



Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Programa de Pós-Graduação em Estatística

Modelos Dinâmicos para Séries Temporais Positivas

Jonas Hendler Carlos

Porto Alegre, Maio de 2022.

CIP - Catalogação na Publicação

Carlos, Jonas Hendler
Modelos Dinâmicos para Séries Temporais Positivas /
Jonas Hendler Carlos. -- 2022.
85 f.
Orientadora: Prof. Dra. Taiane Schaedler Prass.

Coorientador: Prof. Dr. Cleiton Guollo Taufemback.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Instituto de Matemática e
Estatística, Programa de Pós-Graduação em Estatística,
Porto Alegre, BR-RS, 2022.

1. Séries Temporais. 2. Dados Positivos. 3. Modelos
Dinâmicos. 4. Modelos Autoregressivos. 5. Máxima
Verossimilhança Parcial. I. Prass, Prof. Dra. Taiane
Schaedler, orient. II. Taufemback, Prof. Dr. Cleiton
Guollo, coorient. III. Título.

Dissertação submetida por Jonas Hendler Carlos como requisito parcial para a obtenção do título de Mestre em Estatística pelo Programa de Pós-Graduação em Estatística da Universidade Federal do Rio Grande do Sul.

Orientador(a):

Prof. Dra. Taiane Schaedler Prass

Co-orientador(a):

Prof. Dr. Cleiton Guollo Taufemback

Comissão Examinadora:

Prof. Dr. Fernando Arturo Peña Ramírez (UFSM)

Prof. Dr. Guilherme Pumi (PPGEst - UFRGS)

Prof. Dr. Marcio Valk (PPGEst - UFRGS)

Data de Apresentação: 31 de Maio de 2022

*“A menos que modifiquemos a nossa maneira de pensar,
não seremos capazes de resolver os problemas causados
pela forma como nos acostumamos a ver o mundo.”
(Albert Einstein)*

AGRADECIMENTOS

Agradeço a minha família, especialmente a meu pai Alício (in memoriam) que, mesmo com pouca escolaridade, sempre me incentivou a seguir buscando cada vez mais conhecimento. A minha mãe Maria, primeira professora e que foi a responsável pela minha alfabetização e que veio a falecer dias antes de concluir esta versão da dissertação. Assim como meus irmãos Edson e Rosimere, pela presença e importância em todos os momentos da minha vida.

A todos os professores do Programa de Pós-Graduação em Estatística da UFRGS pelos valiosos momentos de aprendizagem, em especial a minha orientadora Prof. Dra. Taiane Schaedler Prass e ao co-orientador Prof. Dr. Cleiton Guollo Taufemback.

Aos professores participantes da banca Prof. Dr. Fernando Arturo Peña Ramírez (UFSM), Prof. Dr. Guilherme Pumi (PPGEst - UFRGS) e Prof. Dr. Marcio Valk (PPGEst - UFRGS).

Ao meu amigo/irmão e brilhante estatístico Juscelino Zemiacki que foi quem me abriu as portas desta maravilhosa área que é a estatística.

Muito Obrigado!

RESUMO

Baseado em trabalhos de [Ferrari and Cribari-Neto \(2004\)](#), [Rocha and Cribari-Neto \(2009\)](#), [Pumi et al. \(2019b\)](#) e [Bourguignon et al. \(2021\)](#), este trabalho propõe a construção de modelos autoregressivos de médias móveis em que a resposta tem distribuição condicional parametrizada através da média e , potencialmente, de um parâmetro (geralmente a dispersão/precisão) não dependente do tempo. O modelo proposto é aplicável nas situações em que a variável de interesse é contínua, pertencente ao intervalo $(0, \infty)$ e possivelmente relacionada com um vetor de covariáveis por meio de uma estrutura de regressão no estilo GLM. A principal vantagem desta parametrização é que ela permite uma interpretação direta da dinâmica da média em função dos regressores. A estimação dos coeficientes do modelo é realizada através do método de máxima verossimilhança parcial (PMLE). Testes de hipótese são conduzidos considerando-se a distribuição assintótica dos estimadores PMLE. Simulações de Monte Carlo são conduzidas para estudar o comportamento dos modelos propostos em amostras finitas. O estudo investiga as propriedades do PMLE, bem como o desempenho preditivo dos modelos dentro (*in-sample*) e fora da amostra (*out-of-sample*), tanto no caso de especificação correta quanto de má especificação da distribuição condicional. Para finalizar, uma aplicação a dados reais é proposta para ilustrar a utilização dos modelos dinâmicos.

ABSTRACT

Based on works by [Ferrari and Cribari-Neto \(2004\)](#), [Rocha and Cribari-Neto \(2009\)](#), [Pumi et al. \(2019b\)](#) and [Bourguignon et al. \(2021\)](#), this work proposes the construction of moving average autoregressive models in which the response has a conditional distribution parameterized through the mean and, potentially, a parameter (usually the dispersion/precision) which is not time dependent. The proposed model is applicable in situations where the variable of interest is continuous, takes values on the interval $(0, \infty)$, and possibly relates to a vector of covariates through a GLM-like regression structure. The main advantage of this parameterization is that it allows a direct interpretation of the dynamics of the mean as a function of the regressors. The estimation of the model coefficients is performed using the partial maximum likelihood (PMLE) method. Hypothesis tests are conducted considering the asymptotic distribution of the PMLE estimators. Monte Carlo simulations are conducted to study the finite samples behavior of the proposed models. The study investigates the properties of PMLE, as well as the *in-sample* and *out-of-sample* predictive performance of the models, both under scenario of correct specification and misspecification of the conditional distribution. Finally, an application to real data is proposed to illustrate the use of dynamic models.

ÍNDICE

1	Introdução	3
2	Conceitos Preliminares	7
2.1	Funções Especiais	7
2.2	Integrais Envolvendo Funções Especiais	9
3	Distribuições de Probabilidade	10
3.1	Distribuição Beta-Prime	11
3.2	Distribuição F (Fisher-Snedecor)	14
3.3	Distribuição Gama	18
3.4	Distribuição Gaussiana Inversa	21
3.5	Distribuição Log-Logística	23
3.6	Distribuição Log-Normal	28
3.7	Distribuição Qui-Quadrado	30
3.8	Distribuição Rayleigh	31
4	Modelo de Regressão para Séries Temporais Positivas	34
4.1	Modelos Clássicos	34
4.1.1	Modelos da Família ARMA	34
4.1.2	Modelos da Família ARMA com Erros do Tipo ARCH	35
4.1.3	Modelos Condicionais do Tipo GARMA	37

ÍNDICE	2
4.2 Modelos Condicionais Para dados Positivos	38
4.2.1 PMLE	41
4.2.2 Análise Residual	42
4.2.3 Previsão	43
5 Simulações de Monte Carlo	44
5.1 Processo Gerador dos Dados	44
5.2 Ajuste dos Modelos e Previsão	45
5.3 Apresentação dos Resultados	46
5.4 Aspectos Computacionais	48
6 Aplicação a Dados Reais	50
6.1 Características da Série Temporal	50
6.2 Seleção de Modelos e Previsão	51
6.3 Resultados	53
7 Conclusões e Trabalhos Futuros	56
Anexos	59
A Artigo Prass et al.	60

CAPÍTULO 1

INTRODUÇÃO

Recentemente, a literatura tem apresentado diversos modelos para séries temporais duplamente limitadas. Algumas dessas propostas de modelos podem ser vistas em [Fokianos and Kedem \(1998\)](#); [Ferrari and Cribari-Neto \(2004\)](#); [Rocha and Cribari-Neto \(2009\)](#); [Bayer et al. \(2017\)](#); [Pumi et al. \(2019b,a, 2021\)](#), entre outros. Esses modelos são baseados na abordagem que hoje é conhecida como GARMA (ARMA Generalizado). O modelo GARMA é uma extensão dos Modelos Lineares Generalizados de [McCullagh and Nelder \(1989\)](#), caracterizando-se pela adição de um termo extra ao preditor linear, o qual incorpora termos autoregressivos (AR) e de médias móveis (MA). A ideia principal por trás da abordagem data do final da década de 70, mas o nome GARMA foi solidificado em [Benjamin et al. \(2003\)](#).

Os modelos GARMA são modelos do tipo *observation-driven*, nomenclatura introduzida em [Cox \(1981\)](#) e que inclui dois componentes principais, o aleatório e o sistemático. Enquanto que o componente aleatório é responsável pelas características distributivas do modelo, geralmente dependente de uma medida de interesse, como a média (condicional) ou mediana, o componente sistemático caracteriza a estrutura de dependência que conduz a medida de interesse, comumente chamado de resposta média. De forma genérica, dado um processo estocástico $\{Y_t\}_{t \in \mathbb{Z}}$, uma coleção de covariáveis $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ e uma filtração $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$, tais modelos são caracterizados por uma estrutura da forma

$$Y_t | \mathcal{F}_{t-1} \sim f(\cdot | \mathcal{F}_{t-1}), \quad \mu_t = h(\mathbf{X}_t, \dots, \mathbf{X}_{t-\ell}, Y_{t-1}, \dots, Y_{t-m}, e_{t-1}, \dots, e_{t-n}), \quad (1.1)$$

onde $\ell, m, n \geq 1$ indicam as defasagens máximas utilizadas, $f(\cdot | \mathcal{F}_{t-1})$ denota a densidade condicional, μ_t é a medida de interesse, $h(\cdot)$ é uma função de ligação, que deve satisfazer certas propriedades para que μ_t esteja bem definida e $\{e_t\}_{t \in \mathbb{Z}}$ denota um termo de erro tipicamente definido de forma iterativa, cuja definição pode variar de modelo para modelo.

A estrutura (1.1) permite grande flexibilidade. Por exemplo, em [Rocha and Cribari-Neto \(2009\)](#) a variável resposta segue uma distribuição Beta, parametrizada em termos da média, e em [Bayer et al. \(2017\)](#) é considerada a distribuição Kumaraswamy, parametrizada em termos da mediana. Ambos os trabalhos consideram uma estrutura do tipo ARMA para o componente sistemático, isto é, assumem que

$$g(\mu_t) = \alpha + \mathbf{X}'_t \boldsymbol{\beta} + \sum_{k=1}^p \phi_k [g(Y_{t-k}) - \mathbf{X}'_{t-k} \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j e_{t-j}, \quad e_t = g(Y_t) - g(\mu_t),$$

onde g é uma função contínua, duas vezes diferenciável, definida no suporte da densidade condicional e tomando valores na reta. A diferença está que o primeiro modela a média condicional, enquanto

que o segundo, a mediana condicional.

Segundo [Box et al. \(2008\)](#), os modelos GARMA apresentam vantagens sobre os modelos clássicos de séries temporais, tais como os modelos da classe ARIMA. Dentre elas está o fato que modelos GARMA são capazes de lidar com séries temporais não gaussianas limitadas sem a necessidade de transformação dos dados, o que torna a interpretação dos modelos muito mais simples. Além disso, a modelagem GARMA permite facilmente a incorporação de características da distribuição de interesse no componente aleatório, mantendo a estrutura de dependência na resposta média condicional. Isso leva a uma estratégia de modelagem muito mais simples do que a abordagem usual, onde particularidades da distribuição são inseridas em modelos lineares por meio do termo de erro. Destaca-se ainda que a inferência condicional é naturalmente acomodada dentro da estrutura, fornecendo uma poderosa ferramenta inferencial.

Embora [Benjamin et al. \(2003\)](#) tenha considerado o caso em que a distribuição por trás do modelo pertence a família exponencial canônica, distribuições fora da família exponencial tem sido exploradas na literatura, como, por exemplo, em [Bayer et al. \(2017\)](#), onde os autores consideram utilizar a distribuição Kumaraswamy para modelar séries temporais duplamente limitadas. A distribuição a ser aplicada e a estrutura do modelo podem variar significativamente dependendo das características dos dados, aplicações e objetivos.

Como exemplos de estruturas distintas utilizadas para incorporar características marcantes dos dados, podemos citar [Ferrari and Cribari-Neto \(2004\)](#); [Rocha and Cribari-Neto \(2009\)](#); [Pumi et al. \(2019b\)](#) e [Pumi et al. \(2021\)](#). Embora os três trabalhos considerem a mesma distribuição condicional (Beta) em seus modelos, a estrutura proposta por [Rocha and Cribari-Neto \(2009\)](#) descreve séries temporais com curta dependência, enquanto que a generalização proposta em [Pumi et al. \(2019b\)](#) é capaz de modelar longa dependência. Por sua vez, [Pumi et al. \(2021\)](#) substituem a estrutura de média móvel da equação (1.1) por um termo da forma $Z_t = h(T^t(U_0))$, denominado processo caótico, onde U_0 é uma variável aleatória tomando valores em $(0, 1)$, $T : [0, 1] \rightarrow [0, 1]$ e $h : (0, 1) \rightarrow \mathbb{R}$. Como apontado pelos autores, alterando a transformação T é possível mudar drasticamente as propriedades dos caminhos e da estrutura de dependência do processo resultante.

No contexto de séries temporais que assumem apenas valores positivos, [Benjamin et al. \(1998\)](#) estuda um modelo GARMA baseado na distribuição Gama, enquanto que, mais recentemente, [Bourguignon et al. \(2021\)](#) introduz um modelo de regressão onde a variável resposta segue uma distribuição Beta-Prime. Em ambos os casos, os modelos propostos permitem a inclusão de covariáveis exógenas para descrever a resposta média. Porém, apenas o primeiro trabalho inclui uma estrutura do tipo (1.1) para lidar com a dependência serial. O objetivo deste trabalho e do artigo (em anexo) intitulado “*Positive Time Series Regression Models*”, de autoria de Taiane S. Prass, Jonas H. Carlos, Cleiton G. Taufemback e Guilherme Pumi, é unir essas duas frentes e propor uma classe de modelos dinâmicos para séries temporais positivas, permitindo a inclusão de covariáveis exógenas (possivelmente dependentes do tempo e aleatórias) e também uma estrutura tipo ARMA para lidar com dependência serial para a resposta média, sem ficar restrito às distribuições pertencentes à família exponencial.

Objetivo

Esta dissertação tem por objetivo propor e estudar os modelos dinâmicos para séries temporais positivas, bem como reparametrizações em termos da média para algumas distribuições conhecidas. Pretende-se ainda analisar o desempenho do estimador de máxima verossimilhança parcial (PMLE) para a estimação dos parâmetros dos modelos obtidos a partir de diferentes distribuições. Também será investigado o desempenho do método nos casos de má especificação da distribuição. Além da estimação dos parâmetros, a dissertação investiga ainda o desempenho dos modelos em termos de previsão dentro (*in-sample*) e fora (*out-of-sample*) da amostra.

Novidades do trabalho

O trabalho propõe uma generalização dos modelos propostos em Benjamin et al. (1998) e Bourguignon et al. (2021). Além de propor alterações na componente sistemática, que permitem maior flexibilidade na modelagem da média condicional, as distribuições consideradas não se restringem à família exponencial. Outra novidade é que, além de utilizar generalizações do componente sistemático, a estimação dos parâmetros do modelo é feita via máxima verossimilhança parcial, em contraste com os trabalhos já existentes na literatura, que utilizam métodos para distribuições da família exponencial. Também são propostas reparametrizações de distribuições conhecidas na literatura, que apresentam grande potencial de aplicação em dados positivos.

Suporte computacional

Na parte computacional, são apresentadas simulações de Monte Carlo considerando diferentes cenários de ajuste do modelo. Posteriormente, é realizada uma aplicação em dados reais. Todas as análises apresentadas nesta dissertação foram implementadas na versão 4.1.2 do R (R Core Team, 2021). As rotinas utilizadas para ajuste e previsão dos modelos fazem parte do pacote PTSR (Prass et al., 2022).

O pacote PTSR está disponível no CRAN (<https://cran.r-project.org/package=PTSR>). Devido ao esforço conjunto do grupo de trabalho, os códigos desenvolvidos e utilizados nas análises preliminares relacionados a este trabalho, deram origem ao pacote. A primeira versão do pacote teve sua publicação em 13 de janeiro de 2022. Até a data de 20 de maio de 2022, já haviam sido realizados 1.460 downloads.

Organização da dissertação

O trabalho é organizado como segue. No Capítulo 2 são apresentados conceitos preliminares englobando notações, funções, definições, relações e resultados a fim de facilitar a compreensão e pesquisa do leitor. No Capítulo 3 são apresentadas as funções de distribuição de probabilidades reparametrizadas na forma $f(\cdot|\mu_t, \varphi)$, bem como algumas de suas propriedades. No Capítulo 4 é apresentado o modelo para dados positivos proposto neste trabalho, bem como os resultados necessários para os próximos capítulos. O Capítulo 5 trata dos estudos de Simulação de Monte Carlo, enquanto que o

Capítulo 6 discute a aplicação do modelo proposto em dados reais de temperatura de São Paulo, coletados na estação meteorológica A701. Por fim, o Capítulo 7 traz as conclusões e discute alternativas de possíveis trabalhos subsequentes.

CAPÍTULO 2

CONCEITOS PRELIMINARES

Com o objetivo de facilitar a compreensão e pesquisa do leitor, neste capítulo são apresentadas as notações utilizadas ao longo deste trabalho. Assim como definições, relações e resultados de outros autores, devidamente citados.

Neste trabalho os símbolos \mathbb{N} , \mathbb{Z} , \mathbb{R} e \mathbb{C} denotam os conjuntos dos números naturais (incluindo o zero), inteiros, reais e complexos, respectivamente. A notação \mathbb{N}^* denota o conjunto dos números naturais sem o zero. Além disso, para $z \in \mathbb{C}$, $\operatorname{Re}(z)$ e $\operatorname{Im}(z)$ denotam, respectivamente, a parte real e a parte imaginária de z .

2.1 Funções Especiais

Função Gama

A função gama, denotada por $\Gamma(\cdot)$, originalmente definida por

$$\Gamma(z) := \int_0^{\infty} t^{z-1} e^{-t} dt, \quad z \in \mathbb{C}, \operatorname{Re}(z) > 0, \quad (2.1)$$

foi posteriormente definida como sendo a continuação analítica de (2.1) para uma função que é holomórfica em todo o plano complexo, exceto no zero e nos inteiros negativos, onde a função tem polos simples (para mais detalhes veja, por exemplo, [Davis, 1959](#)). Tal extensão pode ser obtida combinando-se a equação (2.1) e a identidade

$$\Gamma(z) = \frac{\Gamma(z+1)}{z}, \quad z \in \mathbb{C}, -z \notin \mathbb{N}. \quad (2.2)$$

Outra propriedade importante da função gamma, também denominada fórmula de reflexão de Euler, é a seguinte ([Gradshteyn and Ryzhik, 2015](#), fórmula 8.334.3, página 905)

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}, \quad z \notin \mathbb{Z}.$$

Tal expressão, combinada com (2.2) permite concluir que

$$\Gamma(1-z)\Gamma(1+z) = \frac{\pi z}{\sin(\pi z)}, \quad z \notin \mathbb{Z}. \quad (2.3)$$

Função Digama

A função digama, denotada por $\psi(\cdot)$, é definida como a derivada do logaritmo da função gama, isto é

$$\psi(z) := \frac{d}{dz} \log(\Gamma(z)) = \frac{1}{\Gamma(z)} \frac{d\Gamma(z)}{dz}, \quad z \in \mathbb{C}, \quad -z \notin \mathbb{N}$$

e satisfaz a seguinte relação recursiva

$$\psi(z+1) = \psi(z) + \frac{1}{z}.$$

Função Trigama

A função trigama, denotada por $\psi'(\cdot)$, é definida como a segunda derivada do logaritmo da função gama, isto é

$$\psi'(z) := \frac{d^2}{dz^2} \log(\Gamma(z)) = \frac{d\psi(z)}{dz}, \quad z \in \mathbb{C}, \quad -z \notin \mathbb{N}$$

e satisfaz a seguinte relação recursiva

$$\psi'(z+1) = \psi'(z) - \frac{1}{z^2}.$$

Observação 2.1. Na literatura também é comum a utilização da notação $\psi^{(k)}$ para representar a k -ésima derivada do logaritmo da função gama. Nesse caso, $\psi = \psi^{(0)}$ e $\psi' = \psi^{(1)}$.

Função Beta

A função beta, também chamada de integral de Euler do primeiro tipo, denotada por $B(\cdot, \cdot)$, é usualmente definida por

$$B(\alpha, \beta) := \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt, \quad \alpha, \beta \in \mathbb{C}, \quad \operatorname{Re}(\alpha) > 0, \quad \operatorname{Re}(\beta) > 0, \quad (2.4)$$

Dentre as propriedades da função beta, destacam-se a relação com a função gama dada por

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = B(\beta, \alpha), \quad (2.5)$$

e a identidade

$$\begin{aligned} \frac{1}{B(\alpha, \beta)} \frac{\partial B(\alpha, \beta)}{\partial \alpha} &= \frac{\partial}{\partial \alpha} \log(B(\alpha, \beta)) \\ &= \frac{\partial}{\partial \alpha} \left[\log(\Gamma(\alpha)) + \log(\Gamma(\beta)) - \log(\Gamma(\alpha + \beta)) \right] \\ &= \psi(\alpha) - \psi(\alpha + \beta). \end{aligned} \quad (2.6)$$

A relação (2.5) permite que a função beta seja estendida para $\mathbb{C} \times \mathbb{C}$, de forma análoga à extensão da função gama (Grafakos, 2008), enquanto que a igualdade (2.6) é comumente utilizada no cálculo de esperanças associadas às distribuições Beta e Beta-Prime.

Função Bessel Modificada

As soluções para a equação de Bessel são chamadas de funções de Bessel modificadas (ou funções de Bessel hiperbólicas) de primeiro e segundo tipo. A função Bessel modificada de segundo tipo, denotada por $K_\nu(\cdot)$, $\nu \in \mathbb{C}$, cuja representação integral é dada por (Gradshteyn and Ryzhik, 2015, equação 8.432.1, página 926)

$$K_\nu(z) = \int_0^\infty \cosh(\nu t) e^{-z \cosh(\nu t)} dt, \quad |\operatorname{Im}(z)| < \frac{\pi}{2} \text{ ou } \operatorname{Re}(z) = 0, \nu = 0.$$

Visto que $\cosh(x) = \cosh(-x)$, $x \in \mathbb{R}$, tem-se $K_{-\nu}(z) = K_\nu(z)$. Além disso, (Gradshteyn and Ryzhik, 2015, fórmula 8.468, página 934)

$$K_{n+1/2}(z) = \sqrt{\frac{\pi}{2z}} e^{-z} \sum_{k=0}^n \frac{(n+k)!}{k!(n-k)!(2z)^k}, \quad n \in \mathbb{N}. \quad (2.7)$$

2.2 Integrais Envolvendo Funções Especiais

Diversas esperanças associadas às distribuições estudadas neste trabalho envolvem integrais complexas cujas expressões podem ser encontradas em Gradshteyn and Ryzhik (2015). Por simplicidade, as integrais necessárias para o desenvolvimento do trabalho são listadas abaixo, junto com o número da fórmula e a página onde elas se encontram.

- Fórmula 3.194.3, página 318,

$$\int_0^\infty x^{\alpha-1} (1+\beta x)^{-\nu} dx = \beta^{-\alpha} B(\alpha, \nu - \alpha), \quad |\operatorname{Im}(\beta)| < \pi, \operatorname{Re}(\nu) > 0, \operatorname{Re}(\alpha) > 0. \quad (2.8)$$

- Fórmula 4.293.14, página 563,

$$\int_0^\infty \log(\gamma+x) \frac{x^{\alpha-1}}{(\gamma+x)^\nu} dx = \frac{B(\alpha, \nu - \alpha)}{\gamma^{\nu-\alpha}} [\psi(\nu) - \psi(\nu - \alpha)], \quad 0 < \operatorname{Re}(\alpha) < \operatorname{Re}(\nu). \quad (2.9)$$

- Fórmula 3.381.11, página 349,

$$\int_0^\infty x^\nu e^{-\alpha x^\gamma} dx = \frac{\alpha^{-(\nu+1)/\gamma}}{\gamma} \Gamma\left(\frac{\nu+1}{\gamma}\right), \quad \operatorname{Re}(\nu) > -\frac{1}{2}, \operatorname{Re}(\alpha) > 0, \operatorname{Re}(\gamma) > 0. \quad (2.10)$$

- Fórmula 4.352.1, página 577,

$$\int_0^\infty \log(x) x^{\nu-1} e^{-\alpha x} dx = \frac{\Gamma(\nu)}{\alpha^\nu} [\psi(\nu) - \log(\alpha)], \quad \operatorname{Re}(\nu) > 0, \operatorname{Re}(\alpha) > 0. \quad (2.11)$$

- Fórmula 3.471.9, página 370,

$$\int_0^\infty x^{\nu-1} e^{-\beta/x - \gamma x} dx = 2 \left(\frac{\beta}{\gamma}\right)^{\nu/2} K_\nu(2\sqrt{\beta\gamma}), \quad \operatorname{Re}(\beta) > 0, \operatorname{Re}(\gamma) > 0. \quad (2.12)$$

- Fórmula 3.241.4, página 325,

$$\int_0^\infty \frac{x^{\gamma-1}}{(\alpha + \beta x^\nu)^{n+1}} dx = \frac{(\alpha/\beta)^{\gamma/\nu}}{\nu \alpha^{n+1}} B\left(n+1 - \frac{\gamma}{\nu}, \frac{\gamma}{\nu}\right), \quad 0 < \frac{\gamma}{\nu} < n+1, \alpha \neq 0, \beta \neq 0. \quad (2.13)$$

CAPÍTULO 3

DISTRIBUIÇÕES DE PROBABILIDADE

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo estocástico tal que $P(Y_t \in (0, \infty)) = 1$ e $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ uma filtração. Denota-se por $f(\cdot|\mathcal{F}_{t-1})$ a densidade condicional de Y_t dado \mathcal{F}_{t-1} . Particularmente, há interesse no caso em que $\mu_t = \mathbb{E}(Y_t|\mathcal{F}_{t-1})$ existe e é finita com probabilidade 1 e ainda, que $f(\cdot|\mathcal{F}_{t-1})$ pode ser parametrizada por μ_t e, possivelmente, um parâmetro (ou vetor de parâmetros) φ que não dependa do tempo. Para deixar essa relação clara, define-se $f(\cdot|\mu_t, \varphi) := f(\cdot|\mathcal{F}_{t-1})$ e utiliza-se a seguinte notação

$$Y_t|\mathcal{F}_{t-1} \sim f(\cdot|\mu_t, \varphi), \quad \text{onde } \mu_t = \mathbb{E}(Y_t|\mathcal{F}_{t-1}), \quad \varphi \in \mathbb{R}. \quad (3.1)$$

Alternativamente, ao invés de trabalhar com a média condicional, pode-se considerar outras quantidades de interesse, como por exemplo a moda ([Bourguignon et al., 2020](#)), a mediana condicional ([Bayer et al., 2017](#)) ou, de forma mais geral, outros quantis da distribuição condicional. Nesse caso, é importante que a distribuição possa ser reparametrizada em termos dessa quantidade.

Nos demais capítulos considera-se ainda um conjunto $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ de variáveis exógenas s -dimensionais, possivelmente dependendo do tempo e/ou aleatórias, e assume-se que \mathcal{F}_t é a σ -álgebra gerada por $(\mathbf{X}'_{t+1}, Y_t, \mathbf{X}'_t, Y_{t-1}, \mathbf{X}'_{t-1}, \dots)$, isto é, a σ -álgebra gerada pela informação observada até o tempo t , para todo $t \in \mathbb{Z}$. Sem perda de generalidade, assume-se aqui que \mathbf{X}_{t+1} é conhecido no tempo t , o que sempre válido quando $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ não é aleatório. No caso geral, basta tomar $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ como sendo uma versão transladada do conjunto de covariáveis em questão e a suposição continua válida. O objetivo dessa suposição é simplificar a notação dos modelos definidos no Capítulo 4 onde é proposta uma estrutura do tipo

$$\mu_t = h(\mathbf{X}_t, \dots, \mathbf{X}_{t-\ell}, Y_{t-1}, \dots, Y_{t-m}, e_{t-1}, \dots, e_{t-n}), \quad \ell, m, n \geq 1,$$

onde $h(\cdot)$ é uma função de ligação, que deve satisfazer certas propriedades para que μ_t esteja bem definida e $\{e_t\}_{t \in \mathbb{Z}}$ denota um termo de erro.

