Frank Magalhaes de Pinho

Modelos de espaço de estados não Gaussianos - Distribuições de Caudas Pesadas

Belo Horizonte
2012

Frank Magalhaes de Pinho

Modelos de espaço de estados não Gaussianos - Distribuições de Caudas Pesadas

Tese apresentada ao Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, para a obtenção de Título de Doutor em Estatística, na Área de Séries Temporais.

Orientadora: Glaura da Conceição Franco

Belo Horizonte
2012

Frank Magalhaes de Pinho,
Modelos de espaço de estados não Gaussianos - Distribuições de Caudas Pesadas
136 páginas
Tese (Doutorado) - Instituto de Ciências Exatas da Universidade Federal de Minas Gerais. Departamento de Estatística.

1. Modelos de Espaços de Estados Não-Gaussianos
2. Distribuições de Caudas Pesadas
3. Métodos de Estimação Clássica e Bayesiana
4. Algoritmos de Maximização BFGS, SQP e FSQP.
5. Estimador de Máxima Verossimilhança Penalizada
6. Métodos Bootstrap
7. Volatilidade Estocástica

I. Universidade de Minas Gerais. Instituto de Ciências Exatas. Departamento de Estatística.

Comissão Julgadora:

Prof. Dr. Aureliano A. Bressan (UFMG)

Prof. Dr. Thiago Rezende dos Santos (UFMG)

Prof. Dr. Márcio Polleti Laurini (USP)

Prof. Dr. Ralph Santos Silva (UFRJ)

Profa. Dra. Glaura da Conceição Franco (UFMG)

Dedico este trabalho
a minha maravilhosa esposa Fernanda, pela eterna amizade e companheirismo,
a meus filhos Clara e Felipe, razões do meu viver, amo vocês,
aos meus pais, por toda uma vida de dedicação a mim,
a meus irmãos, pelo apoio incondicional.

Devemos acreditar nisso...

Nasceste no lar que precisavas.

Vestiste o corpo físico que merecias.

Moras onde melhor Deus te proporcionou, de acordo com teu adiantamento.

Possuis os recursos financeiros coerentes com as tuas necessidades, nem mais, nem menos, mas o justo para as tuas lutas terrenas.

Teu ambiente de trabalho é o que elegeste espontaneamente para a tua realização.

Teus parentes e amigos são as almas que atraístes com tua própria afinidade, portanto, teu destino está constantemente sobre teu controle.

Tu escolhes, recolhes, eleges, atrais, buscas, expulsas, modificas, tudo aquilo que te rodeia a existência.

Chico Xavier

Agradecimentos

Agradeço a Deus, à minha esposa Fernanda, aos meus filhos Clara e Felipe, aos meus pais e irmãos, à minha orientadora Glaura, à minha professora de Matemática do ensino fundamental Vanda, aos meus colegas de doutorado, aos professores e técnicos-administrativos do Departamento de Estatística da UFMG e às instituições de fomento a pesquisa CAPES, CNPq e FAPEMIG.

Resumo

Esta tese contém três artigos que ampliam os conhecimentos sobre uma nova família de modelos de espaços de estados proposta por Santos et al. (2010) denominada non-Gaussian state space model (NGSSM). Esta família de modelos é muito interessante porque, além de conter um conjunto significativo de distribuições de probabilidade, tem-se a função de verossimilhança analiticamente, e por consequência há a possibilidade de realizar inferência sobre os parâmetros sem a necessidade de métodos numéricos aproximados, como o filtro de partícula.

No primeiro artigo são propostas outras cinco distribuições de causas pesadas como casos particulares da NGSSM, além das distribuições Weibull e Pareto propostas por Santos et al. (2010). São elas: Log-normal, Log-gama, Fréchet, Lévy, Skew GED. São realizadas simulação Monte Carlo para avaliação dos estimadores clássicos e bayesianos para os modelos de caudas pesadas. Os resultados demonstram, empiricamente, que os estimadores são não viesados assintoticamente e consistentes. Os modelos de caudas pesadas são estimados para as séries dos índices das mais importantes bolsas de valores da América - $S\&P500$, $NASDAQ$, $IBOVESPA$, $INMEX$, $MERVAL$, $IPSA$ - e os resultados são comparados com modelos da família GARCH. O modelo Weibull da NGSSM apresenta melhores resultados para todas as séries estudadas.

No segundo artigo é avaliado o comportamento do estimador de máxima verossimilhança para os parâmetros dos modelos de caudas pesadas quando as séries temporais são pequenas. Observa-se que um dos parâmetros, ω, é sempre sobreestimado, independentemente do modelo e do algorítmo de maximização utilizados. A obtenção de um estimador adequado para ω é fundamental, pois quando este parâmetro é sobreestimado a variabilidade das séries temporais é subestimada. Funções de penalização para a função de verossimilhança são propostas e, por consequência, estimadores de máxima verossimilhança penalizada são propostos e avaliados. Os resultados demonstram que os estimadores propostos apresentam uma redução significativa do viés em relação ao observado pelo estimador de máxima verossimilhança.

No terceiro artigo é avaliado o comportamento do intervalo de confiança assintótico dos parâmetros dos modelos de caudas pesadas quando as séries são pequenas. Observa-se que os intervalos de confiança para o parâmetro ω são inadequados, seja utilizando o estimador de máxima verossimilhança ou o estimador de máxima verossimilhança penalizado. Em razão disto são propostos e avaliados intervalos de confiança bootstrap. Os resultados demonstram que o intervalo de confiança bootstrap com correção de viés obtido a partir do bootstrap paramétrico apresentam taxas de cobertura muito próximas da taxa nominal utilizada no estudo empírico.

Palavras-chave: Distribuições de Caudas Pesadas, Métodos de Estimação Clássica e Bayesiana, Estimador de Máxima Verossimilhança Penalizada, Métodos Bootstrap, Algoritmo de Maximização BFGS, Programação Sequencial Quadrática, Programação Sequencial Quadrática Factível, Volatilidade Estocástica.

Abstract

This thesis contains three papers that expand the knowledge about a new family of state space model proposed by Santos et al. (2010) called non-Gaussian state space model (NGSSM). This family of models is very interesting because, besides containing a significant set of probability distributions, the likelihood function can be written in an exact form. Consequently, there is the possibility of performing inference about the parameters without the need of numerical methods, such as the particle filter.

In the first paper it is shown that besides Weibull and Pareto proposed in the Santos et al. (2010) paper, five other heavy tailed distributions are contained in the NGSSM. They are: Log-normal, log-gamma, Fréchet, Lévy, Skew GED. To evaluate classical and Bayesian estimators for heavy tailed models of the NGSSM Monte Carlo simulations are performed. The results demonstrate empirically that the estimators are not asymptotically biased and they are consistent. The heavy tailed models are estimated for the series of the most important stock exchange indexes of America, such as $S\&P500$, $NASDAQ$, $IBOVESPA$, $INMEX$, $MERVAL$, $IPSA$. The results are compared with the GARCH models and it is observed that the Weibull model of NGSSM shows better results for all time series studied.

In the second paper, it is evaluated the behavior of the maximum likelihood estimator of the parameters of the heavy tailed models when the time series is small. It is observed that the parameter ω is always overestimated, regardless the model and the maximization algorithm used. Obtaining a suitable estimator for ω is critical, because when this parameter is overestimated the variability of the time series is underestimated. Penalty functions are proposed for the likelihood function and, consequently, penalized maximum likelihood estimators are proposed and evaluated. The results demonstrate that the estimators proposed reduce significantly the bias when compared with the bias

obtained by the maximum likelihood estimator.

In the third paper it is evaluated the behavior of the asymptotic confidence interval of the parameters of the heavy tailed models when the time series is small. It is observed that the confidence intervals for the parameter ω are inadequate, either using the maximum likelihood estimator or penalized maximum likelihood estimator. Thus bootstrap confidence intervals are proposed and evaluated. The results show that the bootstrap confidence interval with bias correction obtained from the parametric bootstrap has coverage rates very close to the nominal level used in the empirical study.

Keywords: Heavy Tailed Distributions, Bayesian and Classical Inference, Penalized Maximum Likelihood Estimator, Bootstrap Methods, Bootstrap Confidence Intervals, BFGS Maximization Algorithm, Sequential Quadratic Programming, Feasible Sequential Quadratic Programming, Stochastic Volatility.

Lista de Figuras

5.1 Histograms of the estimates (MLE, BE-Mean and BE-Median) of ω for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500. 54

5.2 Histograms of the estimates (MLE, BE-Mean and BE-Median) of β for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500. 55

5.3 Histograms of the estimates (MLE, BE-Mean and BE-Median) of δ for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500. 56

5.4 The index and the log-return of S&P 500, NASDAQ, INMEX, IBOVESPA, MERVAL and IPSA, in the period from 02/01/2007 to 05/16/2011. 64

6.1 Histograms of 1000 estimates of the MLE, using BFGS, for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) and from the Weibull model with ($\omega = 0.90$; $\beta = 1.0$; $\upsilon = 5.0$), with size 50. 77

6.2 Histograms of 1000 estimates of the MLE, using BFGS, for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) and from the Weibull model with ($\omega = 0.90$; $\beta = 1.0$; $\upsilon = 5.0$), with size 200. 78

6.3 Penalty functions I (at left), IV (at center) and VII (at right) proposed to time series of size 50, 100, 200 and 500. 82

6.4 Percentage of bias and MSE of PMLE over the MLE, by BFGS, for the Log-normal, Log-gamma, Weibull and Skew GED models for $\omega = 0.85$ (at left), $\omega = 0.90$ (at center) and $\omega = 0.95$ (at right). 85

6.5 Percentage of bias and MSE of PMLE over the MLE, by BFGS, for the Pareto, Fréchet and Lévy models for $\omega = 0.85$ (at left), $\omega = 0.90$ (at center) and $\omega = 0.95$ (at right). 86

