &= \frac{\mathbb{P}[S_{t+1} = k \mid \boldsymbol{\psi}_T] \mathbb{P}[S_t = j \mid \boldsymbol{\psi}_T] \mathbb{P}[S_{t+1} = k \mid S_t = j, \boldsymbol{\psi}_T]}{\mathbb{P}[S_{t+1} = k \mid \boldsymbol{\psi}_T]} \end{aligned}$$

E temos que:

$$\mathbb{P}[S_t = j \mid \boldsymbol{\psi}_T] = \sum_{k=1}^M \mathbb{P}[S_t = j, S_{t+1} = k \mid \boldsymbol{\psi}_T]$$

A partir de  $\mathbb{P}[S_T | \boldsymbol{\psi}_T]$  da última iteração no filtro realizado, é possível iterar para  $t = T - 1, T - 2, \dots, 1$ , para calcular as probabilidade suavizadas  $\mathbb{P}[S_t | \boldsymbol{\psi}_T]$  para  $t = T - 1, T - 2, \dots, 1$ . Definindo  $\tilde{\mathbf{h}}_{t+1,T} = (y_{t+1}, y_{t+2}, \dots, y_T)^T \forall T > t$ , temos:

$$\begin{aligned} \mathbb{P}[S_t = j | S_{t+1} = k, \boldsymbol{\psi}_T] &= \mathbb{P}[S_t = j | S_{t+1} = k, \tilde{\mathbf{h}}_{t+1,T}, \boldsymbol{\psi}_T] \\ \mathbb{P}[S_t = j | S_{t+1} = k, \boldsymbol{\psi}_T] &= \frac{f(S_t = j, \tilde{\mathbf{h}}_{t+1,T} | S_{t+1} = k, \boldsymbol{\psi}_T)}{f(\tilde{\mathbf{h}}_{t+1,T} | S_{t+1} = k, \boldsymbol{\psi}_T)} \\ \mathbb{P}[S_t = j | S_{t+1} = k, \boldsymbol{\psi}_T] &= \frac{f(S_t = j | S_{t+1} = k, \boldsymbol{\psi}_T) f(\tilde{\mathbf{h}}_{t+1,T} | S_t = j, S_{t+1} = k, \boldsymbol{\psi}_T)}{f(\tilde{\mathbf{h}}_{t+1,T} | S_{t+1} = k, \boldsymbol{\psi}_T)} \\ \mathbb{P}[S_t = j | S_{t+1} = k, \boldsymbol{\psi}_T] &= \mathbb{P}[S_t = j | S_{t+1} = k, \boldsymbol{\psi}_T] \end{aligned}$$

A equação  $f(\tilde{\mathbf{h}}_{t+1,T} | S_{t+1} = k, S_t = j, \boldsymbol{\psi}_T) = f(\tilde{\mathbf{h}}_{t+1,T} | S_{t+1} = k, \boldsymbol{\psi}_T)$  indica que se  $S_{t+1}$  for conhecido, então  $y_{t+1}$  não terá informação sobre  $S_t$  além das contidas em  $S_{t+1}$  e  $\pi_t$ .

#### 2.4.2 Probabilidades no Estado de Equilíbrio para começar o Filtro

Sendo  $\mathbf{P}$  a matriz de probabilidades de 1ª Ordem em um processo *Markov Switching* com  $M$  estados:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \dots & p_{M1} \\ p_{12} & p_{22} & \dots & p_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M} & p_{2M} & \dots & p_{MM} \end{bmatrix}$$

Com  $\mathbf{i}_M^T \mathbf{P} = \mathbf{i}_M$  com  $\mathbf{i}_M = [1 \ 1 \ \dots \ 1]^T$ . E  $\pi_t$  um vetor  $M \times 1$  probabilidades no estado de equilíbrio:

$$\pi_t = \begin{bmatrix} \mathbb{P}[S_t = 1] \\ \mathbb{P}[S_t = 2] \\ \vdots \\ \mathbb{P}[S_t = M] \end{bmatrix} = \begin{bmatrix} \pi_{1t} \\ \pi_{2t} \\ \vdots \\ \pi_{Mt} \end{bmatrix}$$

Como:

$$\left. \begin{aligned} \pi_{t+1} &= \mathbf{P}\pi_t \\ \pi_{t+1} &= \pi_t \end{aligned} \right\} (\mathbf{I}_M - \mathbf{P})\pi_t = \mathbf{0}_M$$

Então:

$$\begin{bmatrix} I_M - P \\ i_M^T \end{bmatrix} \pi_t = \begin{bmatrix} 0_M \\ 1 \end{bmatrix} \Rightarrow A\pi_t = \begin{bmatrix} 0_M \\ 1 \end{bmatrix}$$

$$\pi_t = (A^T A)^{-1} A^T \begin{bmatrix} 0_M \\ 1 \end{bmatrix}$$

### 2.4.3 Expectativa de duração dos Regimes no modelo *Markov Switching*

A partir da diagonal principal da matriz de probabilidade de transições, é possível calcular a duração esperada ( $D$ ) em cada regime (períodos esperados em que permanecerá em cada regime):

$$D = 1, \text{ se } S_t = j \text{ e } S_{t+1} \neq j; \mathbb{P}[D = 1] = (1 - p_{jj})$$

$$D = 2, \text{ se } S_t = S_{t+1} = j \text{ e } S_{t+2} \neq j; \mathbb{P}[D = 2] = p_{jj}(1 - p_{jj})$$

$$D = 3, \text{ se } S_t = S_{t+1} = S_{t+2} = j \text{ e } S_{t+3} \neq j; \mathbb{P}[D = 3] = p_{jj}^2(1 - p_{jj})$$

$$D = 4, \text{ se } S_t = S_{t+1} = S_{t+2} = S_{t+3} = j \text{ e } S_{t+4} \neq j; \mathbb{P}[D = 4] = p_{jj}^3(1 - p_{jj})$$

⋮

Então:

$$\mathbb{E}(D) = \sum_{j=1}^{\infty} j \mathbb{P}[D = j]$$

$$\mathbb{E}(D) = 1\mathbb{P}[D = 1] + 2\mathbb{P}[D = 2] + 3\mathbb{P}[D = 3] + 4\mathbb{P}[D = 4] + \dots$$

$$\mathbb{E}(D) = 1\mathbb{P}[S_{t+1} \neq j \mid S_t = j] + 2\mathbb{P}[S_{t+1} = j, S_{t+2} \neq j \mid S_t = j]$$

$$+ 3\mathbb{P}[S_{t+1} = S_{t+2} = j, S_{t+3} \neq j \mid S_t = j]$$

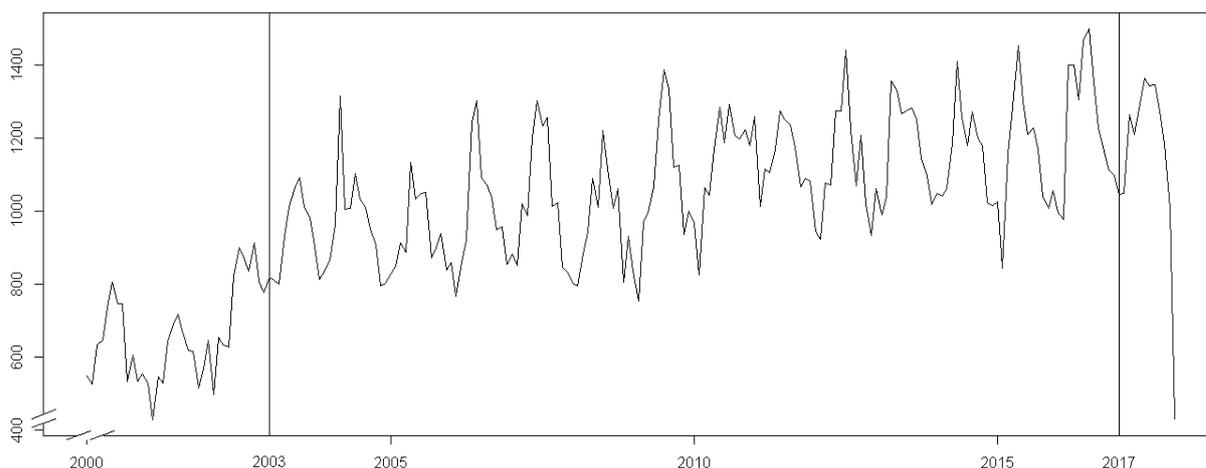
$$+ 4\mathbb{P}[S_{t+1} = S_{t+2} = S_{t+3} = j, S_{t+4} \neq j \mid S_t = j] + \dots$$