No que segue, são apresentadas algumas distribuições com as características mencionadas acima. Destacamos que a distribuição Beta-Prime foi utilizada em [Bourguignon et al. \(2021\)](#), mas com uma parametrização ligeiramente diferente; a distribuição F (Fisher-Snedecor) foi escolhida pelo fato que a distribuição t -Student aparece com frequência na modelagem de log-retornos financeiros e, se $Y \sim t_\nu$ então $Y^2 \sim F(1, \nu)$; a distribuição Gama foi utilizada em [Benjamin et al. \(1998\)](#); a distribuição Qui-quadrado foi escolhida pelo fato que a distribuição Normal aparece com frequência na modelagem de log-retornos financeiros e, se $Y \sim N(0, 1)$, então $Y^2 \sim \chi_1^2$.

Ressalta-se que, na parametrização tradicional, a maioria das distribuições escolhidas não depende diretamente de μ_t . Sendo assim, para cada distribuição, apresentar-se-á a reparametrização necessária para obter o formato desejado, a expressão da densidade na nova parametrização, bem como algumas propriedades da função de log-verossimilhança condicional $\ell_t(\cdot)$, associada ao modelo definido por (3.1). Por simplicidade de notação, a dependência nos valores observados de Y_t ficará implícita, isto é, utiliza-se a seguinte notação

$$\ell_t(\boldsymbol{\gamma}) := \log(f(Y_t|\mu_t, \varphi)), \quad t \in \mathbb{Z},$$

onde $\boldsymbol{\gamma} := (\boldsymbol{\rho}', \varphi)'$ e $\boldsymbol{\rho}$ é um vetor de parâmetros associados a μ_t (veja o Capítulo 4).

3.1 Distribuição Beta-Prime

Uma variável aleatória Y tem distribuição Beta-Prime com parâmetros α e β , denotada por $Y \sim \text{BP}(\alpha, \beta)$, se sua função densidade de probabilidade é dada por

$$f(y|\alpha, \beta) = \frac{y^{\alpha-1}(1+y)^{-(\alpha+\beta)}}{B(\alpha, \beta)}, \quad 0 < y < \infty, \quad \alpha, \beta > 0. \quad (3.2)$$

Propriedades da Distribuição Beta-Prime

Observe que, de (2.8) segue que

$$\mathbb{E}(Y^k) = \frac{1}{B(\alpha, \beta)} \int_0^\infty y^{k+\alpha-1}(1+y)^{-(\alpha+\beta)} dy = \frac{B(\alpha+k, \beta-k)}{B(\alpha, \beta)}, \quad 0 < k < \beta.$$

Então, utilizando-se as relações (2.2) e (2.5), conclui-se que

$$\mathbb{E}(Y) = \frac{\alpha}{\beta-1}, \quad \beta > 1 \quad \text{e} \quad \text{Var}(Y) = \frac{\alpha(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2}, \quad \beta > 2, \quad (3.3)$$

ou seja, a média da distribuição existe sempre que $\beta > 1$. Como este é o cenário de interesse, adota-se essa restrição no momento da reparametrização.

Outras quantidades de interesse para o estudo em questão são $\mathbb{E}(\log(Y))$ e $\mathbb{E}(\log(Y+1))$. Tais quantidades aparecem no cálculo da esperança da derivada da função de log-verossimilhança. Para derivar tais esperanças observe que, de (2.9) conclui-se que

$$\mathbb{E}(\log(Y+1)) = \frac{1}{B(\alpha, \beta)} \int_0^\infty \log(y+1)y^{\alpha-1}(1+y)^{-(\alpha+\beta)} dy = \psi(\alpha+\beta) - \psi(\beta). \quad (3.4)$$

Além disso,

$$\frac{\partial}{\partial \alpha} \left[y^{\alpha-1}(1+y)^{-(\alpha+\beta)} \right] = \log(y)(1+y)^{-(\alpha+\beta)} y^{\alpha-1} - \log(1+y)(1+y)^{-(\alpha+\beta)} y^{\alpha-1}$$

implica que

$$\log(y)(1+y)^{-(\alpha+\beta)} y^{\alpha-1} = \log(1+y)(1+y)^{-(\alpha+\beta)} y^{\alpha-1} + \frac{\partial}{\partial \alpha} \left[y^{\alpha-1}(1+y)^{-(\alpha+\beta)} \right].$$

Combinando esse resultado com (2.6) e (3.4), obtém-se

$$\begin{aligned}
\mathbb{E}(\log(Y)) &= \frac{1}{B(\alpha, \beta)} \int_0^\infty \log(y) y^{\alpha-1} (1+y)^{-(\alpha+\beta)} dy \\
&= \frac{1}{B(\alpha, \beta)} \left[\int_0^\infty \log(1+y) (1+y)^{-(\alpha+\beta)} y^{\alpha-1} dy + \int_0^\infty \frac{\partial}{\partial \alpha} y^{\alpha-1} (1+y)^{-(\alpha+\beta)} dy \right] \\
&= \mathbb{E}(\log(Y+1)) + \frac{1}{B(\alpha, \beta)} \frac{\partial}{\partial \alpha} \int_0^\infty y^{\alpha-1} (1+y)^{-(\alpha+\beta)} dy \\
&= \mathbb{E}(\log(Y+1)) + \frac{1}{B(\alpha, \beta)} \frac{\partial}{\partial \alpha} B(\alpha, \beta) \\
&= \psi(\alpha) - \psi(\beta).
\end{aligned} \tag{3.5}$$

Reparametrização

Assumindo $\beta > 1$ e considerando-se a reparametrização

$$\mu := \frac{\alpha}{\beta - 1} \quad \text{e} \quad \varphi := \beta - 1, \tag{3.6}$$

obtém-se $\alpha = \mu\varphi$, $\beta = \varphi + 1$. Consequentemente, utilizando-se a relação (2.5), (3.2) pode ser reescrita como

$$f(y|\mu, \varphi) = \frac{\Gamma(\varphi(\mu+1)+1)}{\Gamma(\varphi\mu)\Gamma(\varphi+1)} y^{\varphi\mu-1} (1+y)^{-\varphi(\mu+1)}, \quad 0 < y < \infty, \quad \mu, \varphi > 0, \tag{3.7}$$

cuja notação correspondente é $Y \sim \text{BP}(\mu, \varphi)$. De (3.3) e (3.6) tem-se ainda que

$$\text{Var}(Y) = \frac{\mu(\mu+1)}{\varphi-1}, \quad \varphi > 1.$$

A Figura 3.1 mostra o comportamento da função densidade de probabilidade da distribuição Beta-Prime, reparametrizada em termos de μ e φ , conforme (3.7). Nessa figura considera-se $\mu = 2.5$ fixo e $\varphi \in \{0.5, 1, 2, 10, 50\}$.

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim \text{BP}(\mu_t, \varphi)$ conclui-se que

$$\begin{aligned}
\ell_t(\gamma) &= (\varphi\mu_t - 1) \log(Y_t) - \varphi(\mu_t + 1) \log(1 + Y_t) \\
&\quad - \log(\Gamma(\varphi\mu_t)) - \log(\Gamma(\varphi + 1)) + \log(\Gamma(\varphi(\mu_t + 1) + 1)).
\end{aligned}$$

Então, definindo

$$Y_t^* := \log(Y_t) - \log(1 + Y_t) \quad \text{e} \quad \mu_t^* := \psi(\varphi\mu_t) - \psi(\varphi(\mu_t + 1) + 1),$$

obtém-se

$$\frac{\partial \ell_t(\gamma)}{\partial \mu_t} = \varphi \log(Y_t) - \varphi \log(1 + Y_t) - \varphi \psi(\varphi\mu_t) + \varphi \psi(\varphi(\mu_t + 1) + 1) = \varphi(Y_t^* - \mu_t^*),$$

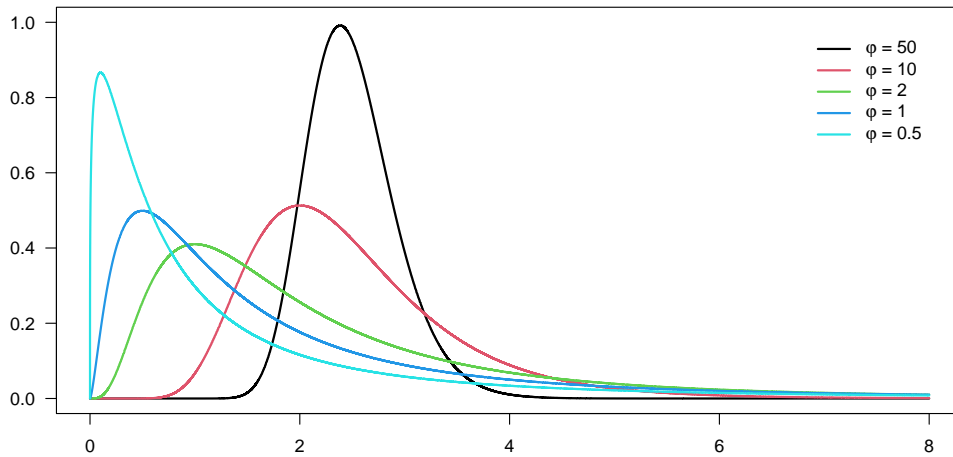


Figura 3.1: Distribuição Beta-Prime para $\mu = 2.5$ e diferentes valores para φ .

e

$$\begin{aligned} \frac{\partial \ell_t(\gamma)}{\partial \varphi} &= \mu_t \log(Y_t) - (\mu_t + 1) \log(1 + Y_t) - \psi(\varphi \mu_t) \mu_t - \psi(\varphi + 1) + \psi(\varphi(\mu_t + 1) + 1)(\mu_t + 1) \\ &= \mu_t(Y_t^* - \mu_t^*) - \log(1 + Y_t) - \psi(\varphi + 1) + \psi(\varphi(\mu_t + 1) + 1). \end{aligned}$$

Ainda, observando-se que

$$\frac{\partial Y_t^*}{\partial \mu_t} = \frac{\partial Y_t^*}{\partial \varphi} = 0, \quad \frac{\partial \mu_t^*}{\partial \mu_t} = \varphi \psi'(\varphi \mu_t) - \varphi \psi'(\varphi(\mu_t + 1) + 1)$$

e

$$\frac{\partial \mu_t^*}{\partial \varphi} = \mu_t \psi'(\varphi \mu_t) - (\mu_t + 1) \psi'(\varphi(\mu_t + 1) + 1),$$

conclui-se que

$$\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} = \varphi^2 \left[\psi'(\varphi(\mu_t + 1) + 1) - \psi'(\varphi \mu_t) \right],$$

$$\begin{aligned} \frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} &= -\mu_t \frac{\partial \mu_t^*}{\partial \varphi} - \psi'(\varphi + 1) + (\mu_t + 1) \psi'(\varphi(\mu_t + 1) + 1) \\ &= (\mu_t + 1)^2 \psi'(\varphi(\mu_t + 1) + 1) - \mu_t^2 \psi'(\varphi \mu_t) - \psi'(\varphi + 1) \end{aligned}$$

e

$$\begin{aligned} \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} &= Y_t^* - \mu_t^* - \varphi \frac{\partial \mu_t^*}{\partial \varphi} \\ &= Y_t^* - \mu_t^* - \varphi \mu_t \psi'(\varphi \mu_t) + \varphi(\mu_t + 1) \psi'(\varphi(\mu_t + 1) + 1). \end{aligned}$$

Fazendo-se a substituição (3.6), os resultados em (3.4) e (3.5) implicam que $\mathbb{E}(Y_t^* | \mathcal{F}_{t-1}) = \mu_t^*$,

$$\mathbb{E} \left(\frac{\partial \ell_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1} \right) = 0 \quad \text{e} \quad \mathbb{E} \left(\frac{\partial \ell_t(\gamma)}{\partial \varphi} \middle| \mathcal{F}_{t-1} \right) = 0,$$

como esperado. Além disso,

$$\mathbb{E} \left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) = \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2}, \quad \mathbb{E} \left(\frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} \middle| \mathcal{F}_{t-1} \right) = \frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2},$$

pois as derivadas envolvidas são \mathcal{F}_{t-1} -mensuráveis, e

$$\mathbb{E} \left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} \middle| \mathcal{F}_{t-1} \right) = \varphi(\mu_t + 1) \psi'(\varphi(\mu_t + 1) + 1) - \varphi \mu_t \psi'(\varphi \mu_t).$$

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição Beta-Prime utiliza, respectivamente, as funções `dbetapr` e `rbetapr` implementadas no pacote `extraDistr` (Wolodzko, 2020). O pacote `extraDistr` implementa a versão generalizada da distribuição Beta-Prime, cuja densidade é dada por

$$f(y) = \frac{1}{\sigma B(\alpha, \beta)} \left[\left(\frac{y}{\sigma} \right)^{\alpha-1} \left(1 + \frac{y}{\sigma} \right)^{-\alpha-\beta} \right], \quad 0 < y < \infty, \quad \alpha, \beta, \sigma > 0.$$

O parâmetro σ é denominado parâmetro de escala. Quando $\sigma = 1$, obtém-se a expressão (3.2).

3.2 Distribuição F (Fisher-Snedecor)

Uma variável aleatória Y tem distribuição F com d_1 e d_2 graus de liberdade, denotada por $Y \sim F(d_1, d_2)$, se sua função de densidade de probabilidade é dada por

$$f(y|d_1, d_2) = \frac{1}{B(d_1/2, d_2/2)} \left(\frac{d_1}{d_2} \right)^{d_1/2} y^{d_1/2-1} \left(1 + \frac{d_1}{d_2} y \right)^{-(d_1+d_2)/2}, \quad 0 < y < \infty, \quad d_1, d_2 > 0. \quad (3.8)$$

Propriedades da Distribuição F (Fisher-Snedecor)

Pode ser mostrado que (Taboga, 2012, página 427),

$$\mathbb{E}(Y^k) = \left(\frac{d_2}{d_1} \right)^k \frac{\Gamma(d_1/2 + k) \Gamma(d_2/2 - k)}{\Gamma(d_1/2) \Gamma(d_2/2)}, \quad d_2 > 2k.$$

Então, utilizando-se as relações (2.2) e (2.5), conclui-se que

$$\mathbb{E}(Y) = \frac{d_2}{d_2 - 2}, \quad d_2 > 2, \quad \text{e} \quad \text{Var}(Y) = \frac{2d_2^2(d_1 + d_2 - 2)}{d_1(d_2 - 2)^2(d_2 - 4)}, \quad d_2 > 4,$$

ou seja, a média da distribuição existe sempre que $d_2 > 2$. Sendo assim, deve-se adotar essa restrição no momento da reparametrização.

No que segue, são apresentados resultados encontrados na literatura relacionados com algumas transformações da variável aleatória Y , onde $Y \sim F(d_1, d_2)$. Tais resultados são utilizados na sequência, para o cálculo das esperanças das derivadas da função da log-verossimilhança. Em todos os casos é considerada a parametrização tradicional da distribuição. Se $Y \sim F(d_1, d_2)$, então

- **Distribuição Z de Fisher** (veja Aroian, 1941)

$$\frac{\log(Y)}{2} \sim \text{FisherZ}(d_1, d_2).$$

A função geradora de momentos associada a essa distribuição é dada por

$$M(s) = \left(\frac{d_2}{d_1}\right)^{s/2} \frac{\Gamma((d_1 + s)/2)\Gamma((d_2 - s)/2)}{\Gamma(d_1/2)\Gamma(d_2/2)}, \quad \{s \in \mathbb{R} : -s - d_1, s - d_2 \notin \mathbb{N}\}.$$

Então, observando que

$$\frac{dM(s)}{ds} = \frac{1}{2} \left(\frac{d_2}{d_1}\right)^{s/2} \frac{\Gamma((d_1 + s)/2)\Gamma((d_2 - s)/2)}{\Gamma(d_1/2)\Gamma(d_2/2)} \left[\log\left(\frac{d_2}{d_1}\right) + \psi\left(\frac{d_1 + s}{2}\right) - \psi\left(\frac{d_2 - s}{2}\right) \right],$$

conclui-se que

$$\mathbb{E}(\log(Y)) = 2 \left. \frac{dM(s)}{ds} \right|_{s=0} = \log\left(\frac{d_2}{d_1}\right) + \psi\left(\frac{d_1}{2}\right) - \psi\left(\frac{d_2}{2}\right). \quad (3.9)$$

• **Distribuição Beta-Prime**

$$\frac{d_1}{d_2}Y \sim \text{BP}\left(\frac{d_1}{2}, \frac{d_2}{2}\right) \implies \mathbb{E}\left(\log\left(\frac{d_1}{d_2}Y + 1\right)\right) = \psi\left(\frac{d_1}{2} + \frac{d_2}{2}\right) - \psi\left(\frac{d_2}{2}\right), \quad d_2 > 2, \quad (3.10)$$

onde a primeira relação segue de imediato, basta observar que

$$P\left(\frac{d_1}{d_2}Y \leq y\right) = P\left(Y \leq \frac{d_2}{d_1}y\right) \implies f_{\frac{d_1}{d_2}Y}(y) = \frac{d_2}{d_1} f_Y\left(\frac{d_2}{d_1}y\right),$$

e a última igualdade segue da expressão (3.4).

• **Distribuição Beta**

$$\frac{Yd_1/d_2}{1 + Yd_1/d_2} \sim \text{Beta}\left(\frac{d_1}{2}, \frac{d_2}{2}\right) \implies \mathbb{E}\left(\frac{Yd_1/d_2}{1 + Yd_1/d_2}\right) = \frac{d_1}{d_1 + d_2}. \quad (3.11)$$

Para provar essa relação basta observar que, para todo $a > 0$,

$$Z := \frac{Ya}{1 + Ya} < 1 \implies P\left(\frac{Ya}{1 + Ya} \leq y\right) = P\left(Y \leq \frac{y}{a(1 - y)}\right), \quad 0 < y < 1,$$

o que implica

$$f_Z(y) = \frac{1}{a(1 - y)^2} f_Y\left(\frac{y}{a(1 - y)}\right). \quad (3.12)$$

Para finalizar, basta tomar $a = d_1/d_2$ e observar que, após um pouco de álgebra, a expressão (3.12) se torna

$$f_Z(y) = \frac{1}{B(d_1/2, d_2/2)} y^{d_1/2-1} (1 - y)^{d_2/2-1}, \quad 0 < y < 1.$$

A esperança da distribuição Beta segue de (2.4) e das relações (2.2) e (2.5).

Reparametrização

Assumindo $d_2 > 2$ e considerando-se a reparametrização

$$\mu := \frac{d_2}{d_2 - 2} \quad \text{e} \quad \varphi := d_1, \quad (3.13)$$

obtém-se $d_1 = \varphi$ e $d_2 = 2\mu/(\mu - 1)$. Consequentemente, utilizando-se a relação (2.5), (3.8) pode ser reescrita como

$$f(y|\mu, \varphi) = \frac{\Gamma(\varphi/2 + \mu/(\mu - 1))}{\Gamma(\varphi/2)\Gamma(\mu/(\mu - 1))} \left(\frac{\varphi(\mu - 1)}{2\mu}\right)^{\varphi/2} y^{\varphi/2-1} \left(1 + \frac{y\varphi(\mu - 1)}{2\mu}\right)^{-(\varphi/2 + \mu/(\mu - 1))}, \quad (3.14)$$

$0 < y < \infty$, $\varphi > 0$ e $\mu > 1$, cuja notação correspondente é $Y \sim F(\mu, \varphi)$. Tem-se ainda que

$$\text{Var}(Y) = \frac{\mu^2(\mu\varphi - \varphi + 2)}{(2 - \mu)\varphi}, \quad 1 < \mu < 2.$$

A Figura 3.2 mostra o comportamento da função densidade de probabilidade da distribuição F (Fisher-Snedecor), reparametrizada em termos de μ e φ , conforme (3.14). Nessa figura considera-se $\mu = 1.5$ (esquerda) e 2.5 (direita) fixo e $\varphi \in \{0.5, 1, 2, 5, 50\}$. No primeiro caso $\text{Var}(Y) < \infty$ e no segundo caso ela não existe.

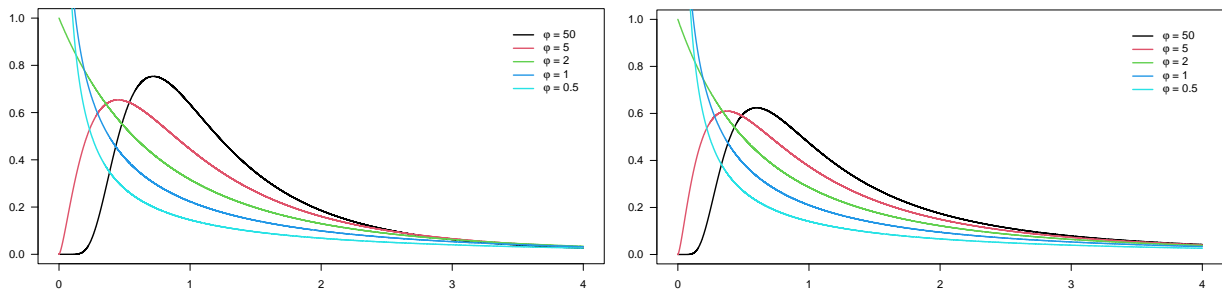


Figura 3.2: Distribuição F com $\mu = 1.5$ (esquerda) e 2.5 (direita) e diferentes valores para φ .

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim F(\mu_t, \varphi)$ tem-se

$$\begin{aligned} \ell_t(\gamma) &= \left(\frac{\varphi}{2} - 1\right) \log(Y_t) - \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1}\right) \log\left(\frac{Y_t \varphi(\mu_t - 1)}{2\mu_t} + 1\right) + \log\left(\Gamma\left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1}\right)\right) \\ &\quad - \log\left(\Gamma\left(\frac{\varphi}{2}\right)\right) - \log\left(\Gamma\left(\frac{\mu_t}{\mu_t - 1}\right)\right) + \frac{\varphi}{2} \left[\log\left(\frac{\varphi}{2}\right) + \log\left(\frac{\mu_t - 1}{\mu_t}\right)\right]. \end{aligned}$$

Então, definindo

$$\rho_t := \frac{Y_t \varphi(\mu_t - 1)}{2\mu_t}, \quad (3.15)$$

e observando que

$$\frac{\partial \rho_t}{\partial \mu_t} = \frac{Y_t \varphi}{2\mu_t^2} = \frac{\rho_t}{\mu_t(\mu_t - 1)} \quad \text{e} \quad \frac{\partial \rho_t}{\partial \varphi} = \frac{Y_t(\mu_t - 1)}{2\mu_t} = \frac{\rho_t}{\varphi}, \quad (3.16)$$

obtem-se

$$\begin{aligned} \frac{\partial \ell_t(\gamma)}{\partial \mu_t} &= \frac{\log(\rho_t + 1)}{(\mu_t - 1)^2} - \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{1}{(\rho_t + 1)} \frac{\partial \rho_t}{\partial \mu_t} \\ &\quad - \frac{1}{(\mu_t - 1)^2} \left[\psi \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi \left(\frac{\mu_t}{\mu_t - 1} \right) \right] + \frac{\varphi}{2\mu_t(\mu_t - 1)} \\ &= \frac{\log(\rho_t + 1)}{(\mu_t - 1)^2} - \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{1}{\mu_t(\mu_t - 1)} \frac{\rho_t}{(\rho_t + 1)} \\ &\quad - \frac{1}{(\mu_t - 1)^2} \left[\psi \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi \left(\frac{\mu_t}{\mu_t - 1} \right) \right] + \frac{\varphi}{2\mu_t(\mu_t - 1)} \end{aligned}$$

e

$$\begin{aligned} \frac{\partial \ell_t(\gamma)}{\partial \varphi} &= \frac{\log(Y_t)}{2} - \frac{1}{2} \log(\rho_t + 1) - \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{1}{(\rho_t + 1)} \frac{\partial \rho_t}{\partial \varphi} \\ &\quad + \frac{1}{2} \left[\psi \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi \left(\frac{\varphi}{2} \right) + \log \left(\frac{\varphi}{2} \right) + \log \left(\frac{\mu_t - 1}{\mu_t} \right) + 1 \right] \\ &= \frac{\log(Y_t)}{2} - \frac{1}{2} \log(\rho_t + 1) - \left(\frac{1}{2} + \frac{\mu_t}{\varphi(\mu_t - 1)} \right) \frac{\rho_t}{(\rho_t + 1)} \\ &\quad + \frac{1}{2} \left[\psi \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi \left(\frac{\varphi}{2} \right) + \log \left(\frac{\varphi}{2} \right) + \log \left(\frac{\mu_t - 1}{\mu_t} \right) + 1 \right]. \end{aligned}$$

De (3.16) conclui-se ainda que

$$\begin{aligned} \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} &= \frac{1}{(\mu_t - 1)^2(\rho_t + 1)} \frac{\partial \rho_t}{\partial \mu_t} - \frac{2 \log(1 + \rho_t)}{(\mu_t - 1)^3} + \frac{1}{\mu_t(\mu_t - 1)} \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{1}{(\rho_t + 1)^2} \frac{\partial \rho_t}{\partial \mu_t} \\ &\quad - \frac{\rho_t}{(\rho_t + 1)} \left[\frac{\varphi}{2} \left(\frac{1}{\mu_t^2} - \frac{1}{(\mu_t - 1)^2} \right) - \frac{2}{(\mu_t - 1)^3} \right] + \frac{1}{(\mu_t - 1)^4} \left[\psi' \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \right. \\ &\quad \left. - \psi' \left(\frac{\mu_t}{\mu_t - 1} \right) \right] + \frac{2}{(\mu_t - 1)^3} \left[\psi \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi \left(\frac{\mu_t}{\mu_t - 1} \right) \right] - \frac{\varphi(2\mu_t - 1)}{2\mu_t^2(\mu_t - 1)^2} \\ &= \left(\frac{2\mu_t + 1}{\mu_t(\mu_t - 1)^3} + \frac{\varphi(2\mu_t - 1)}{2\mu_t^2(\mu_t - 1)^2} \right) \frac{\rho_t}{(\rho_t + 1)} + \frac{1}{\mu_t^2(\mu_t - 1)^2} \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{\rho_t}{(\rho_t + 1)^2} \\ &\quad - \frac{2 \log(1 + \rho_t)}{(\mu_t - 1)^3} + \frac{1}{(\mu_t - 1)^4} \left[\psi' \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi' \left(\frac{\mu_t}{\mu_t - 1} \right) \right] - \frac{\varphi(2\mu_t - 1)}{2\mu_t^2(\mu_t - 1)^2} \\ &\quad + \frac{2}{(\mu_t - 1)^3} \left[\psi \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi \left(\frac{\mu_t}{\mu_t - 1} \right) \right]. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} &= -\frac{1}{2(\rho_t + 1)} \frac{\partial \rho_t}{\partial \varphi} + \frac{\mu_t}{\varphi^2(\mu_t - 1)} \frac{\rho_t}{(\rho_t + 1)} - \left(\frac{1}{2} + \frac{\mu_t}{\varphi(\mu_t - 1)} \right) \frac{1}{(\rho_t + 1)^2} \frac{\partial \rho_t}{\partial \varphi} \\ &\quad + \frac{1}{4} \left[\psi' \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi' \left(\frac{\varphi}{2} \right) \right] + \frac{1}{2\varphi} \\ &= \left(-\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{\rho_t}{\varphi^2(\rho_t + 1)} - \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{\rho_t}{\varphi^2(\rho_t + 1)^2} \\ &\quad + \frac{1}{4} \left[\psi' \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) - \psi' \left(\frac{\varphi}{2} \right) \right] + \frac{1}{2\varphi}. \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} &= \frac{1}{(\mu_t - 1)^2} \frac{1}{(\rho_t + 1)} \frac{\partial \rho_t}{\partial \varphi} - \frac{1}{2\mu_t(\mu_t - 1)} \frac{\rho_t}{(\rho_t + 1)} + \frac{1}{2\mu_t(\mu_t - 1)} \\
&\quad - \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{1}{\mu_t(\mu_t - 1)} \frac{1}{(\rho_t + 1)^2} \frac{\partial \rho_t}{\partial \varphi} - \frac{1}{2(\mu_t - 1)^2} \psi' \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \\
&= \left[\frac{1}{\varphi(\mu_t - 1)^2} - \frac{1}{2\mu_t(\mu_t - 1)} \right] \frac{\rho_t}{(\rho_t + 1)} - \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) \frac{1}{\mu_t(\mu_t - 1)} \frac{\rho_t}{\varphi(\rho_t + 1)^2} \\
&\quad - \frac{1}{2(\mu_t - 1)^2} \psi' \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1} \right) + \frac{1}{2\mu_t(\mu_t - 1)}.
\end{aligned}$$

Note que, de (3.13) e (3.15) segue que

$$\rho_t = \frac{Y_t \varphi(\mu_t - 1)}{2\mu_t} = \frac{d_1}{d_2} Y_t \quad \text{e} \quad \frac{\rho_t}{1 + \rho_t} = \frac{Y_t d_1/d_2}{1 + Y_t d_1/d_2}.$$

Logo, usando a relação (3.13), de (3.9) obtém-se

$$\mathbb{E}(\log(Y_t) | \mathcal{F}_{t-1}) = -\log\left(\frac{\mu_t - 1}{\mu_t}\right) - \log\left(\frac{\varphi}{2}\right) + \psi\left(\frac{\varphi}{2}\right) - \psi\left(\frac{\mu_t}{\mu_t - 1}\right),$$

e, de (3.10) e (3.11) segue que

$$\mathbb{E}(\log(\rho_t + 1) | \mathcal{F}_{t-1}) = \psi\left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1}\right) - \psi\left(\frac{\mu_t}{\mu_t - 1}\right)$$

e

$$\mathbb{E}\left(\frac{\rho_t}{1 + \rho_t} \middle| \mathcal{F}_{t-1}\right) = \frac{\varphi}{2} \left(\frac{\varphi}{2} + \frac{\mu_t}{\mu_t - 1}\right)^{-1}.$$

Portanto,

$$\mathbb{E}\left(\frac{\partial \ell_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1}\right) = 0 \quad \text{e} \quad \mathbb{E}\left(\frac{\partial \ell_t(\gamma)}{\partial \varphi} \middle| \mathcal{F}_{t-1}\right) = 0,$$

como esperado.