6.6 Boxplot of the 1000 estimates (MLE, PMLE I, PMLE IV and PMLE VII) for $\omega = 0.95$, by BFGS and SQP, for time series of size 50 and for Log-normal, Pareto, Weibull and Skew GED models. 87

6.7 Boxplot of the 1000 estimates (MLE, PMLE I, PMLE IV and PMLE VII) for $\omega = 0.95$, by BFGS and SQP, for time series of size 50 and for Log-normal, Pareto, Weibull and Skew GED models. 88

7.1 Penalty functions IV proposed to time series of size 50, 100, 200 and 500. 105

7.2 Parametric Bootstrap - Asymptotic confidence interval and bootstrap confidence interval by PMLE for the estimates of vector parameter $\boldsymbol{\varphi}$ of the Log-normal, Log-gamma, Weibull and Fréchet models. 116

7.3 Parametric Bootstrap - Asymptotic confidence interval and bootstrap confidence inverval by PMLE for the estimates of vector parameter $\boldsymbol{\varphi}$ of the Pareto, Lévy and Skew GED models. 117

Lista de Tabelas

4.1 Modelos de espaços de estados 32

5.1 Monte Carlo study for the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$). 57

5.2 Monte Carlo study for the Log-gamma model with ($\omega = 0.90$; $\beta = 1.0$; $\alpha = 5.0$). 58

5.3 Monte Carlo study for the Fréchet model with ($\omega = 0.90$; $\beta = 1.0$; $\alpha = 5.0$). 58

5.4 Monte Carlo study for the Lévy model with ($\omega = 0.90$; $\beta = 1.0$). 59

5.5 Monte Carlo study for the Skew GED model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$; $\kappa = 1.0$). 59

5.6 Monte Carlo study for the Pareto model with ($\omega = 0.90$; $\beta = 1.0$). . . . 60

5.7 Monte Carlo study for the Weibull model with ($\omega = 0.90$; $\beta = 1.0$; $\upsilon = 5.0$). 60

5.8 Fitted models for the North and South American stock indexes. 65

5.9 Parameter estimates of the Weibull models for the volatility of the indexes. 65

6.1 Distributions in the NGSSM 74

6.2 Percentage of times that the maximum likelihood estimates of parameter ω is 1.00 in 1000 Monte Carlo simulations using BFGS, SQP and FSQP algorithms. 79

6.3 Values of n_1 and n_2 for the penalized function $\upsilon(\omega, n_1, n_2)$. 81

6.4 Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Log-normal and Log-gamma models). 89

6.5 Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Pareto and Weibull models). 90

6.6 Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Fréchet and Lévy models). 91
6.7 Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Skew GED model). 92
6.8 Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for φ by BFGS and SQP (Log-normal and Log-gamma models). 93
6.9 Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for φ by BFGS and SQP (Pareto and Weibull models). 94
6.10 Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for φ by BFGS and SQP (Fréchet and Lévy models). 95
6.11 Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for φ by BFGS and SQP (Skew GED model). 96
6.12 95% Asymptotic confidence interval of MLE by BFGS 3 differents PMLE using BFGS for time series of size 50. 97

7.1 Cases of the NGSSM 103
7.2 Parametric Bootstrap - bootstrap estimates, range and coverage rate by MLE. 113
7.3 Parametric Bootstrap - bootstrap estimates, range and coverage rate by PMLE. 114
7.4 Bootstrap on standardized Pearson residual - bootstrap estimates, range and coverage rate by PMLE. 115

Sumário

1 Introdução **1**

I Revisão de Literatura **7**

2 Conceitos de Processos Estocásticos e Séries Temporais **9**

3 Classe de Distribuições de Caudas Pesadas e *Outliers* **13**

3.1 Classes de distribuições de caudas pesadas 13

3.1.1 A classe de distribuições de cauda longa 16

3.1.2 A classe de distribuições subexponencial 16

3.1.3 A classe de distribuições de variação regular 17

3.1.4 A classe de distribuições de variação dominada 18

3.1.5 Relações entre as classes de distribuições de cauda pesada 18

3.2 Distribuições resistentes e propensas a *outliers* 19

3.2.1 Distribuições resistentes a *outliers* 19

3.2.2 Distribuições propensas a *outliers* 22

3.2.3 Classificação das distribuições de probabilidade relacionada a sensibilidade a *outliers* 24

4 Modelos de Espaços de Estados **27**

4.1 Origem dos modelos de espaços de estados 28
4.2 Modelo de tendência linear local – MTL 28
4.3 Modelo estrutural básico – MEB 29
4.4 Modelo de espaços de estados – MEE 31
4.4.1 Representação do MNL pelo MEE 34
4.4.2 Representação do MTL pelo MEE 34
4.5 Modelos de Espaços de Estados Não-Gaussianos 34

II Artigos Científicos 36

5 Modelling Volatility Using State Space Models with Heavy Tailed Distributions 37

5.1 Introduction 38
5.2 A non-Gaussian state space model 40
5.2.1 Inference procedure 42
5.3 Heavy tailed distributions in the NGSSM 46
5.3.1 Log-normal model 47
5.3.2 Log-gamma model 47
5.3.3 Fréchet model 48
5.3.4 Lévy model 49
5.3.5 Skew GED model 49
5.3.6 Pareto model 50
5.3.7 Weibull model 51
5.4 Monte Carlo study 51
5.4.1 Empirical distribution of the estimators 53
5.4.2 Point and interval estimation 53
5.5 Application to South and North American stock exchange indexes 62

5.6 Conclusion 66

6 Penalized Likelihood for a Non Gaussian State Space Model Considering Heavy Tailed Distributions 71

6.1 Introduction 72

6.2 A non-Gaussian state space model 73

6.3 Penalized likelihood function for the NGSSM 74

6.3.1 Maximum Likelihood Estimator (MLE) 75

6.3.2 Penalized Maximum Likelihood Estimator 78

6.4 Monte Carlo study 82

6.5 Conclusion 84

7 Bootstrapping Non Gaussian State Space Models 99

7.1 Introduction 100

7.2 A non-Gaussian state space model 101

7.3 Bootstrap methods 105

7.3.1 Bootstrap schemes 107

7.3.2 Bootstrap confidence intervals 108

7.4 Monte Carlo study 110

7.5 Conclusion 112

8 Considerações Finais 123

Referências Bibliográficas 126

Capítulo 1

Introdução

Na literatura, tem-se uma quantidade significativa de modelos que são desenvolvidos baseados em determinadas suposições, tais como normalidade, homoscedasticidade e independência dos erros, entretanto existe um número siginificativo de conjuntos de dados que descrevem problemas reais nas organizações, na economia, nos mercados financeiros, em fenômenos naturais, que são incompatíveis com essas suposições.

Sob o contexto de séries temporais, a hipótese de independência dos erros é raramente satisfeita, não obstante a suposição de normalidade e homoscedasticidade são frequentemente inapropriadas para séries em diversos campos de aplicação, mas em especial para séries econômicas e financeiras. A modelagem via espaço de estados, também denominado por modelos dinâmicos quando os métodos de estimação utilizados são bayesianos e modelos estruturais, quando a abordagem frequentista é utilizada, é o tema central que será proposto pesquisar neste projeto de pesquisa. Em particular, propõe-se obter novos resultados para uma família de modelos dinâmicos proposta por Santos et al. (2010), denominada *Non-Gaussian State Space Model* (NGSSM). Esta abordagem possibilita o tratamento de séries temporais que extrapolam as restrições descritas acima e é uma generalização dos resultados apresentados por Smith & Miller (1986), que definem um modelo dinâmico com equação de evolução exata para qualquer série temporal com distribuição exponencial e às transformações um a um dessas séries, permitindo assim a integração analítica dos estados e a obtenção da verossimilhança preditiva.

Santos et al. (2010) apresentaram a NGSSM e as equações de evolução exata, com a restrição de que apenas a componente de nível da série seja estocástica, ou seja, as demais componentes (tendência, sazonalidade, ciclicidade e ponto de mudança) são determinísticas, e portanto, seus efeitos podem ser capturados no modelo por meio de covariáveis.

Após a proposta de Santos et al. (2010) apresenta-se um conjunto considerável de questões que devem ser investigadas a fim de se avaliar os métodos adequados de estimação dos parâmentros dos modelos desta família, avaliar a real contribuição desta família para o universo de aplicações práticas em séries temporais, bem como avaliar as possíveis extensões desta família. Desta forma, esta pesquisa tem como finalidade responder algumas perguntas que devem ser formuladas para a melhor compreensão sobre esta nova família de modelos. As principais questões são:

1. Quais são as distribuições de probabilidade que estão contidas nesta família de distribuições? Em especial, quais são as distribuições de probabilidade de caudas pesadas que estão contidas nesta família de distribuições?

2. Quais são os métodos de inferência (clássico e bayesianos) adequados e mais eficientes para estimar os parâmetros dos modelos da NGSSM?

3. Quais os estimadores intervalares mais adequados?

4. Quais os refinamentos necessários aos métodos de estimação que apresentam resultados insatisfatórios?

5. Quais são as séries temporais, e em que área do conhecimento, em que a modelagem por meio da NGSSM apresentam resultados melhores do que os demais

modelos já propostos na literatura?

Com a finalidade de contribuir de maneira efetiva com o desenvolvimento da ciência, e em particular com uma melhor compreensão sobre esta nova família de modelos proposta por Santos et al. (2010), este trabalho propõe-se obter respostas para os questionamentos apresentados acima. Desta forma, pode-se estabelecer os seguintes objetivos geral e específicos a serem atingidos na pesquisa:

Objetivo Geral

Ampliar o conhecimento sobre os NGSSM quanto às distribuições que estão contidas, quanto aos métodos de estimação dos parâmetros e quanto a sua aplicabilidade a conjuntos de dados reais.