$$\mathbb{E}(D) = 1(1 - p_{jj}) + 2p_{jj}(1 - p_{jj}) + 3p_{jj}^2(1 - p_{jj}) + 4p_{jj}^3(1 - p_{jj}) + \dots$$

$$\mathbb{E}(D) = \frac{1}{1 - p_{jj}}$$

### 3 BASE DE DADOS

Os dados de internações por doenças respiratórias para pessoas com 60 anos ou mais da cidade de São Paulo utilizados neste trabalho foram extraídos do Portal da Saúde - DataSUS<sup>2</sup>, sítio do Governo Federal com dados para o Município de São Paulo. Por meio dele, acessamos os dados mensais de internações de janeiro de 2000 a dezembro 2017.



Fonte: DataSUS Elaboração Própria

Figura 2: Número mensal de internações de pessoas com mais de 60 anos em São Paulo

Ao analisar a série histórica disponível, observamos que há uma quebra estrutural em 2002, levando a uma alteração no nível da série; além de algum problema de falta de dados no final de 2017. Então decidimos trabalhar com os dados mensais entre janeiro de 2003 a dezembro de 2016. Seguem abaixo as estatísticas descritivas da série de número de internações utilizada neste estudo.

Tabela 1: Estatísticas Descritivas do Número de internações mensais de pessoas com mais de 60 anos devido a doenças respiratórias em São Paulo

<b>Número de observações</b>	<b>Média</b>	<b>Desvio Padrão</b>	<b>Erro Padrão</b>		
168	1075.315	171.114	13.202		
<b>Mínimo</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>3º Quartil</b>	<b>Máximo</b>	
753	948	1061	1210	1501	

<sup>2</sup><http://www2.datasus.gov.br/DATASUS/index.php?area=02>

## 4 METODOLOGIA APLICADA

Com o objetivo de modelar as internações por doenças respiratórias de pessoas com mais de 60 anos no município de São Paulo, este trabalho seguiu a metodologia apresentada na revisão bibliográfica que estimou um modelo GLM com distribuição de Poisson e posterior aplicação do modelo de *Markov Switching* para capturar as trocas de regimes.

### 4.1 Modelo com efeitos sazonais mensais

Primeiramente foi estimado um modelo GLM com variáveis explicativas de tendência e *dummies* sazonais de mês sendo que o número de hospitalizações ( $hosp_t$ ) segue distribuição Poisson com média  $\mu_t$ :

$$\begin{aligned} \ln(\mu_t) = & \alpha + \beta_1 t + \beta_2 fevereiro_t + \beta_3 março_t + \beta_4 abril_t \\ & + \beta_5 maio_t + \beta_6 junho_t + \beta_7 julho_t + \beta_8 agosto_t \\ & + \beta_9 setembro_t + \beta_{10} outubro_t + \beta_{11} novembro_t \\ & + \beta_{12} dezembro_t + e_t \end{aligned}$$

Na equação acima:  $t$  é índice que determina o período vigente. O coeficiente  $\alpha$  indica o coeficiente linear;  $\beta_1$  o coeficiente angular; e os meses indicam *dummies*, sendo 1 se o mês corresponda ao mês da *dummy*, e 0 caso contrário.

Posteriormente foi estimado um *Markov Switching* para verificar a presença de 2 regimes na série:

$$\begin{aligned} \ln(\mu_t) = & \alpha_{S_t} + \beta_{1,S_t} t + \beta_{2,S_t} fevereiro_t + \beta_{3,S_t} março_t + \beta_{4,S_t} abril_t \\ & + \beta_{5,S_t} maio_t + \beta_{6,S_t} junho_t + \beta_{7,S_t} julho_t + \beta_{8,S_t} agosto_t \\ & + \beta_{9,S_t} setembro_t + \beta_{10,S_t} outubro_t + \beta_{11,S_t} novembro_t \\ & + \beta_{12,S_t} dezembro_t + e_t, \quad t = 1, 2, \dots, T \end{aligned}$$

Na equação acima  $S_t$  é índice que determina o regime vigente. Então, os coeficiente variam de acordo com o regime vigente:  $\alpha_{S_t}$  indica o coeficiente linear em cada regime;  $\beta_{1,S_t}$  o coeficiente angular; e os meses indicam *dummies*, sendo 1 se o mês corresponda ao mês da *dummy*, e 0 caso contrário. Assim temos que:

$$\begin{cases} t = 1, 2, \dots, T \\ S_t = 0 \text{ ou } 1 \end{cases}$$

$$\beta_{S_t} = \beta_0(1 - S_t) + \beta_1 S_t$$

## 4.2 Modelo com variável defasada

Também foram estimados modelos com tendência e uma defasagem de 1 período das hospitalizações ( $hosp_{t-1}$ ). O modelo GLM com distribuição de Poisson:

$$\ln(\mu_t) = \alpha + \beta_1 t + \beta_2 hosp_{t-1} + e_t, \quad t = 1, 2, \dots, T$$

Assim, o coeficiente  $\alpha_{S_t}$  indica o coeficiente linear;  $\beta_1$  o coeficiente angular; e o  $\beta_2$  será o coeficiente em relação ao efeito da variável de internações com 1 defasagem.

Ao estimar um *Markov Switching*, temos:

$$\ln(\mu_t) = \alpha_{S_t} + \beta_{1,S_t} t + \beta_{2,S_t} hosp_{t-1} + e_t$$

Mantendo a notação anterior,  $t$  e  $S_t$  são índices que determinam, respectivamente, o período e o regime vigente. O coeficiente  $\alpha_{S_t}$  indica o coeficiente linear;  $\beta_{1,S_t}$  o coeficiente angular; e o  $\beta_{2,S_t}$  será o coeficiente em relação ao efeito da variável de internações com 1 defasagem.

$$\begin{cases} t = 1, 2, \dots, T \\ S_t = 0 \text{ ou } 1 \end{cases}$$

$$\beta_{S_t} = \beta_0(1 - S_t) + \beta_1 S_t$$

Portanto, sob o regime 0 o parâmetro estimado será  $\beta_0$ . Analogamente, sob o regime 1 o parâmetro estimado será  $\beta_1$ . Se as quebras estruturais  $S_t$  fossem conhecidas, seria apenas um modelo GLM com variáveis *dummies*. Entretanto, neste caso será preciso estimar  $S_t$ , e a função de log-verossimilhança será dada por:

$$\ln L = \sum_{t=1}^T \ln(f(y_t | S_t))$$

$$f(y_t | S_t) = \frac{e^{-e^{x_t \beta_{S_t}}} (e^{x_t \beta_{S_t}})^y}{y!}$$

Para realizar estas estimações, tanto o GLM como o *Markov Switching*, foi utilizada a *library* MSwM.

## 5 RESULTADOS

### 5.1 Modelo com efeitos sazonais mensais

Nesta primeira subsecção, são apresentados os resultados para os modelos que possuem como variáveis explicativas: a tendência e as 11 *dummies* de mês. O modelo tem como base o mês de janeiro, foi retirada a *dummy* referente a este mês para não incorrer em problema de multicolinearidade.