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição F (Fisher-Snedecor) utiliza, respectivamente, as funções `df` e `rf` implementadas no pacote `stats` (R Core Team, 2021), que faz parte do pacote básico do R. O pacote `stats` implementa tanto a distribuição F não-central, cuja densidade é dada por

$$f(y) = \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \frac{e^{-\lambda/2} (\lambda/2)^k}{k! B(d_1/2 + k, d_2/2)} \left(\frac{d_1}{d_2}\right)^{d_1/2+k} y^{d_1/2-1+k} \left(1 + \frac{d_1}{d_2} y\right)^{-(d_1+d_2)/2-k},$$

onde $\lambda, d_1, d_2, y > 0$, quanto a F central cuja densidade é dada por (3.8). No pacote `stats`, a distribuição F não-central possui um algoritmo diferente da F central que é selecionado somente quando o parâmetro de não centralidade λ (`ncp` nas funções `df` e `rf`) é definido como sendo diferente de zero. Neste trabalho consideramos apenas a F central.

3.3 Distribuição Gama

Existem duas parametrizações amplamente utilizadas na literatura para a distribuição Gama. As duas parametrizações clássicas e a relação entre elas são apresentadas no que segue.

Parametrização 1

Uma variável aleatória Y tem distribuição Gama com parâmetros α e θ , denotada por $Y \sim \text{Gama}(\alpha, \theta)$, se sua função de densidade de probabilidade é dada por

$$f(y|\alpha, \theta) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-y/\theta}, \quad 0 < y < \infty, \quad \alpha, \theta > 0. \quad (3.17)$$

Nessa parametrização α e θ são denominados, respectivamente, parâmetro de forma (*shape*) e de escala (*scale*).

Parametrização 2

Uma variável aleatória Y tem distribuição Gama com parâmetros α e β , denotada por $Y \sim \text{Gama}(\alpha, \beta)$, se sua função de densidade de probabilidade é dada por

$$f(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \quad 0 < y < \infty, \quad \alpha, \beta > 0. \quad (3.18)$$

O parâmetro α possui a mesma interpretação da parametrização 1, enquanto que β é denominado parâmetro de taxa (*rate*), ou escala inversa, e satisfaz $\beta = 1/\theta$.

Propriedades da Distribuição Gama

Observe que, de (2.10) e (3.17) segue que, na parametrização 1,

$$\mathbb{E}(Y^k) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1+k} e^{-y/\theta} dy = \theta^k \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}, \quad k > 0.$$

Então, utilizando a relação (2.2), conclui-se que

$$\mathbb{E}(Y) = \alpha\theta \quad \text{e} \quad \text{Var}(Y) = \alpha\theta^2, \quad \alpha, \theta > 0.$$

De forma análoga, para a parametrização 2 conclui-se que

$$\mathbb{E}(Y^k) = \frac{\Gamma(\alpha+k)}{\beta^k \Gamma(\alpha)}, \quad k > 0, \quad \mathbb{E}(Y) = \frac{\alpha}{\beta} \quad \text{e} \quad \text{Var}(Y) = \frac{\alpha}{\beta^2}. \quad (3.19)$$

Outra relação de interesse neste trabalho é a esperança do logaritmo de Y . Utilizando-se a expressão (2.11), conclui-se que, para a parametrização 1,

$$\mathbb{E}(\log(Y)) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} \int_0^\infty \log(y) y^{\alpha-1} e^{-y/\theta} dy = \psi(\alpha) + \log(\theta)$$

e, de forma análoga, para a parametrização 2,

$$\mathbb{E}(\log(Y)) = \psi(\alpha) - \log(\beta). \quad (3.20)$$

Reparametrização

Assumindo $\alpha, \theta > 0$, e considerando-se a reparametrização para a expressão (3.17)

$$\mu := \alpha\theta \quad \text{e} \quad \varphi := \alpha,$$

obtém-se $\alpha = \varphi$ e $\theta = \mu/\alpha$. Alternativamente, assumindo $\alpha, \beta > 0$, e considerando-se a seguinte reparametrização para a expressão (3.18)

$$\mu := \frac{\alpha}{\beta} \quad \text{e} \quad \varphi := \alpha, \tag{3.21}$$

obtém-se $\alpha = \varphi$ e $\beta = \varphi/\mu$. Consequentemente, tanto (3.17), quanto (3.18) podem ser reescritas como

$$f(y|\mu, \varphi) = \left(\frac{\varphi}{\mu}\right)^\varphi \frac{1}{\Gamma(\varphi)} y^{\varphi-1} e^{-y\varphi/\mu}, \quad 0 < y < \infty, \quad \mu, \varphi > 0, \tag{3.22}$$

cuja notação correspondente é $Y \sim \text{Gama}(\mu, \varphi)$. Tem-se ainda que

$$\text{Var}(Y) = \frac{\mu^2}{\varphi}, \quad \varphi > 0.$$

A Figura 3.3 mostra o comportamento da função densidade de probabilidade da distribuição Gama, reparametrizada em termos de μ e φ , conforme (3.22). Nessa figura considera-se $\mu = 2.5$ fixo e $\varphi \in \{0.5, 1, 2, 15, 30\}$.

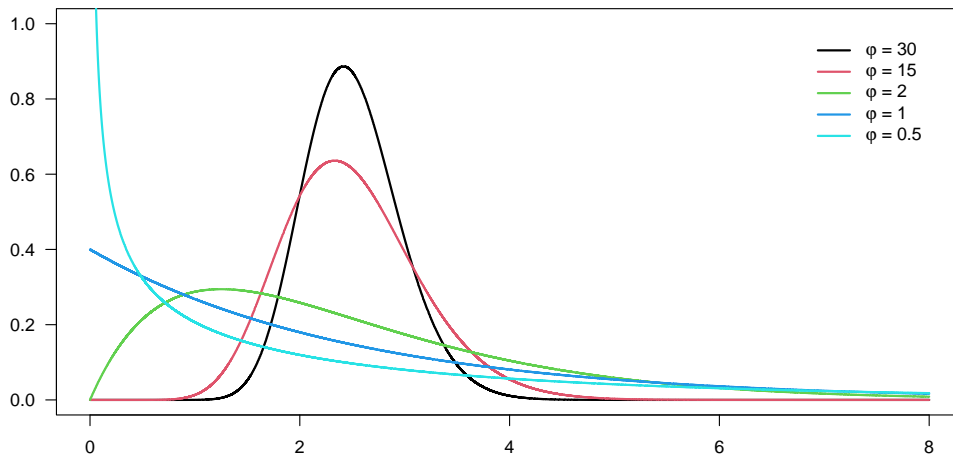


Figura 3.3: Distribuição Gama para $\mu = 2.5$ e diferentes valores para φ .

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim \text{Gama}(\mu_t, \varphi)$ conclui-se que

$$l_t(\gamma) = \varphi \log(\varphi) - \log(\Gamma(\varphi)) - \varphi \log(\mu_t) + (\varphi - 1) \log(Y_t) - \frac{Y_t \varphi}{\mu_t},$$

de onde segue que

$$\frac{\partial l_t(\gamma)}{\partial \mu_t} = \frac{Y_t \varphi}{\mu_t^2} - \frac{\varphi}{\mu_t}, \quad \frac{\partial l_t(\gamma)}{\partial \varphi} = 1 + \log(\varphi) - \psi(\varphi) - \log(\mu_t) + \log(Y_t) - \frac{Y_t}{\mu_t},$$

e ainda,

$$\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} = \frac{\varphi(\mu_t - 2Y_t)}{\mu_t^3}, \quad \frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} = \frac{1}{\varphi} - \psi'(\varphi) \quad \text{e} \quad \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} = \frac{Y_t - \mu_t}{\mu_t^2}.$$

Então, dado que $\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \mu_t$ e, pelas expressões (3.21) (3.20),

$$\mathbb{E}(\log(Y_t) | \mathcal{F}_{t-1}) = \psi(\varphi) - \log(\varphi) + \log(\mu_t),$$

segue que

$$\mathbb{E}\left(\frac{\partial \ell_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1}\right) = 0 \quad \text{e} \quad \mathbb{E}\left(\frac{\partial \ell_t(\gamma)}{\partial \varphi} \middle| \mathcal{F}_{t-1}\right) = 0,$$

como esperado, e ainda,

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1}\right) = -\frac{\varphi}{\mu_t^2}, \quad \mathbb{E}\left(\frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} \middle| \mathcal{F}_{t-1}\right) = \frac{1}{\varphi} - \psi'(\varphi) \quad \text{e} \quad \mathbb{E}\left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} \middle| \mathcal{F}_{t-1}\right) = 0.$$

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição Gama utiliza, respectivamente, as funções `dgamma` e `rgamma` implementadas no pacote `stats` (R Core Team, 2021), que faz parte do pacote básico do R. As funções implementadas no pacote `stats` permitem que ambas parametrizações sejam utilizadas. A parametrização 1 é selecionada quando atribui-se valores aos argumentos `shape`, que corresponde ao parâmetro α , e `scale`, que corresponde ao parâmetro θ . Neste caso, automaticamente, as funções definem `scale = 1/rate`. Para obter-se a parametrização 2 basta atribuir valores aos parâmetros `shape` e `rate`, que corresponde ao parâmetro β , deixando o parâmetro `scale` livre.

3.4 Distribuição Gaussiana Inversa

Uma variável aleatória Y tem distribuição Gaussiana Inversa com parâmetros μ e λ , denotada por $Y \sim \text{IG}(\mu, \lambda)$, se sua função de densidade de probabilidade é dada por

$$f(y|\mu, \lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left\{-\frac{\lambda(y - \mu)^2}{2y\mu^2}\right\}, \quad 0 < y < \infty, \quad \mu, \lambda > 0. \quad (3.23)$$

Propriedades da Distribuição Gaussiana Inversa

Observe que, utilizando (2.12) tem-se

$$\begin{aligned} \mathbb{E}(Y^k) &= \sqrt{\frac{\lambda}{2\pi}} \int_0^\infty y^{k-3/2} \exp\left\{-\frac{\lambda(y - \mu)^2}{2y\mu^2}\right\} dy \\ &= \sqrt{\frac{\lambda}{2\pi}} e^{\lambda/\mu} \int_0^\infty y^{k-3/2} \exp\left\{-\frac{\lambda}{2\mu^2}y - \frac{\lambda}{2y}\right\} dy = \sqrt{\frac{\lambda}{2\pi}} e^{\lambda/\mu} \mu^{k-1/2} K_{k-1/2}(\lambda/\mu). \end{aligned}$$

Então, observando-se que $K_{-1/2} = K_{1/2}$, $K_{3/2} = K_{1+1/2}$ e aplicando-se (2.7), conclui-se que

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{\mu^3}{\lambda}, \quad \mu, \lambda > 0,$$

ou seja, a média da distribuição existe sempre que $\mu, \lambda > 0$. Além disso, tem-se ainda que (veja Shuster, 1968)

$$\frac{(Y - \mu)^2}{Y\mu^2\varphi} \sim \chi_1^2. \quad (3.24)$$

Reparametrização

Assumindo $\mu, \lambda > 0$ e considerando-se a reparametrização

$$\mu := \mu \quad \text{e} \quad \varphi := \frac{1}{\lambda},$$

obtem-se $\lambda = 1/\varphi$ e, conseqüentemente a expressão (3.23) pode ser reescrita como

$$f(y|\mu, \varphi) = \sqrt{\frac{1}{2\pi y^3 \varphi}} \exp\left\{-\frac{(y - \mu)^2}{2y\mu^2\varphi}\right\}, \quad 0 < y < \infty, \quad \mu, \varphi > 0, \quad (3.25)$$

cuja notação correspondente é $Y \sim \text{IG}(\mu, \varphi)$. Tem-se ainda que:

$$\text{Var}(Y) = \mu^3 \varphi, \quad \varphi > 0.$$

A Figura 3.4 mostra o comportamento da função densidade de probabilidade da distribuição Gaussiana Inversa, reparametrizada em termos de μ e φ , conforme (3.25). Nessa figura considera-se $\mu = 2.5$ fixo e $\varphi \in \{0.01, 0.5, 1, 2, 3\}$.

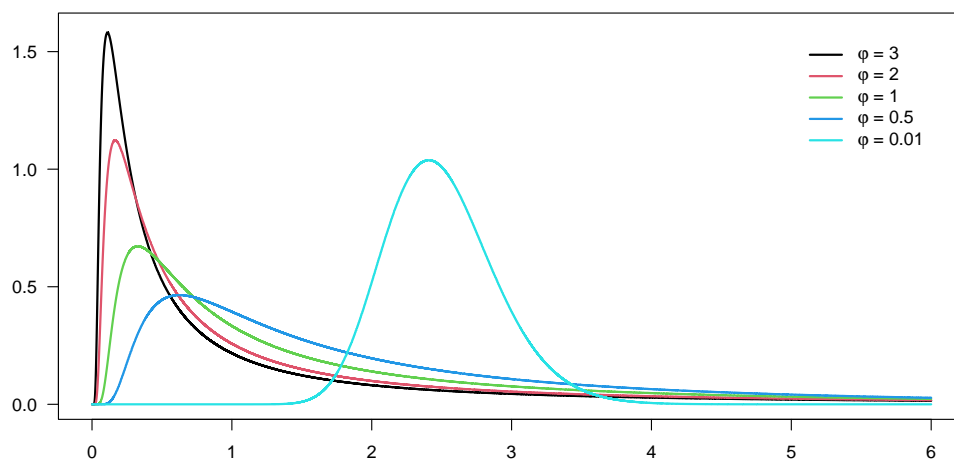


Figura 3.4: Distribuição Gaussiana Inversa para $\mu = 2.5$ e diferentes valores para φ .

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim \text{IG}(\mu_t, \varphi)$ tem-se

$$\ell_t(\gamma) = -\frac{1}{2} [\log(\varphi) + \log(2\pi) + 3 \log(Y_t)] - \frac{(Y_t - \mu_t)^2}{2Y_t\mu_t^2\varphi},$$

de onde segue que

$$\begin{aligned} \frac{\partial \ell_t(\gamma)}{\partial \mu_t} &= \frac{(Y_t - \mu_t)}{\mu_t^3 \varphi}, & \frac{\partial \ell_t(\gamma)}{\partial \varphi} &= -\frac{1}{2\varphi} + \frac{(Y_t - \mu_t)^2}{2Y_t \mu_t^2 \varphi^2}, \\ \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} &= \frac{2\mu_t - 3Y_t}{\mu_t^4 \varphi}, & \frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} &= \frac{1}{2\varphi^2} - \frac{(Y_t - \mu_t)^2}{Y_t \mu_t^2 \varphi^3} \quad \text{e} \quad \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} = \frac{3Y_t - 2\mu_t}{\mu_t^4 \varphi^2}. \end{aligned}$$

Agora, observando-se que $\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \mu_t$ e, por (3.24),

$$\mathbb{E}\left(\frac{(Y_t - \mu_t)^2}{Y_t \mu_t^2 \varphi} \middle| \mathcal{F}_{t-1}\right) = 1,$$

conclui-se que

$$\mathbb{E}\left(\frac{\partial \ell_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1}\right) = 0 \quad \text{e} \quad \mathbb{E}\left(\frac{\partial \ell_t(\gamma)}{\partial \varphi} \middle| \mathcal{F}_{t-1}\right) = 0,$$

como esperado, e

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1}\right) = -\frac{1}{\mu_t^3 \varphi}, \quad \mathbb{E}\left(\frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} \middle| \mathcal{F}_{t-1}\right) = -\frac{1}{2\varphi^2} \quad \text{e} \quad \mathbb{E}\left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} \middle| \mathcal{F}_{t-1}\right) = \frac{1}{\mu_t^3 \varphi^2}.$$

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição Gaussiana Inversa utiliza, respectivamente, as funções `dinvGauss` e `rinvGauss` implementadas no pacote `SuppDists` (Wheeler, 2020). As funções implementadas no pacote `SuppDist` correspondem à parametrização (3.23).

3.5 Distribuição Log-Logística

Uma variável aleatória Y tem distribuição Log-Logística com parâmetros α e β , denotada por $Y \sim \text{LL}(\alpha, \beta)$, se sua função densidade de probabilidade é dada por

$$f(y|\alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{y}{\alpha}\right)^{\beta-1} \left[1 + \left(\frac{y}{\alpha}\right)^\beta\right]^{-2}, \quad 0 < y < \infty, \quad \alpha, \beta > 0. \quad (3.26)$$

Nessa parametrização α e β são denominados, respectivamente, parâmetro de escala (*scale*) e de forma (*shape*).

Propriedades da Distribuição Log-Logística

Observe que, de (2.13) segue que

$$\mathbb{E}(Y^k) = \beta \alpha^\beta \int_0^\infty \frac{y^{\beta+k-1}}{(\alpha^\beta + y^\beta)^2} dy = \alpha^k B\left(1 - \frac{k}{\beta}, 1 + \frac{k}{\beta}\right) = \frac{\alpha^k k \pi}{\beta \sin(k\pi/\beta)}, \quad k < \beta,$$

onde a última igualdade é consequência de (2.3) e (2.5). Então, utilizando-se as relações (2.2) e (2.5), conclui-se que

$$\mathbb{E}(Y) = \frac{\alpha\pi}{\beta \sin(\pi/\beta)}, \quad \beta > 1, \quad \text{Var}(Y) = \alpha^2 \left[\frac{2\pi}{\beta \sin(\pi/\beta)} - \left(\frac{\pi}{\beta \sin(\pi/\beta)} \right)^2 \right] \quad \text{e} \quad \beta > 2,$$

ou seja, a média da distribuição existe sempre que $\beta > 1$. Portanto, deve-se adotar essa restrição no momento da reparametrização.

Outras relações interessantes são que, se $Y \sim \text{LL}(\alpha, \beta)$, então

$$\theta Y \sim \text{LL}(\theta\alpha, \beta) \quad \text{e} \quad Y^\theta \sim \text{LL}(\alpha^\theta, \beta/\theta) \quad \theta > 0. \quad (3.27)$$

A prova desse resultado é imediata, bastando observar que

$$P(Y \leq y) = \frac{1}{1 + (y/\alpha)^{-\beta}}, \quad P(\theta Y \leq y) = P(Y \leq y/\theta) \quad \text{e} \quad P(Y^\theta \leq y) = P(Y \leq y^{1/\theta}).$$

Como consequência imediata de (3.27) tem-se que, se $Y \sim \text{LL}(\alpha, \beta)$, então $\alpha^{-\beta} Y^\beta \sim \text{LL}(1, 1)$. Essa propriedade e os resultados que seguem são importantes para o cálculo de esperanças envolvendo derivadas da função de log-verossimilhança condicional.

Defina $Z := \alpha^{-\beta} Y^\beta$, então $Z \sim \text{LL}(1, 1)$ e assim,

$$\mathbb{E}\left(\frac{1}{1+Z}\right) = \int_0^\infty \frac{1}{(1+z)^3} dz = -\frac{1}{2(z+1)^2} \Big|_0^\infty = \frac{1}{2}. \quad (3.28)$$

Além disso, observando que

$$t = \frac{1}{1+z} \quad \text{implica} \quad dt = -\frac{1}{(1+z)^2} dz \quad \text{e} \quad z = \frac{1-t}{t},$$

conclui-se que

$$\mathbb{E}\left(\frac{Z}{(1+Z)^2}\right) = \int_0^\infty \frac{z}{(1+z)^4} dz = \int_0^1 t(1-t) dt = \frac{1}{6}, \quad (3.29)$$

$$\mathbb{E}(\log(Z)) = \int_0^\infty \frac{\log(z)}{(1+z)^2} dz = \int_0^1 \log(1-t) dt - \int_0^1 \log(t) dt = 0, \quad (3.30)$$

$$\mathbb{E}\left(\frac{\log(Z)}{1+Z}\right) = \int_0^\infty \frac{\log(z)}{(1+z)^3} dz = \int_0^1 t \log(1-t) dt - \int_0^1 t \log(t) dt = -\frac{1}{2} \quad (3.31)$$

e

$$\mathbb{E}\left(\frac{\log(Z)}{(1+Z)^2}\right) = \int_0^\infty \frac{\log(z)}{(1+z)^4} dz = \int_0^1 t^2 \log(1-t) dt - \int_0^1 t^2 \log(t) dt = -\frac{1}{2}, \quad (3.32)$$

onde a última igualdade é obtida via integração por partes, com $u = \log(1-t)$ para a primeira parcela, $u = \log(t)$ para a segunda parcela e $dv = t^k dt$ em ambos os casos, para $k = 1$ e 2 , respectivamente, em (3.31) e (3.32).

Observando-se ainda que

$$\frac{z}{(1+z)^4} = \frac{1}{(1+z)^3} - \frac{1}{(1+z)^4},$$

tem-se, por (3.31) e (3.32),

$$\mathbb{E}\left(\frac{Z \log(Z)}{(1+Z)^2}\right) = \int_0^\infty \frac{z \log(z)}{(1+z)^4} dz = \int_0^\infty \frac{\log(z)}{(1+z)^3} dz - \int_0^\infty \frac{\log(z)}{(1+z)^4} dz = 0. \quad (3.33)$$

Finalmente,

$$\mathbb{E}\left(\frac{Z[\log(Z)]^2}{(1+Z)^2}\right) = \int_0^\infty \frac{z[\log(z)]^2}{(1+z)^4} dz = \left[\frac{1}{3(1+z)} + \frac{(3+z)[z \log(z)]^2}{6(1+z)^3} - \frac{\log(z) \log(1+z)}{3} + \frac{z \log(z)}{3(1+z)^2} - \frac{\text{Li}_2(-z)}{3} \right] \Big|_0^\infty = \frac{\pi^2 - 6}{18}, \quad (3.34)$$

onde $\text{Li}_2(\cdot)$ é a função polilogaritmo de ordem 2, também conhecida como dilogaritmo, que é definida por

$$\text{Li}_2(z) = - \int_0^z \frac{\ln(1-t)}{t} dt, \quad z \in \mathbb{C},$$

e satisfaz (Morris, 1979)

$$\text{Li}_2(0) = 0 \quad \text{e} \quad \text{Li}_2(-z) \sim -\frac{\pi^2}{6} - \frac{1}{2}[\log(z)]^2, \quad \text{quando } z \in \mathbb{R}, \quad z \rightarrow \infty.$$

A obtenção de (3.34) envolve substituição de variáveis e múltiplas integrações por partes. Devido à extensão dos cálculos o passo a passo foi omitido.

Reparametrização

Assumindo $\beta > 1$ e considerando a reparametrização

$$\mu := \frac{\alpha\pi}{\beta \sin(\pi/\beta)} \quad \text{e} \quad \varphi := \beta, \quad (3.35)$$

obtemos $\alpha = \mu\varphi\pi^{-1} \sin(\pi/\varphi)$, $\beta = \varphi$ e, conseqüentemente, (3.26) pode ser reescrita como

$$f(y|\mu, \varphi) = \frac{\varphi}{y} \left(\frac{y\pi}{\mu\varphi \sin(\pi/\varphi)} \right)^\varphi \left(1 + \left[\frac{y\pi}{\mu\varphi \sin(\pi/\varphi)} \right]^\varphi \right)^{-2}, \quad 0 < y < \infty, \quad \mu > 0, \quad \varphi \geq 1, \quad (3.36)$$

cuja notação correspondente é $Y \sim \text{LL}(\mu, \varphi)$. Tem-se ainda que

$$\text{Var}(Y) = \left[\frac{\mu\varphi}{\pi} \sin\left(\frac{\pi}{\varphi}\right) \right]^2 \left[\frac{2\pi}{\beta \sin(\pi/\varphi)} - \left(\frac{\pi}{\varphi \sin(\pi/\varphi)} \right)^2 \right], \quad \varphi > 2,$$

A Figura 3.5 mostra o comportamento da função densidade de probabilidade da distribuição Log-Logística, reparametrizada em termos de μ e φ , conforme (3.36). Nessa figura considera-se $\mu = 2.5$ fixo e $\varphi \in \{1.3, 2, 3, 5, 10\}$.