Objetivos Específicos

1. Desenvolver novos casos particulares para a NGSSM;
2. Implementar em Ox os casos particulares já existentes e os em desenvolvimento da NGSSM e gerar séries temporais desta família de distribuições;
3. Implementar os estimadores clássicos e bayesianos para os parâmetros da NGSSM;
4. Avaliar o comportamento do estimador de máxima verossimilhança (MLE);
5. Avaliar o comportamento dos estimadores bayesianos;
6. Propor uma função de penalização para a função de verossimilhança e avaliar o comportamento do estimador de máxima verossimilhança penalizado (PMLE);
7. Propor e avaliar o comportamento de métodos bootstrap e intervalos bootstrap;
8. Avaliar as aplicações desta família em conjunto de dados reais em que esta família apresente resultados melhores que os demais modelos existentes na literatura.

Esta tese contém três artigos que ampliam os conhecimentos sobre uma nova família de modelos de espaços de estados proposta por Santos et al. (2010) denominada non-Gaussian state space model (NGSSM). Esta família de modelos é muito interessante porque, apesar de conter um conjunto significativo de distribuições de probabilidade, tem-se a função de verossimilhança analiticamente, e por consequência há a possibilidade de realizar inferência sobre os parâmetros sem a necessidade de métodos numéricos aproximados, como o filtro de partícula.

Este trabalho em sua Parte I tem-se uma revisão de literatura:

- No Capítulo 2 tem-se uma revisão dos conceitos básicos sobre processos estocásticos e series temporais.
- No Capítulo 3 tem-se uma revisão dos conceitos e definições de classes de distribuições de caudas pesadas e outliers.
- No Capítulo 4 apresenta-se os modelos de espaços de estados gaussianos básicos uma introdução dos modelos de espaços de estados não Gaussianos.

Em sua Parte II tem-se três artigos desenvolvidos que abordam os questionamentos descritos anteriormente nesta seção e apresentam respostas às mesmas.

- No Capítulo 5 tem-se o primeiro artigo intitulado Modelling Volatility Using State Space Models with Heavy Tailed Distributions. Neste artigo demonstra-se que outras cinco distribuições de causas pesadas também são casos particulares da NGSSM, além das distribuições Weibull e Pareto propostas por Santos et al. (2010). As distribuição são: Log-normal, Log-gama, Fréchet, Lévy, Skew GED. Para avaliação dos estimadores clássicos e bayesianos para os sete modelos de caudas pesadas são realizadas simulação Monte Carlo e os resultados demonstram que os estimadores são não viesados assintoticamente e consistentes. Ainda neste artigo, os modelos de caudas pesadas são estimados para as séries dos índices de bolsas de valores da América com maior índice de negociabilidade, são eles

$S\&P500$, $NASDAQ$, $IBOVESPA$, $INMEX$, $MERVAL$, $IPSA$. Os resultados estimados para as distribuições de caudas pesadas da NGSSM são comparados com modelos da família GARCH e verifica-se que o modelo Weibull da NGSSM apresenta melhores resultados para todas as séries estudadas.

- No Capítulo 6 tem-se o segundo artigo intitulado Penalized Likelihood for a Non Gaussian State Space Model Considering Heavy Tailed Distributions. Neste artigo propõe-se novos estimadores para os parâmetros dos modelos de caudas pesadas da NGSSM quando o modelo é estimado para séries temporais com poucas observações. Este estimador proposto tem por finalidade corrigir preventivamente o viés do estimador de máxima verossimilhança observado empiricamente, por meio de simulação Monte Carlo, para séries temporais pequenas. Observa-se que o parâmetro ω é sempre sobreestimado, independentemente do modelo de cauda pesada e do algorítmo de maximização utilizados. A obtenção de um estimador adequado para o parâmetro ω é excencial à qualidade do ajuste do modelo, bem como sua utilidade prática, uma vez que quando este parâmetro é sobreestimado a variabilidade das séries temporais é subestimada. Funções de penalização para a função de verossimilhança são propostas e, por consequência, estimadores de máxima verossimilhança penalizada são propostos e suas propriedades são avaliadas por meio de simulação Monte Carlo. Os resultados demonstram que os estimadores propostos apresentam uma redução significativa do viés em relação ao observado pelo estimador de máxima verossimilhança.

- No Capítulo 7 tem-se o terceiro artigo intitulado Bootstrapping Non Gaussian State Space Models. Neste artigo é avaliado o comportamento do intervalo de confiança assintótico dos parâmetros dos modelos de caudas pesadas quando as séries são pequenas. Observa-se que os intervalos de confiança para o parâmetro ω são inadequados, seja utilizando o estimador de máxima verossimilhança ou o estimador de máxima verossimilhança penalizado proposto no segundo artigo no Capítulo 6. Em razão disto são propostos intervalos de confiança bootstrap e suas propriedades são avaliadas por meio de simulação Monte Carlo. Os resultados demonstram que o intervalo de confiança bootstrap com correção de viés obtido a partir do bootstrap paramétrico apresentam taxas de cobertura muito próximas da taxa nominal utilizada no estudo empírico.

Parte I

Revisão de Literatura

Capítulo 2

Conceitos de Processos Estocásticos e Séries Temporais

Os diversos modelos apresentados na literatura utilizados para descrever séries temporais são processos estocásticos, ou seja, processos controlados por leis probabilísticas. Para uma melhor compreensão dos conceitos que serão abordados sobre os modelos de espaços de estados para séries temporais faz-se necessário apresentar algumas definições básicas da teoria de probabilidades, dentre as quais os conceitos de elemento aleatório, vetor aleatório, processo estocástico e séries temporais.

A definição 1.1, dada por Shiryaev (1989), define formalmente um elemento aleatório.

Definição 1.1. Seja $(\Omega, \mathcal{F})$ e $(E, \mathcal{E})$ espaços mensuráveis. Diz-se que uma função $Y = Y(\omega)$, definida em Ω e assume valores em E, é $\mathcal{F}/\mathcal{E} - \mathit{mensurável}$ ou é um elemento aleatório se $\{\omega : Y(\omega) \in B\} \in \mathcal{F}$, para todo $B \in \mathcal{E}$.

Para o caso particular em que $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a definição de elemento aleatório é a mesma de variável aleatória. Na literatura $\mathcal{B}(\mathbb{R})$ é conhecida como a $\sigma - \acute{a}lgebra$ de Borel.

Para o caso particular em que $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$o elemento aleatório $Y(\omega)$ é um ponto aleatório e pode ser representado por $Y(\omega) = (Y_1(\omega), \cdots, Y_n(\omega))$, onde $Y_k = \pi_k \circ X$ se π_k é a projeção de $\mathbb{R}^n$ na $k-ésima$ coordenada do eixo. Portanto, para $B \in \mathcal{B}(\mathbb{R})$ e desde que $\mathbb{R}\times\cdots\times\mathbb{R}\times B\times\mathbb{R}\times\cdots\times\mathbb{R} \in \mathcal{B}(\mathbb{R}^n)$, tem-se que $\{\omega : Y_k(\omega) \in B\} = \{\omega : Y_1(\omega) \in \mathbb{R}, \cdots, Y_{k-1}(\omega) \in \mathbb{R}, Y_k(\omega) \in B, Y_{k+1}(\omega) \in \mathbb{R}, \cdots, Y_n(\omega) \in \mathbb{R}\}$, o que implica que $\{\omega : Y_k(\omega) \in B\} = \{\omega : Y(\omega) \in (\mathbb{R}\times\cdots\times\mathbb{R}\times B\times\mathbb{R}\times\cdots\times\mathbb{R}) \in\} = \mathcal{F}$.

A definição 1.2, dada por Shiryaev (1989), define formalmente um vetor aleatório.

Definição 1.2. Um conjunto ordenado $(Y_1(\omega), \cdots, Y_n(\omega))$ de variáveis aleatórias é denotado por vetor aleatório $n - dimensional$.

Apropriando-se desta definição tem-se que $Y(\omega) = Y_1(\omega), \cdots, Y_n(\omega)$ com valores em $\mathbb{R}^n$ é um vetor aleatório $n - dimensional$, portanto se $B_k \in \mathcal{B}(\mathbb{R})$, $k = 1, \cdots, n$, então:

$$\{\omega : Y(\omega) \in B_1 \times \cdots \times B_{k-1} \times B_k \times B_{k+1} \times \cdots \times B_n\} = \prod_{k=1}^{n} \{\omega : Y_k(\omega) \in B_k\} \in \mathcal{F}.$$

Para o caso particular em que $(E, \mathcal{E}) = \left(\mathbb{R}^T, \mathcal{B}\left(\mathbb{R}^T\right)\right)$, onde o tempo T é um sub conjunto da reta real, o elemento aleatório $Y = Y(\omega)$ pode ser apresentado como $Y = (Y_t)_{t\in T}$ com $Y_t = \pi_t \circ X$, e é denotado por uma função aleatória com domínio do tempo T.

A definição 1.3, dada por Shiryaev (1989), define formalmente um processo estocástico.

Definição 1.3. Seja T um subconjunto da reta real, o vetor aleatório $Y = (Y_t)_{t\in T}$ é denotado por processo aleatório ou processo estocástico com domínio do tempo T.

Pode-se entender um processo estocástico como uma família de variáveis aleatórias com índices extraídos de um subconjunto T. Para o caso particular em que $T = \{1, 2, \cdots\}$ denota-se $T = \{Y_1, Y_2, \cdots\}$ por um processo estocástico com tempo discreto.

Para o caso particular em que $T = [0, 1], (-\infty, +\infty), [-\infty, +\infty), \cdots$, denota-se $Y = (Y_t)_{t\in T}$ por um processo estocástico com tempo contínuo.

Neste trabalho os modelos apresentados e desenvolvidos serão de processos estocásticos com tempo discreto.

Ressalta-se ainda que um processo estocástico $Y = (Y_t)_{t \in T} = Y = (Y_t(\omega))_{t \in T}$ é função de duas variáveis, do tempo $t \in T$ e de ω. Para um tempo t fixado, tem-se apenas uma variável aleatória.

Para ω fixado, a definição 1.4, dada por Shiryaev (1989), define formalmente uma série temporal.