As estimativas dos parâmetros<sup>3</sup> estão na Tabela 2, na primeira coluna estão as estimativas do modelo GLM com distribuição de Poisson; na segunda coluna estão os resultados para o primeiro regime do modelo de *Markov Switching*; e na terceira coluna os estimadores para o segundo regime. Para cada variável temos o valor do coeficiente em cima, e entre parênteses o erro padrão. No final da tabela, são apresentados os valores do Critério de Informação de Akaike (AIC).

---

<sup>3</sup>Os resultados completos da estimação estão presentes no Anexo.

Tabela 2: Estimativas e erros padrão para os modelos GLM e *Markov Switching* GLM

	GLM	<i>Markov Switching</i> GLM	
		Regime 0	Regime 1
Coefficiente linear	6.640*** (0.011)	6.620*** (0.012)	6.895*** (0.031)
Coefficiente angular	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
fevereiro	-0.064*** (0.013)	0.034. (0.018)	-0.409. (0.033)
março	0.103*** (0.012)	0.079*** (0.013)	0.086* (0.034)
abril	0.131*** (0.012)	0.258*** (0.017)	-0.177*** (0.030)
maio	0.240*** (0.012)	0.318*** (0.015)	-0.079* (0.031)
junho	0.256*** (0.012)	0.324*** (0.014)	-0.079* (0.032)
julho	0.253*** (0.012)	0.218*** (0.015)	0.067* (0.031)
agosto	0.225*** (0.012)	0.221*** (0.013)	0.067* (0.033)
setembro	0.133*** (0.012)	0.198*** (0.018)	-0.174*** (0.032)
outubro	0.104*** (0.012)	0.098*** (0.015)	-0.082* (0.037)
novembro	0.018 (0.012)	0.137*** (0.017)	-0.301*** (0.030)
dezembro	0.003 (0.012)	-0.003 (0.015)	-0.131** (0.043)

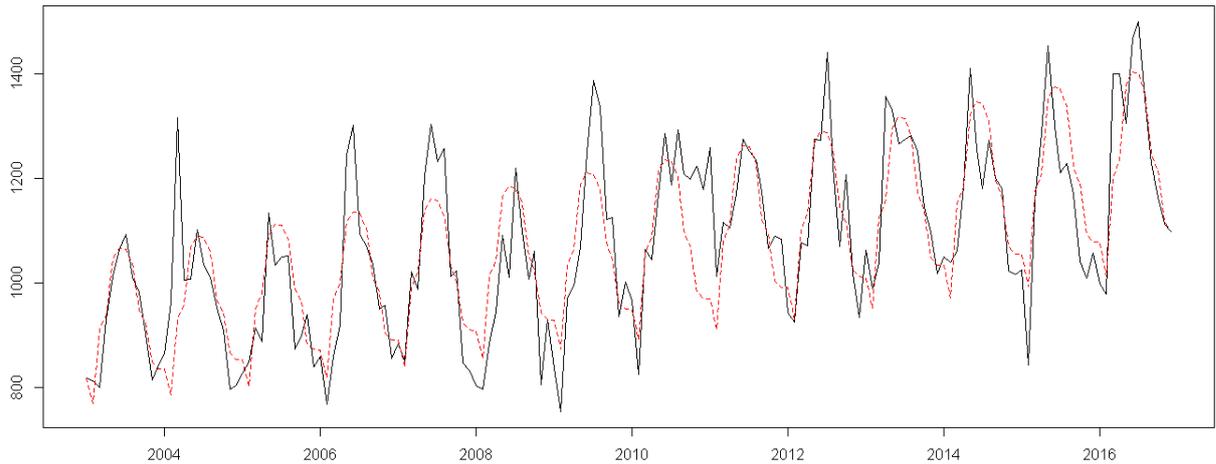
Valor-p: \*\*\* se  $p < 0.001$ ; \*\* se  $p < 0.01$ ; \* se  $p < 0.05$ ; . se  $p < 0.1$

	GLM	<i>Markov Switching</i>
AIC	2697	2082

É interessante notar na Tabela 2 que os coeficientes das *dummies* para os meses de novembro e dezembro não apresentam significância estatística no modelo GLM, entretanto se tornam significantes no modelo de *Markov Switching*. Assim como o esperado, o modelo de MS apresenta um valor para a estatística AIC menor que o modelo GLM.

Comparando os dois regimes do modelo MS, observamos que o regime 0 apresenta, em geral, coeficientes maiores para para *dummies* mensais (com exceção para o mês de março). Enquanto o regime 1 possui um maior coeficiente linear.

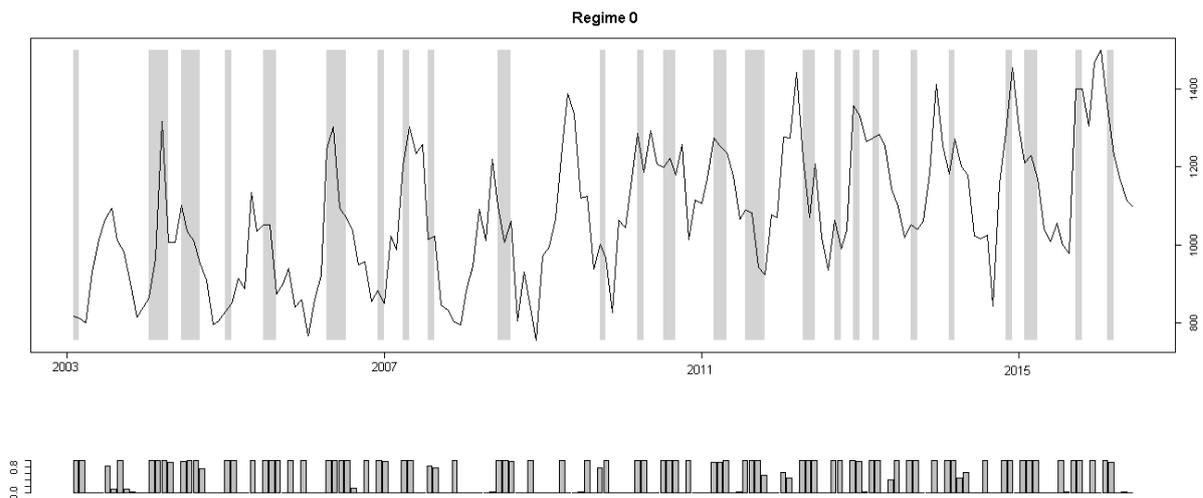
Na Figura 3, é possível comparar os dados da série original de internações (linha contínua em preto) com a série estimada pelo modelo GLM com distribuição de Poisson (linha tracejada em vermelho). Podemos observar que o modelo, conforme o esperado, capta os movimentos de tendência e sazonalidade propostos pela especificação.



Fonte: DataSUS Elaboração Própria

Figura 3: Número mensal de internações observado e predito pelo pelo modelo GLM com efeitos sazonais

Enquanto na Figura 4 apresentamos a série de internações em contraste com o regime 0 na região escura; sendo assim, a região clara corresponde ao regime 1.