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim \text{LL}(\mu_t, \varphi)$ conclui-se que

$$\begin{aligned} \ell_t(\gamma) &= \log(\varphi) - \log(Y_t) + \log\left(\left[\frac{Y_t\pi}{\mu_t\varphi \sin(\pi/\varphi)}\right]^\varphi\right) - 2\log\left(1 + \left[\frac{Y_t\pi}{\mu_t\varphi \sin(\pi/\varphi)}\right]^\varphi\right) \\ &= \log(\varphi) - \log(Y_t) + \log(\delta_t) - 2\log(1 + \delta_t), \end{aligned}$$

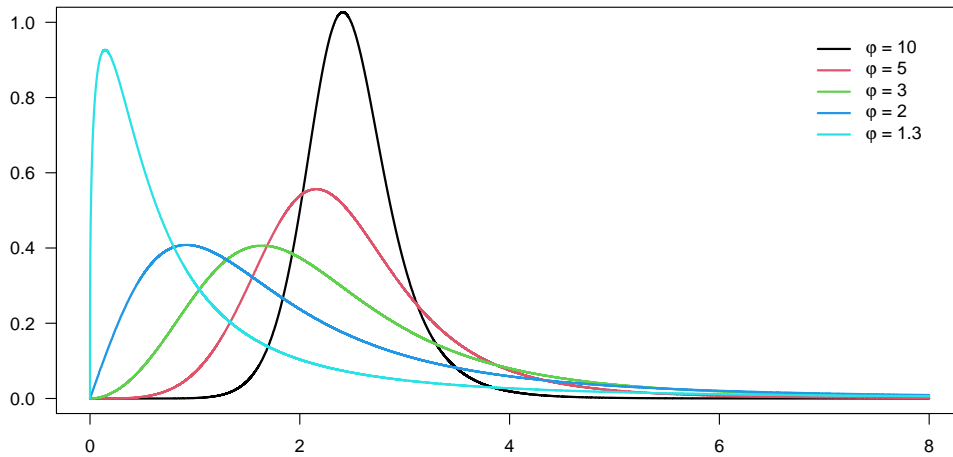


Figura 3.5: Distribuição Log-Logística para $\mu = 2.5$ e diferentes valores para φ .

onde

$$\delta_t := \left(\frac{Y_t \pi}{\mu_t \varphi \sin(\pi/\varphi)} \right)^\varphi. \quad (3.37)$$

Observe que

$$\frac{\partial \delta_t}{\partial \mu_t} = -\frac{\varphi}{\mu_t} \left(\frac{Y_t \pi}{\mu_t \varphi \sin(\pi/\varphi)} \right)^\varphi = -\frac{\varphi \delta_t}{\mu_t}$$

e

$$\frac{\partial \delta_t}{\partial \varphi} = \frac{1}{\varphi} \left(\frac{Y_t \pi}{\mu_t \varphi \sin(\pi/\varphi)} \right)^\varphi \left[\log \left(\left[\frac{Y_t \pi}{\mu_t \varphi \sin(\pi/\varphi)} \right]^\varphi \right) - \varphi + \pi \cot \left(\frac{\pi}{\varphi} \right) \right] = \frac{\delta_t \omega_t}{\varphi},$$

onde

$$\omega_t := \log(\delta_t) - \varphi + \pi \cot \left(\frac{\pi}{\varphi} \right). \quad (3.38)$$

Logo,

$$\frac{\partial \ell_t(\gamma)}{\partial \mu_t} = \frac{1 - \delta_t}{\delta_t(1 + \delta_t)} \frac{\partial \delta_t}{\partial \mu_t} = \frac{\varphi(\delta_t - 1)}{\mu_t(1 + \delta_t)} = \frac{\varphi}{\mu_t} \left(1 - \frac{2}{1 + \delta_t} \right)$$

e

$$\frac{\partial \ell_t(\gamma)}{\partial \varphi} = \frac{1}{\varphi} + \frac{1 - \delta_t}{\delta_t(1 + \delta_t)} \frac{\partial \delta_t}{\partial \varphi} = \frac{1}{\varphi} - \frac{\omega_t}{\varphi} \left(1 - \frac{2}{1 + \delta_t} \right).$$

Observando ainda que

$$\frac{\partial \omega_t}{\partial \varphi} = \frac{1}{\delta_t} \frac{\partial \delta_t}{\partial \varphi} - 1 + \left(\frac{\pi}{\varphi \sin(\pi/\varphi)} \right)^2 = \frac{\omega_t}{\varphi} + \left(\frac{\pi}{\varphi \sin(\pi/\varphi)} \right)^2 - 1,$$

conclui-se que

$$\begin{aligned} \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} &= -\frac{\varphi}{\mu_t^2} \left(1 - \frac{2}{1 + \delta_t} \right) + \frac{\varphi}{\mu_t} \frac{2}{(1 + \delta_t)^2} \frac{\partial \delta_t}{\partial \mu_t} \\ &= \frac{\varphi}{\mu_t^2} \left(\frac{2}{1 + \delta_t} - \frac{2\varphi \delta_t}{(1 + \delta_t)^2} - 1 \right) = -\frac{1}{\mu_t} \frac{\partial \ell_t(\gamma)}{\partial \mu_t} - \frac{2\varphi^2}{\mu_t^2} \frac{\delta_t}{(1 + \delta_t)^2}, \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} &= -\frac{1}{\varphi^2} + \frac{\omega_t}{\varphi^2} \left(1 - \frac{2}{1 + \delta_t}\right) - \frac{1}{\varphi} \left(1 - \frac{2}{1 + \delta_t}\right) \frac{\partial \omega_t}{\partial \varphi} - \frac{\omega_t}{\varphi} \frac{2}{(1 + \delta_t)^2} \frac{\partial \delta_t}{\partial \varphi} \\
&= -\frac{1}{\varphi^2} + \frac{\omega_t}{\varphi^2} \left(1 - \frac{2}{1 + \delta_t}\right) - \frac{1}{\varphi} \left(1 - \frac{2}{1 + \delta_t}\right) \left[\frac{\omega_t}{\varphi} + \left(\frac{\pi}{\varphi \sin(\pi/\varphi)} \right)^2 - 1 \right] - \frac{2}{\varphi^2} \frac{\delta_t \omega_t^2}{(1 + \delta_t)^2} \\
&= -\frac{1}{\varphi^2} - \frac{1}{\varphi} \left(1 - \frac{2}{1 + \delta_t}\right) \left[\left(\frac{\pi}{\varphi \sin(\pi/\varphi)} \right)^2 - 1 \right] - \frac{2}{\varphi^2} \frac{\delta_t \omega_t^2}{(1 + \delta_t)^2} \\
&= -\frac{1}{\varphi^2} - \frac{\mu_t}{\varphi^2} \left[\left(\frac{\pi}{\varphi \sin(\pi/\varphi)} \right)^2 - 1 \right] \frac{\partial \ell_t(\gamma)}{\partial \mu_t} - \frac{2}{\varphi^2} \frac{\delta_t \omega_t^2}{(1 + \delta_t)^2}
\end{aligned}$$

e

$$\begin{aligned}
\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} &= \frac{1}{\mu_t} \left(1 - \frac{2}{1 + \delta_t}\right) + \frac{\varphi}{\mu_t} \frac{2}{(1 + \delta_t)^2} \frac{\partial \delta_t}{\partial \varphi} \\
&= \frac{1}{\mu_t} \left(1 - \frac{2}{1 + \delta_t}\right) + \frac{2\delta_t \omega_t}{\mu_t (1 + \delta_t)^2} = \frac{1}{\varphi} \frac{\partial \ell_t(\gamma)}{\partial \mu_t} + \frac{2\delta_t \omega_t}{\mu_t (1 + \delta_t)^2}.
\end{aligned}$$

De (3.27), (3.35) e (3.37) conclui-se que $\delta_t \sim \text{LL}(\alpha = 1, \beta = 1)$. Ressaltamos que δ_t não pode ser reparametrizado conforme (3.36) pois a esperança de δ_t não existe. Então, (3.28), (3.30), (3.31) e (3.38) implicam que

$$\mathbb{E} \left(\frac{\partial \ell_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1} \right) = 0 \quad \text{e} \quad \mathbb{E} \left(\frac{\partial \ell_t(\gamma)}{\partial \varphi} \middle| \mathcal{F}_{t-1} \right) = 0.$$

E assim, (3.28) e (3.29) implicam

$$\mathbb{E} \left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) = -\frac{2\varphi^2}{\mu_t^2} \mathbb{E} \left(\frac{\delta_t}{(1 + \delta_t)^2} \middle| \mathcal{F}_{t-1} \right) = -\frac{\varphi^2}{3\mu_t^2},$$

(3.29), (3.33) e (3.34) implicam

$$\begin{aligned}
\mathbb{E} \left(\frac{\partial^2 \ell_t(\gamma)}{\partial \varphi^2} \middle| \mathcal{F}_{t-1} \right) &= -\frac{1}{\varphi^2} - \frac{2}{\varphi^2} \mathbb{E} \left(\frac{\delta_t [\log(\delta_t) - \varphi + \pi \cot(\pi/\varphi)]^2}{(1 + \delta_t)^2} \middle| \mathcal{F}_{t-1} \right) \\
&= -\frac{1}{\varphi^2} \left[\frac{1}{3} + \frac{\pi^2}{9} + \frac{1}{3} \left(\pi \cot\left(\frac{\pi}{\varphi}\right) - \varphi \right)^2 \right]
\end{aligned}$$

e

$$\mathbb{E} \left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t \partial \varphi} \middle| \mathcal{F}_{t-1} \right) = \frac{2}{\mu_t} \mathbb{E} \left(\frac{\delta_t [\log(\delta_t) - \varphi + \pi \cot(\pi/\varphi)]}{(1 + \delta_t)^2} \middle| \mathcal{F}_{t-1} \right) = \frac{1}{3\mu_t} \left[\pi \cot\left(\frac{\pi}{\varphi}\right) - \varphi \right].$$

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição Log-Logística utiliza, respectivamente, as funções `dllogis` e `rllogis` implementadas no pacote `actuar` (Dutang et al., 2008). As funções implementadas no pacote `actuar` correspondem à parametrização (3.26).

3.6 Distribuição Log-Normal

Uma variável aleatória Y que tem distribuição Log-Normal com parâmetros η e σ^2 , denotada por $Y \sim \text{LogNormal}(\eta, \sigma^2)$, se sua função densidade de probabilidade é dada por

$$f(y|\eta, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log(y) - \eta)^2}{2\sigma^2}\right\}, \quad 0 < y < \infty, \quad \eta \in \mathbb{R}, \quad \sigma^2 > 0. \quad (3.39)$$

Propriedades da Distribuição Log-Normal

Note que $Y \sim \text{LogNormal}(\eta, \sigma^2)$ se, e somente se, $\log(Y) \sim N(\eta, \sigma^2)$. Sendo assim, $Y = e^{\log(Y)}$ e $\log(Y) \sim N(\eta, \sigma^2)$ implica que

$$\mathbb{E}(Y^k) = \mathbb{E}(e^{k \log(Y)}) = \exp\left\{k\eta + \frac{k^2\sigma^2}{2}\right\}, \quad \eta \in \mathbb{R}, \quad \sigma^2 > 0,$$

onde a última igualdade segue da expressão da função geradora de momentos da distribuição normal. Então,

$$\mathbb{E}(Y) = e^{\eta + \sigma^2/2} \quad \text{e} \quad \text{Var}(Y) = (e^{\sigma^2} - 1)e^{2\eta + \sigma^2}, \quad \eta \in \mathbb{R}, \quad \sigma^2 > 0.$$

Além disso, do fato que $\log(Y) \sim N(\eta, \sigma^2)$ conclui-se ainda que

$$\mathbb{E}(\log(Y)) = \eta \quad \text{e} \quad \mathbb{E}([\log(Y)]^2) = \eta^2 + \sigma^2. \quad (3.40)$$

Reparametrização

Considerando-se a reparametrização

$$\mu := e^{\eta + \sigma^2/2} \quad \text{e} \quad \varphi := \sigma, \quad (3.41)$$

obtem-se $\sigma^2 = \varphi^2$ e $\eta = \log(\mu) - \varphi^2/2$. Consequentemente, a equação (3.39) pode ser reescrita como

$$f(y|\mu, \varphi) = \frac{1}{y\sqrt{2\pi\varphi^2}} \exp\left\{-\frac{(\log(y) - \log(\mu) + \varphi^2/2)^2}{2\varphi^2}\right\}, \quad 0 < y < \infty, \quad \mu > 0, \quad \varphi > 0, \quad (3.42)$$

cujas notação correspondente é $Y \sim \text{LogNormal}(\mu, \varphi)$. Tem-se ainda que

$$\text{Var}(Y) = (e^{\varphi^2} - 1)\mu^2, \quad \varphi > 0. \quad (3.43)$$

Pela expressão (3.43) conclui-se que $\text{Var}(Y) \rightarrow \infty$, quando $\varphi \rightarrow \infty$. Entretanto, na prática, valores relativamente baixos de φ podem levar a instabilidade computacional. Por exemplo, no R, $\mu = 1$ e $\varphi = 27$ já produz um resultado para $(e^{\varphi^2} - 1)\mu^2$ que é considerado infinito pela máquina. Do ponto de vista de simulação, o fato da variância crescer muito rápido com φ pode levar a instabilidade na geração de amostras da distribuição.

A Figura 3.6 mostra o comportamento da função densidade de probabilidade da distribuição Log-Normal, reparametrizada em termos de μ e φ , conforme (3.42). Nessa figura considera-se $\mu = 2.5$ fixo e $\varphi \in \{0.16, 0.2, 0.5, 1, 1.5\}$.

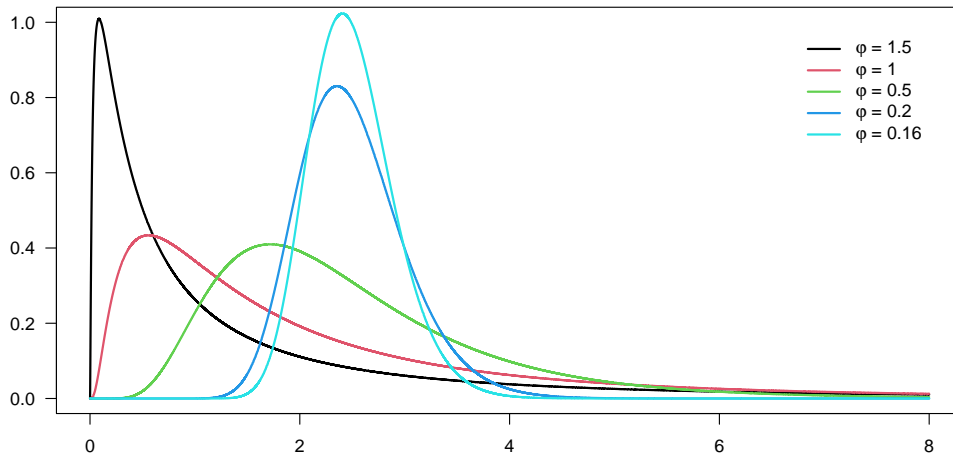


Figura 3.6: Distribuição Log-Normal para $\mu = 2.5$ e diferentes valores para φ .

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim \text{LogNormal}(\mu_t, \varphi)$ conclui-se que

$$l_t(\gamma) = -\log(Y_t) - \log(\varphi) - \frac{1}{2} \log(2\pi) - \frac{1}{2\varphi^2} \left(\log(Y_t) - \log(\mu_t) + \frac{\varphi^2}{2} \right)^2.$$

Segue que,

$$\frac{\partial l_t(\gamma)}{\partial \mu_t} = \frac{\log(Y_t) - \log(\mu_t) + \varphi^2/2}{\mu_t \varphi^2} \quad \text{e} \quad \frac{\partial l_t(\gamma)}{\partial \varphi} = \frac{(\log(Y_t) - \log(\mu_t))^2 - (\varphi^2/2)^2 - \varphi^2}{\varphi^3}.$$

Além disso,

$$\frac{\partial^2 l_t(\gamma)}{\partial \mu_t^2} = -\frac{\log(Y_t) - \log(\mu_t) + \varphi^2/2 + 1}{\mu_t^2 \varphi^2}, \quad \frac{\partial^2 l_t(\gamma)}{\partial \mu_t \partial \varphi} = -\frac{2(\log(Y_t) - \log(\mu_t))}{\mu_t \varphi^3}$$

e

$$\frac{\partial^2 l_t(\gamma)}{\partial \varphi^2} = -\frac{3(\log(Y_t) - \log(\mu_t))^2 + (\varphi^2/2)^2 - \varphi^2}{\varphi^4}.$$

De (3.40) e (3.41) conclui-se

$$\mathbb{E}(\log(Y_t)) = \log(\mu_t) - \frac{\varphi^2}{2}, \quad \mathbb{E}([\log(Y_t)]^2) = \left(\log(\mu_t) - \frac{\varphi^2}{2} \right)^2 + \varphi^2$$

e

$$\mathbb{E}([\log(Y_t) - \log(\mu_t)]^2 | \mathcal{F}_{t-1}) = \varphi^2 + \frac{\varphi^4}{4}.$$

Segue que

$$\mathbb{E}\left(\frac{\partial l_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1}\right) = 0 \quad \text{e} \quad \mathbb{E}\left(\frac{\partial l_t(\gamma)}{\partial \varphi} \middle| \mathcal{F}_{t-1}\right) = 0,$$

como esperando, e ainda,

$$\mathbb{E}\left(\frac{\partial^2 l_t(\gamma)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1}\right) = -\frac{1}{\mu_t^2 \varphi^2}, \quad \mathbb{E}\left(\frac{\partial^2 l_t(\gamma)}{\partial \varphi^2} \middle| \mathcal{F}_{t-1}\right) = -\frac{\varphi^2 + 2}{\varphi^2} \quad \text{e} \quad \mathbb{E}\left(\frac{\partial^2 l_t(\gamma)}{\partial \mu_t \partial \varphi} \middle| \mathcal{F}_{t-1}\right) = \frac{1}{\mu_t \varphi}.$$

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição Log-Normal utiliza, respectivamente, as funções `dlnorm` e `rlnorm` implementadas no pacote `stats` (R Core Team, 2021) que faz parte do pacote básico do R. O pacote `stats` implementa a versão generalizada da distribuição Log-Normal, cuja densidade é análoga à expressão (3.39).

3.7 Distribuição Qui-Quadrado

Uma variável aleatória Y tem distribuição Qui-Quadrado com $\nu > 0$ graus de liberdade, denotada por $Y \sim \chi_\nu^2$, se sua função densidade de probabilidade é dada por

$$f(y|\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2}, \quad 0 < y < \infty, \quad \nu > 0. \quad (3.44)$$

Propriedades da Distribuição Qui-Quadrado

Observe que, de (3.18) e (3.44) segue que se, $Y \sim \chi_\nu^2$, então $Y \sim \Gamma(\alpha, \beta)$, com $\alpha = \nu/2$ e $\beta = 1/2$. Logo, de (3.19) tem-se

$$\mathbb{E}(Y^k) = 2^k \frac{\Gamma(k + \nu/2)}{\Gamma(\nu/2)}.$$

Então, utilizando-se a relação (2.2), conclui-se que

$$\mathbb{E}(Y) = \nu \quad \text{e} \quad \text{Var}(Y) = \nu^2, \quad \nu > 0.$$

Reparametrização

Como a distribuição Qui-Quadrado já é parametrizada pela média, a notação da distribuição passa a ser representada por μ ao invés de ν . Consequentemente, (3.44) pode ser reescrita como

$$f(y|\mu) = \frac{1}{2^{\mu/2}\Gamma(\mu/2)} y^{\mu/2-1} e^{-y/2}, \quad 0 < y < \infty, \quad \mu > 0. \quad (3.45)$$

cuja notação correspondente é $Y \sim \chi_\mu^2$. Tem-se ainda que

$$\text{Var}(Y) = \mu^2.$$

A Figura 3.7 mostra o comportamento da função densidade de probabilidade da distribuição Qui-Quadrado, reparametrizada em termos de μ , conforme (3.45). Nessa figura considera-se $\mu \in \{0.5, 1, 2, 5, 10\}$.

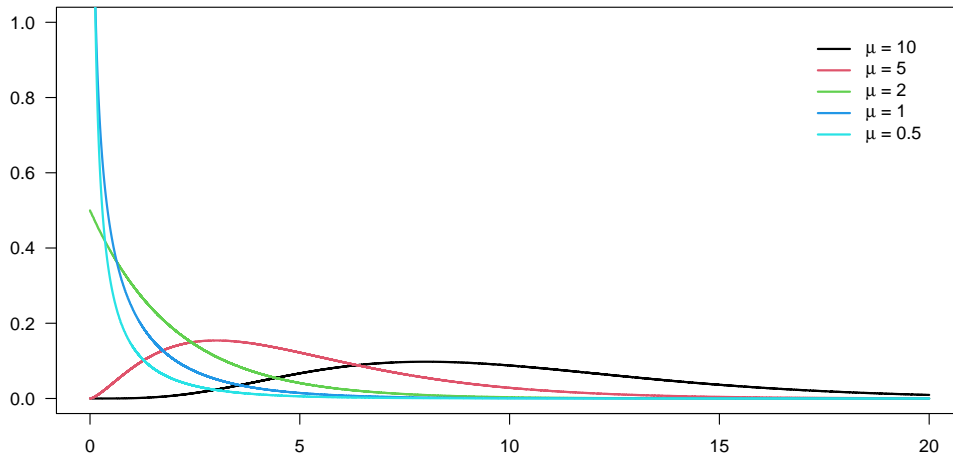


Figura 3.7: Distribuição Qui-Quadrado para diferentes valores de μ .

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim \chi_{\mu_t}^2$ conclui-se que

$$\ell_t(\gamma) = -\frac{\mu_t}{2} \log(2) - \log\left(\Gamma\left(\frac{\mu_t}{2}\right)\right) + \left(\frac{\mu_t}{2} - 1\right) \log(y_t) - \frac{Y_t}{2},$$

$$\frac{\partial \ell_t(\gamma)}{\partial \mu_t} = \frac{1}{2} \left[\log\left(\frac{Y_t}{2}\right) - \psi\left(\frac{\mu_t}{2}\right) \right] \quad \text{e} \quad \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} = -\frac{1}{4} \psi'\left(\frac{\mu_t}{2}\right).$$

Além disso, (2.11) e o fato que a segunda derivada é \mathcal{F}_{t-1} mensurável implicam, respectivamente, que

$$\mathbb{E}\left(\frac{\partial \ell_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1}\right) = 0 \quad \text{e} \quad \mathbb{E}\left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1}\right) = -\frac{1}{4} \psi'\left(\frac{\mu_t}{2}\right).$$

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição Qui-Quadrado utiliza, respectivamente, as funções `dchisq` e `rchisq` implementadas no pacote `stats` (R Core Team, 2021) que faz parte do pacote básico do R. O pacote `stats` implementa a versão generalizada da distribuição Qui-Quadrado, cuja densidade é análoga à expressão (3.44).

3.8 Distribuição Rayleigh

Uma variável aleatória Y tem distribuição Rayleigh com parâmetro de escala σ , denotada por $Y \sim \text{Rayleigh}(\sigma)$, se sua função densidade de probabilidade é dada por

$$f(y|\sigma) = \frac{y}{\sigma^2} \exp\left\{-\frac{y^2}{2\sigma^2}\right\}, \quad 0 < y < \infty, \quad \sigma > 0. \quad (3.46)$$

Propriedades da Distribuição Rayleigh

De (2.10) segue que

$$\mathbb{E}(Y^k) = \frac{1}{\sigma^2} \int_0^\infty y^{k+1} e^{-y^2/(2\sigma^2)} dy = \sigma^k 2^{k/2} \Gamma\left(\frac{k}{2} + 1\right), \quad k > 0. \quad (3.47)$$

Então, observando-se que $\Gamma(1/2) = \sqrt{\pi}$ e utilizando-se a igualdade (2.2) conclui-se que

$$\mathbb{E}(Y) = \sigma \sqrt{\frac{\pi}{2}}, \quad \mathbb{E}(Y^2) = 2\sigma^2 \quad \text{e} \quad \text{Var}(Y) = \frac{(4 - \pi)\sigma^2}{2}. \quad (3.48)$$

Reparametrização

Observe que, fazendo

$$\mu := \sigma \sqrt{\frac{\pi}{2}}, \quad (3.49)$$

tem-se que $\sigma = \mu \sqrt{2/\pi}$. Consequentemente, a equação (3.46) pode ser reescrita como

$$f(y|\mu) = \frac{y\pi}{2\mu^2} \exp\left\{\frac{-y^2\pi}{4\mu^2}\right\}, \quad (3.50)$$

cuja notação correspondente é $Y \sim \text{Rayleigh}(\mu)$. De (3.48) e (3.49) tem-se ainda que

$$\text{Var}(Y) = \frac{(4 - \pi)\mu^2}{\pi}.$$

A Figura 3.8 mostra o comportamento da função densidade de probabilidade da distribuição Rayleigh, reparametrizada em termos de μ , conforme (3.50). Nessa figura considera-se $\mu \in \{0.8, 1, 2, 5, 10\}$.

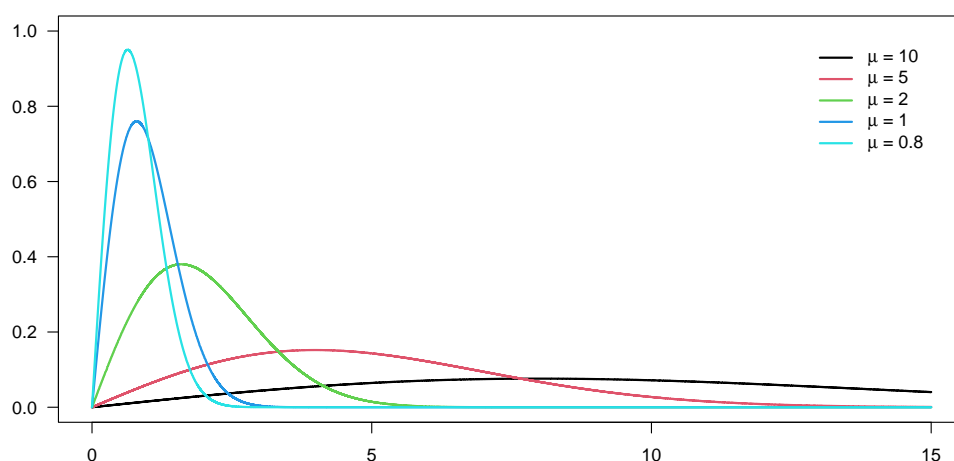


Figura 3.8: Distribuição Rayleigh para diferentes valores de μ .

Verossimilhança Condicional

Supondo que $Y_t | \mathcal{F}_{t-1} \sim \text{Rayleigh}(\mu_t)$ conclui-se que

$$\ell_t(\gamma) = \log(Y_t) + \log(\pi) - \log(2) - 2 \log(\mu_t) - \frac{Y_t^2 \pi}{4\mu_t^2},$$

de onde segue que

$$\frac{\partial \ell_t(\gamma)}{\partial \mu_t} = \frac{\pi Y_t^2 - 4\mu_t^2}{2\mu_t^3} \quad \text{e} \quad \frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} = \frac{4\mu_t^2 - 3\pi Y_t^2}{2\mu_t^4}.$$

Além disso, (3.47) e (3.49) implicam que

$$\mathbb{E} \left(\frac{\partial \ell_t(\gamma)}{\partial \mu_t} \middle| \mathcal{F}_{t-1} \right) = 0 \quad \text{e} \quad \mathbb{E} \left(\frac{\partial^2 \ell_t(\gamma)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) = -\frac{4}{\mu_t^2}.$$

Implementação no R

A implementação da densidade de probabilidade e do gerador de números aleatórios para a versão reparametrizada da distribuição Rayleigh utiliza, respectivamente, as funções `drayleigh` e `rrayleigh` implementadas no pacote `extraDistr` (Wolodzko, 2020). O pacote `extraDistr` implementa a versão generalizada da distribuição Rayleigh de forma análoga à apresentada na equação (3.46).

CAPÍTULO 4

MODELO DE REGRESSÃO PARA SÉRIES TEMPORAIS POSITIVAS

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo estocástico e \mathcal{F}_t a σ -álgebra gerada pela informação disponível até o tempo t . No que segue apresentamos uma breve descrição de algumas abordagens existentes na literatura para a modelagem de $\{Y_t\}_{t \in \mathbb{Z}}$. Em seguida, descrevemos os modelos estudados neste trabalho e sua relação com os modelos clássicos apresentados.

4.1 Modelos Clássicos

Nesta seção são apresentados os modelos clássicos para séries temporais, que inspiraram os modelos dinâmicos propostos neste trabalho. Os modelos da família ARMA, os modelos ARMA com erros do tipo ARCH (ARMA-ARCH) e suas generalizações, são definidos a partir de uma estrutura que baseia-se em hipóteses sobre a distribuição dos resíduos do modelo. Resultados envolvendo a distribuição condicional podem ser derivadas a partir dessa definição. Já para os modelos do tipo GARMA, o foco está em descrever o comportamento condicional do processo. Na maioria dos casos, resultados relacionados à distribuição incondicional são desconhecidos ou difíceis de obter sem a inclusão de hipóteses adicionais.

4.1.1 Modelos da Família ARMA

O processo $\{Y_t\}_{t \in \mathbb{Z}}$ segue um modelo ARMA(p, q) se ele é a solução estacionária de

$$\phi(L)(Y_t - \mu) = \theta(L)\varepsilon_t, \quad t \in \mathbb{Z}, \quad (4.1)$$

onde $\mu = \mathbb{E}(Y_t)$, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco, com média zero,

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p \quad \text{e} \quad \theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q. \quad (4.2)$$

Observe que a definição acima não menciona nada sobre a estrutura condicional de $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Sendo assim, é comum assumir que

$$\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0 \quad \text{e} \quad \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2. \quad (4.3)$$

Em outras palavras, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco, com média zero, tal que ε_{t+h} é não correlacionado com \mathcal{F}_t , para todo $h > 0$. É importante observar que essa hipótese só faz sentido quando $\{Y_t\}_{t \in \mathbb{Z}}$ é causal.

Os modelos $AR(p)$ e $MA(q)$ são casos particulares do modelo $ARMA(p, q)$, obtidos quando $p = 0$ e $q = 0$, respectivamente. As condições para a existência de uma solução estacionária (única) para as equações (4.1) e de causalidade são as mesmas de um processo $AR(p)$, enquanto que as condições de invertibilidade são as mesmas de um processo $MA(q)$. Para mais detalhes sobre esses resultados veja, por exemplo, [Brockwell and Davis \(1991\)](#); [van der Vaart \(2010\)](#).

Observe agora que, definindo-se

$$\mu_t := \phi(1)\mu + \sum_{k=1}^p \phi_k Y_{t-k} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (4.4)$$

a expressão (4.1) pode ser reescrita como

$$Y_t = \mu_t + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (4.5)$$

onde μ_t é \mathcal{F}_{t-1} mensurável. Além disso, sob as hipóteses em (4.3),

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \mu_t \quad \text{e} \quad \text{Var}(Y_t | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2, \quad t \in \mathbb{Z}. \quad (4.6)$$

Alternativamente, assumindo-se que $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é uma sequência de variáveis aleatórias i.i.d., com média zero e distribuição comum F_ε , obtém-se (4.6) e também

$$F_Y(y | \mathcal{F}_{t-1}) := P(Y_t \leq y | \mathcal{F}_{t-1}) = P(\varepsilon_t \leq y - \mu_t | \mathcal{F}_{t-1}) = F_\varepsilon(y - \mu_t), \quad t \in \mathbb{Z}.$$

O ponto fraco dos modelos aqui descritos é que a variância condicional não varia com o tempo, tais como log-retornos, visto que uma das características mais marcantes desse tipo de série temporal é a presença de volatilidade.