Definição 1.4. Seja $Y = (Y_t)_{t \in T}$ um processo estocástico. Para cada $\omega \in \Omega$ fixado, a função $(Y_t(\omega))_{t \in T}$ é denotado por uma realização, ou uma trajetória, ou ainda uma série temporal do processo estocástico correspondente ao resultado ω.

Neste trabalho denotar-se-á as séries temporais $(Y_t(\omega))_{t \in T}$ por Y_t^1, Y_t^2, e assim por diante, para $t \in T$, para o processo estocástico $(Y_t)_{t \in T}$.

Pode-se entender uma série temporal como o conjunto de obervações para análise, ou seja, é uma parte da trajetória ou uma realização do processo dentre as muitas ou não enumeráveis realizações que poderiam ter sido observadas.

Em algumas áreas do conhecimento (Agronomia e Física, por exemplo), pode-se desenvolver experimentos que permitem observar algumas realizações do processo estocástico, ou seja, tem-se repetições do mesmo processo para análise.

Em diversas áreas do conhecimento (Economia e Astrologia, por exemplo), na maioria das vezes não é possível fazer experimentações. Esta limitação restringe ao pesquisador a observação de apenas uma única realização do processo, ou seja, tem-se apenas uma série temporal para análise.

Tem-se a especificação de um processo estocástico quando se conhece as funções de distribuição finito dimensionais do processo. Shiryaev (1989) a define por:

Definição 1.5. Seja $Y = (Y_t)_{t \in T}$ um processo estocástico. A medida de probabilidade P_Y em $\left(\mathbb{R}^T, \mathcal{B}\left(\mathbb{R}^T\right)\right)$ é $P_Y = P\left\{\omega : Y(\omega) \in B\right\}, B \in \mathcal{B}\left(\mathbb{R}^T\right)$, e é denotada por distribuição de probabilidade de Y. As probabilidades $P_{t_1, \cdots, t_n} \equiv P\left\{\omega : (Y_{t_1}, \cdots, Y_{t_n}) \in B\right\}$

com $t_i \in T$, $t_1 < t_2 < \cdots < t_n$, são denotadas por probabilidades finito dimensionais. As funções $F_{t_1,\cdots,t_n}(Y_1, \cdots, Y_n) \equiv P\{\omega : Y_{t_1} \leq y_1, \cdots, Y_{t_n} \leq y_n\}$ com $t_i \in T$, $t_1 < t_2 < \cdots < t_n$, são denotadas por funções de distribuições finito dimensionais.

Apropriando-se desta definição para $n = 1$, tem-se a distribuição *unidimensional* da variável aleatória $Y = Y_{t_1}$, $t_1 \in T$, para $n = 2$, tem-se a distribuição *bidimensional* da variável aleatória $Y = (Y_{t_1}, Y_{t_2})$, $t_1, t_2 \in T$, para $n = k$, tem-se a distribuição $k - dimensional$ da variável aleatória $Y = (Y_{t_1}, Y_{t_2}, \cdots, Y_{t_k})$, $t_1, t_2, \cdots, t_k \in T$.

Capítulo 3

Classe de Distribuições de Caudas Pesadas e *Outliers*

Neste capítulo, será apresentada as classificações das distribuições de probabilidades, encontradas na literatura, em relação as caudas e as suas relações com a propensão ou resistência a ocorrência de *outliers*.

3.1 Classes de distribuições de caudas pesadas

A definição da classe de distribuições de caudas pesadas está intrinsecamente associada ao comportamento das caudas da distribuição de probabilidade, mais especificamente, associada à velocidade do decaimento a zero da cauda da distribuição em relação à velocidade do decaimento a zero da cauda da distribuição exponencial, que apresenta um decaimento rápido.

A discussão sobre estas classes baseiam-se na cauda da direita da distribuição de probabilidade, entretanto, pode-se estender os resultados para a cauda a esquerda. Denota-se-á por $f(\bullet)$ a função de densidade, $F(\bullet)$ a função de distribuição, onde $F(\bullet) < 1$, para todo y finito, $F(\infty) = 1$, $\overline{F}(\bullet) = 1 - F(\bullet)$ a função relacionada à cauda a direita da distribuição e $\widehat{F}(\bullet)$ a função geradora de momento, onde $\widehat{F}(s) = \int_{-\infty}^{+\infty} e^{-sy} dF(y)$.

A função de densidade e/ou a função relacionada à cauda a direita da distribuição, de todas as distribuições, citadas neste trabalho são apresentadas em Embrechts et al. (1997) e/ou em Casella & Berger (2002).

A característica principal, que inclusive define as distribuições de caudas pesadas, é a de não apresentar função geradora de momentos. Para uma melhor compreensão desta característica faz-se necessário, inicialmente, definir a classe de distribuições de cauda leve.

Definição 2.1. Diz-se que uma função de distribuição F pertence à classe de distribuições de cauda leve a direita se para algum $\varepsilon > 0$ tem-se que $\overline{F}(y) = O(e^{-\varepsilon y})$, ou seja, $limsup_{y\to\infty} \frac{\overline{F}(y)}{e^{-\varepsilon y}} < \infty$.

Santana (2008) demonstra a relação entre o comportamento da cauda de uma função de distribuição com a existência da função geradora de momentos por meio da proposição a seguir.

Proposição 2.1. Seja a função de distribuição F com função geradora de momento $\widehat{F}\hat{}$, então $\overline{F}(y) = O(e^{-\varepsilon y})$ para algum $\varepsilon > 0$, se e somente se, $\widehat{F}(s)$ é finita para algum $s > 0$.

Demonstração. Inicialmente supõe-se que $\overline{F}(y) = O(e^{-\varepsilon y})$ para algum $\varepsilon > 0$, então existe $M > 0$, $y_0 > 0$, tal que, para todo $y \geq y_0$,$\left|\overline{F}(y)\right| \leq Me^{-\varepsilon y}$. Assim, para $0 < s < \varepsilon$, tem-se que

$$\widehat{F}(s) = \int_0^\infty P\left(e^{-\varepsilon y} > y\right) dy = \int_0^{e^{sy_0}} \overline{F}\left(\frac{ln(y)}{s}\right) dy + \int_{e^{sy_0}}^\infty \overline{F}\left(\frac{ln(y)}{s}\right) dy \leq$$

$$\leq e^{sy_0} + \int_{e^{sy_0}}^\infty Me^{-\frac{\varepsilon}{s} ln(y)} dy \leq e^{\varepsilon y_0} + \int_{e^{sy_0}}^\infty Mye^{-\frac{\varepsilon}{s}} dy = e^{\varepsilon y_0} + M\frac{s}{\varepsilon - s} e^{-\varepsilon y_0}.$$

Portanto, tem-se que $\widehat{F}(s) < \infty$ para $0 < s < \varepsilon$. Supõe-se agora que $\widehat{F}(s) < \infty$ para algum $s > 0$, então pela desigualdade de Chebyschev tem-se que

$$\overline{F}(y) = P(Y > y) = P\left(e^{\varepsilon Y} > e^{\varepsilon y}\right) \leq \frac{E\left(e^{\varepsilon Y}\right)}{e^{\varepsilon y}} = \frac{\widehat{F}(s)}{e^{\varepsilon y}} < \infty.$$

Logo, tem-se que $limsup_{y\to\infty}\frac{\overline{F}(y)}{e^{-\varepsilon y}} \leq \widehat{F}(s) < \infty$, e, portanto, conclui-se que $\overline{F}(y) = O\left(e^{-\varepsilon y}\right)$.■

A partir da Proposição 2.1 e pela Definição 2.1, pode-se concluir que as distribuições de cauda leve têm função geradora de momento. Logo, algumas distribuições de probabilidade conhecidas, que por terem função geradora de momento, estão contidas nesta classe, tais como[1]: Bernoulli, Binomial, Uniforme Discreta e Contínua, Geométrica, Hipergeométrica, Binomial Negativa, Poisson, Beta, Gama (Qui Quadrado e Exponencial por serem casos particulares), Exponencial Dupla, Logística, Weibull (restrito ao parâmetro $\gamma \geq 1$).

A classe de distribuições de caudas pesadas é definida pela função relacionada a cauda à direita da distribuição e $\widehat{F}(\bullet)$ não ser um $O\left(e^{-\varepsilon y}\right)$ e por conseqüência não ter função geradora de momentos finita, portanto, as distribuições de probabilidade que enquadram-se nesta situação não têm função geradora de momentos definidas.

Segue a definição formal da classe de distribuições de cauda pesada.

Definição 2.2. Diz-se que uma função de distribuição F pertence à classe de distribuições de cauda pesada à direita se a função geradora de momentos não é finita, ou seja, $\widehat{F}(s) = \infty$, para todo $s > 0$. (notação: $F \in \mathcal{K}$)

A partir da Definição 2.2 pode-se elencar algumas distribuições de probabilidade conhecidas, que por não terem função geradora de momentos, estão contidas nesta classe, tais como[2]: Loggama, Lognormal, Pareto, t-Student, F-Snedecor, Cauchy e as

[1]Segundo Casella & Berger (2002) as distribuições de probabilidade citadas têm função geradora de momentos.

[2]Segundo Embrechts et al. (1997) as distribuições do Valor Extremo não têm função geradora de momentos e segundo Casella & Berger (2002) as demais distribuições de probabilidade citadas não têm função geradora de momentos.

distribuições do Valor Extremos dos tipos I, II e III – Gumbel, Fréchet e Weibull (restrito a $0 < \gamma < 1$), respectivamente.

Embrechts et al. (1997) apresentam algumas propriedades específicas das distribuições de probabilidade que estão contidas na classe de distribuições de cauda pesada e, baseado nestas propriedades específicas, classificam-as nas seguintes classes: classe de cauda longa, classe subexponencial, classe de variação regular e a classe de variação dominada.

3.1.1 A classe de distribuições de cauda longa

Esta classe apresenta denominações distintas na literatura, Embrechts et al. (1997) a denomina classe de distribuição de cauda longa e Teugels (1975) a denomina classe de distribuição de variação lenta. Neste trabalho utilizar-se-á a primeira denominação, uma vez que a segunda denominação será utilizada posteriormente para outra classe de distribuições. A Definição 2.3 referente a classe de cauda longa é baseada em Embrechts & Godie (1980).