Fonte: DataSUS Elaboração Própria

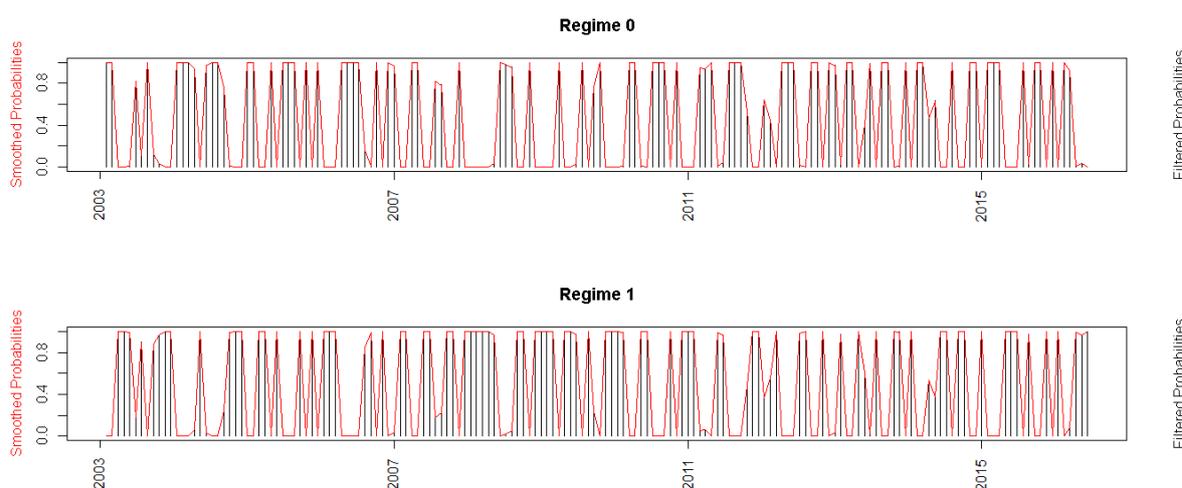
Figura 4: Quantidade de internações em relação ao Regime 0 (estimado pelo modelo *Markov Switching* com efeitos sazonais)

A tabela 3 apresenta a matriz de transição entre regimes: cada linha corresponde a um regime inicial, enquanto em cada coluna são apresentadas as probabilidades de regime para o próximo período, de forma a totalizar 1 na soma da linha. Ou seja, o quando a série esta no regime 0, ela tem chance 0.55 de permanecer neste regime, contra 0.45 de trocar de regime; enquanto ao estar no regime 1 ela possui uma probabilidade de 0.74 de retornar ao regime 0.

Tabela 3: Matriz de Probabilidades de Transição no modelo *Markov Switching* com efeitos sazonais**Probabilidade de Transição:**

	Regime 0	Regime 1
Regime 0	0.55	0.45
Regime 1	0.74	0.26

Já a figura 5 apresenta os gráficos de probabilidade suavizada para cada um dos regimes. Neste caso, como são apenas 2 regimes, as probabilidades são complementares.



Fonte: DataSUS Elaboração Própria

Figura 5: Probabilidades Suavizadas pelo modelo *Markov Switching* com efeitos sazonais

Tanto a tabela 3 quanto a figura 5 indicam que os regimes não são tão bem identificados, pois as probabilidades de permanecer no mesmo regime são baixas e há frequentes alterações de regimes.

## 5.2 Modelo com variável defasada

Esta subsecção terá como foco os modelos que possuem como variáveis explicativas: a tendência e a variável de interações com 1 defasagem. Assim como nos na subsecção anterior, serão apresentados as estimativas dos parâmetros<sup>4</sup> na tabela 4. Na primeira coluna estão as estimativas do modelo GLM com distribuição de Poisson; na segunda coluna estão os resultados para o regime 0 do modelo de *Markov Switching*; e na terceira coluna os estimadores para o regime 1. Para cada variável temos o valor do coeficiente em cima, e entre parênteses o erro padrão. No final da tabela, são apresentados os valores do Critério de Informação de Akaike (AIC).

<sup>4</sup>Os resultados completos da estimação estão presentes no Anexo.

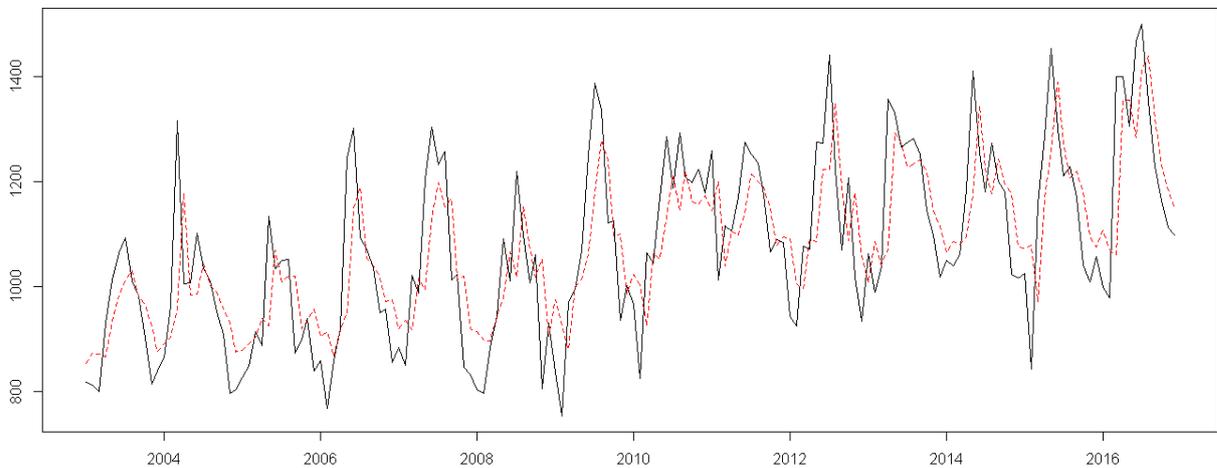
Analisando a tabela 4, é interessante notar que as variáveis de tendência e a defasagem das internações possuem coeficientes bem próximos tanto no GLM como no *Markov Switching*. E conforme o esperado, o modelo *Markov Switching* apresenta um valor para a estatística AIC menor que o modelo GLM.

Tabela 4: Estimativas e erros padrão para os modelos GLM e *Markov Switching* GLM

	GLM	<i>Markov Switching</i>	
		Regime 0	Regime 1
Coefficiente linear	2.470*** (0.119)	2.411*** (0.368)	3.956*** (0.258)
Coefficiente angular	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
$\ln(hosp_{-1})$	0.636*** (0.018)	0.637*** (0.055)	0.436*** (0.038)
Valor-p: *** se $p < 0.001$ ; ** se $p < 0.01$ ; * se $p < 0.05$ ; . se $p < 0.1$			

	GLM	<i>Markov Switching</i>
AIC	3393	2370

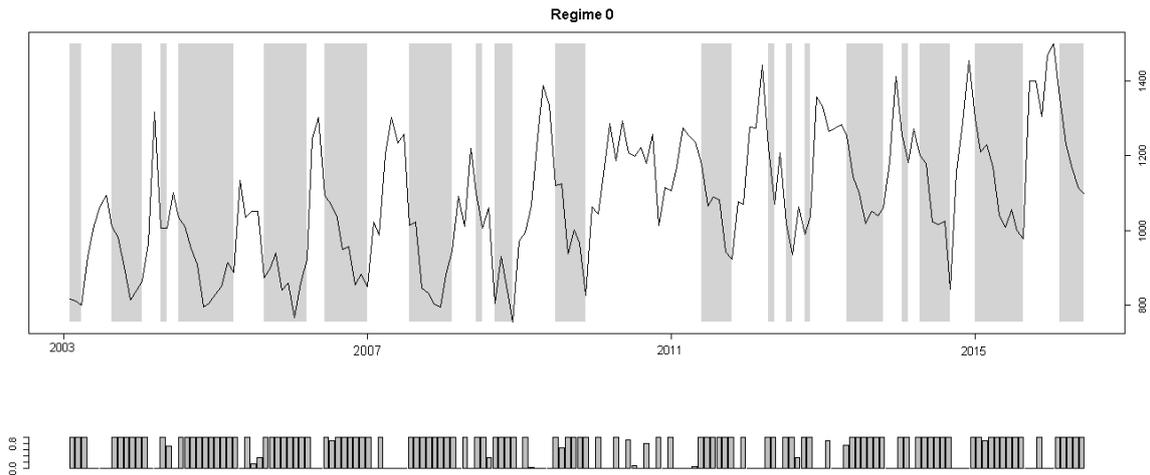
A Figura 6 apresenta os dados da série original de internações (linha contínua em preto) com a série estimada pelo modelo GLM com distribuição de Poisson (linha tracejada em vermelho).