Vale ressaltar ainda que os modelos ARMA sazonais (SARMA) podem ser reescritos como um caso particular dos modelos ARMA, utilizando-se defasagens específicas. Portanto, os modelos ARMA são capazes de descrever um tipo particular de sazonalidade estocástica. Uma generalização dos modelos ARMA, capaz de incorporar tendências e sazonalidades determinísticas são os modelos ARMAX. Um processo $\{Y_t\}_{t \in \mathbb{Z}}$ segue um modelo ARMAX(p, q) se ele pode ser escrito como

$$\phi(L)(Y_t - \mu) = \mathbf{X}'_t \boldsymbol{\beta} + \theta(L)\varepsilon_t, \quad t \in \mathbb{Z},$$

onde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_s)' \in \mathbb{R}^p$, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ é um conjunto de variáveis exógenas s -dimensionais, $\phi(\cdot)$ e $\theta(\cdot)$ são dados em (4.2). Frequentemente, transformações de $\{Y_t\}_{t \in \mathbb{Z}}$ são utilizadas em combinação com os modelos ARMAX para modelagem de dados limitados. Embora essa abordagem seja amplamente utilizadas, os modelos resultantes são de difícil interpretação.

4.1.2 Modelos da Família ARMA com Erros do Tipo ARCH

Assuma que

$$\begin{aligned} Y_t &= \mu_t + \varepsilon_t, \\ \varepsilon_t &= \sigma_t Z_t, \quad t \in \mathbb{Z}, \end{aligned} \quad (4.7)$$

onde μ_t e σ_t são \mathcal{F}_{t-1} -mensuráveis e $\{Z_t\}_{t \in \mathbb{Z}}$ é um processo i.i.d ou então, um ruído branco condicionalmente independente do passado, com $\mathbb{E}(Z_t) = 0$ e $\text{Var}(Z_t) = 1$. Note que, nesse contexto,

$$\mathbb{E}(Z_t | \mathcal{F}_{t-1}) = \mathbb{E}(Z_t) = 0 \quad \text{e} \quad \mathbb{E}(Z_t^2 | \mathcal{F}_{t-1}) = \mathbb{E}(Z_t^2) = 1,$$

logo,

$$\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) = \mathbb{E}(\sigma_t Z_t | \mathcal{F}_{t-1}) = \sigma_t \mathbb{E}(Z_t | \mathcal{F}_{t-1}) = \sigma_t \times 0 = 0, \quad t \in \mathbb{Z}, \quad (4.8)$$

o que implica que $\{\varepsilon_t, \mathcal{F}_{t-1}\}_{t \in \mathbb{Z}}$ é um *martingale difference*, sempre que $\mathbb{E}(|\sigma_t|) < \infty$. Ressalta-se que a condição $\mathbb{E}(|\sigma_t|) < \infty$ garante que $\mathbb{E}(\varepsilon_t) = \mathbb{E}(|\sigma_t Z_t|) < \infty$, o que é uma condição necessária na definição de *martingale difference*. Além disso, observe que

$$\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \mathbb{E}(\sigma_t^2 Z_t^2 | \mathcal{F}_{t-1}) - [\mathbb{E}(\sigma_t Z_t | \mathcal{F}_{t-1})]^2 = \sigma_t^2 \mathbb{E}(Z_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 \times 1 = \sigma_t^2, \quad t \in \mathbb{Z} \quad (4.9)$$

e, se $\mathbb{E}(\sigma_t^2) < \infty$ então, para $h \geq 0$

$$\begin{aligned} \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) &= \mathbb{E}(\varepsilon_t \varepsilon_{t+h}) - \mathbb{E}(\varepsilon_t) \mathbb{E}(\varepsilon_{t+h}) = \mathbb{E}(\mathbb{E}(\varepsilon_t \varepsilon_{t+h} | \mathcal{F}_{t+h-1})) - \mathbb{E}(\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1})) \mathbb{E}(\varepsilon_{t+h}) \\ &= \mathbb{E}(\varepsilon_t \mathbb{E}(\varepsilon_{t+h} | \mathcal{F}_{t+h-1})) - \mathbb{E}(\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1})) \mathbb{E}(\varepsilon_{t+h}) = \mathbb{E}(\varepsilon_t \times 0) - 0 \times \mathbb{E}(\varepsilon_{t+h}) = 0 \end{aligned}$$

e, de forma análoga, condicionando-se em \mathcal{F}_{t-1} , $\text{Cov}(\varepsilon_t, \varepsilon_{t-h}) = 0$. Segue que

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+h}) = 0, \quad \text{para todo } t, h \in \mathbb{Z},$$

ou seja, se $\mathbb{E}(\sigma_t^2) < \infty$, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco.

Da definição (4.7) e de (4.8) e (4.9) segue que

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \mu_t \quad \text{e} \quad \text{Var}(Y_t | \mathcal{F}_{t-1}) = \sigma_t^2, \quad t \in \mathbb{Z}.$$

Em particular, quando $\mu_t = 0$ o processo $\{Y_t\}_{t \in \mathbb{Z}}$ coincide com $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. É fácil ver ainda que

$$F_Y(y | \mathcal{F}_{t-1}) := P(Y_t \leq y | \mathcal{F}_{t-1}) = P\left(Z_t \leq \frac{y - \mu_t}{\sigma_t} \middle| \mathcal{F}_{t-1}\right) = F_Z\left(\frac{y - \mu_t}{\sigma_t}\right), \quad t \in \mathbb{Z}.$$

Um exemplo clássico de modelo com essa estrutura é o ARCH(p), onde

$$Y_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{k=1}^p \alpha_k Y_{t-k}^2, \quad t \in \mathbb{Z}.$$

Nesse caso, $\mu_t = 0$, para todo $t \in \mathbb{Z}$. Por sua vez, o modelo ARMA(p_1, q_1)-GARCH(p_2, q_2), para um processo com média (incondicional) $\mu = 0$, é definido por

$$\begin{aligned} Y_t &= \sum_{k=1}^{p_1} \phi_k Y_{t-k} + \sum_{j=1}^{q_1} \theta_j \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{k=1}^{p_2} \alpha_k Y_{t-k}^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}. \end{aligned}$$

Esse modelo engloba a estrutura ARMA para a média condicional, análogo a (4.4), e a estrutura GARCH para a variância condicional.

Diversas estruturas para σ_t foram propostas na literatura para garantir que essa variável seja positiva com probabilidade 1, sem a necessidade de impor restrições complexas nos valores dos parâmetros, como ocorre com os modelos ARCH e GARCH. Uma lista vasta de modelos da família ARCH pode ser encontrada em [Bollerslev \(2010\)](#). O ponto fraco dos modelos descritos nessa seção é a estrutura ARMA considerada para a média condicional μ_t , que pode não ser apropriada quando os dados são limitados, a menos que sejam impostas restrições severas nos parâmetros dos modelos.

4.1.3 Modelos Condicionais do Tipo GARMA

Os modelos descritos nas seções anteriores, partem de hipóteses sobre a distribuição do ruído $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Outra maneira de modelar os dados é partindo da hipótese que $Y_t | \mathcal{F}_{t-1}$ possui uma dada distribuição, com densidade condicional $f(\cdot | \mu_t, \varphi)$, parametrizada através da média condicional μ_t , isto é,

$$Y_t | \mathcal{F}_{t-1} \sim f(\cdot | \mu_t, \varphi), \quad \text{onde } \mu_t = \mathbb{E}(Y_t | \mathcal{F}_{t-1}), \quad \varphi \in \mathbb{R}.$$

Embora as duas formulações partem de hipóteses distintas, muitas propriedades dos modelos clássicos definidos anteriormente se refletem nos modelos condicionais.

No contexto condicional é sempre válido que Y_t pode ser escrito como em (4.5), basta observar que

$$Y_t = \mathbb{E}(Y_t | \mathcal{F}_{t-1}) + [Y_t - \mathbb{E}(Y_t | \mathcal{F}_{t-1})] = \mu_t + e_t,$$

onde $\mu_t := \mathbb{E}(Y_t | \mathcal{F}_{t-1})$ e $e_t := Y_t - \mathbb{E}(Y_t | \mathcal{F}_{t-1})$. Também é válido que,

- se $\mathbb{E}(|Y_t|) < \infty$, $\{e_t, \mathcal{F}_{t-1}\}_{t \in \mathbb{Z}}$ é um *martingale difference*;
- se $\mathbb{E}(Y_t^2) < \infty$ e não depende de t , $\{e_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco;
- para todo $t \in \mathbb{Z}$,

$$P(e_t \leq x | \mathcal{F}_{t-1}) = P(Y_t \leq x + \mu_t | \mathcal{F}_{t-1}),$$

ou seja, a distribuição condicional de e_t é uma versão transladada da distribuição condicional de Y_t .

A demonstração desses resultados segue os mesmos passos já apresentados nas seções anteriores.

No contexto condicional é possível modelar tanto processos que tomam valores na reta (veja, por exemplo, [Benaduce, 2021](#); [Maior and Cysneiros, 2018](#)), quanto processos que assumem apenas valores em um intervalo limitado. Nas últimas décadas, vários modelos para séries temporais duplamente limitadas foram propostos na literatura. Em geral, esses modelos são baseados na abordagem que hoje é conhecida como modelos de média móvel autorregressiva generalizada (GARMA). A abordagem GARMA incorpora uma estrutura dependente do tempo na estrutura de modelos lineares generalizados (GLM). Embora essa ideia tenha surgido pro volta dos anos 70, o nome GARMA foi solidificado em [Benjamin et al. \(2003\)](#).

Dentre os trabalhos que consideram modelos condicionais para dados limitados, podemos citar [Rocha and Cribari-Neto \(2009\)](#) que introduz o modelo β ARMA, constituído sobre a classe de regressão beta anteriormente descrita por [Ferrari and Cribari-Neto \(2004\)](#), propondo um modelo dinâmico para séries temporais limitadas no intervalo $(0, 1)$. Esse modelo que segue uma estrutura GARMA onde

o componente aleatório é especificado por uma distribuição beta. Assim como em [Benjamin et al. \(2003\)](#), seu preditor linear modela a média condicional através de uma função de ligação.

Alternativamente aos modelos β ARMA, [Bayer et al. \(2017\)](#) apresenta o modelo KARMA (Kumaraswamy autoregressivo de médias móveis), introduzindo uma classe de modelos dinâmicos para séries temporais duplamente limitadas com distribuição condicional Kumaraswamy. Nesse modelo, a mediana condicional é modelada por uma estrutura dinâmica contendo termos ARMA. Já o modelo β ARFIMA, em [Pumi et al. \(2019b\)](#), estabelece uma generalização do modelo β ARMA considerando uma estrutura ARFIMA(p, d, q), capaz de acomodar uma estrutura de longa dependência no componente sistemático. No que segue, tratamos dos modelos condicionais para dados positivos, que é o foco deste trabalho.

4.2 Modelos Condicionais Para dados Positivos

A modelagem de séries temporais positivas é um assunto pouco abordado na literatura. Entretanto, esse é um problema com grande potencial de aplicação prática. Por exemplo, no contexto de séries temporais financeiras pode ser interessante modelar o quadrado dos log-retornos ao invés do log-retornos $\{r_t\}_{t \in \mathbb{Z}}$ em si. Uma possível abordagem é assumir que $\mathbb{E}(r_t | \mathcal{F}_{t-1}) = 0$, $Y_t := r_t^2$ e $Y_t | \mathcal{F}_{t-1}$ possui uma dada distribuição, parametrizada através da média. Note que isso implica que $\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \mathbb{E}(r_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2$, que representa a volatilidade condicional.

[Benjamin et al. \(1998\)](#) mostra que essa abordagem pode ser utilizada no contexto de modelos GARCH. Mais especificamente, o autor observa que, se $\{r_t\}_{t \in \mathbb{Z}}$ segue um modelo GARCH(p, q) tal que $r_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$ e $Y_t = r_t^2$, então $Y_t | \mathcal{F}_{t-1} \sim h_t \chi_t^2$, que é equivalente a dizer que $Y_t | \mathcal{F}_{t-1} \sim \text{Gamma}(\mu_t, \varphi)$, com $\mu_t = \sigma_t^2$ e $\varphi = 2$. Embora este artigo seja voltado para séries temporais, o autor considera apenas as distribuições da família exponencial. Mesmo não envolvendo séries temporais, [Bourguignon and Gallardo \(2020\)](#) e [Bourguignon et al. \(2021\)](#) propõe modelos de regressão, considerando as distribuições Gama Inversa e Beta-Prime, respectivamente, onde a média μ_i e o parâmetro de precisão φ_i satisfazem as seguintes relações funcionais,

$$g_1(\mu_i) = \mathbf{X}_i' \boldsymbol{\beta} \quad e \quad g_2(\varphi_i) = \mathbf{Z}_i' \boldsymbol{\nu}, \quad i = 1, \dots, n,$$

onde ambos, a média e o parâmetro de precisão φ_i são parametrizados em termos de covariáveis exógenas $\{\mathbf{X}_i\}_{i=1}^n$ e $\{\mathbf{Z}_i\}_{i=1}^n$.

Neste trabalho e no artigo em anexo, a proposta é generalizar/estender essas duas abordagens e modelar a média através da seguinte equação, mantendo o parâmetro de precisão fixo,

$$\eta_t := g_1(\mu_t) = \alpha + \mathbf{X}_t' \boldsymbol{\beta} + \sum_{k=1}^p \phi_k [g_2(Y_{t-k}) - I_X \mathbf{X}_{t-k}' \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j e_{t-j}, \quad (4.10)$$

$$e_t := Y_t - \mu_t,$$

onde $g_1 : (a^*, b^*) \rightarrow \mathbb{R}$ é uma função de ligação bijetora duas vezes diferenciável, (a^*, b^*) é o intervalo onde μ_t toma valores, $g_2 : (0, \infty) \rightarrow \mathbb{R}$ é uma função de ligação (não necessariamente diferenciável), $\alpha \in \mathbb{R}$ é o intercepto, $\boldsymbol{\beta} := (\beta_1, \dots, \beta_s)'$ é um vetor de parâmetros s -dimensional associado às covariáveis, I_X é uma função indicadora que assume valor 1 no caso em que os regressores devem ser

incluídos na parte autoregressiva do modelo e 0, caso contrário, $\phi := (\phi_1, \dots, \phi_p)'$ e $\theta := (\theta_1, \dots, \theta_q)'$ são os vetores de parâmetros de dimensões p e q , respectivamente. Nessa proposta, está sendo considerado para a estrutura do μ_t a possibilidade de ter as funções de ligação $g_1 \neq g_2$ e incluir ou não as covariáveis regressoras na estrutura AR, enquanto que na literatura existente, $g_1 = g_2$ e $I_X = 1$ sempre que $p > 0$. Além disso, neste trabalho o modelo não é restrito às distribuições da família exponencial.

Alguns exemplos de covariáveis determinísticas que podem aparecer na prática são $X_t = t$, nesse caso, $s = 1$ e X_t descreve uma tendência linear ou, de forma mais geral, $\mathbf{X}_t = (X_{1,t}, \dots, X_{s,t})'$ com $X_{k,t} = t^k$, para $k = 1, \dots, s$ e $s > 1$, caso em que as covariáveis descrevem uma tendência polinomial. Outras possibilidades são a utilização das funções $\sin(\cdot)$ e $\cos(\cdot)$ ou combinações das mesmas, por exemplo, tomando-se $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$, com $X_{1,t} = \cos(\omega_1 t)$ e $X_{2,t} = \sin(\omega_2 t)$. Nesse caso, $s = 2$ e as covariáveis descrevem um comportamento sazonal. Nas Figuras 4.1 e 4.2, tem-se alguns exemplos de séries temporais positivas que foram geradas utilizando-se essas covariáveis.

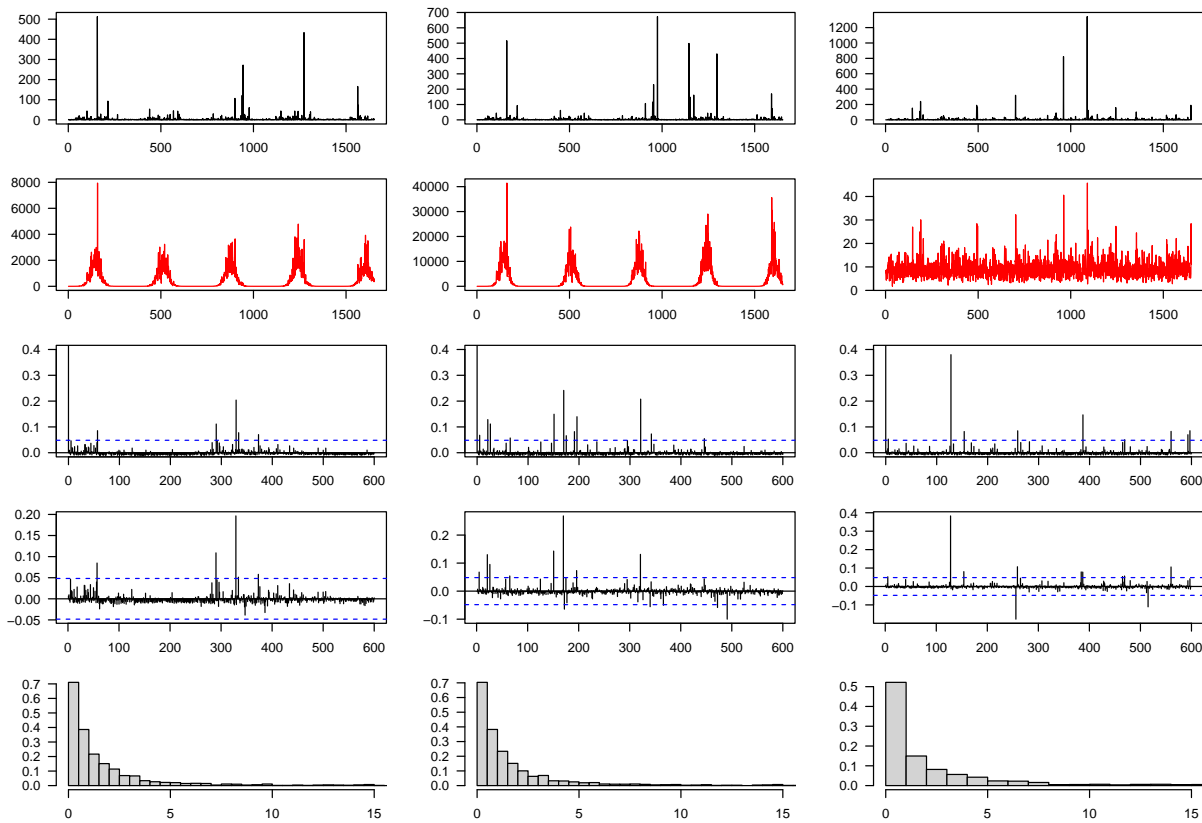


Figura 4.1: Série temporal $\{Y_t\}_{t=1}^n$ (primeira linha), com distribuição condicional F. De cima para baixo: a média condicional $\{\mu_t\}_{t=1}^n$, as funções de autocorrelação e autocorrelação parcial e o histograma. Da esquerda para a direita, considerando $I_X = 1$, $I_X = 0$ e sem a inclusão de covariáveis.

Nas Figuras 4.1 e 4.2, são apresentadas séries temporais positivas $\{Y_t\}_{t=1}^n$ (primeira linha), com tamanho $n = 1.652$, gerados a partir da expressão (4.10), com distribuição condicional F e Gaussiana Inversa, respectivamente. São apresentados ainda, de cima para baixo, a partir da segunda linha, o gráfico da média condicional $\{\mu_t\}_{t=1}^n$, da função de autocorrelação amostral (o intervalo no eixo-y foi reduzido para melhor visualização), da função de autocorrelação parcial amostral e do histograma (o intervalo no eixo-x foi reduzido para melhor visualização). Em ambas as figuras utilizou-se $\mathbf{X}_t =$

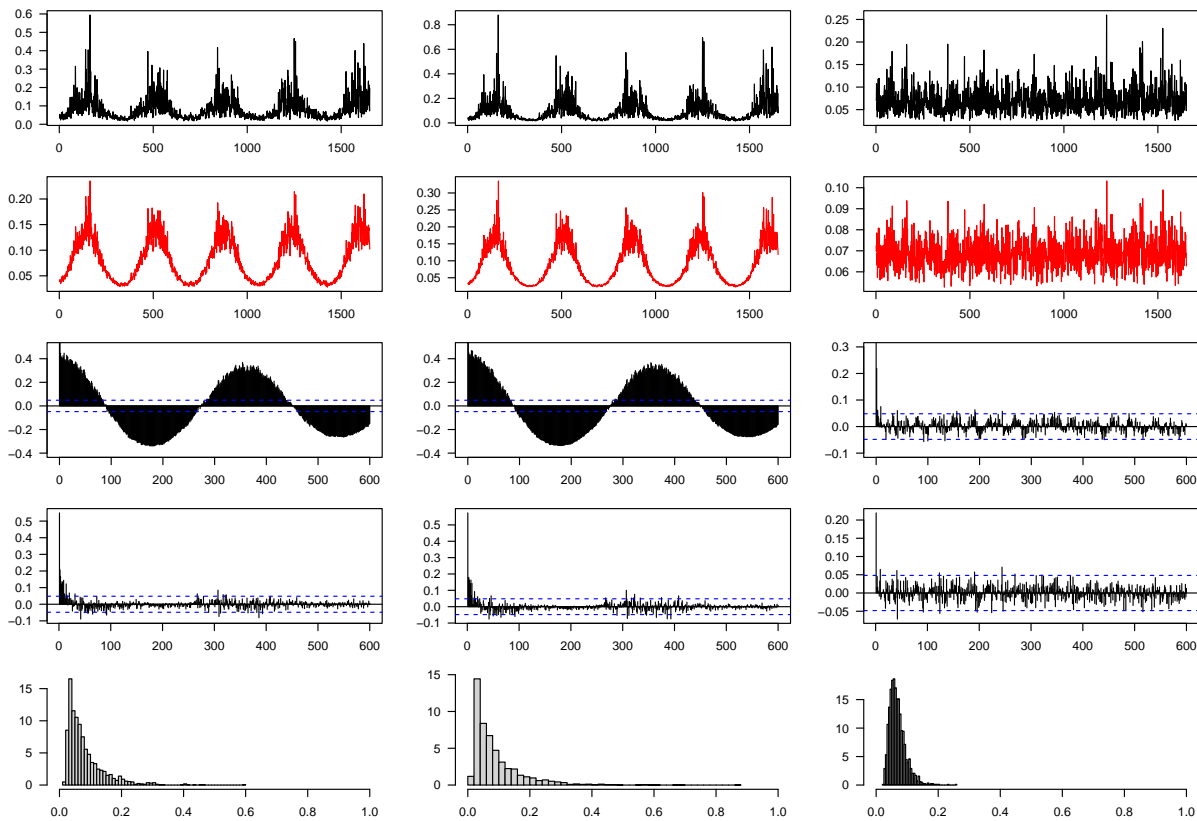


Figura 4.2: Série temporal $\{Y_t\}_{t=1}^n$ (primeira linha), com distribuição condicional Gaussiana Inversa. De cima para baixo: a média condicional $\{\mu_t\}_{t=1}^n$, as funções de autocorrelação e autocorrelação parcial e o histograma. Da esquerda para a direita, considerando $I_X = 1$, $I_X = 0$ e sem a inclusão de covariáveis.

$(\cos(\omega_1 t), \sin(\omega_1 t))'$ com $\omega_1 = \omega_2 = 2\pi/365$, exceto para a série temporal na terceira coluna, que foi gerada sem a inclusão de covariáveis. Além disso, em ambas as figuras, a segunda série temporal foi gerada sem a presença do componente de regressão da parte AR, isto é, tomando-se $I_X = 0$. Para a Figura 4.1, considerou-se como funções de ligação $g_1(x) = \log(x+1)$ e $g_2(x) = \log(x)$ e como ordem do modelo, $p = 1$ e $q = 0$. Além disso, os valores dos parâmetros foram fixados como sendo $\varphi = 2$, $\alpha = 2$, $\beta = (-3.5, -6.5)'$ e $\phi_1 = 0.25$. Já para a Figura 4.2, $g_1(x) = g_2(x) = \log(x)$, $p = q = 1$, $\varphi = 2$, $\alpha = -2$, $\beta = (-0.35, -0.65)'$, $\phi_1 = 0.25$ e $\theta_1 = 0.35$.

A série temporal $\{Y_t\}_{t=1}^n$, apresentada na primeira linha da Figura 4.1, assemelha-se a uma série de retornos financeiros ao quadrado. Devido à grande variação dos valores da série, não é possível visualizar de forma clara a presença da componente $\{X_t\}_{t=1}^n$. A segunda linha da figura representa a média μ_t . Para as duas primeiras séries, pode-se observar claramente a presença do comportamento sazonal, o que não ocorre com a terceira série temporal, como esperado. Observa-se ainda que os valores de $\{\mu_t\}_{t=1}^n$ são muito maiores do que os valores de $\{Y_t\}_{t=1}^n$. Isso pode ser explicado pelo fato de que a variância condicional de Y_t existe apenas quando $1 < \mu_t < 2$ e valores muito grandes de μ_t tendem a gerar valores de Y_t próximos de zero. A terceira e quarta linha da figura representam a AFC e PACF, respectivamente. Diferente do que ocorre nos modelos ARMA, o comportamento cíclico da média e a dependência devido à componente autoregressiva não se refletem nesses gráficos. Já na última linha dessa figura, tem-se o histograma que é uma estimativa da densidade incondicional. Embora não seja esperado, para esta simulação, o histograma reflete as características da distribuição

condicional apresentada na Figura 3.2.

A primeira linha da Figura 4.2 apresenta a série temporal $\{Y_t\}_{t=1}^n$ enquanto que a segunda linha representa a média $\{\mu_t\}_{t=1}^n$. Para as duas primeiras colunas, pode-se observar claramente a presença da componente sazonal tanto na série temporal, na média condicional ou pela ACF. Além disso, o comportamento da ACF indica que a componente sazonal é do tipo determinística. O comportamento observado é muito semelhante ao que ocorre nos modelos ARMA. A quarta linha dessa figura representam a PACF. Observa-se que, para as duas primeiras colunas, a PACF parece decair rapidamente, como ocorre nos modelos ARMA. Para a terceira coluna, a PACF no *lag* 1 se destaca enquanto que os demais valores estão muito próximos de zero. Já na última linha dessa figura, tem-se o histograma que é uma estimativa da densidade incondicional. Assim como no exemplo anterior, embora não seja esperado, para esta simulação, o histograma também reflete as características da distribuição condicional apresentada na Figura 3.4.

4.2.1 PMLE

Denote por $\gamma := (\alpha, \beta', \phi', \theta', \varphi)'$ o vetor $(p + q + s + 2)$ -dimensional de parâmetros do modelo e por $\Omega \subseteq \mathbb{R}^{p+q+s+2}$ o espaço paramétrico. A estimação dos parâmetro pode ser feita através do método de máxima verossimilhança parcial. Nessa abordagem, o estimador de máxima verossimilhança parcial (PMLE) é obtido ao maximizar o logaritmo da função de verossimilhança parcial dada por

$$\ell(\gamma) = \sum_{t=1}^n \ell_t(\gamma), \quad \ell_t(\gamma) := \log(f(Y_t | \mu_t, \varphi)),$$

de forma que o estimador de máxima verossimilhança parcial de γ é dado por

$$\hat{\gamma} = \underset{\gamma \in \Omega}{\operatorname{argmax}}(\ell(\gamma)). \quad (4.11)$$

As propriedades assintóticas do PMLE, intervalos de confiança para os parâmetros e testes de hipóteses utilizando a distribuição assintótica são discutidas no artigo em anexo.

Em muitos casos, γ não pode ser obtido analiticamente, sendo necessário utilizar métodos de otimização numérica para resolver (4.11). Dentre os métodos mais utilizados na literatura podemos citar os algoritmos de Nelder-Mead e o L-BFGS-B. Algumas informações sobre esses métodos são apresentadas a seguir.

Nelder-Mead

A técnica proposta por Nelder and Mead (1965) é uma modificação do método criado por Spendley et al. (1962). O algoritmo Nelder-Mead consiste em um método numérico amplamente utilizado para encontrar o mínimo ou máximo de uma função objetivo em um espaço multidimensional. É aplicado frequentemente a problemas de otimização não-linear para os quais as derivadas não podem ser encontradas. A eficiência do método baseia-se na capacidade de se adaptar às curvaturas das funções, mas que por vezes, pode convergir para pontos não estacionários, principalmente em problemas de alta dimensionalidade. A ausência da teoria da convergência é frequentemente refletida na prática como

uma quebra numérica do algoritmo, mesmo para funções suaves e bem-comportadas. Além disso, este método não permite que a otimização dos parâmetros seja realizada com restrição de intervalo de valores, e possui baixa velocidade de convergência.