Definição 2.3. Diz-se que uma função de distribuição F pertence à classe de distribuições de cauda longa se $lim_{y\to\infty}\frac{\overline{F}(y-x)}{\overline{F}(y)} = 1$, para todo $y \in \mathbb{R}$, $x \in \mathbb{R}^{+}$. (notação: $F \in \mathcal{L}$)

3.1.2 A classe de distribuições subexponencial

A classe de distribuições subexponencial foi introduzida por Chystiakov (1964) e Chover et al. (1972). É a classe mais conhecida e explorada na literatura, dentre as classes de cauda pesada, em razão de sua maior aplicabilidade nas diversas áreas do conhecimento por conter distribuições de probabilidade adequadas à modelagem de dados de problemas reais. A definição da classe subexponencial apresentada a seguir é baseada em Goldie & Klüppelberg (1998).

Definição 2.4. Sejam $(Y_j)_{j\in\mathbb{N}}$ variáveis aleatórias positivas, independentes e identicamente distribuídas com função de distribuição F, e $\overline{F^{*n}}(y) = 1 - F^{*n} = P(Y_1 + \cdots + Y_n > y)$ a cauda da $n - ésima$ convolução de F. Diz-se que uma função de distribuição F pertence à classe de distribuições subexponencial se uma das duas condições equivalentes ocorrer: (notação: $F \in \mathcal{S}$)

1. $lim_{y\to\infty} \frac{\overline{F^{*n}}(y)}{\overline{F}(y)} = n, \forall y \in \mathbb{R}^+, n \geq 2$;
2. $lim_{y\to\infty} \frac{P(Y_1+\cdots+Y_n>y)}{P(max(Y_1+\cdots+Y_n>y))} = 1, \forall y \in \mathbb{R}^+, n \geq 2$.

Embrechts & Godie (1980) demonstram que ambas as condições apresentadas na definição são equivalentes, Embrechts et al. (1997) cita a Pareto, Burr, Loggama, Weibull, Lognormal, Benktander tipo I, Benktander tipo II, "Quase" Exponencial, as distribuições estáveis truncadas como distribuições pertencentes a esta classe e Junior (2007) cita além das anteriores a Cauchy.

Teugels (1975), Embrechts & Godie (1980), Klüppelberg (1988), Embrechts et al. (1997), Yakymiv (1997), Goldie & Klüppelberg (1998), Junior (2007) e Santana (2008), dentre vários outras publicações, apresentam uma vasta discussão sobre propriedades e aplicações da classe de distribuições subexponencial.

3.1.3 A classe de distribuições de variação regular

Junior (2007) cita trabalhos anteriores para apresentar uma definição para a classe de distribuições de cauda de variação regular baseada na função de densidade. Também apresenta outra definição baseada na função relacionada a cauda à direita da distribuição $\overline{F}$, mas diferente da apresentada por Embrechts et al. (1997), e denota a classe por cauda de variação regular estendida. A definição da classe de variação regular apresentada a seguir é baseada em Embrechts et al. (1997).

Definição 2.5. Diz-se que uma função de distribuição F em $(0, \infty)$ pertence à classe de distribuições de cauda de variação regular se existir α, onde $0 \leq \alpha < \infty$ tal que $lim_{y\to\infty} \frac{\overline{F}(yx)}{\overline{F}(y)} = x^{-\alpha}, \forall y \in \mathbb{R}, x \in \mathbb{R}^+$. (notação: $F \in \mathcal{R}$)

Se $\overline{F} \in \mathcal{R}_{-\alpha}$ diz-se que a função relacionada à cauda a direita da distribuição $\overline{F}$ é de variação regular com expoente, ou $\alpha - variante$ no infinito.

Há dois casos particulares importantes nesta classe. O primeiro caso é estabelecido para $\alpha = 0$, assim $\overline{F} \in \mathcal{R}_0$ e denota-se a classe de distribuição por cauda de variação lenta. Neste caso tem-se que o $lim_{y\to\infty}\frac{\overline{F}(yx)}{\overline{F}(y)} = 1$. O segundo caso é estabelecido para $\alpha = \infty$, assim $\overline{F} \in \mathcal{R}_{-\infty}$ e denota-se a classe de distribuição por cauda de variação rápida. Neste caso tem-se que se $x > 1$ o $lim_{y\to\infty}\frac{\overline{F}(yx)}{\overline{F}(y)} = 0$, e se $0 < x < 1$ o $lim_{y\to\infty}\frac{\overline{F}(yx)}{\overline{F}(y)} = \infty$.

Embrechts et al. (1997) e Bingham et al. (1987) apresentam algumas propriedades e aplicações desta classe de distribuições, Embrechts et al. (1997) citam a Pareto, Burr, Loggama, Weibull e as distribuições estáveis truncadas como distribuições pertencentes a esta classe e Junior (2007) cita além das anteriores a Cauchy.

3.1.4 A classe de distribuições de variação dominada

A definição da classe de cauda de variação dominada apresentada a seguir é baseada em Santana (2008).

Definição 2.6. Diz-se que uma função de distribuição F pertence à classe de distribuições de variação dominada se $lim_{y\to\infty}\frac{\overline{F}(yx)}{\overline{F}(y)} < \infty, \forall y \in \mathbb{R}, x \in (0,1)$. (notação: $F \in \mathcal{D}$)

Embrechts et al. (1997) e Junior (2007) apresentam a definição para a classe consiterando um caso particular, sem perda de generalidade, onde $x = \frac{1}{2}$, conseqüentemente, faz-se necessário que $lim_{y\to\infty}\frac{\overline{F}(\frac{y}{2})}{\overline{F}(y)} < \infty$.

Embrechts et al. (1997) demonstram que a classe de distribuições de cauda de variação regular está contida nesta classe.

3.1.5 Relações entre as classes de distribuições de cauda pesada

Embrechts & Omey (1984) e Klüppelberg (1988) demonstram em detalhes as relações que seguem abaixo entre as classes de distribuições de cauda pesada:

1. $\mathcal{R} \subset \mathcal{S} \subset \mathcal{L} \subset \mathcal{K}$ e $\mathcal{R} \subset \mathcal{D}$;
2. $\mathcal{L} \cap \mathcal{D} \subset \mathcal{S}$;
3. $\mathcal{D} \nsubseteq \mathcal{S}$ e $\mathcal{S} \nsubseteq \mathcal{D}$;
4. $\mathcal{S} \neq \mathcal{L}$;

onde $\mathcal{R}$ é a classe de cauda de variação regular, $\mathcal{S}$ é a classe subexponencial, $\mathcal{L}$ é a classe de cauda longa, $\mathcal{K}$ é a classe de cauda pesada e $\mathcal{D}$ é a classe de cauda de variação dominada.

Junior (2007) apresenta duas relações adicionais em decorrência de definir distribuições de cauda de variação regular e de cauda de variação regular estendida:

1. $\mathcal{R} \subset \mathcal{R}_{estendida}$;
2. $\mathcal{R}_{estendida} \subset \mathcal{D}$.

3.2 Distribuições resistentes e propensas a *outliers*

Utilizar-se-á as definições de distribuições resistentes a outliers e distribuições propensas a *outliers* estabelecidas por Neyman & Scott (1971). Estas definições, segundo Green (1974), são aplicáveis à famílias de distribuições e não à distribuições individualmente. Foram também demonstradas por Green (1976) algumas relações entre as definições e as funções relativas à cauda da família de distribuições as densidades da família de distribuições.

3.2.1 Distribuições resistentes a *outliers*

Seguem as definições de distribuições absolutamente e relativamente resistentes a *outliers* segundo Neyman & Scott (1971), onde considerar-se-á que $\{Y_n\}_{n\in\mathbb{N}}$ são variáveis aleatórias independentes e identicamente distribuidas e $\left\{Y_{(n)}\right\}_{n\in\mathbb{N}}$ as estatísticas de ordem.

Definição 2.7. Diz-se que uma função de distribuição F é absolutamente resistente a *outliers* – ARO se, para todo $\varepsilon > 0$, $lim_{n\to\infty} P\left(Y_{(n)} - Y_{(n-1)} > \varepsilon\right) = 0$.

Definição 2.8. Diz-se que uma função de distribuição F é relativamente resistente a *outliers* – RRO se, para todo $\varepsilon > 0$, $lim_{n\to\infty} P\left(\frac{Y_{(n)}}{Y_{(n-1)}} > \varepsilon\right) = 0$.

A interpretação natural destas definições é de que à medida que o tamanho da amostra, de uma variável aleatória proveniente de distribuições resistentes a *outliers* aumenta, espera-se que as observações maiores em magnitude estejam cada vez mais próximas entre si e, portanto, não se espera que ocorram *outliers*. Junior (2007) demonstra por meio de simulação da função distribuição empírica que a família de distribuição Normal é ARO e RRO. Há uma complexidade em avaliar se uma determinada família de distribuições é resistente a *outliers*, uma vez que as definições de Neyman & Scott (1971) estão baseadas na distribuição de $Y_{(n)} - Y_{(n-1)}$ e $\frac{Y_{(n)}}{Y_{(n-1)}}$. Em razão disto, Green (1976) apresentou e demonstrou dois teoremas que relacionam as definições às funções relativas às caudas da família de distribuições e um teorema que relaciona as definições à densidade da família de distribuições. Seguem os teoremas.

Teorema 2.1. Diz-se que uma função de distribuição F é absolutamente resistente a *outliers* – ARO se, e somente se, para todo $\varepsilon > 0$, $lim_{y\to\infty} \frac{\overline{F}(y+\varepsilon)}{\overline{F}(y)} = 0$.

Teorema 2.2. Diz-se que uma função de distribuição F é relativamente resistente a *outliers* – ARO se, e somente se, para todo $k > 1$, $lim_{y\to\infty} \frac{\overline{F}(ky)}{\overline{F}(y)} = 0$.