Fonte: DataSUS Elaboração Própria

Figura 6: Número mensal de internações observado e predito pelo pelo modelo GLM com variável defasada

Enquanto na Figura 7 apresentamos a série de internações em contraste com o regime 0 na região escura; sendo assim, a região clara corresponde ao regime 1. Em comparação com o modelo *Markov Switching* com efeitos sazonais, este modelo apresenta menos transições entre os regimes, indicando uma maior estabilidade e uma melhor identificação. O modelo MS identificou os regimes pela sazonalidade, em que os períodos de diminuição de internações estão no regime 0; enquanto os períodos de alta estão no contidos no regime 1.



Fonte: DataSUS Elaboração Própria

Figura 7: Quantidade de internações em relação ao Regime 0 (estimado pelo modelo *Markov Switching* com variável defasada)

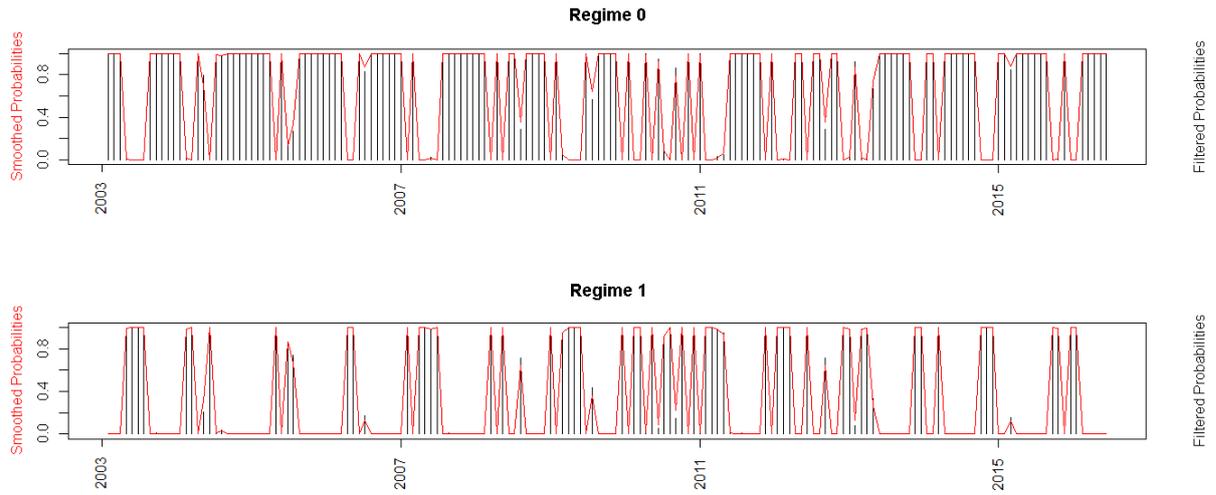
A Tabela 5 apresenta a matriz de probabilidades de transição entre regimes: conforme o esperado, o regime 0 apresenta uma maior estabilidade, pois possui 0.73 de chance de permanecer neste regime; justificando a figura 7.

Tabela 5: Matriz de Probabilidades de Transição no modelo *Markov Switching* com variável defasada

**Probabilidade de Transição:**

	Regime 0	Regime 1
Regime 0	0.73	0.27
Regime 1	0.53	0.47

A Figura 8 apresenta os gráficos de probabilidade suavizada para cada um dos regimes. É interessante notar que há um predomínio do regime 0, que é explicado pelas probabilidades de transição da Tabela 5. Na tabela vemos que independente do regime a série possui uma maior probabilidade de permanecer/ou mudar para o regime 0. Neste caso, como são apenas 2 regimes, as probabilidades são complementares.



Fonte: DataSUS Elaboração Própria

Figura 8: Probabilidades Suavizadas pelo modelo *Markov Switching* com variável defasada

## 6 CONSIDERAÇÕES FINAIS

O modelos estimados para estudar a quantidade mensal de internações hospitalares de pessoas com mais de 60 anos devido a doenças respiratórias apresentaram bons resultados, com as séries estimadas registrando boa aderência à série original.

Os modelos que levam em conta os fatores sazonais apresentaram um resultado mais satisfatório do que os modelos com variável defasada. A explicação mais plausível é a forte sazonalidade nas doenças respiratórias, devido ao clima frio e seco do inverno que intensifica os efeitos da poluição.

Seguindo o critério de informação Akaike, os modelos com regimes são preferíveis aos modelos apenas um regime. Além disso, é interessante notar que os modelos com efeitos sazonais também são preferíveis ao respectivo modelo com variável defasada, isto é, comparando os modelos GLM entre si; e ao comparar os modelos *Markov Switching* GLM.

É importante notar que o modelo com variável defasada apresenta uma maior estabilidade, indicando que ele consegue identificar melhor os regimes da série. Entretanto, ele capta o padrão sazonal aos identificar os regimes.

## Referências

- Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 149–192.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605.
- Farley, J. U. and Hinich, M. J. (1970). A test for a shifting slope coefficient in a linear model. *Journal of the American Statistical Association*, 65(331):1320–1329.
- Goldfeld, S. M. and Quandt, R. E. (1973). A markov model for switching regressions. *Journal of econometrics*, 1(1):3–15.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.
- Kim, Chang-Jin; Nelson, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, volume 1 of *MIT Press Books*. The MIT Press.
- Kim, H.-J. and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, 76(3):409–423.
- Kim, I. and Maddala, G. (1991). Multiple structural breaks and unit roots in the nominal and real exchange rates. *Unpublished manuscript, University of Florida, Department of Economics*.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.
- Ploberger, W., Krämer, W., and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40(2):307–318.
- Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American statistical association*, 53(284):873–880.
- Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American statistical Association*, 55(290):324–330.
- Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American statistical association*, 67(338):306–310.

## A ANEXO

### A.1 Resultados Completos dos Modelos com efeitos sazonais

Tabela 6: Modelo GLM Poisson com efeitos sazonais

**Call:**

```
glm(formula = hosp trend + fevereiro + março + abril + maio
+ junho + julho + agosto + setembro + outubro + novembro
+ dezembro, family = "poisson")
```

**Deviance Residuals:**

Min	1Q	Median	3Q	Max
-5.225	-1.77	-0.464	1.358	11.957

<b>Coefficients:</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
Coeficiente linear	6.64E+00	1.05E-02	634.51	< 2e-16 ***
Coeficiente angular	1.77E-03	4.87E-05	36.33	< 2e-16 ***
fevereiro	-6.36E-02	1.25E-02	-5.09	3.7e-07 ***
março	1.03E-01	1.20E-02	8.61	< 2e-16 ***
abril	1.31E-01	1.19E-02	11	< 2e-16 ***
maio	2.40E-01	1.16E-02	20.61	< 2e-16 ***
junho	2.56E-01	1.16E-02	22.13	< 2e-16 ***
julho	2.53E-01	1.16E-02	21.8	< 2e-16 ***
agosto	2.25E-01	1.16E-02	19.34	< 2e-16 ***
setembro	1.33E-01	1.19E-02	11.22	< 2e-16 ***
outubro	1.04E-01	1.20E-02	8.68	< 2e-16 ***
novembro	1.75E-02	1.22E-02	1.43	0.15
dezembro	2.57E-03	1.23E-02	0.21	0.83

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4537.0 on 167 degrees of freedom

Residual deviance: 1191.3 on 155 degrees of freedom

AIC: 2697

Number of Fisher Scoring iterations: 4

Tabela 7: Modelo *Markov Switching* GLM Poisson com efeitos sazonais**Markov Switching Model****Call:**

```
msmFit(object = model, k = 2, sw = rep(TRUE, 13), family = "poisson",
control = list(parallel = FALSE))
```