L-BFGS-B (Limited-memory Broyden–Fletcher–Goldfarb–Shanno Bound-constrained)

O Algoritmo L-BFGS-B de [Byrd et al. \(1995\)](#) é uma extensão do algoritmo L-BFGS de [Liu and Nocedal \(1989\)](#) para o tratamento dos limites inferiores e superiores dos parâmetros do problema de otimização. Ambos os algoritmos são generalizações do algoritmo BFGS de [Fletcher \(1970\)](#). O método BFGS é um método quase-Newton de busca local que usa valores da função e gradientes para criar uma imagem da superfície a ser otimizada. O L-BFGS-B é um algoritmo de memória limitada, capaz de resolver grandes problemas de otimização não lineares onde a matriz Hessiana original é inviável de ser obtida e restrições de valores devem ser satisfeitas. Diferentemente do método de Nelder-Mead, o L-BFGS-B depende do cálculo das derivadas da função objetivo.

4.2.2 *Análise Residual*

Ferramentas úteis na análise residual são a análise da ACF/PACF e o teste de Ljung-Box, conforme descritos no artigo em anexo. Entretanto, para os modelos propostos neste trabalho, os testes mais apropriados são aqueles que verificam a hipótese de que os resíduos são um *martingale difference*. Isso porque na definição dos modelos condicionais não são feitas hipóteses sobre a existência de momentos de ordem maior do que 1. No que segue é apresentada uma descrição breve sobre o teste utilizado neste trabalho.

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ uma série temporal estacionária tomando valores na reta. Sob a hipótese de *martingale difference* (MDH - *martingale difference hypothesis*) deve-se ter ([Charles et al., 2011](#))

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \mu_t, \quad \text{ou, equivalente,} \quad \mathbb{E}((Y_t - \mu_t)\omega(\mathcal{F}_{t-1})) = 0,$$

onde \mathcal{F}_t é a σ -álgebra gerada pela informação disponível até o tempo t e $\omega(\cdot)$ é uma função de ponderação. Aqui, $\omega(\mathcal{F}_{t-1})$ representa uma transformação linear ou não linear no passado. Dependendo da escolha desta função de ponderação, os testes são classificados como sendo baseados em medidas lineares ou não lineares de dependência.

Segundo [Charles et al. \(2011\)](#), geralmente constata-se que o teste *automatic variance ratio* (AVR) de [Kim \(2009\)](#), que é uma extensão do trabalho anterior de [Choi \(1999\)](#), apresenta grande eficiência para uma gama de modelos lineares e não lineares, sem distorção de tamanho. Além disso, embora o teste AVR não seja projetado para detectar dependência não linear, ele se mostra robusto para uma gama de modelos não lineares. Dado que, muitas vezes há incerteza na prática se a dependência é linear ou não linear, o uso do teste AVR junto com os testes DL (*Dominguez and Lobato test*) e GS (*generalized spectral test*) é fortemente recomendado e ainda sugerido que, na implementação dos testes para amostras finitas, seja utilizado o método conhecido como *wild bootstrap* (veja [Charles et al., 2011](#), e referências ali contidas).

A estatística de teste do AVR é dada por

$$\hat{V}(k) = 1 + 2 \sum_{i=1}^{k-1} \left(1 - \frac{i}{k}\right) \hat{\rho}(i),$$

onde $\hat{\rho}(\cdot)$ denota a função de autocorrelação amostral. Uma descrição do método do teste AVR é encontrada na seção 2.1 de [Charles et al. \(2011\)](#). No Software R, o teste AVR é implementado na função `AutoBoot.test` do pacote `vrtest` ([Kim, 2014](#)), sendo que no argumento `wild` da função, pode-se escolher entre três opções de distribuições, Normal, Rademacher ou Mammen.

4.2.3 Previsão

Após obter $\hat{\gamma}$, esses valores podem ser substituídos em (4.10) para obter-se as previsões dentro da amostra (*in-sample*), denotadas por $\{\hat{\mu}_t\}_{t=1}^n$, e também as previsões h passos à frente (*out-of-sample*), denotado por $\hat{\mu}_{n+h} = \hat{\mu}_n(h)$, da média condicional do modelo. Deve-se supor ainda que as covariáveis \mathbf{X}_t , para $t = n + 1, \dots, n + h$, existam ou possam ser obtidas.

A obtenção das previsões dá-se da seguinte forma. Iniciando em $t = 1$, definimos recursivamente,

$$\hat{\eta}_t = \hat{\alpha} + \hat{\mathbf{X}}_t' \hat{\boldsymbol{\beta}} + \sum_{i=1}^p \hat{\phi}_i [g_{1,2}(\hat{Y}_{t-i}) - I_X \hat{\mathbf{X}}_{t-i}' \hat{\boldsymbol{\beta}}] + \sum_{k=1}^q \hat{\theta}_k \hat{e}_{t-k}, \quad t \geq 1, \quad (4.12)$$

com

$$\hat{Y}_t = \begin{cases} 0, & p = 0, t < 1, \\ \frac{1}{p} \sum_{i=1}^p Y_i, & p > 0, t < 1, \\ Y_t, & 1 \leq t \leq n, \\ \hat{\mu}_t, & t > n, \end{cases} \quad \hat{\mathbf{X}}_t' = \begin{cases} 0, & p I_X = 0, t < 1, \\ \frac{1}{p} \sum_{i=1}^p \mathbf{X}_i', & p I_X > 0, t < 1, \\ \mathbf{X}_t, & t \geq 1, \end{cases}$$

$$\hat{\mu}_t = g_1^{-1}(\hat{\eta}_t), \quad t \geq 1, \quad \text{e} \quad \hat{e}_t = \begin{cases} \hat{Y}_t - \hat{\mu}_t, & 1 \leq t \leq n, \\ 0, & \text{caso contrário.} \end{cases}$$

Para acessar a qualidade de ajuste e previsão dos modelos consideram-se as medidas denominadas erro absoluto médio (MAE), erro quadrático médio (MSE) e erro percentual absoluto médio (MAPE), respectivamente definidas por

$$\text{MAE} = \frac{1}{n_p} \sum_{t \in T} |Y_t - \hat{Y}_t|, \quad \text{MSE} = \frac{1}{n_p} \sum_{t \in T} (Y_t - \hat{Y}_t)^2 \quad \text{e} \quad \text{MAPE} = \frac{1}{n_p} \sum_{t \in T} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|, \quad (4.13)$$

onde n_p é o número de previsões realizadas e T é o conjunto de índices para o qual foram feitas as n_p previsões. Tais medidas podem ser utilizadas tanto para a avaliar as previsões dentro quanto fora da amostra. Ressaltamos ainda que, substituindo-se Y_t e \hat{Y}_t por γ_i e $\hat{\gamma}_i$, respectivamente, e fazendo as devidas alterações no conjunto de índices, as mesmas medidas podem ser utilizadas para avaliar o desempenho do PMLE.

CAPÍTULO 5

SIMULAÇÕES DE MONTE CARLO

Neste capítulo é apresentado um estudo de simulação de Monte Carlo para investigar o comportamento dos modelos propostos neste trabalho em amostras finitas. São consideradas três distribuições, dentre aquelas estudadas no Capítulo 3: Beta-Prime, Gama e Log-Normal. A escolha dessas distribuições baseou-se em um estudo preliminar que indicou que essas três opções eram as que melhor se adaptavam aos dados escolhidos para a aplicação apresentada no Capítulo 6.

O estudo tem por objetivo comparar o desempenho do estimador de máxima verossimilhança parcial (PMLE) para a estimação dos parâmetros dos modelos obtidos a partir das três distribuições utilizadas. Pretende-se também investigar o desempenho do método no caso de má especificação da distribuição. Desta forma, é possível avaliar se os parâmetros ajustados quando consideram-se distribuições distintas da distribuição geradora dos dados são próximos daqueles obtidos quando a distribuição utilizada para obtenção do PMLE coincide com a distribuição que deu origem aos dados. Além da estimação dos parâmetros, o estudo investiga também o desempenho dos modelos em termos de previsão dentro e fora da amostra, tanto no contexto de especificação correta quanto de má especificação. As rotinas utilizadas nesta simulação e citadas neste capítulo fazem parte do pacote PTSR (Prass et al., 2022). Uma descrição detalhada dos argumentos a serem passados para as funções citadas a seguir pode ser encontrada no manual do pacote.

5.1 Processo Gerador dos Dados

Nesta seção são descritas as configurações e os passos adotados para a geração das séries temporais positivas, consideradas neste estudo. As séries temporais simuladas foram geradas com parâmetros semelhantes aos obtidos nos modelos ajustados na série temporal real analisada no Capítulo 6. Os critérios utilizados para a geração das séries temporais foram os seguintes:

- Cada série temporal foi gerada com tamanho $N = 1.752$ observações sendo que os $b = 100$ primeiros valores foram descartados (*burn-in*).
- As séries temporais $\{Y_t\}_{t=-b+1}^n$ foram geradas com a função `ptsr.sim` que implementa a expressão (4.10).
- Para a distribuição condicional $f(\cdot|\mu_t, \varphi)$, correspondente à $Y_t|\mathcal{F}_{t-1}$, foram consideradas as dis-

tribuições Beta-Prime, Gama e Log-Normal, de acordo com suas respectivas reparametrizações em (3.7), (3.22) e (3.42). Os geradores de números aleatórios correspondentes a tais reparametrizações estão implementadas nas funções `r.betap`, `r.gamma` e `r.logNorm`, respectivamente.

- Nos três cenários considerados definiu-se como covariáveis $\mathbf{X}_t = (X_{t,1}, X_{t,2})'$, onde $X_{t,1} = \cos(2\pi t/365)$ e $X_{t,2} = \sin(2\pi t/365)$, e utilizou-se $I_X = 1$. Tais covariáveis foram escolhidas de forma a imitar o comportamento cíclico observado nos dados reais.
- Os cenários considerados, em termos de ordem dos modelos (p e q), valores de parâmetros e funções de ligação g_1 e g_2 (*links*), são apresentados na Tabela 5.1, junto com as respectivas distribuições condicionais. Devido ao grande número de configurações possíveis, um estudo mais detalhado se mostrou inviável, sendo assim, optou-se por considerar alguns dos melhores cenários (em termos preditivos) obtidos na aplicação prática.
- Para cada cenário considerado, foram realizadas 1.000 replicações.

Tabela 5.1: Valores dos parâmetros dos modelos, as respectivas distribuições e funções de ligação utilizadas na geração dos dados.

Parâmetros	$Y_t \mathcal{F}_{t-1} \sim f(\cdot \mu_t, \varphi)$		
	Beta-Prime	Gama	Log-Normal
α	2.4496	0.4538	0.9381
β_1	-0.0354	-0.0682	-0.0726
β_2	-0.0601	-0.0900	-0.1132
ϕ_1	0.0274	1.1121	1.0057
ϕ_2	-	-0.4303	-0.4555
ϕ_3	-	0.3695	0.1397
ϕ_4	-	-0.2008	-
θ_1	0.0229	-0.0052	-
θ_2	-	-0.0062	-
θ_3	-	-0.0140	-
θ_4	-	-0.0040	-
θ_5	-	0.0045	-
φ	154.9584	74.0205	0.0882
g_1	log	log	log
g_2	<i>identidade</i>	log	log

5.2 Ajuste dos Modelos e Previsão

Nesta seção são apresentadas as configurações adotadas para a estimação dos parâmetros dos modelos e para as previsões dentro (*in-sample*) e fora da amostra (*out-of-sample*). Para cada um dos três modelos considerados foram adotados os seguintes critérios.

- Dadas as observações $\{(Y_t, \mathbf{X}_t)\}_{t=1}^{1.632}$, os primeiros 1.319 valores foram utilizados para a estimação dos parâmetros e os últimos 333 valores foram reservados para estimar os erros de

previsão fora da amostra.

- A estimação dos parâmetros foi realizada utilizando o método de máxima verossimilhança parcial, implementado na função `ptsr.fit`. Nessa função o PMLE é obtido através da expressão (4.11), via otimização numérica.
- Considerou-se a mesma ordem (p e q), funções de ligação e número de covariáveis do processo gerador dos dados em questão.
- O PMLE foi obtido considerando-se três distribuições, Beta-Prime, Gama e Log-Normal. Uma delas coincide com a do processo gerador dos dados e as outras duas correspondem à má especificações do modelo.
- A estimação dos parâmetros foi realizada utilizando uma combinação de dois métodos de otimização. Primeiramente utilizou-se Nelder-Mead para obter uma solução inicial. Tal solução foi então utilizada como valores iniciais para o método L-BFGS-B. Ressalta-se que o método Nelder-Mead não aceita restrições nos parâmetros, enquanto que no método L-BFGS-B, é possível criar intervalos nos quais os parâmetros devem ser procurados. O código foi adaptado de forma que, caso um dos algoritmos falhasse, apenas o outro seria utilizado. Além disso, caso a falha seja no Nelder-Mead, o L-BFGS-B é inicializado com os valores padrões utilizados no Nelder-Mead.
- Para aplicação do Nelder-Mead, todos os parâmetros foram inicializados com valor igual a zero, com exceção de φ cuja inicialização depende da distribuição adotada. Para este estudo considerou-se como valores iniciais, $\varphi \in \{60, 60, 0.003\}$ para as distribuições Beta-Prime, Gama e Log-Normal, respectivamente.
- Para o método L-BFGS-B, utilizou-se as derivadas numéricas da função de verossimilhança.
- Foi estabelecido que o número máximo de iterações (comum aos dois métodos) fosse de 5.000. Além disso, para o método L-BFGS-B, fixou-se $\varphi \in (0, \infty)$ enquanto que para os demais parâmetros considerou-se o intervalo $(-\infty, \infty)$.
- Para comparar a eficiência dos modelos, foram obtidas as previsões dentro da amostra (*in-sample*) e fora da amostra (*out-of-sample*). As previsões foram obtidas utilizando a função `predict` (função genérica do R que internamente chama a função `predict.ptsr`) que implementa a equação 4.12.
- Para a previsão fora da amostra, dois cenários foram considerados, h previsões 1 passo à frente, para $h = 333$, e uma previsão a h passos à frente, para $h = 1, \dots, 333$.
- Em todos os casos, a medida utilizada para analisar o desempenho preditivo dos modelos foi o MAPE, definido em (4.13).

5.3 Apresentação dos Resultados

Nesta seção são apresentados resultados relacionados a estimação dos parâmetros dos modelos obtidos a partir das três distribuições utilizadas e avaliado o desempenho do método nos casos de má especificação da distribuição. Além disso, também são apresentados os resultados do desempenho dos modelos nas previsões dentro e fora da amostra. Lembramos que, para o ajuste dos modelos, foram

utilizados os $n - 333$ valores da série temporal real e os 333 últimos valores foram reservados para a previsão sendo, h previsões 1 passo à frente, para $h = 333$, e uma previsão a h passos à frente, para $h = 1, \dots, 333$.

Na Tabela 5.1 observa-se que, como esperado, para o parâmetro φ , as estimativas obtidas com a distribuição Log-Normal diferem significativamente das estimativas obtidas com as distribuições Beta-Prime e Gama. As duas últimas por sua vez apresentam estimativas próximas umas das outras. Tendo em vista que a variância condicional da variável aleatória Y_t depende diretamente do parâmetro φ e essa relação não é a mesma para as três distribuições utilizadas, não era esperado que os valores estimados do parâmetro φ , no caso de má especificação, ficassem próximos dos valores utilizados nas simulações. No caso em que os dados foram gerados a partir da Gama e estimados utilizando-se a Beta-Prime, a semelhança das estimativas poderia ser explicada pelo fato de $\text{Var}(Y_t | \mathcal{F}_{t-1})$ para tais distribuições ser dada, respectivamente, por $\mu_t(\mu_t - 1)/(\varphi - 1)$ e μ_t^2/φ , cujo comportamento assintótico quando $\varphi \rightarrow \infty$ é o mesmo. Por outro lado, quando os dados foram gerados utilizando-se a distribuição Beta-Prime, as estimativas obtidas com a Gama (média igual a 68.1678) ficaram bem abaixo dos valores estimados com a Beta-Prime (média igual a 144.6687), que por sua vez ficaram próximos do valor verdadeiro (154.9584).

Já na Tabela 5.2 são apresentados os resultados das simulações. Nessa tabela, as colunas intituladas “Simulado” indicam o verdadeiro valor dos parâmetros, conforme indicado na Tabela 5.1, enquanto que nas colunas “BP”, “G” e “LN”, constam os valores estimados, obtidos considerando-se, respectivamente, as distribuições Beta-Prime, Gama e Log-Normal. A verdadeira distribuição condicional é indicada na linha “Gerados”. Os valores do MAPE para a previsão dentro da amostra estão descritos na linha “MAPE (*in*)”, enquanto que os valores correspondentes às previsões h passos à frente e um passo à frente, fora da amostra, são apresentados nas linhas “MAPE (*out - h*)” e “MAPE (*out - 1*)”, respectivamente. Todos os valores apresentados na tabela são médias aritméticas das respectivas estatísticas, tomadas sobre as 1.000 replicações, exceto os valores entre parêntesis que correspondem aos desvios-padrões das estimativas.

Em relação aos demais valores na Tabela 5.2, observa-se que, quando a série temporal simulada foi gerada a partir da distribuição Log-Normal, todos os três ajustes apresentaram resultados muito similares, tanto para os valores médios dos parâmetros, que ficaram muito próximos dos valores verdadeiros, quanto para as medidas de qualidade de ajuste. Ressalta-se que o modelo escolhido para gerar as séries temporais a partir dessa distribuição (ou seja, com $p = 3$ e $q = 0$ e coeficientes dados na Tabela 5.1) corresponde ao modelo que, na aplicação à dados reais, apresentou o melhor “MAPE (*out - 1*)”. Um resultado que chama a atenção é que, nas simulações correspondentes a esse cenário, todas as três distribuições apresentaram o mesmo nível de eficiência (em termos de previsão) quando avaliado pelo “MAPE (*out - 1*)”.

Já para a série temporal simulada a partir da distribuição Beta-Prime, os valores médios dos parâmetros não apresentam diferenças significativas entre si, tão pouco com os valores verdadeiros. No entanto, uma maior variabilidade foi observada quando a distribuição Gama foi utilizada (de forma incorreta) para obtenção das estimativas. Um fato interessante é que, nesse cenário, o MAPE relativo às previsões h passo à frente, foi ligeiramente menor (em média) para a distribuição Gama do que para a própria Beta-Prime. Inclusive, o modelo escolhido para gerar as séries temporais a partir da distribuição Beta-Prime (ou seja, com $p = 1$ e $q = 1$ e coeficientes dados na Tabela 5.1) foi o que apresentou o melhor “MAPE (*out - h*)” para essa distribuição na aplicação à dados reais.

Por fim, quando a série temporal simulada foi gerada a partir da distribuição Gama, as estimativas não ficam nem próximas entre si, nem próximas dos valores reais dos parâmetros. Além desta discrepância apresentada entre as diferentes distribuições, há ainda uma volatilidade maior nos valores das estimativas evidenciada pela maior magnitude dos desvios-padrões. Avaliando as medidas de qualidade de ajuste, é observado um desempenho inferior e significativo em relação aos demais cenários considerados, principalmente nas previsões fora da amostra. Interessante observar que neste cenário, a má especificação da distribuição Beta-Prime obteve todas as medidas de ajuste melhores que a própria distribuição Gama (distribuição corretamente especificada). Nesse ponto não é possível afirmar com certeza os motivos desses resultados, pois apenas um cenário foi analisado, mas a complexidade do modelo associado à distribuição Gama pode ser o motivo dos resultados insatisfatórios obtidos. Para investigar essa possibilidade seria necessário um estudo mais detalhado envolvendo outros tamanhos de amostra e também diferentes ordens p e q .

5.4 Aspectos Computacionais

As simulações foram realizadas em um PC equipado com 16GB de RAM e processador Intel(R) Core(TM) i7-3770 (3.40GHz, 4 núcleos e 8 processadores lógicos) rodando Windows 10 Pro. As simulações foram implementadas na versão 4.1.2 do R ([R Core Team, 2021](#)), usando o pacote `doParallel` ([Corporation and Weston, 2022](#)). Para cada um dos três cenários analisados, a execução em paralelo foi realizada considerando-se 3 núcleos (um para cada distribuição).

Cada replicação consiste em estimar α , β_i , para $i \in \{1, 2\}$, ϕ_k , para $k \in \{1, p\}$, θ_j , para $j \in \{0, q\}$, φ , além da obtenção das medidas de desempenho preditivo para cada cenário. Enquanto que, para executar as 1.000 replicações da distribuição Beta-Prime foram necessárias 9 horas e 18 minutos, para a distribuição Gama foram necessárias 50 horas e 43 minutos. Por fim, para a distribuição Log-Normal, foram necessárias 23 horas e 15 minutos.

Tabela 5.2: Resultados da Simulação.

Gerados Parâmetro	Beta-Prime (μ, φ)			Gama (μ, φ)			Log-Normal (μ, φ)			
	Simulado	BP	G	Simulado	BP	G	Simulado	BP	G	LN
α	2.4496	2.4205 (0.2228)	2.9699 (0.7163)	2.4254 (0.2083)	0.4538	0.7861 (0.6464)	0.9513 (1.1810)	0.8776 (0.2387)	0.9540 (0.0665)	0.9549 (0.0661)
β_1	-0.0354	-0.0326 (0.0151)	-0.0613 (0.0540)	-0.0343 (0.0172)	-0.0682	-0.0242 (0.0173)	-0.0089 (0.0659)	0.1516 (0.1915)	-0.0520 (0.0112)	-0.0521 (0.0113)
β_2	-0.0601	-0.0570 (0.0273)	-0.1234 (0.0951)	-0.0598 (0.0238)	-0.0900	-0.1162 (0.0163)	-0.1104 (0.0353)	-0.1171 (0.1870)	-0.1247 (0.0113)	-0.1248 (0.0113)
ϕ_1	0.0274	0.0288 (0.0105)	0.0027 (0.0340)	0.0285 (0.0099)	1.1121	0.9730 (0.3113)	0.9384 (0.4235)	0.1370 (0.5191)	1.0009 (0.0278)	1.0015 (0.0276)
ϕ_2					-0.4303	-0.3956 (0.2931)	-0.3700 (0.3070)	-0.1980 (0.2037)	-0.4542 (0.0363)	-0.4545 (0.0361)
ϕ_3					0.3695	0.3169 (0.2691)	0.2861 (0.1303)	0.0525 (0.2635)	0.1371 (0.0274)	0.1374 (0.0275)
ϕ_4					-0.2008	-0.1541 (0.1393)	-0.1579 (0.1303)	0.1053 (0.2190)		
θ_1	0.0229	0.0219 (0.0073)	0.0296 (0.0131)	0.0228 (0.0017)	-0.0052	0.0012 (0.0149)	0.0023 (0.0192)	0.0797 (0.0477)		
θ_2					-0.0062	-0.0013 (0.0132)	-0.0008 (0.0179)	0.0933 (0.0590)		
θ_3					-0.0140	-0.0094 (0.0115)	-0.0078 (0.0151)	0.0805 (0.0569)		
θ_4					-0.0040	-0.0029 (0.0055)	-0.0010 (0.0086)	0.0511 (0.0377)		
θ_5					0.0045	0.0038 (0.0032)	0.0050 (0.0047)	0.0220 (0.0153)		
φ	154.9584	144.6687 (35.8360)	68.1678 (24.3051)	0.0846 (0.0209)	74.0205	61.7378 (2.1515)	61.2715 (1.8869)	0.3606 (0.1479)	133.1383 (8.7073)	127.9836 (8.9750)
MAPE (<i>im</i>)		0.0680	0.0830	0.0678		0.0925	0.0987	0.4056	0.0702	0.0702
MAPE (<i>out - h</i>)		0.1229	0.1130	0.1210		0.1619	0.2084	1.6564	0.1123	0.1122
MAPE (<i>out - 1</i>)		0.0946	0.1725	0.0941		0.1361	0.1667	0.3052	0.0906	0.0906

CAPÍTULO 6

APLICAÇÃO A DADOS REAIS

Neste capítulo é apresentada a análise de uma série temporal real correspondente à dados de temperatura coletados diariamente pela estação A701 (-23.496294, -46.620088) da cidade de São Paulo. Os dados em questão foram obtido em <https://mapas.inmet.gov.br/>. As distribuições condicionais empregadas e suas propriedades são descritas no Capítulo 3. O modelo considerado bem como os métodos de estimação de parâmetros, procedimentos para análise residual, método de previsão e medidas de qualidade de ajuste, utilizado neste capítulo, são descritos no Capítulo 4.

6.1 Características da Série Temporal

Nesta seção é apresentada uma análise das características da série temporal da temperatura (°C) coletada na estação A701. Essa análise busca avaliar se a serie temporal apresenta sazonalidade, tendências, ou outras características relevantes para a modelagem e, conseqüentemente, auxiliar na identificação do modelo mais adequado.

A Figura 6.1 apresenta o gráfico da série temporal $\{Y_t\}_{t=1}^n$, composta por $n = 1.652$ observações, que representa a temperatura (°C) média diária no período de 23 de maio de 2017 a 29 de novembro de 2021. Pela Figura 6.1 é possível observar a presença de sazonalidade na série, o que era esperado, uma vez que condições climáticas tendem a se comportar de forma similar ao longo dos anos.

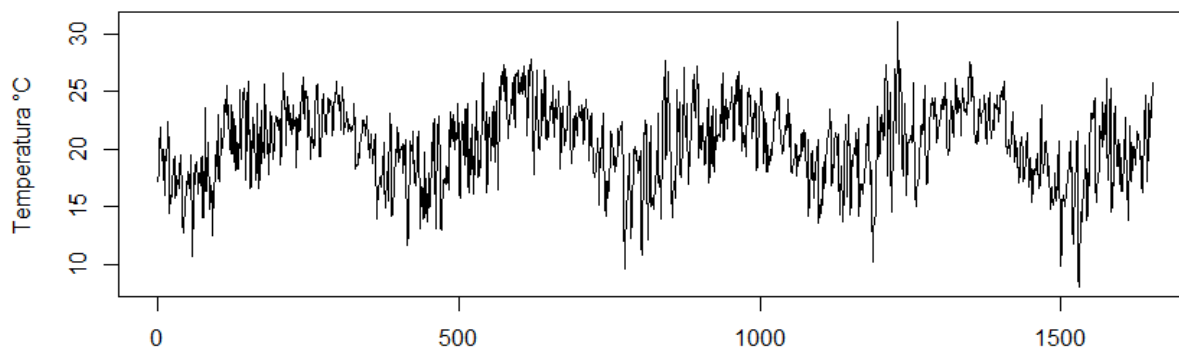


Figura 6.1: Temperatura diária (em °C), coletada na estação A701 de São Paulo, no período de 23/05/2017 a 29/11/2021.

A Figura 6.2 apresenta o histograma (à esquerda) junto com uma estimativa da função densidade de probabilidade obtida via kernel (linha azul), a função de autocorrelação amostral (ACF - no centro) e a função de autocorrelação parcial amostral (PACF - à direita), referentes à $\{Y_t\}_{t=1}^n$. Observa-se que, para essa série temporal, a distribuição incondicional aparenta ser aproximadamente simétrica, com média próxima de 20.4°C. Embora essa característica seja interessante, ela não nos auxilia na escolha da distribuição incondicional para o modelo. A ACF evidencia a existência de sazonalidade enquanto que a PACF sugere a utilização das defasagens de *lag* 1 e 2 no modelo para os dados. O decaimento lento da ACF, em conjunto com o comportamento cíclico observado no gráfico da série temporal indicam que a sazonalidade é do tipo determinística.

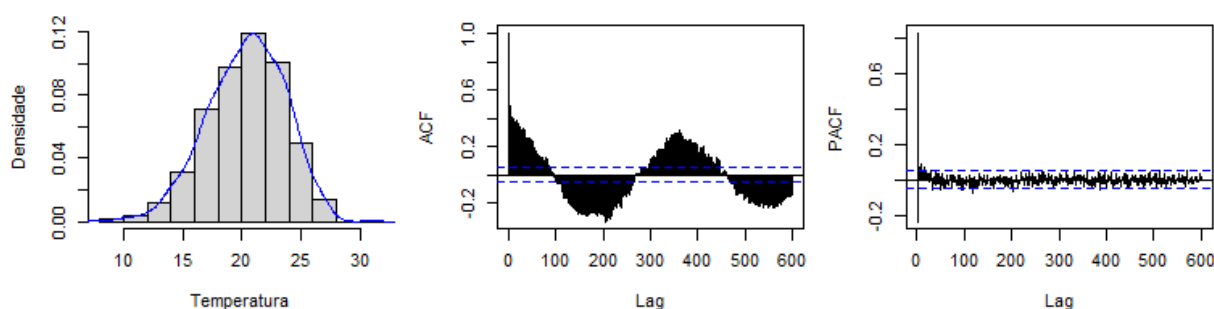


Figura 6.2: Histograma (à esquerda) e função densidade de probabilidade (linha azul), função de autocorrelação amostral (centro) e função de autocorrelação amostral parcial (à direita), correspondentes à série temporal da temperatura.