Teorema 2.3. Se a densidade f existe então a função de distribuição F é absolutamente resistente a *outliers* – ARO se a condição 1 é satisfeita e é relativamente resistente a *outliers* – RRO se a condição 2 é satisfeita. As condições são:

1. $lim_{y\to\infty} \frac{f(y+\varepsilon)}{f(y)} = 0$ para todo $\varepsilon > 0$;

2. $lim_{y\to\infty} \frac{f(ky)}{f(y)} = 0$ para todo $k > 1$.

Nos **exemplo 2.1** e **2.2** verificar-se-á, por meios dos **teoremas 2.1**, **2.2** e **2.3**, se as famílias de distribuições Exponencial e Normal são ARO e RRO.

Exemplo 2.1. Para a família de distribuição Exponencial tem-se que $\overline{F}\left(y|\,\lambda\right) = e^{-\lambda y} I_{y\geq 0}$, $\lambda > 0$. Logo, para $\varepsilon > 0$ e $k > 1$:

$$lim_{y\to\infty}\frac{\overline{F}\left(y+\varepsilon|\,\lambda\right)}{\overline{F}\left(y|\,\lambda\right)} = lim_{y\to\infty}\frac{e^{-\lambda(y+\varepsilon)}}{e^{-\lambda y}} = lim_{y\to\infty}e^{-\lambda(y+\varepsilon)+\lambda y} = lim_{y\to\infty}e^{-\lambda\varepsilon} = e^{-\lambda\varepsilon}\neq 0;$$

$$lim_{y\to\infty}\frac{\overline{F}\left(ky|\,\lambda\right)}{\overline{F}\left(y|\,\lambda\right)} = lim_{y\to\infty}\frac{e^{-\lambda ky}}{e^{-\lambda y}} = lim_{y\to\infty}e^{-\lambda ky+\lambda y} = lim_{y\to\infty}e^{-\lambda y(k-1)} = 0.$$

Portanto, conclui-se que a família de distribuição Exponencial não é ARO, mas é RRO. ■

Exemplo 2.2. Para a família de distribuição Normal tem-se a função de densidade $f\left(y|\,\mu,\sigma\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}y^2\right\} I_{-\infty<y<+\infty}$, $-\infty < \mu < +\infty$, $0 < \sigma^2 < +\infty$. Sem perda de generalidade, considerar-se-á $\mu = 0$. Logo, para $\varepsilon > 0$ e $k > 1$:

$$lim_{y\to\infty}\frac{f\left(y+\varepsilon|\,\mu,\sigma\right)}{f\left(y|\,\mu,\sigma\right)} = lim_{y\to\infty}\frac{exp\left\{-\frac{1}{2\sigma^2}\left(y+\varepsilon\right)^2\right\}}{exp\left\{-\frac{1}{2\sigma^2}y^2\right\}} = lim_{y\to\infty}exp\left\{-\frac{1}{2\sigma^2}\left(y+\varepsilon\right)^2 - y^2\right\}$$

$$= lim_{y\to\infty}exp\left\{-\frac{1}{2\sigma^2}\left(y^2+2y\varepsilon+\varepsilon^2-y^2\right)\right\} = lim_{y\to\infty}exp\left\{-\frac{2y\varepsilon+\varepsilon^2}{2\sigma^2}\right\} = 0;$$

$$lim_{y\to\infty}\frac{f\left(ky|\,\mu,\sigma\right)}{f\left(y|\,\mu,\sigma\right)} = lim_{y\to\infty}\frac{exp\left\{-\frac{1}{2\sigma^2}\left(ky\right)^2\right\}}{exp\left\{-\frac{1}{2\sigma^2}y^2\right\}} = lim_{y\to\infty}exp\left\{-\frac{1}{2\sigma^2}\left(ky\right)^2 - y^2\right\}$$

$$= lim_{y\to\infty} exp\left\{-\frac{k^2y^2 - y^2}{2\sigma^2}\right\} = lim_{y\to\infty} exp\left\{-y^2\frac{k^2-1}{2\sigma^2}\right\} = 0.$$

Portanto, conclui-se que a família de distribuição Normal é ARO e RRO.■

3.2.2 Distribuições propensas a *outliers*

Seguem as definições de distribuições absolutamente e relativamente propensas a *outliers* segundo Neyman & Scott (1971).

Definição 2.9. Diz-se que uma função de distribuição F é absolutamente propensas a *outliers* – APO se existirem $\varepsilon > 0$, $\delta > 0$, n_0 inteiro, tal que $lim_{n\to\infty} P\left(Y_{(n)} - Y_{(n-1)} > \varepsilon\right) \geq \delta$, para todo $n \geq n_0$.

Definição 2.10. Diz-se que uma função de distribuição F é relativamente propensas a *outliers* – RPO se existirem $\varepsilon > 0$, $\delta > 0$, n_0 inteiro, tal que $lim_{n\to\infty} P\left(\frac{Y_{(n)}}{Y_{(n-1)}} > \varepsilon\right) \geq \delta$, para todo $n \geq n_0$.

A interpretação natural destas definições é de que à medida que o tamanho da amostra, de uma variável aleatória proveniente de distribuições propensas a *outliers* aumenta, espera-se que haja observações maiores em magnitude que apresentem diferença significativa em relação às demais e, portanto, se espera que ocorra *outliers*.

Junior (2007) demonstra por meio de simulação da função distribuição empírica que a família de distribuição Cauchy é APO e RPO. Há uma complexidade em avaliar se uma determinada família de distribuições é propensa a *outliers* uma vez que as definições de Neyman & Scott (1971) estão baseadas na distribuição de$Y_{(n)} - Y_{(n-1)}$ e $\frac{Y_{(n)}}{Y_{(n-1)}}$. Em razão disto, Green (1976) apresentou e demonstrou dois teoremas que relacionam as definições às funções relativas às caudas da família de distribuições e um teorema que relaciona as definições à densidade da família de distribuições. Seguem os teoremas.

Teorema 2.4. Diz-se que uma função de distribuição F é absolutamente propensa a *outliers* – APO se, e somente se, existirem $\varepsilon > 0$, $\delta > 0$, tal que $\frac{\overline{F}(y+\varepsilon)}{\overline{F}(y)} \geq \delta$ para todo y finito.

Teorema 2.5. Diz-se que uma função de distribuição F é relativametne propensa a *outliers* – RPO se, e somente se, existirem $k > 1$, $\delta > 0$, tal que $\frac{\overline{F}(ky)}{\overline{F}(y)} \geq \delta$ para todo y finito.

Teorema 2.6. Se a densidade f existe então a função de distribuição F é absolutamente propensa a outliers – APO se a condição 1 é satisfeita e é relativamente resistente a outliers – RPO se a condição 2 é satisfeita. As condições são:

1. Existem $\varepsilon > 0$, $\delta > 0$ e y_0, tal que $\frac{f(y+\varepsilon)}{f(y)} \geq \delta$, para todo $y \geq y_0$;

2. Existem $k > 1$, $\delta > 0$ e y_0, tal que $\frac{f(ky)}{f(y)} \geq \delta$, para todo $y \geq y_0$.

Junior (2007) demonstra, por meio dos **teoremas 2.4**, **2.5** e **2.6**, que as famílias de distribuição Gama e Exponencial Dupla são APO, mas não são RPO, a Logística é APO e a distribuição t-Student é APO e RPO.

No **exemplos 2.3** e **2.4** verificar-se-á, por meios dos **teoremas 2.4**, **2.5** e **2.6**, se as famílias de distribuições Pareto e Weibull são APO e RPO.

Exemplo 2.3. Para a família de distribuição Weibull tem-se que $\overline{F}\,(y|\,\beta,\gamma) = e^{-\left(\frac{y}{\beta}\right)^{\gamma}} I_{y\geq 0}$, $\beta > 0$, $0 < \gamma < 1$. Logo:

$$\frac{\overline{F}\,(y+\varepsilon|\,\beta,\gamma)}{\overline{F}\,(y|\,\beta,\gamma)} = \frac{e^{-\left(\frac{y+\varepsilon}{\beta}\right)^{\gamma}}}{e^{-\left(\frac{y}{\beta}\right)^{\gamma}}} = e^{-\left(\frac{y+\varepsilon}{\beta}\right)^{\gamma}+\left(\frac{y}{\beta}\right)^{\gamma}} = e^{(\beta)^{-\gamma}[y^{\gamma}-(y+\varepsilon)^{\gamma}]}$$

$$\geq e^{\beta^{-1}[y-(y+\varepsilon)]} \geq e^{\frac{\varepsilon}{\beta}} \Rightarrow \frac{\overline{F}\,(y+\varepsilon|\,\beta,\gamma)}{\overline{F}\,(y|\,\beta,\gamma)} \geq \delta, \forall y \geq y_0;$$

$$\frac{\overline{F}\,(ky|\,\beta,\gamma)}{\overline{F}\,(y|\,\beta,\gamma)} = \frac{e^{-\left(\frac{ky}{\beta}\right)^{\gamma}}}{e^{-\left(\frac{y}{\beta}\right)^{\gamma}}} = e^{-\left(\frac{ky}{\beta}\right)^{\gamma}+\left(\frac{y}{\beta}\right)^{\gamma}} = e^{(1-k^{\gamma})\left(\frac{y}{\beta}\right)^{\gamma}}$$

$$\Rightarrow lim_{y\to\infty}\frac{\overline{F}\left(ky|\,\beta,\gamma\right)}{\overline{F}\left(y|\,\beta,\gamma\right)}=0.$$

Portanto, conclui-se que a família de distribuição Weibull é APO, mas não é RPO.■