<b>AIC</b>	<b>BIC</b>	<b>logLik</b>
2082	2297	-1015

**Coefficients:**

<b>Regime 0</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
Coefficiente linear(S)	6.620	0.012	551.67	< 2e-16 ***
Coefficiente angular(S)	0.002	0.000	Inf	< 2e-16 ***
fevereiro(S)	0.034	0.018	1.89	0.059 .
março(S)	0.079	0.013	6.08	1.2e-09 ***
abril(S)	0.258	0.017	15.18	< 2e-16 ***
maio(S)	0.318	0.015	21.20	< 2e-16 ***
junho(S)	0.324	0.014	23.14	< 2e-16 ***
julho(S)	0.218	0.015	14.53	< 2e-16 ***
agosto(S)	0.221	0.013	17.00	< 2e-16 ***
setembro(S)	0.198	0.018	11.00	< 2e-16 ***
outubro(S)	0.098	0.015	6.53	6.4e-11 ***
novembro(S)	0.137	0.017	8.06	6.7e-16 ***
dezembro(S)	-0.003	0.015	-0.20	0.841

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

<b>Regime 1</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
Coefficiente linear(S)	6.895	0.031	222.42	< 2e-16 ***
Coefficiente angular(S)	0.002	0.000	Inf	< 2e-16 ***
fevereiro(S)	-0.409	0.033	-12.39	< 2e-16 ***
março(S)	0.086	0.034	2.53	0.0114 *
abril(S)	-0.177	0.03	-5.9	3.6e-09 ***
maio(S)	-0.079	0.031	-2.55	0.0108 *
junho(S)	-0.079	0.032	-2.47	0.0135 *
julho(S)	0.067	0.031	2.16	0.0307 *
agosto(S)	0.067	0.033	2.03	0.0424 *
setembro(S)	-0.174	0.032	-5.44	5.4e-08 ***
outubro(S)	-0.082	0.037	-2.22	0.0267 *
novembro(S)	-0.301	0.03	-10.03	< 2e-16 ***
dezembro(S)	-0.131	0.043	-3.05	0.0023 **

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

<b>Transition probabilities:</b>	<b>Regime 0</b>	<b>Regime 1</b>
Regime 0	0.55	0.45
Regime 1	0.74	0.26

## A.2 Análises dos Resíduos do Modelo *Markov Switching* Poisson com efeitos sazonais

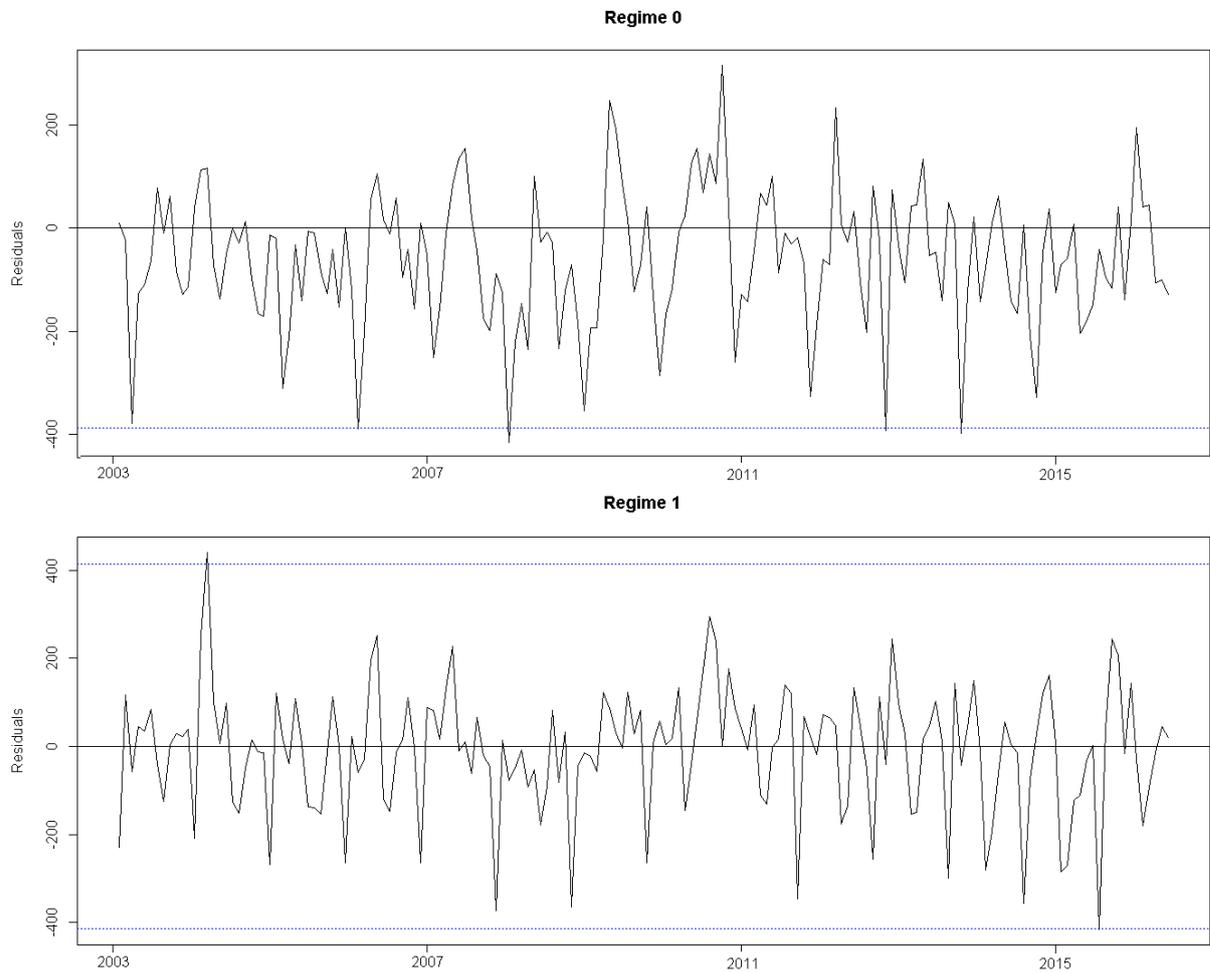


Figura 9: Resíduos do Modelo *Markov Switching* GLM Poisson com efeitos sazonais

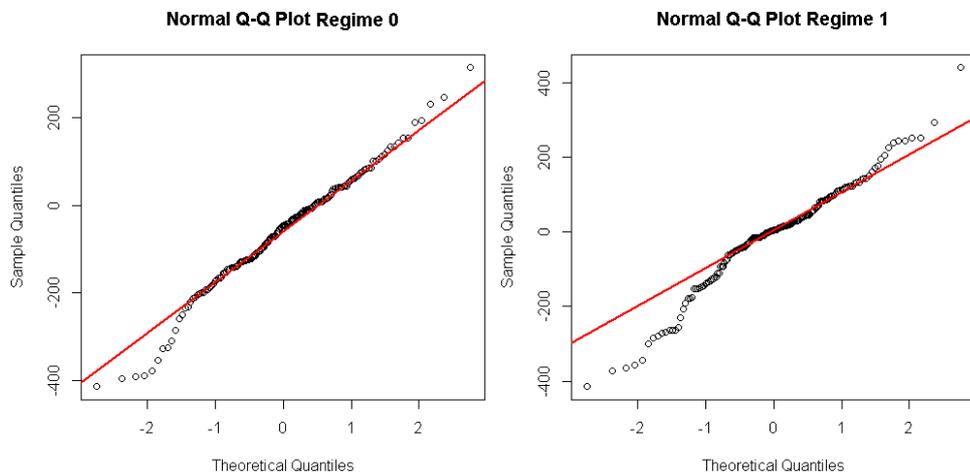


Figura 10: Normal Q-Q Plot do Modelo *Markov Switching* GLM Poisson com efeitos sazonais