6.2 Seleção de Modelos e Previsão

Nesta seção são apresentados os procedimentos adotados para a identificação de potenciais modelos para descrever o comportamento da série temporal analisada, bem como os critérios adotados para comparação dos modelos concorrentes. Tais critérios englobam análise de significância dos parâmetros, análise residual e medidas do poder preditivo dos modelos (dentro e fora da amostra). Sendo assim, dada a série temporal $\{Y_t\}_{t=1}^{1,632}$, os primeiros 1.319 valores foram utilizados para a estimação dos parâmetros e os últimos 333 valores foram reservados para comparar as previsão fora da amostra.

Com base na análise gráfica apresentada na Seção 6.1, optou-se por utilizar um modelo com covariáveis determinísticas para modelar o comportamento sazonal da média. Devido ao comportamento senoidal do gráfico da série temporal, as covariáveis selecionadas foram

$$X_{t,1} = \cos\left(\frac{2\pi t}{365}\right) \quad \text{e} \quad X_{t,2} = \sin\left(\frac{2\pi t}{365}\right),$$

de forma que $\mathbf{X}_t = (X_{t,1}, X_{t,2})'$. Um resultado análogo pode ser obtido utilizando-se apenas uma dessas funções, porém, com a inclusão do termo de fase.

Para cada modelo considerado, a estimativa $\hat{\gamma}$ para o vetor de parâmetros γ foi obtida via método máxima verossimilhança parcial, conforme descrito na Seção 4.2.1. Para tal utilizou-se a função `ptsr.fit`. Os seguintes critérios foram adotados:

- Foram considerados todos os possíveis modelos com $0 \leq p \leq 6$ e $0 \leq q \leq 6$. Quando $p > 0$

fixou-se ainda $I_X = 1$.

- Para a distribuição condicional $f(\cdot|\mu_t, \varphi)$, correspondente à $Y_t|\mathcal{F}_{t-1}$, foram consideradas as distribuições Beta-Prime, Gama e Log-Normal, de acordo com suas respectivas reparametrizações em (3.7), (3.22) e (3.42). As demais distribuições descritas no Capítulo 3 foram consideradas apenas em um estudo preliminar.
- Tendo em vista que, para as distribuições condicionais consideradas $\mu_t \in (0, \infty)$, em todos os cenários fixou-se a função de ligação g_1 como sendo a função logaritmo. Para a função de ligação g_2 considerou-se a função identidade e a função logaritmo.
- Em termos de inicialização de parâmetros e algoritmos de otimização, utilizou-se exatamente os mesmos critérios adotados no estudo de simulação descrito no Capítulo 5.

Para cada modelo ajustado variando-se a distribuição condicional, a função de ligação g_2 e os valores de p e q , foram analisados os seguintes critérios:

- Significância dos parâmetros p -valor: O cálculo do p -valor dos parâmetros dos modelos foi baseado na distribuição assintótica do PMLE. A matriz de covariâncias foi aproximada pela inversa da Hessiana numérica. Para o modelo ser considerado candidato, todos os parâmetros, excluindo-se φ , devem ser significativos. Devido à grande quantidade de modelos concorrentes, optou-se por esse critério ao invés de reajustar um modelo com defasagens específicas.
- Existência de raiz menor ou igual a 1 (em módulo) no polinômio autoregressivo $\phi(\cdot)$: As raízes do polinômio $\phi(\cdot)$ foram obtidas utilizando-se a função `polyroot`. Verificou-se se todas elas (em módulo) eram maiores do que 1. O modelo foi considerado satisfatório se esse critério fosse satisfeito. Ressalta-se que não foi analisado quão distantes as raízes (em módulo) estavam de 1. A adoção desse critério é baseada puramente em estudos empíricos. Até o momento não existem resultados teóricos para os modelos em questão, que justifiquem sua adoção.
- Teste de Ljung-Box: Os resíduos do modelo, definidos como $\hat{\epsilon}_t = Y_t - \hat{\mu}_t$ (consulte a Seção 4.2.3), foram testados para a presença de correlação serial, utilizando-se o teste de Ljung-Box. Para realização do teste considerou-se como defasagem máxima, $l \in \{20, 40\}$. Para o modelo ser considerado satisfatório a hipótese nula não pode ser rejeitada.
- Teste para *martingale difference hypothesis* (MDH): Utilizou-se o teste AVR (consulte a Seção 4.2.2) com *wild bootstrap* e as três opções de distribuição: Normal, Mammen e Rademacher. Para o modelo ser considerado satisfatório, a hipótese nula não pode ser rejeitada.

Para todos os testes de hipóteses adotou-se como nível de significância $\alpha = 0.05$. A título de complemento, para cada modelo foram reportados ainda os valores dos critérios AIC (*Akaike information criterion*), BIC (*Bayesian information criterion*), HQ (*Hannan-Quinn information criterion*) e da Log-Verossimilhança, onde

$$\text{AIC} = 2k - 2\hat{\ell}, \quad \text{BIC} = k \log(n) - 2\hat{\ell} \quad \text{e} \quad \text{HQ} = k \log(\log(n)) - 2\hat{\ell},$$

onde n é o número de observações da série temporal, k é o número de parâmetros estimados do modelo e $\hat{\ell}$ é o valor da função de log-verossimilhança avaliada no PMLE.

Para comparar a eficiência dos modelos, foram obtidas as previsões dentro (*in-sample*) e fora da amostra (*out-of-sample*). No segundo caso, dois cenários foram considerados, h previsões 1 passo à frente, para $h = 333$, e uma previsão a h passos à frente, para $h = 1, \dots, 333$. Em todos os casos, a medida utilizada para analisar o desempenho dos modelos foi o MAPE. As previsões foram obtidas utilizando a função `predict` que implementa a equação (4.12). A título de complemento foram reportados ainda o MAE e o MSE.

6.3 Resultados

Nesta seção são apresentados os resultados referentes aos ajustes dos modelos dinâmicos à série temporal de dados reais. Ao todo, foram ajustados 294 modelos, sendo 98 para cada uma das distribuições: Beta-Prime, Gama e Log-Normal. A seguinte metodologia foi adotada para eliminar os casos que não são considerados satisfatórios, conforme descrito na Seção 6.2:

- Exclusão de todos os casos nos quais os testes de Ljung-Box e *martingale difference* apresentaram p -valor < 0.05 .
- Exclusão de todos os casos em que pelo menos um dos parâmetros, excluindo-se φ , foi considerado não significativo ao nível de 5%, ou seja, p -valor > 0.05 .
- Exclusão dos ajustes nos quais alguma das demais métricas (MAE, MSE, MAPE, AIC, BIC, HQ, Log-Verossimilhança) apresentaram valores NaN (o que indica que ocorreu *overflow*) ou Inf e também dos casos em que o polinômio autoregressivo apresenta raiz unitária.

Analisando os critérios acima observou-se que, dos 294 modelos ajustados,

- 97 não passaram nos testes de Ljung-Box e 82 não passaram em alguns dos testes de *Martingale difference*. Todas as vezes que a hipótese de *martingale difference* é rejeitada, os resíduos também não passam no teste de Ljung-Box.
- Apenas 97 apresentaram todos os coeficientes significativos. Dentre esses, apenas 35 não foram excluídos levando-se em consideração a análise residual.
- Nenhum NaN ou Inf foi encontrado para os valores do MAE, MSE e MAPE dentro da amostra e para o AIC, BIC, HQ e Log-Verossimilhança, o que indica que não houve problemas relacionados ao cálculo de $\hat{\mu}_t$ e dos valores da densidade condicional, dentro da amostra.
- Para o MAE, MSE e MAPE fora da amostra foram encontrados 3 casos com NaN e 8 casos com Inf, nenhum dos problemas estava associado às previsões 1 passo à frente.
- Dentre os 252 modelos com $p > 0$, apenas um deles resultou em $\phi(z) = 0$ para algum $z \in \mathbb{C}$ tal que $|z| \leq 1$. Para esse modelo, nenhuma medida foi igual a NaN ou Inf, porém, para as previsões h passos à frente, $\text{MAE} > 2.4 \times 10^7$, $\text{MSE} > 1.04 \times 10^6$ e $\text{MAPE} > 1.77 \times 10^{17}$ e o modelo não é satisfatório em termos de análise residual.

Tabela 6.1: Estatísticas correspondentes os modelos finais selecionados.

Distribuição	Beta-Prime(μ, φ)		Gama(μ, φ)		Log-Normal(μ, φ)	
	log identidade	log log	log identidade	log log	log identidade	log log
g_1	1	3	3	4	3	3
g_2	1	0	0	5	0	2
p	<0.0001	<0.0001	<0.0001	0.0273	<0.0001	<0.0001
q	0.6094	0.7479	0.8858	0.7532	0.9064	0.8081
p -valor*	0.8594	0.8716	0.9462	0.8580	0.9538	0.9055
Ljung-Box (20)	0.7820	0.8000	0.8100	0.8020	0.7860	0.7780
Ljung-Box (40)	0.7780	0.7980	0.8240	0.8060	0.8140	0.7800
Martingale (Mammen)	0.7860	0.7820	0.7900	0.8140	0.7800	0.7560
Martingale (Rademacher)	1.3010	1.2672	1.3011	1.2597	1.3030	1.2649
MAE (in)	3.0644	2.9393	3.0759	2.9419	3.0787	2.9338
MSE (in)	0.0644	0.0625	0.0642	0.0620	0.0644	0.0624
MAPE (in)	2.0022	2.0675	2.0086	2.1678	2.0004	2.0743
MAE (out - h)	6.7427	7.0368	6.8230	7.1698	6.7590	7.0863
MSE (out - h)	0.1106	0.1162	0.1110	0.1239	0.1106	0.1168
MAPE (out - h)	1.6541	1.2800	1.3080	1.3897	1.3092	1.2811
MAE (out - 1)	4.7813	3.1805	3.2849	3.5614	3.3083	3.1626
MSE (out - 1)	0.0966	0.0696	0.0704	0.0750	0.0705	0.0696
MAPE (out - 1)	5409.9580	5319.0180	5491.1010	5239.9520	5347.4150	5279.7170
AIC	5441.0660	5355.3100	5527.3930	5510.8800	5383.7070	5316.0090
BIC	5409.7900	5318.8220	5490.9040	5443.1150	5347.2180	5279.5200
HQ	-2698.9790	-2652.5090	-2738.5500	-2708.7400	-2666.7070	-2632.8580
Log-Verossimilhança						

Nota: *Maior p -valor dentre os p -valores dos coeficientes, excluindo-se φ . Para cada uma das distribuições, o negrito indica o menor valor do MAPE (dentro e fora da amostra).

Após adotar os critérios de exclusão acima restaram 33 candidatos. Dentre esses, para cada uma das três distribuições, foram selecionados os modelos com melhores valores para o critério MAPE para as previsões dentro (*in-sample*) e fora da amostra (*out-of-sample*), totalizando 9 modelos finais. As estatísticas associadas aos 9 modelos selecionados são apresentadas na Tabela 6.1. Nessa tabela, p -valor refere-se ao maior p -valor, dentre os p -valores dos coeficientes, excluindo-se φ . Além disso, para cada uma das distribuições, o menor valor de MAPE (dentro e fora da amostra) é indicado em negrito.

Por fim, dentre os 9 modelos obtidos, foram escolhidos 3 modelos para apresentação de resultados mais detalhados. Tais modelos correspondem aos cenários considerados nas simulações de Monte Carlo. Para a distribuição Beta-Prime, o modelo selecionado é aquele que apresentou melhor poder preditivo em termos de previsões 1 passo à frente. Para a distribuição Gama, o modelo apresentado é aquele com melhor poder preditivo dentro da amostra (*in-sample*), enquanto que, para a distribuição Log-Normal, o modelo selecionado é aquele cujas previsões h passos à frente, para $h \in \{1, \dots, 333\}$, apresentaram o menor MAPE. Os valores dos parâmetros desses modelos são apresentados na Tabela 6.2, junto com os respectivos desvios padrão.

Tabela 6.2: Parâmetros de 3 dentre os 9 modelos finais selecionados, suas respectivas distribuições condicionais e funções de ligação.

Parâmetros	Beta-Prime		Gama		Log-Normal	
	Estimativa	dp	Estimativa	dp	Estimativa	dp
α	2.4496	0.0276	0.4538	0.0047	0.9381	0.0647
β_1	-0.0354	0.0051	-0.0682	0.0146	-0.0726	0.0110
β_2	-0.0601	0.0059	-0.0900	0.0148	-0.1132	0.0112
ϕ_1	0.0274	0.0013	1.1121	<0.0001	1.0057	0.0273
ϕ_2	-	-	-0.4303	<0.0001	-0.4555	0.0368
ϕ_3	-	-	0.3695	<0.0001	0.1397	0.0273
ϕ_4	-	-	-0.2008	0.0021	-	-
θ_1	0.0229	0.0013	-0.0052	0.0002	-	-
θ_2	-	-	-0.0062	0.0002	-	-
θ_3	-	-	-0.0140	<0.0001	-	-
θ_4	-	-	-0.0040	0.0002	-	-
θ_5	-	-	0.0045	0.0002	-	-
φ	154.9584	6.0665	74.0205	2.8770	0.0882	0.0017
g_1	log		log		log	
g_2	<i>identidade</i>		log		log	

Nota: Para todos os parâmetros das distribuições Beta-Prime e da Log-Normal, p -valor < 0.0001, enquanto que, para a distribuição Gama, p -valor < 0.03.

CAPÍTULO 7

CONCLUSÕES E TRABALHOS FUTUROS

Nesta dissertação, foi proposta uma classe de modelos dinâmicos autoregressivos de médias móveis para séries temporais positivas, em que a resposta tem distribuição condicional parametrizada através da média. Oito distribuições foram reparametrizadas em termos da média μ e de um parâmetro φ . Diferentemente da literatura atual, o modelo proposto considera funções de ligação para a média e para as defasagens da série temporal (g_1 e g_2 , respectivamente) não necessariamente iguais. Além disso, possibilita a inclusão, ou não, dos regressores na parte AR do modelo. Dado que as distribuições condicionais não se restringem à família exponencial, o método proposto para a estimação dos parâmetros do modelo foi o método de máxima verossimilhança parcial (PMLE).

Simulações de Monte Carlo foram conduzidas para avaliar o comportamento dos modelos em amostras finitas. Foi analisado o desempenho do PMLE e o poder preditivo do modelo tanto para o caso em que a estimação dos parâmetros foi realizada considerando-se a verdadeira distribuição condicional quanto no contexto de má especificação da distribuição condicional. Avaliou-se o desempenho do modelo para previsões dentro (*in-sample*) e fora (*out-of-sample*) da amostra. Para finalizar, uma aplicação a dados reais foi conduzida a fim de ilustrar a utilização do modelo dinâmico proposto.

Potencial trabalho futuro envolve estudos mais detalhados associados à distribuição Gama, que apresentou comportamento de maior complexidade. Além disso, avaliar quais tipos de dados são propícios para a aplicação do modelo considerando-se as demais distribuições estudadas, mas não utilizadas nas simulações e aplicações desta dissertação.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aroian, L.A., 1941. A Study of R. A. Fisher's z Distribution and the Related F Distribution. *The Annals of Mathematical Statistics* 12, 429 – 448.
- Bayer, F.M., Bayer, D.M., Pumi, G., 2017. Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology* 555, 385–396.
- Benaduce, H., 2021. SYMARFIMA: A dynamical model for conditionally symmetric time series with long range dependence mean structure. *Mestrado em estatística. Universidade Federal do Rio Grande do Sul. Porto Alegre.*
- Benjamin, M., Rigby, R., Stasinopoulos, D., 1998. Fitting non-gaussian time series models, in: Payne, R., Green, P. (Eds.), *COMPSTAT: Proceedings in Computational Statistics*. Heidelberg: Physica-Verlag, pp. 191–196.
- Benjamin, M., Rigby, R., Stasinopoulos, D., 2003. Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98, 214–223.
- Bollerslev, T., 2010. Glossary to ARCH (GARCH), in: Bollerslev, T., Russell, J., Watson, M. (Eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford University Press, pp. 137–163.
- Bourguignon, M., Gallardo, D.I., 2020. Reparameterized inverse gamma regression models with varying precision. *Statistica Neerlandica* 74, 611–627. doi:<https://doi.org/10.1111/stan.12221>.
- Bourguignon, M., Leão, J., Gallardo, D.I., 2020. Parametric modal regression with varying precision. *Biometrical Journal* 62, 202–220. doi:<https://doi.org/10.1002/bimj.201900132>.
- Bourguignon, M., Santos-Neto, M., de Castro, M., 2021. A new regression model for positive random variables with skewed and long tail. *METRON* 79, 33–55. doi:<https://doi.org/10.1007/s40300-021-00203-y>.
- Box, G., Jenkins, G.M., Reinsel, G., 2008. *Time series analysis: forecasting and control*. Hardcover, John Wiley and Sons.
- Brockwell, P.J., Davis, R.A., 1991. *Time Series: Theory and Methods*. 2 ed., Springer Science & Business Media.
- Byrd, R., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing* 16, 1190–1208. doi:[10.1137/0916069](https://doi.org/10.1137/0916069).
- Charles, A., Darné, O., Kim, J.H., 2011. Small sample properties of alternative tests for martingale difference hypothesis. *Economics Letters* 110, 151–154.

- Choi, I., 1999. Testing the random walk hypothesis for real exchange rates. *Journal of Applied Econometrics* 14, 293–308. URL: <http://www.jstor.org/stable/223180>.
- Corporation, M., Weston, S., 2022. doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.17.
- Cox, D.R., 1981. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.
- Davis, P.J., 1959. Leonhard euler's integral: A historical profile of the gamma function: In memoriam: Milton abramowitz. *The American Mathematical Monthly* 66, 849 – 869.
- Dutang, C., Goulet, V., Pigeon, M., 2008. actuar: An r package for actuarial science. *Journal of Statistical Software* 25, 38.
- Ferrari, S.L.P., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31, 799–815.
- Fletcher, R., 1970. A new approach to variable metric algorithms. *The Computer Journal* 13, 317–322. URL: <https://doi.org/10.1093/comjnl/13.3.317>, doi:10.1093/comjnl/13.3.317, arXiv:<https://academic.oup.com/comjnl/article-pdf/13/3/317/988678/130317.pdf>.
- Fokianos, K., Kedem, B., 1998. Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis* 67, 277–296.
- Gradshteyn, I.S., Ryzhik, I.M., 2015. Table of integrals, series, and products. 8 ed., Academic Press.
- Grafakos, L., 2008. Classical Fourier Analysis. 2 ed., Springer, New York, NY.
- Kim, J.H., 2009. Automatic variance ratio test under conditional heteroskedasticity. *Finance Research Letters* 6, 179–185.
- Kim, J.H., 2014. vrtest: Variance Ratio tests and other tests for Martingale Difference Hypothesis. R package version 0.97.
- Liu, D., Nocedal, J., 1989. On the limited memory bfgs method for large scale optimization. *Mathematical Programming* 45, 503–528. doi:10.1007/BF01589116.
- Maio, V.Q.S., Cysneiros, F.J.A., 2018. SYMARMA: a new dynamic model for temporal data on conditional symmetric distribution. *Statistical Papers* 59, 75–97. URL: https://ideas.repec.org/a/spr/stpapr/v59y2018i1d10.1007_s00362-016-0753-z.html, doi:10.1007/s00362-016-0753-z.
- McCullagh, P., Nelder, J., 1989. Generalized Linear Models. 2nd ed., Chapman and Hall.
- Morris, R., 1979. The dilogarithm function of a real argument. *Mathematics of Computation* 33, 778 – 787.
- Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. *The Computer Journal* 7, 308–313. doi:10.1093/comjnl/7.4.308.
- Prass, T.S., Carlos, J.H., Taufemback, C.G., 2022. PTSR: Positive Time Series Regression. R package version 0.1.2.

- Pumi, G., Oliveira, C.R., Bayer, F.M., 2019a. Kumaraswamy regression model with Aranda-Ordaz link function. *TEST* .
- Pumi, G., Prass, T.S., Souza, R.R.a., 2021. A dynamic model for double-bounded time series with chaotic-driven conditional averages. *Scandinavian Journal of Statistics* 48, 68–86.
- Pumi, G., Valk, M., Bisognin, C., Bayer, F.M., Prass, T.S., 2019b. Beta autoregressive fractionally integrated moving average models. *Journal of Statistical Planning and Inference* 200, 196–212.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. *Test* 18, 529–545.
- Shuster, J., 1968. On the inverse gaussian distribution function. *Journal of the American Statistical Association* 63, 1514–1516. doi:[10.1080/01621459.1968.10480942](https://doi.org/10.1080/01621459.1968.10480942).
- Spendley, W., Hext, G.R., Himsforth, F.R., 1962. Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics* 4, 441–461. doi:[10.1080/00401706.1962.10490033](https://doi.org/10.1080/00401706.1962.10490033).
- Taboga, M., 2012. *Lectures on Probability Theory and Mathematical Statistics*. 2 ed., CreateSpace Independent Publishing Platform.
- van der Vaart, A.W., 2010. Time series. Lecture notes for courses “Tijdreeksen”, “Time Series” and “Financial Time Series” held at Vrije Universiteit Amsterdam, 1995-2010.
- Wheeler, B., 2020. SuppDists: Supplementary Distributions. R package version 1.1-9.5.
- Wolodzko, T., 2020. extraDistr: Additional Univariate and Multivariate Distributions. R package version 1.9.1.

ANEXO A

ARTIGO PRASS ET AL.

Autores: Taiane Schaedler Prass, Jonas Hendler Carlos, Cleiton Guolo Taufemback and Guilherme Pumi

Título: Positive Time Series Regression Models

Revista: <https://arxiv.org/abs/2201.03667>

Ano: 2022

Positive Time Series Regression Models

Taiane Schaedler Prass^{a,*} Jonas Hendler Carlos^a Cleiton Guolo Taufemback^a
and Guilherme Pumi^a

Abstract

In this paper we discuss dynamic ARMA-type regression models for time series taking values in $(0, \infty)$. In the proposed model, the conditional mean is modeled by a dynamic structure containing autoregressive and moving average terms, time-varying regressors, unknown parameters and link functions. We introduce the new class of models and discuss partial maximum likelihood estimation, hypothesis testing inference, diagnostic analysis and forecasting.

Keywords: Positive time series; Dynamic models; ARMA-type, GLM.

Mathematics Subject Classification (2000): 62M10 · 62F12 · 62J12 · 62J99.

1 Introduction

In the last decades, several models for double bounded time series have been proposed in the literature (Fokianos and Kedem, 1998; Ferrari and Cribari-Neto, 2004; Rocha and Cribari-Neto, 2009; Bayer et al., 2017; Pumi et al., 2019b,a, among others). These models are usually based on the approach nowadays known as generalized autoregressive moving average models (GARMA). The GARMA approach is based on embedding a time dependent structure into the generalized linear models (GLM) framework. The main idea behind the approach dates back to the late 70's, but the name GARMA was solidified in Benjamin et al. (2003).

Technically, GARMA models can be categorized as an observation-driven model (Cox, 1981) and, as such, includes two main components, the random and the systematic components. On one hand, the random component is responsible for the distributional features of the model, usually depending on a measure of interest, such as the (conditional) mean or median. For instance, in Rocha and Cribari-Neto (2009) the response variable follows a beta distribution, parametrized in terms of its mean, while Bayer et al. (2017) consider a Kumaraswamy distribution parameterized in terms of its median. On the other hand, the systematic component prescribes the dependence structure driving the measure of interest, often called the mean response. For instance, in Rocha and Cribari-Neto (2009) and Bayer et al. (2017), both models consider an ARMA-like structure for the systematic component. The difference is that in the former this structure models the conditional mean, while in the later, the conditional median.

GARMA modeling presents several advantages over classical time series models, such as the class of ARIMA models (Box et al., 2008). For instance, GARMA models are tailored to handle bounded non-gaussian time series without the necessity of data transformations, or other adaptations to accommodate such features. Although Benjamin et al. (2003) only consider the case in which the underlying distribution a member of the (canonical) exponential

*Corresponding author. E-mail: taiane.prass@ufrgs.br.

^aInstituto de Matemática e Estatística and Programa de Pós-Graduação em Estatística - Universidade Federal Rio Grande do Sul.

This Version: May 6, 2022

family, several distributions outside the exponential family have been considered in the literature, as for instance, in Bayer et al. (2017), where the authors consider the Kumaraswamy distribution to model double bounded time series. The distribution applied and the particular structure of the model may vary drastically depending on the characteristics of the data, the application and its goals.

The GARMA framework also allows wide variety of distributional features to be embedded into the model's random component, while retaining any desired dependence structure on the conditional mean response. This leads to a much simpler modeling strategy than the usual approach of inserting distributional features through the error term in linear models also allowing for non-gaussianity to be readily accommodated. Furthermore, conditional inference is naturally accommodated within the framework, providing a powerful inferential tool.

Considering the case where the time series assume only positive values, Benjamin et al. (1998) studies a GARMA model based on the Gamma distribution, while, more recently, Bourguignon et al. (2021) introduces a regression model where the response variable is beta prime distributed. In terms of structure, in both cases exogenous covariates are allowed in the mean response, but the later also includes an ARMA-like structure to handle serial dependence. In this work our goal is to unite these two fronts and propose a class of positive time series regression (PTSR) models allowing the dynamical part of the model to include exogenous (possibly time dependent and random) covariates and also an ARMA-like structure to handle serial dependence for the mean response, in the lines of Benjamin et al. (1998); Rocha and Cribari-Neto (2009); Bayer et al. (2017); Pumi et al. (2019b). Moreover, we shall not restrict ourselves to distributions belonging to the exponential family.

The paper is organized as follows. In the next section we introduce the proposed PTSR model. In Section 3 we introduce a partial maximum likelihood approach for parameter inference in PTSR models and derive closed formulas for the related score vector and Fisher information matrix. Section 4 we develop the asymptotic theory related to the proposed partial maximum likelihood estimator and from the asymptotic results we derive confidence intervals and hypothesis testing for the model's parameters. In Section 5 we discuss diagnostic analysis and forecast for the proposed model.

2 Proposed Model

Let $\{Y_t\}_{t \in \mathbb{Z}}$ be a stochastic process such that $P(Y_t \in (0, \infty)) = 1$ and let $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ be a set of s -dimensional exogenous covariates, possibly time dependent and random. Let \mathcal{F}_t denote the σ -field representing the history of the model known to the researcher up to time t , that is, the sigma-field generated by $(\mathbf{X}'_{t+1}, Y_t, \mathbf{X}'_t, Y_{t-1}, \mathbf{X}'_{t-1}, \dots)$. Notice that it is assumed that \mathbf{X}_{t+1} is known at time t . This is always true when $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is non-random. In the general case, one can assume that $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is a shifted version of a set of covariates and the assumption holds.

Denote by $f(\cdot | \mathcal{F}_{t-1})$ the conditional density of Y_t given \mathcal{F}_{t-1} . In this work we only consider distributions for which $\mu_t = \mathbb{E}(Y_t | \mathcal{F}_{t-1})$ exists and is finite with probability 1 and that $f(\cdot | \mathcal{F}_{t-1})$ can be parameterized in terms of μ_t and, possibly, a parameter φ that is not time-dependent. To make this relationship clear, throughout the text we shall set $f(\cdot | \mu_t, \varphi) :=$

$f(\cdot|\mathcal{F}_{t-1})$ and use the following notation

$$Y_t|\mathcal{F}_{t-1} \sim f(\cdot|\mu_t, \varphi), \quad \text{where } \mu_t = \mathbb{E}(Y_t|\mathcal{F}_{t-1}), \quad \varphi \in \mathbb{R}. \quad (1)$$

We propose to model μ_t using a dynamic GLM-like structure of the form

$$\eta_t := g_1(\mu_t) = \alpha + \mathbf{X}'_t \boldsymbol{\beta} + \sum_{k=1}^p \phi_k [g_2(Y_{t-k}) - I_X \mathbf{X}'_{t-k} \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j e_{t-j}, \quad (2)$$

$$e_t := Y_t - \mu_t,$$

where $g_1 : (a^*, b^*) \rightarrow \mathbb{R}$ is a twice differentiable, one-to-one link function, where (a^*, b^*) is the range of μ_t , $g_2 : (0, \infty) \rightarrow \mathbb{R}$ is a link function (not necessarily differentiable), $\alpha \in \mathbb{R}$ is an intercept, $\boldsymbol{\beta} := (\beta_1, \dots, \beta_s)'$ is an s -dimensional vector of parameter associated to the covariates, I_X is an indicator function which assumes the value 1 if the regressors must be included in the AR recursion and 0, otherwise, $\boldsymbol{\phi} := (\phi_1, \dots, \phi_p)'$ and $\boldsymbol{\theta} := (\theta_1, \dots, \theta_q)'$ are p and q -dimensional vectors of parameters, respectively.