Exemplo 2.4. Para a família de distribuição de Pareto tem-se que $f\left(y|\,\alpha,\beta\right)=\frac{\beta\alpha^{\beta}}{y^{\beta+1}}I_{y\geq\alpha}$, $\alpha,\beta>0$. Logo, existem $\varepsilon>0$, $\delta>0$, $k>1$ e y_0, tal que:

$$\frac{\overline{F}\left(y+\varepsilon|\,\alpha,\beta\right)}{\overline{F}\left(y|\,\alpha,\beta\right)}=\frac{\frac{\beta\alpha^{\beta}}{(y+\varepsilon)^{\beta+1}}}{\frac{\beta\alpha^{\beta}}{y^{\beta+1}}}=\frac{y^{\beta+1}}{\left(y+\varepsilon\right)^{\beta+1}}=\left(\frac{y+\varepsilon}{y}\right)^{-(\beta+1)}=\left(1+\frac{\varepsilon}{y}\right)^{-(\beta+1)}$$

$$\Rightarrow lim_{y\to\infty}\frac{\overline{F}\left(y+\varepsilon|\,\alpha,\beta\right)}{\overline{F}\left(y|\,\alpha,\beta\right)}=1\Rightarrow lim_{y\to\infty}\frac{\overline{F}\left(y+\varepsilon|\,\alpha,\beta\right)}{\overline{F}\left(y|\,\alpha,\beta\right)}\geq\delta,\forall y\geq y_0;$$

$$\frac{\overline{F}\left(ky|\,\alpha,\beta\right)}{\overline{F}\left(y|\,\alpha,\beta\right)}=\frac{\frac{\beta\alpha^{\beta}}{(ky)^{\beta+1}}}{\frac{\beta\alpha^{\beta}}{y^{\beta+1}}}=\frac{y^{\beta+1}}{\left(ky\right)^{\beta+1}}=k^{-(\beta+1)}\geq\delta,\forall y\geq y_0.$$

Portanto, conclui-se que a família de distribuição de Pareto é APO e RPO. ■

3.2.3 Classificação das distribuições de probabilidade relacionada a sensibilidade a *outliers*

Green (1976) propõe uma classificação em classes das distribuições de probabilidade relacionada à sua resistência/propensão, absoluta/relativa à outliers e classifica algumas distribuições. As classes são:

- **Classe I** Distribuições que são ARO e RRO (Normal, por exemplo);
- **Classe II** Distribuições que é RRO, mas não é ARO (Poisson, por exemplo);
- **Classe III** Distribuições que são APO e RRO;
- **Classe IV** Distribuições que são APO, mas não é RRO (Gama, por exemplo);

- **Classe V** Distribuições que são APO e RPO (Cauchy, por exemplo);

- **Classe VI** Distribuições que não são APO nem RPO.

Chapter 4

Modelos de Espaços de Estados

O MEE apresenta duas denominações na literatura – modelo estrutural (abordagem clássica) e modelo linear dinâmico – MLD (abordagem bayesiana).

A idéia central destes modelos é a de decompor a série temporal $Y = \{Y_t\}_{t \in T}$ em componentes não observáveis determinísticas ou estocásticas. Pode-se elencar como as principais componentes que compõem uma série temporal:

1. Nível (μ_t): refere-se ao piso ou nível que a série se desenvolve ao longo do tempo;

2. Tendência (β_t): refere-se ao sentido que a série se desenvolve, seja de crescimento ou decrescimento, ao longo do tempo;

3. Sazonalidade (γ_t): refere-se a padrões semelhantes recorrentes de baixa e média periodicidade que uma série temporal apresenta ao longo do tempo. A periodicidade é normalmente semanal, mensal, trimestral, quadrimestral ou anual;

4. Ciclicidade (δ_t): refere-se a padrões semelhantes recorrentes de alta periodicidade que uma série temporal apresenta ao longo do tempo. A periodicidade pode ser em alguns anos ou décadas;

5. Erro ou distúrbio (ε_t): refere-se a componente estocástica.

Desta forma, a série pode ser definida por meio da equação

$$Y_t = \mu_t + \beta_t + \gamma_t + \delta_t + \varepsilon_t \tag{4.1}$$

onde supõe-se que $\varepsilon_t \sim \left(0, \sigma_\varepsilon^2\right)$ e são independentes entre si.

4.1 Origem dos modelos de espaços de estados

Os primeiros trabalhos que surgiram na literatura, com o objetivo de decompor a série temporal em componentes não observáveis (especificamente para o nível, tendência e sazonalidade), foram desenvolvidos por Holt (1957), com a proposição das técnicas de alisamento exponencial de uma série temporal e Winters (1960), que estende as técnicas de alisamento exponencial e as aplica à previsão de vendas de curto prazo.

Kalman (1960) e Kalman & Bucy (1961) introduziram o MEE para solucionar problemas reais na engenharia, pressupondo que as componentes não observáveis evoluíam no tempo de acordo com um processo linear Markoviano e que a componente estocástica tem distribuição gaussiana.

Nas próximas três seções seguintes serão apresentados alguns modelos particulares que estão contidos no MEE e em seguida a representação formal e geral do MEE.

4.2 Modelo de tendência linear local – MTL

O modelo de tendência linear local é também denotado na literatura como modelo linear dinâmico de segunda ordem. Este modelo é o MNL com a inserção de uma componente de tendência.

A característica básica deste modelo é a presença de uma componente de tendência estocástica β_t, ou seja, a tendência da série pode variar ao longo do tempo t.

Esta característica propicia uma flexibilidade importante, pois torna o modelo mais geral, e portanto, gerador de um conjunto maior de séries temporais. Desta forma, pode-se inferir que o MTL explica melhor e um conjunto maior de séries temporais reais que apresentam mudanças em seu nível e em sua tendência ao longo do tempo.

O MTL é dado por

$$y_t = \mu_t + \varepsilon_t,\, \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \tag{4.2}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t,\, \eta_t \sim N\left(0, \sigma_\eta^2\right), \tag{4.3}$$

$$\beta_t = \beta_{t-1} + \xi_t,\, \xi_t \sim N\left(0, \sigma_\xi^2\right), \tag{4.4}$$

para $t = 1, \ldots, n$, onde μ_t é o nível não observado no tempo t, β_t é a tendência não observada no tempo t, ε_t é o distúrbio das observações no tempo t, η_t é o distúrbio do nível no tempo t, ξ_t é a componente aleatória da tendência denotada por erro ou distúrbio da tendência no tempo t.

Assume-se que ε_t, η_t, ξ_t são não correlacionados e são normalmente distribuídos com média zero e variâncias constantes σ_ε^2, σ_η^2 e σ_ξ^2, respectivamente.

A equação 4.2 é a equação das observações e as equações 4.3 e 4.4 são as equações dos estados[1].

Commandeur & Koopman (2007) ressaltam a vantagem do MTL em modelar a tendência de séries temporais, por apresentar uma componente de tendência estocástica, em relação a um modelo de regressão clássico, que apresenta uma componente determinística.

4.3 Modelo estrutural básico – MEB

Há séries que apresentam algum tipo de periodicidade recorrente, por exemplo, ano a ano, portanto, estas séries apresentam altas correlações em defasagens de tempo sazonais.

[1]Equações de nível e tendência, respectivamente.

O modelo estrutural básico é o MTL com a inserção de uma componente sazonal estocástica β_t, ou seja, a sazonalidade da série, se existir, é captada no modelo e pode variar ao longo do tempo t. Esta característica do modelo permite uma maior adequação às séries temporais que apresentam periodicidade recorrente.

O período sazonal, denotado por s, pode ser semanal para dados diários ($s = 7$), mensal para dados diários ($s = 30$), trimestral ou quadrimestral para dados mensais ($s = 3$, $s = 4$), ou, mais comumente, mensal para dados anuais ($s = 12$).

Harvey (1989) apresenta duas maneiras de se modelar a sazonalidade. Na primeira, equação 4.8, a componente sazonal é representada por variáveis *dummy* e na segunda, equação 4.9, a componente sazonal é representada por funções trigonométricas.

O MEB é dado por

$$y_t = \mu_t + \beta_t + \varepsilon_t,\ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \tag{4.5}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t,\ \eta_t \sim N\left(0, \sigma_{\eta^2}\right), \tag{4.6}$$

$$\beta_t = \beta_{t-1} + \xi_t,\ \xi_t \sim N\left(0, \sigma_\xi^2\right), \tag{4.7}$$

$$\beta_t = \sum_{j=1}^{s-1} \beta_{t-j} + w_t,\ w_t \sim N\left(0, \sigma_w^2\right), \tag{4.8}$$

ou

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{t,j}, \tag{4.9}$$

onde

$$\begin{bmatrix} \gamma_t \\ \gamma_t^* \end{bmatrix} = \rho \begin{bmatrix} cos\lambda_c & sen\lambda_c \\ -sen\lambda_c & cos\lambda_c \end{bmatrix} + \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} w_t \\ w_t^* \end{bmatrix},$$

$0 \leq \rho < 1$, $\lambda_c = \lambda_j = \frac{2\pi j}{s}$, $j = 1,2,\ldots,[s/2]$. Para $t = 1,\ldots,n$, onde ψ_t é um ciclo, μ_t é o nível não observado no tempo t, β_t é a tendência não observada no tempo t, β_t é a sazonalidade não observada no tempo t, ε_t é o distúrbio das observações no tempo t,

η_t é o distúrbio do nível no tempo t, ξ_t é o distúrbio da tendência no tempo t e w_t é o distúrbio da sazonalidade no tempo t.

Assume-se que ε_t, η_t, ξ_t e w_t são não correlacionados e são normalmente distribuídos com média zero e variâncias constantes σ_{ε^2}, σ_η^2, σ_ξ^2 e σ_w^2, respectivamente.

A equação 4.5 é a equação das observações e as equações 4.6, 4.7, 4.8 e 4.9 são as equações dos estados[2].

Na Tabela 4.1 abaixo segue uma síntese dos modelos de espaços de estados apresentados anteriormente bem como outros três modelos destinados a modelagem de ciclicidade não detalhados anteriormente neste trabalho.

Commandeur & Koopman (2007) apresentam outras formulações dos modelos de espaços por meio da inserção de covariáveis na equação de observação e/ou nas equações de estados, entretanto estas formulações não serão apresentadas e detalhadas neste trabalho.