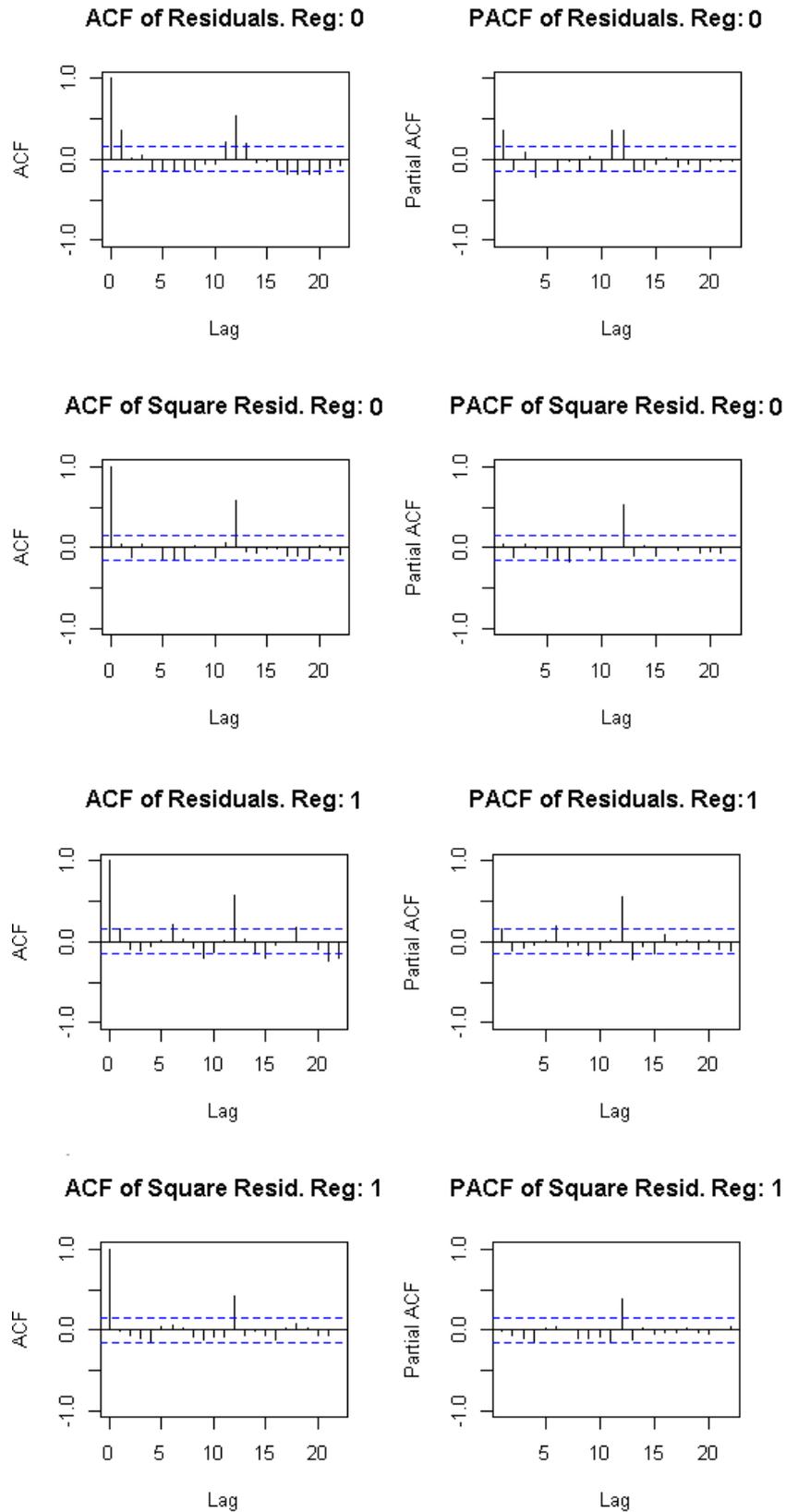


Figura 11: ACF/PACF dos Resíduos e dos Quadrados do Resíduo do Modelo *Markov Switching* GLM Poisson com efeitos sazonais

### A.3 Resultados Completos dos Modelos com variável defasada

Tabela 8: Modelo GLM Poisson com variável defasada

**Call:**

**glm(formula = hosp trend+ ln(hosp(-1)), family = "poisson")**

**Deviance Residuals:**

Min	1Q	Median	3Q	Max
-8.17	-2.081	-0.519	1.818	10.804

<b>Coefficients:</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
Coefficiente linear	2.470	0.119	20.7	<2e-16 ***
Coefficiente angular	0.001	0.000	10.4	<2e-16 ***
ln( <i>hosp</i> <sub>-1</sub> )	0.636	0.018	36.0	<2e-16 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4537.0 on 167 degrees of freedom

Residual deviance: 1893.8 on 165 degrees of freedom

AIC: 3379

Number of Fisher Scoring iterations: 4

Tabela 9: Modelo *Markov Switching* GLM Poisson com variável defasada**Markov Switching Model****Call:**

```
msmFit(object = model2, k = 2, sw = rep(TRUE, 3), family = "poisson",
control = list(parallel = FALSE))
```

<b>AIC</b>	<b>BIC</b>	<b>logLik</b>
2368	2417	-1178

**Coefficients:**

<b>Regime 0</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
Coefficiente linear(S)	2.411	0.368	6.55	<5.7e-11 ***
Coefficiente angular(S)	0.001	0.000	Inf	<2e-16 ***
$\ln(hosp_{-1})(S)$	0.637	0.055	11.58	<2e-16 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

<b>Regime 1</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
Coefficiente linear(S)	3.956	0.258	15.30	<2e-16 ***
Coefficiente angular(S)	0.436	0.038	11.50	<2e-16 ***
$\ln(hosp_{-1})(S)$	0	0	NA	NA

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

<b>Transition probabilities:</b>	<b>Regime 0</b>	<b>Regime 1</b>
Regime 0	0.73	0.27
Regime 1	0.53	0.47