Remark 2.1. For most distributions $(a^*, b^*) = (a, b)$. A special case where this property does not hold is the F distribution where $Y_t \in (0, \infty)$ and $\mu_t > 1$.

In (2), $g(\mu_t) = g(\mathbb{E}(Y_t|\mathcal{F}_{t-1}))$ follows a linear model incorporating the covariates and an adjacent ARMA(p, q)-like structure responsible for modeling a possible serial correlation in the conditional mean. In the context of financial time series, Y_t can be viewed as the squared returns while μ_t is the conditional volatility. The proposed model is observation-driven specified by the random component (1) and the systematic component (2). This approach is closely related to other GARMA-like models for non-Gaussian time series, such as the β ARMA (Rocha and Cribari-Neto, 2009), KARMA (Bayer et al., 2017), β ARFIMA (Pumi et al., 2019b) and others (see also Kedem and Fokianos, 2002).

The main difference between the model proposed here and those in the literature is that we consider the error term in the nominal level, namely, $e_t := Y_t - \mu_t$, instead of the predictive level $e_t := g_1(Y_t) - \mathbb{E}(Y_t|\mathcal{F}_{t-1})$ considered in the aforementioned works. This difference is mainly due to the fact that β ARMA and KARMA are models for double bounded time series, while the proposed model considers positive responses. However, it entails an advantage that will be important later: provided that $\mathbb{E}(|Y_t|) < \infty$, the sequence $\{e_t, \mathcal{F}_{t-1}\}_{t \in \mathbb{Z}}$ forms a martingale difference sequence. The nature of the positive response is also reflected in the autoregressive component in (2), which is measured in the transformed scale $g_2(Y_t)$. By choosing different g_2 one can either match the linear structure scale η_t or keep the original scale (by using the identity function).

3 Partial Likelihood Inference

Parameter estimation can be carried out by partial maximum likelihood approach. Let $\{(Y_t, \mathbf{X}_t)\}_{t=1}^n$ be a sample from a PTSR model under specification (1) and (2). Denote by $\boldsymbol{\gamma} := (\alpha, \boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\theta}', \varphi)'$ the $(p + q + s + 2)$ -dimensional parameter vector and let $\Omega \subseteq \mathbb{R}^{p+q+s+2}$ be the parameter space. The partial maximum likelihood estimators (PMLE) are obtained

upon maximizing the logarithm of the partial likelihood function given by

$$\ell(\boldsymbol{\gamma}) = \sum_{t=1}^n \ell_t(\boldsymbol{\gamma}), \quad \ell_t(\boldsymbol{\gamma}) := \log(f(Y_t|\mu_t, \varphi)), \quad (3)$$

so that the partial maximum likelihood estimator of $\boldsymbol{\gamma}$ is given by

$$\hat{\boldsymbol{\gamma}} = \operatorname{argmax}_{\boldsymbol{\gamma} \in \Omega} (\ell(\boldsymbol{\gamma})).$$

In most cases, $\boldsymbol{\gamma}$ cannot be analytically obtained and we have to rely on numerical optimization of the partial log-likelihood or upon solving the so-called normal equations.

3.1 Score Vector

From (3) the derivative of the log-likelihood $\ell(\boldsymbol{\gamma})$ with respect to γ_j is given by

$$\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_j} = \sum_{t=1}^n \left[\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \gamma_i} + \frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \varphi} \frac{\partial \varphi}{\partial \gamma_i} \right] = \sum_{t=1}^n \left[\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t} \frac{1}{g'_1(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_i} + \frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \varphi} \frac{\partial \varphi}{\partial \gamma_i} \right],$$

so that the score vector $U(\boldsymbol{\gamma}) = (U_{\boldsymbol{\rho}}(\boldsymbol{\gamma})', U_{\varphi}(\boldsymbol{\gamma}))'$, with $\boldsymbol{\rho} := (\alpha, \boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\theta}')$, can be written as

$$U_{\boldsymbol{\rho}}(\boldsymbol{\gamma}) = D'_{\boldsymbol{\rho}} T \mathbf{h}_1 \quad \text{and} \quad U_{\varphi}(\boldsymbol{\gamma}) = \mathbf{1}'_n \mathbf{h}_2, \quad (4)$$

where $D_{\boldsymbol{\rho}}$ is the matrix for which the (i, j) th elements is given by $[D_{\boldsymbol{\rho}}]_{i,j} = \partial \eta_i / \partial \rho_j$, T is a diagonal matrix given by

$$T_1 = \operatorname{diag} \left\{ \frac{\partial \mu_1}{\partial \eta_t}, \dots, \frac{\partial \mu_n}{\partial \eta_n} \right\} = \operatorname{diag} \left\{ \frac{1}{g'_1(\mu_1)}, \dots, \frac{1}{g'_1(\mu_n)} \right\},$$

$\mathbf{1}_n = (1, \dots, 1) \in \mathbb{R}^n$, \mathbf{h}_1 and \mathbf{h}_2 are the vectors defined by

$$\mathbf{h}_1 = \left(\frac{\partial \ell_1(\boldsymbol{\gamma})}{\partial \mu_1}, \dots, \frac{\partial \ell_n(\boldsymbol{\gamma})}{\partial \mu_n} \right)' \quad \text{and} \quad \mathbf{h}_2 = \left(\frac{\partial \ell_1(\boldsymbol{\gamma})}{\partial \varphi}, \dots, \frac{\partial \ell_n(\boldsymbol{\gamma})}{\partial \varphi} \right)'.$$

Notice that both $D_{\boldsymbol{\rho}}$ and T depend only on the structure defined by (2) and will always be the same, independently on the choice of the underlying conditional distribution. Moreover, since $e_t = Y_t - \mu_t$, the following relationship holds

$$\frac{\partial e_t}{\partial \gamma_i} = -\frac{\partial \mu_t}{\partial \gamma_i} = -\frac{\partial g_1^{-1}(\mu_t)}{\partial \gamma_i} = -\frac{1}{g'_1(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_i},$$

and implies that

$$\begin{aligned} \frac{\partial \eta_t}{\partial \alpha} &= 1 - \sum_{j=1}^q \frac{\theta_j}{g'_1(\mu_{t-j})} \frac{\partial \eta_{t-j}}{\partial \alpha}; \\ \frac{\partial \eta_t}{\partial \beta_i} &= X_{ti} - I_X \sum_{i=1}^p \phi_i X_{(t-i)j} - \sum_{j=1}^q \frac{\theta_j}{g'_1(\mu_{t-j})} \frac{\partial \eta_{t-j}}{\partial \beta_i}, \quad i \in \{1, \dots, s\}; \end{aligned}$$

$$\begin{aligned}\frac{\partial \eta_t}{\partial \phi_i} &= g_1(Y_{t-i}) - I_X \mathbf{X}'_{t-i} \boldsymbol{\beta} - \sum_{j=1}^q \frac{\theta_j}{g'_1(\mu_{t-j})} \frac{\partial \eta_{t-j}}{\partial \phi_i}, \quad i \in \{1, \dots, p\}; \\ \frac{\partial \eta_t}{\partial \theta_i} &= e_{t-i} - \sum_{j=1}^q \frac{1}{g'_1(\mu_{t-j})} \frac{\partial \eta_{t-j}}{\partial \theta_i}, \quad i \in \{1, \dots, q\}.\end{aligned}$$

Now, upon observing that

$$\mathbb{E} \left(\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_j} \right) = \sum_{t=1}^n \mathbb{E} \left(\mathbb{E} \left(\left[\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t} \frac{1}{g'_1(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_i} + \frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \varphi} \frac{\partial \varphi}{\partial \gamma_i} \right] \middle| \mathcal{F}_{t-1} \right) \right)$$

and using the fact that $1/g'_1(\mu_t)$, $\partial \eta_t / \partial \gamma_i$ and $\partial \varphi / \partial \gamma_i$ are \mathcal{F}_{t-1} -measurable, one concludes that

$$\mathbb{E} \left(\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t} \middle| \mathcal{F}_{t-1} \right) = \mathbb{E} \left(\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \varphi} \middle| \mathcal{F}_{t-1} \right) = 0, \quad \text{implying} \quad \mathbb{E} \left(\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_j} \right) = 0.$$

3.2 Conditional information matrix

In this section we derive the Fisher conditional information matrix, which will be useful later on deriving the asymptotic properties of the partial maximum likelihood estimator for the proposed model.

Let $H_t(\boldsymbol{\gamma})$ be defined by

$$H_t(\boldsymbol{\gamma}) = - \frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'},$$

and observe that

$$H(\boldsymbol{\gamma}) = - \frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = - \sum_{t=1}^n \frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = \sum_{t=1}^n H_t(\boldsymbol{\gamma}).$$

Also, observe that both, $H(\boldsymbol{\gamma})$ and $\ell(\boldsymbol{\gamma})$ depend on n , however, for simplicity and since no confusion will arise, we omit this dependence from the notation.

Let $I_n(\boldsymbol{\gamma}) := \mathbb{E}(H(\boldsymbol{\gamma}))$ be the information matrix corresponding to the sample of size n and $I^{(n)}(\boldsymbol{\gamma})$ is the negative expectation of the hessian $H_t(\boldsymbol{\gamma})$ averaged over all observations, that is,

$$I^{(n)}(\boldsymbol{\gamma}) = - \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right).$$

Hence,

$$I^{(n)}(\boldsymbol{\gamma}) = - \frac{1}{n} \mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right) \quad \text{and} \quad I_n(\boldsymbol{\gamma}) = n I^{(n)}(\boldsymbol{\gamma}).$$

Now, observe that

$$I^{(n)}(\boldsymbol{\gamma}) = - \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left(\mathbb{E} \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \middle| \mathcal{F}_{t-1} \right) \right) = \frac{1}{n} \mathbb{E}(K_n(\boldsymbol{\gamma}))$$

with

$$K_n(\boldsymbol{\gamma}) := - \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \middle| \mathcal{F}_{t-1} \right).$$

The matrix $K_n(\boldsymbol{\gamma})$ is known as the conditional information matrix corresponding to the sample

of size n and its (i, j) th element is given by

$$[K_n(\boldsymbol{\gamma})]_{i,j} = - \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}_{t-1} \right).$$

Under some regularity conditions (see Section 4),

$$\frac{1}{n} H(\boldsymbol{\gamma}) - I^{(n)}(\boldsymbol{\gamma}) \xrightarrow{P} 0 \quad \text{and} \quad \frac{1}{n} K_n(\boldsymbol{\gamma}) - I^{(n)}(\boldsymbol{\gamma}) \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty. \quad (5)$$

Furthermore, $I^{(n)}(\boldsymbol{\gamma}) \rightarrow I(\boldsymbol{\gamma})$, where

$$I(\boldsymbol{\gamma}) = \lim_{n \rightarrow \infty} I^{(n)}(\boldsymbol{\gamma}) = \lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right)$$

which is the analogous of the $I_1(\boldsymbol{\gamma})$ matrix for i.i.d. samples.

In order to derive K_n for the model defined by (1) and (2), observe that the first derivative of the log-likelihood $\ell_t := \ell_t(\boldsymbol{\gamma})$ with respect to γ_j can be written as

$$\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \gamma_j} = \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \gamma_j} + \frac{\partial \ell_t}{\partial \varphi} \frac{\partial \varphi}{\partial \gamma_j}$$

so that

$$\begin{aligned} \frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} &= \left[\frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \gamma_j} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \mu_t}{\partial \gamma_j} \right) + \frac{\partial^2 \ell_t}{\partial \mu_t \partial \varphi} \frac{\partial \varphi}{\partial \gamma_j} + \frac{\partial \ell_t}{\partial \varphi} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \varphi}{\partial \gamma_j} \right) \right] \frac{\partial \mu_t}{\partial \gamma_i} \\ &\quad + \left[\frac{\partial^2 \ell_t}{\partial \varphi \partial \mu_t} \frac{\partial \mu_t}{\partial \gamma_j} + \frac{\partial \ell_t}{\partial \varphi} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \mu_t}{\partial \gamma_j} \right) + \frac{\partial^2 \ell_t}{\partial \varphi^2} \frac{\partial \varphi}{\partial \gamma_j} + \frac{\partial \ell_t}{\partial \varphi} \frac{\partial}{\partial \varphi} \left(\frac{\partial \varphi}{\partial \gamma_j} \right) \right] \frac{\partial \varphi}{\partial \gamma_i}. \end{aligned}$$

Since,

$$\frac{\partial \mu_t}{\partial \gamma_k}, \quad \frac{\partial \varphi}{\partial \gamma_k}, \quad \frac{\partial}{\partial \mu_t} \left(\frac{\partial \mu_t}{\partial \gamma_k} \right), \quad \frac{\partial}{\partial \mu_t} \left(\frac{\partial \varphi}{\partial \gamma_k} \right), \quad \frac{\partial}{\partial \varphi} \left(\frac{\partial \mu_t}{\partial \gamma_k} \right), \quad \text{and} \quad \frac{\partial}{\partial \varphi} \left(\frac{\partial \varphi}{\partial \gamma_k} \right)$$

are all \mathcal{F}_{t-1} -measurable, it follows that

$$\mathbb{E} \left(\frac{\partial \ell_t}{\partial \mu_t} \middle| \mathcal{F}_{t-1} \right) = 0 \quad \text{and} \quad \mathbb{E} \left(\frac{\partial \ell_t}{\partial \varphi} \middle| \mathcal{F}_{t-1} \right) = 0.$$

Hence

$$\begin{aligned} [K_n]_{i,j} &= \sum_{t=1}^n \left\{ \left[\mathbb{E} \left(\frac{\partial^2 \ell_t}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \gamma_j} + \mathbb{E} \left(\frac{\partial^2 \ell_t}{\partial \mu_t \partial \varphi} \middle| \mathcal{F}_{t-1} \right) \frac{\partial \varphi}{\partial \gamma_j} \right] \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \gamma_i} \right. \\ &\quad \left. + \left[\mathbb{E} \left(\frac{\partial^2 \ell_t}{\partial \varphi \partial \mu_t} \middle| \mathcal{F}_{t-1} \right) \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \gamma_j} + \mathbb{E} \left(\frac{\partial^2 \ell_t}{\partial \varphi^2} \middle| \mathcal{F}_{t-1} \right) \frac{\partial \varphi}{\partial \gamma_j} \right] \frac{\partial \varphi}{\partial \gamma_i} \right\} \end{aligned}$$

and the conditional Fisher information matrix for $\boldsymbol{\gamma}$ is then given by

$$K_n(\boldsymbol{\gamma}) := \begin{pmatrix} K_{\boldsymbol{\rho}, \boldsymbol{\rho}} & K_{\boldsymbol{\rho}, \varphi} \\ K_{\boldsymbol{\lambda}, \boldsymbol{\rho}} & K_{\varphi, \varphi} \end{pmatrix}, \quad (6)$$

with

$$K_{\rho,\rho} = D'_\rho T_1 E_\mu T_1 D_\rho, \quad K_{\rho,\varphi} = K'_{\varphi,\rho} = D'_\rho T_1 E_{\mu\varphi} \mathbf{1}_n \quad \text{and} \quad K_{\varphi,\varphi} = \mathbf{1}'_n E_\nu \mathbf{1}_n$$

where D_ρ , T_1 and $\mathbf{1}_n$ are the matrices and the vector defined in (4) and E_μ , $E_{\mu\nu}$ and E_ν are diagonal matrices for which the (t, t) th element is given by

$$[E_\mu]_{t,t} = -\mathbb{E}\left(\frac{\partial^2 \ell_t}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1}\right), \quad [E_{\mu\nu}]_{t,t} = -\mathbb{E}\left(\frac{\partial^2 \ell_t}{\partial \mu_t \partial \varphi} \middle| \mathcal{F}_{t-1}\right) \quad \text{and} \quad [E_\nu]_{t,t} = -\mathbb{E}\left(\frac{\partial^2 \ell_t}{\partial \varphi^2} \middle| \mathcal{F}_{t-1}\right).$$

4 Asymptotic theory and hypothesis testing

A rigorous asymptotic theory for the PMLE in the context of GARMA-like models the underlying distribution belongs to the canonical exponential family can be found in Fokianos and Kedem (1998, 2004). Although the exponential family is broad enough to be useful in practice, this is still a limitation that must be observed in practice. For PTSR models, when the underlying distribution belongs to the canonical exponential family, the model falls into the context of Fokianos and Kedem (2004) and the asymptotic theory for the PMLE follows under assumptions A1 to A4 there stated. Under those conditions, there exists a non-random information matrix, denoted by $I(\gamma)$, such that

$$\frac{K_n(\gamma)}{n} \xrightarrow[n \rightarrow \infty]{P} I(\gamma),$$

holds (in probability), $I(\gamma)$ is positive definite and invertible matrix in an open neighborhood of the true parameter γ_0 . It can also be shown that the probability that a locally unique maximum partial likelihood estimator exists in a neighborhood of γ_0 tends to one. Furthermore, the estimator is consistent

$$\hat{\gamma}_n \xrightarrow[n \rightarrow \infty]{P} \gamma_0$$

asymptotically normal

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow[n \rightarrow \infty]{d} N_{p+q+s+2}(\mathbf{0}, I(\gamma_0)^{-1}), \quad (7)$$

and (5) holds.

For distributions that are not member of the canonical exponential family, a general asymptotic theory for the PMLE in the context of GARMA-like models is not available. We speculate that, under assumptions closely related to A1 to A4 in Fokianos and Kedem (2004), the proofs presented in Fokianos and Kedem (1998) can be adapted to provide asymptotic results similar to (7), in a case by case fashion. However, we shall not pursue this matter here.

4.1 Confidence intervals and hypothesis testing inference

Construction of asymptotic confidence intervals/regions and test statistics for hypothesis testing can be obtained using (7). Let $\{Y_t\}_{t=1}^n$ be a sample from a PTSR model, γ_i denote the i th component of the true parameter vector γ and let $\hat{\gamma}_i$ be its PMLE obtained from the sample. Let $I(\hat{\gamma})^{ij}$ denote the (i, j) th element of the inverse of the conditional information matrix (6) evaluated at $\hat{\gamma} \in \mathbb{R}^{p+q+s+2}$.

From (7), we have

$$\frac{\hat{\gamma}_i - \gamma_i}{\sqrt{I(\hat{\gamma})^{ii}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Hence a $100(1 - \alpha)\%$, $0 < \alpha < 1/2$, asymptotic confidence interval for γ_i is given by

$$\left[\hat{\gamma}_i - z_{1-\alpha/2} \sqrt{I(\hat{\gamma})^{ii}}; \hat{\gamma}_i + z_{1-\alpha/2} \sqrt{I(\hat{\gamma})^{ii}} \right],$$

where z_δ is the δ -quantile of the standard normal distribution.

From (7) one can also derive asymptotic test statistics for hypothesis testing. Let γ_i^0 be a given hypothesized value for the true parameter γ_i and consider the test

$$\mathcal{H}_0 : \gamma_i = \gamma_i^0 \quad \text{against} \quad \mathcal{H}_1 : \gamma_i \neq \gamma_i^0.$$

An asymptotic version for the signed square root of Wald's statistic can be obtained from (7) by considering

$$Z = \frac{\hat{\gamma}_i - \gamma_i^0}{\sqrt{I(\hat{\gamma})^{ii}}}.$$

Under \mathcal{H}_0 , the distribution of Z is approximately standard normal for large n . For details and a proof of this claim see Pawitan (2001) and Fahrmeir (1987).

Versions for other well-known statistics such as the likelihood ratio, Rao's score, Wald's and the gradient statistics to perform more general hypothesis testing inference can also be derived from (7) in similar fashion. In large samples and under the null hypothesis, such statistics are approximately distributed as in the traditional i.i.d. case. More generally, for $k < p + q + s + 2$, let $T : \mathbb{R}^{p+q+s+2} \rightarrow \mathbb{R}^k$ be a vector valued transformation such that its jacobian $\mathbf{J}(\boldsymbol{\gamma})$ exists, is of full rank k and it is a continuous function of $\boldsymbol{\gamma}$ in an open subset of Ω . To test a composite hypothesis of the form

$$\mathcal{H}_0 : T(\boldsymbol{\gamma}) = \mathbf{0} \quad \text{versus} \quad \mathcal{H}_1 : T(\boldsymbol{\gamma}) \neq \mathbf{0}.$$

we can use the traditional Wald's statistic, given by

$$W = nT(\hat{\boldsymbol{\gamma}})' [\mathbf{J}(\hat{\boldsymbol{\gamma}})' I^{-1}(\hat{\boldsymbol{\gamma}}) \mathbf{J}(\hat{\boldsymbol{\gamma}})]^{-1} T(\hat{\boldsymbol{\gamma}}).$$

Under \mathcal{H}_0 its distribution converges to a chi-square with the usual degrees of freedom.

5 Diagnostic analysis and forecasting

Model selection criteria

Diagnostics in the context of PTSR models follow the usual procedures of GLM theory.

Model selection among several competing models may be based on the usual information criteria such as Akaike's (AIC), Schwartz's (SIC) and Hannan Quinn's (HQ) information

criteria, respectively defined by

$$\text{AIC} = -2\hat{\ell} + 2k, \quad \text{SIC} = -2\hat{\ell} + \log(n)k, \quad \text{and} \quad \text{HQ} = -2\hat{\ell} + \log(\log(n))k,$$

where k denotes the number of parameter in the fitted model and $\hat{\ell}$ is the log-likelihood function (3) evaluated at the PMLE. As usual, these criteria should be applied in combination with residual analysis, discussed next.

Residuals

Residuals are extremely important in assessing the quality of fit from a model. There are several types of residuals that can be computed given a model. The most commonly applied ones are the simple residual given by $r_t = Y_t - \mu_t$ and the so-called quantile residuals defined by

$$e_t^{(q)} = \Phi^{-1}(F(Y_t | \mathcal{F}_{t-1})),$$

where Φ^{-1} denotes the standard normal quantile function. In the present framework, if the model is correctly specified, then r_t should behave as a martingale difference (with respect to \mathcal{F}_{t-1}), while the quantile residuals should follow a standard normal distribution. These simple results are often applied in the construction of goodness-of-fit tests.

The literature related to testing the martingale difference hypothesis has grown significantly in the last decade and several tests and computational packages are available to perform such tests. For instance, Kim (2009) proposed the so-called wild bootstrap automatic variance ratio test. Domínguez and Lobato (2003) proposed an approach based on the Cramer von Mises and Kolmogorov-Smirnov statistics to test the martingale difference hypothesis, which is called the Domínguez-Lobato test. Finally, another approach based on the generalized spectral distribution function is presented in Escanciano and Velasco (2006). We refer the reader to the aforementioned papers for details. See also Charles et al. (2011), where the authors discuss finite sample performance of these methods. Such tests are readily available in most softwares. For instance in R (R Core Team, 2021), they are available in the package `vrtest` (Kim, 2014).

In the present scenario, after we perform parameter estimation, we can obtain an estimate for the simple residuals. Then a martingale difference test can be applied to the estimated simple residuals, resulting in a goodness-of-fit test. As long as the second moment of the fitted model is finite, one can also apply a white noise test to the estimated simple residuals as, in this scenario, a martingale difference is, unconditionally, a white noise.

When the model is correctly specified, the quantile residual should follow a standard normal distribution. Hence, testing the estimated quantile residual for normality can be used a goodness-of-fit test. Another useful diagnostic tool is as follows. When n is sufficiently large, the distribution of the residuals sample autocorrelation function at lag h , $\hat{\rho}(h)$, is approximately normal with zero mean and constant variance $1/(n-m)$ (Kedem and Fokianos, 2002; Anderson, 1942; Box et al., 2008). The plots of the residual ACF with horizontal lines at $\pm 1.96/\sqrt{(n-m)}$ can be useful for assessing whether the residuals display white noise behavior (Kedem and Fokianos, 2002). The traditional Ljung-Box test (Ljung and Box, 1978) based on the residual, to test the null hypothesis $\mathcal{H}_0 : \rho(1) = \dots = \rho(l) = 0$, for some $l > 0$,

the following test statistic can be used

$$Q = n(n+2) \sum_{i=1}^l \frac{\hat{\rho}(i)^2}{n-i}.$$

Under the null hypothesis and large n , Q is approximately chi-squared distributed with l degrees of freedom.

5.1 Forecasting

Upon applying the partial maximum likelihood estimators in (2), we can obtain the in-sample forecast, denoted by $\{\hat{\mu}_t\}_{t=1}^n$, and the h steps ahead predicted values (out-of-sample forecast) for the conditional mean of a PTSR model, which we denote by $\hat{\mu}_{n+h} = \hat{\mu}_n(h)$. We shall assume that the covariates \mathbf{X}_t , for $t = n+1, \dots, n+h$, are available or can be obtained.

Starting at $t = 1$, we sequentially set

$$\hat{\eta}_t = \hat{\alpha} + \hat{\mathbf{X}}_t' \hat{\boldsymbol{\beta}} + \sum_{i=1}^p \hat{\phi}_i [g_{1,2}(\hat{Y}_{t-i}) - I_X \hat{\mathbf{X}}_{t-i}' \hat{\boldsymbol{\beta}}] + \sum_{k=1}^q \hat{\theta}_k \hat{e}_{t-k},$$

with

$$\hat{Y}_t = \begin{cases} 0, & p = 0, t < 1, \\ \frac{1}{p} \sum_{i=1}^p Y_i, & p > 0, t < 1, \\ Y_t, & 1 \leq t \leq n, \\ \hat{\mu}_t, & t > n, \end{cases} \quad \hat{\mathbf{X}}_t' = \begin{cases} 0, & p I_X = 0, t < 1, \\ \frac{1}{p} \sum_{i=1}^p \mathbf{X}_i', & p I_X > 0, t < 1, \\ \mathbf{X}_t, & t \geq 1, \end{cases}$$

$$\hat{\mu}_t = g_1^{-1}(\hat{\eta}_t), t \geq 1, \quad \text{and} \quad \hat{e}_t = \begin{cases} \hat{Y}_t - \hat{\mu}_t, & 1 \leq t \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

References

- Anderson, R. L., 1942. Distribution of the serial correlation coefficient. *The Annals of Mathematical Statistics* 13 (1), 1–13.
- Bayer, F. M., Bayer, D. M., Pumi, G., 2017. Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology* 555, 385–396.
- Benjamin, M., Rigby, R., Stasinopoulos, D., 1998. Fitting non-Gaussian time series models. *COMPSTAT Proceedings in Computational Statistics Heidelberg: Physica-Verlag*, 191–196.
- Benjamin, M., Rigby, R., Stasinopoulos, D., 2003. Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98 (461), 214–223.
- Bourguignon, M., Santos-Neto, M., de Castro, M., 2021. A new regression model for positive random variables with skewed and long tail. *METRON* 79, 33–55.
- Box, G., Jenkins, G. M., Reinsel, G., June 2008. *Time series analysis: forecasting and control*. Hardcover, John Wiley & Sons.

- Charles, A., Darné, O., Kim, J. H., 2011. Small sample properties of alternative tests for martingale difference hypothesis. *Economics Letters* 110 (2), 151–154.
- Cox, D. R., 1981. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.
- Domínguez, M. A., Lobato, I. N., 2003. Testing the martingale difference hypothesis. *Econometric Reviews* 22 (4), 351–377.
- Escanciano, J. C., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* 134 (1), 151–185.
- Fahrmeir, L., 1987. Asymptotic testing theory for generalized linear models. *Statistics* 18 (1), 65–76.
- Ferrari, S. L. P., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 (7), 799–815.
- Fokianos, K., Kedem, B., 1998. Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis* 67 (2), 277–296.
- Fokianos, K., Kedem, B., 2004. Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis* 25 (2), 173–197.
- Kedem, B., Fokianos, K., 2002. *Regression Models for Time Series Analysis*. John Wiley & Sons.
- Kim, J. H., 2009. Automatic variance ratio test under conditional heteroskedasticity. *Finance Research Letters* 6 (3), 179–185.
- Kim, J. H., 2014. *vrtest: Variance Ratio tests and other tests for Martingale Difference Hypothesis*. R package version 0.97.
URL <https://CRAN.R-project.org/package=vrtest>
- Ljung, G. M., Box, G. E. P., 1978. On a measure of lack of fit in time series models. *Biometrika* 65 (2), pp. 297–303.
- Pawitan, Y., 2001. *All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford Science publications.
- Pumi, G., Oliveira, C. R., Bayer, F. M., 2019a. Kumaraswamy regression model with Aranda-Ordaz link function. *TEST*.
- Pumi, G., Valk, M., Bisognin, C., Bayer, F. M., Prass, T. S., 2019b. Beta autoregressive fractionally integrated moving average models. *Journal of Statistical Planning and Inference* 200, 196–212.
- R Core Team, 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL <https://www.R-project.org/>
- Rocha, A. V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. *Test* 18 (3), 529–545.