## A.4 Análises dos Resíduos do Modelo *Markov Switching* Poisson com variável defasada

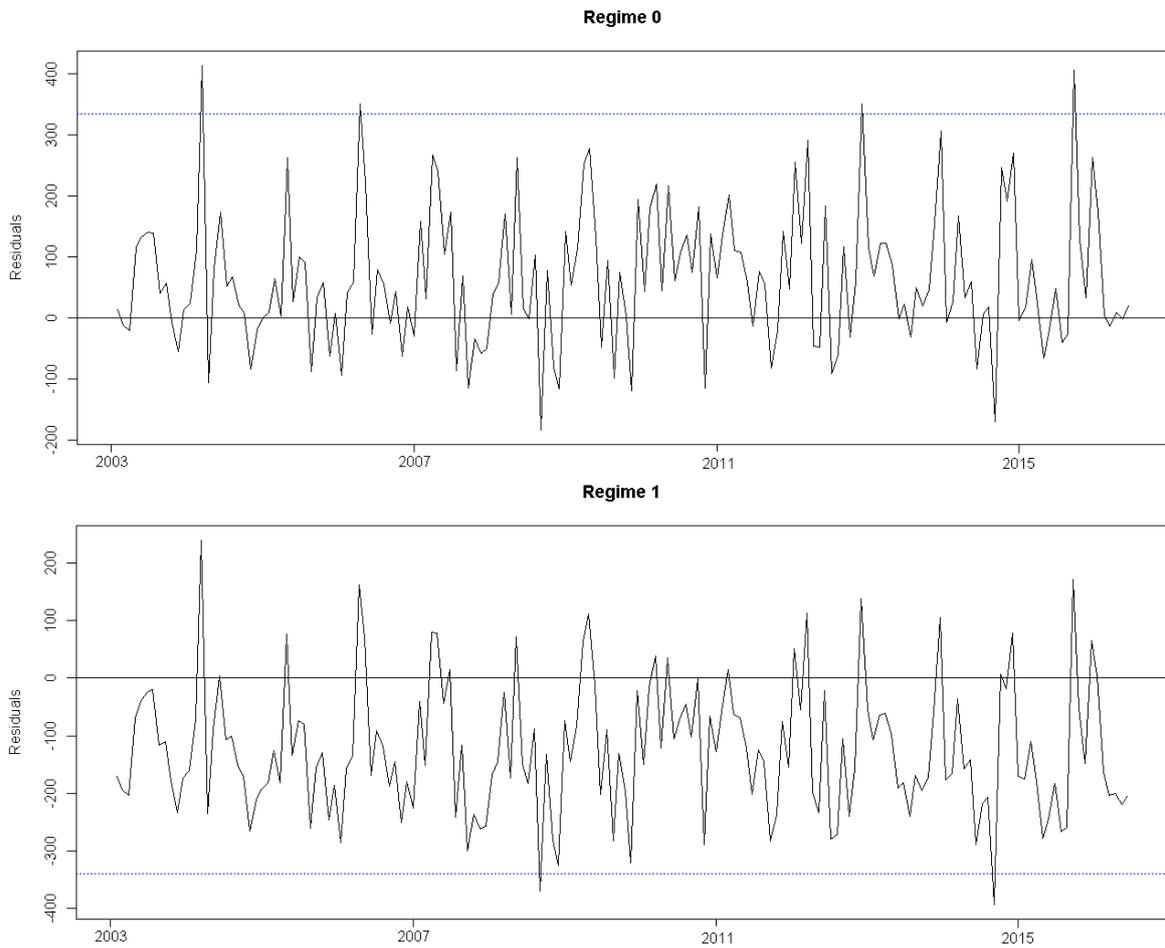


Figura 12: Resíduos do Modelo *Markov Switching* GLM Poisson com variável defasada

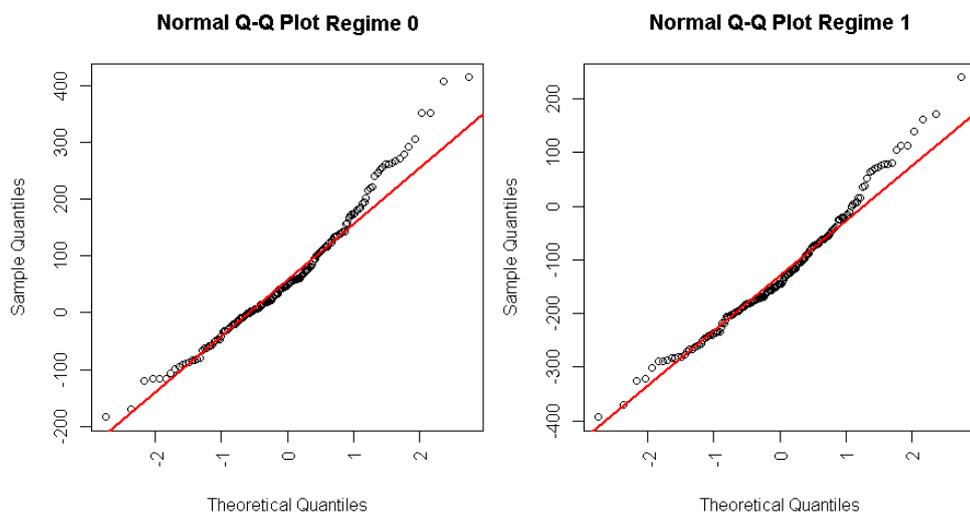


Figura 13: Normal Q-Q Plot do Modelo *Markov Switching* GLM Poisson com variável defasada

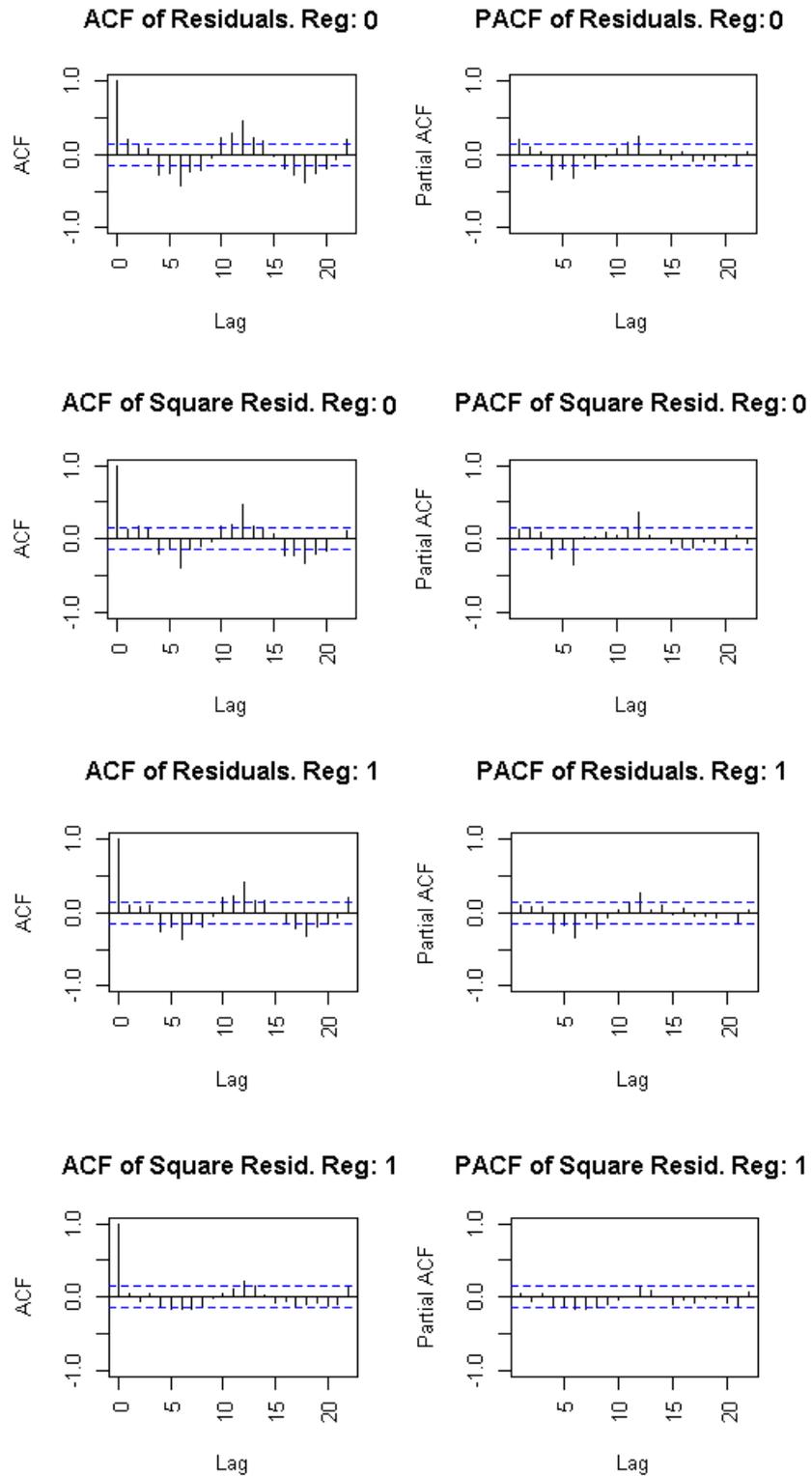


Figura 14: ACF/PACF dos Resíduos e dos Quadrados do Resíduo do Modelo *Markov Switching* GLM Poisson com variável defasada