4.4 Modelo de espaços de estados – MEE

O MEE é muito flexível e permite representar várias estruturas para séries temporais, tais como incorporar variáveis explicativas, funções ou variáveis indicadores para a inclusão de quebra estrutural, componentes de tendência, sazonalidade, ciclicidade, estruturas não lineares e não gaussianas, dentre outras.

O MEE univariado[3] é dado por

$$\mathbf{y_t} = \mathbf{Z_t'}\alpha_\mathbf{t} + \mathbf{d_t} + \varepsilon_\mathbf{t},\ \varepsilon_\mathbf{t} \sim N\left(0, \mathbf{H_t}\right), \tag{4.10}$$

$$\alpha_\mathbf{t} = \mathbf{T_t}\alpha_\mathbf{t-1} + \mathbf{c_t} + \mathbf{R_t}\eta_\mathbf{t},\ \eta_\mathbf{t} \sim N\left(0, \mathbf{Q_t}\right), \tag{4.11}$$

para $t = 1, \ldots, n$, onde $\varepsilon_\mathbf{t}$ é o vetor $n \times 1$ dos distúrbios das observações, no tempo t e

[2]Equações de nível, tendência e sazonalidade, respectivamente.
[3]Representação do MEE extraída de Harvey (1989).

Table 4.1: Modelos de espaços de estados

$^{(M)}$**MODELO** $^{(C)}$**COMPONENTE**	**ESPECIFICAÇÃO**
$^{(C)}$Passeio aleatório	$\mu_t = \mu_{t-1} + \eta_t$
$^{(C)}$Passeio aleatório$_{com\,drift}$	$\mu_t = \mu_{t-1} + \beta + \eta_t$
$^{(M)}$Nível Local	$Y_t = \mu_t + \varepsilon_t$ $\mu_t = \mu_{t-1} + \eta_t$
$^{(C)}$Tendência estocástica	$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$ $beta_t = beta_{t-1} + \xi_t$
$^{(M)}$Tendência Linear Local	$y_t = \mu_t + varepsilon_t$ $\mu_t = \mu_{t-1} + beta_{t-1} + \eta_t$ $beta_t = beta_{t-1} + \xi_t$
$^{(C)}$Ciclo estocástico	$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} cos\lambda_c & sen\lambda_c \\ -sen\lambda_c & cos\lambda_c \end{bmatrix} + \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} {}_t \\ {}^*_t \end{bmatrix}$ ψ_t é o ciclo, $0 \leq \rho < 1$ e $0 \leq \lambda_c <$ p
$^{(M)}$Ciclo	$y_t = \mu + \psi_t + varepsilon_t$, ψ_t é o ciclo estocástico
$^{(M)}$Tendência e Ciclo	$y_t = \mu_t + \psi_t + varepsilon_t,$ $\mu_t = \mu_{t-1} + beta_{t-1} + \eta_t$ $beta_t = beta_{t-1} + \xi_t$
$^{(M)}$Tendência Cíclica	$y_t = \mu_t + varepsilon_t,$ $\mu_t = \mu_{t-1} + \psi_{t-1} + beta_{t-1} + \eta_t$ $beta_t = beta_{t-1} + \xi_t$
$^{(C)}$Ciclo não estacionário	$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} cos\lambda_c & sen\lambda_c \\ -sen\lambda_c & cos\lambda_c \end{bmatrix} + \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} {}_t \\ {}^*_t \end{bmatrix}$ ψ_t é o ciclo, $\rho = 1$ e $\lambda_c = \lambda_j = \frac{2}{s}$, $j = 1{,}2, \ldots, [s/2]$
$^{(C)}$Sazonalidade$_{variável\,dummy}$	$beta_t = \sum_{j=1}^{s-1} beta_{t-j} +_t$
$^{(C)}$Sazonalidade$_{função\,trigonométrica}$	$beta_t = \sum_{j=1}^{[s/2]} beta_{t,j}$ $beta_t$ é o ciclo, $\rho = 1$ e $0 \leq \lambda_c <$ p
$^{(M)}$Estrutural Básico	$y_t = \mu_t + beta_t + varepsilon_t,$ $\mu_t = \mu_{t-1} + beta_{t-1} + \eta_t$ $beta_t = beta_{t-1} + \xi_t$ $beta_t = \sum_{j=1}^{s-1} beta_{t-j} +_t$ ou $beta_t = \sum_{j=1}^{[s/2]} beta_{t,j}$

Fonte: Adaptado de Harvey (1989).

$\eta_{\mathbf{t}}$ é o vetor $g \times 1$ dos distúrbios do estado, no tempo t.

A equação 4.10 é a equação das observações e a equação 4.11 é a equação dos estados.

Assume-se que $\varepsilon_{\mathbf{t}}$ e $\eta_{\mathbf{t}}$ são não correlacionados e são normalmente distribuídos com média zero e variâncias constantes $\mathbf{H_t}$ e matriz de covariâncias constantes $\mathbf{Q_t}$, respectivamente.

As matrizes do sistema $\mathbf{Z_t}$, $\mathbf{T_t}$ e $\mathbf{R_t}$, de ordens $n \times m$, $m \times m$ e $m \times g$, respectivamente, são determinísticas e conhecidas, entretanto, podem apresentar elementos desconhecidos que podem ser estimados.

A matriz $\mathbf{Z_t}$ desempenha papel semelhante ao da matriz de desenho no modelo de regressão da variável independente, a matriz $\mathbf{T_t}$ é denotada por matriz de evolução do estado.

O $\alpha_{\mathbf{t}}$ é o vetor $m \times 1$ de estados ou vetor de sistema do modelo, $\mathbf{d_t}$ e $\mathbf{c_t}$, de ordens $n \times 1$ e $m \times 1$, são covariáveis inseridas nas equações de observações e de estado, respectivamente. Segundo Harvey (1989), em geral, os elementos de $\alpha_{\mathbf{t}}$ são não observáveis, entretanto, pressupõe-se que sejam gerados a partir de um processo de Markov de primeira ordem.

O MEE tem como pressupostos que o vetor de estado inicial $\alpha_{\mathbf{0}} \sim N\left(\mathbf{a_0}, \mathbf{P_0}\right)$ e que $\varepsilon_{\mathbf{t}}$ e $\eta_{\mathbf{t}}$ são não correlacionados entre si e não correlacionados com o estado inicial, ou seja, $E\left(\varepsilon_{\mathbf{t}}\eta_{\mathbf{s}}^{'}\right) = 0$, $E\left(\varepsilon_{\mathbf{t}}\alpha_{\mathbf{0}}^{'}\right) = 0$ e $E\left(\eta_{\mathbf{t}}\alpha_{\mathbf{0}}^{'}\right) = 0$, para todo $t, s = 1, \ldots, n$.

Diz-se que o MEE é invariante no tempo ou homogêneo no tempo quando $\mathbf{Z_t}$, $\mathbf{T_t}$, $\mathbf{R_t}$, $\mathbf{d_t}$, $\mathbf{c_t}$, $\mathbf{H_t}$ e $\mathbf{Q_t}$ são constantes no tempo. Um caso particular desse tipo de modelo são os modelos estacionários. Para este modelo Harvey (1989) apresenta ainda o tratamento de dados faltantes, o tratamento para séries observadas em tempo contínuo, o tratamento para séries quando não há periodicidade nas observações, ou seja, há irregularidade temporal das observações, bem como o MEE multivariado.

4.4.1 Representação do MNL pelo MEE

O MNL pode ser facilmente representado pelo MEE definindo-se as quantidades

$$\mathbf{Z_t'} = 1, \alpha_\mathbf{t} = \mu_t, \mathbf{d_t} = 0, \varepsilon_\mathbf{t} = \varepsilon_t, \mathbf{H_t} = \sigma_\varepsilon^2,$$

$$\mathbf{T_t} = 1, \mathbf{c_t} = 0, \mathbf{R_t} = 1, \eta_\mathbf{t} = \eta_t, \mathbf{Q_t} = \sigma_\eta^2.$$

4.4.2 Representação do MTL pelo MEE

O MTL pode ser representado pelo MEE definindo-se as quantidades

$$\mathbf{Z_t'} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \alpha_\mathbf{t} = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}, \mathbf{d_t} = 0, \varepsilon_\mathbf{t} = \varepsilon_t, \mathbf{H_t} = \sigma_\varepsilon^2,$$

$$\mathbf{T_t} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \mathbf{c_t} = 0, \mathbf{R_t} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \eta_\mathbf{t} = \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}, \mathbf{Q_t} = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}.$$

Desta forma tem-se que

$$\mathbf{y_t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t,\ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right),$$

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}, \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}\right).$$

4.5 Modelos de Espaços de Estados Não-Gaussianos

Nelder & Wedderburn (1972) propuseram a Famlia de Modelos Lineares Generalizados (MLG), propiciando a unificação em uma classe de vários modelos já existentes de forma isolada. A idéia central desses modelos consiste em permitir que se tenha várias opções para a distribuição da variável-resposta, permitindo ainda que a mesma pertença a família exponencial de distribuições, e por consequências todas as boas propriedades desta família.

No contexto de séries temporais, a estrutura de correlação das observações não pode ser desprezada. Nesse sentido, uma estrutura mais geral, denominada por Modelos Lineares Dinâmicos Generalizados (MLDG), foi proposta por West et al. (1985), gerando a partir de então um significativo interesse nestes modelos devido à sua aplicabilidade em diversas áreas do conhecimento.

Vários trabalhos foram publicados sobre estes modelos, dentre os quais pode-se citar o de Gamerman & West (1987), Grunwald et al. (1993), Fahrmeir (1987), Fruhwirth-Schnatter (1994), Lindsey & Lambert (1995), Gamerman (1991), Gamerman (1998), Chiogna & Gaetan (2002), Hemming & Shaw (2002) e Godolphin & Triantafyllopoulos (2006).

Há na literatura ainda outros trabalhos que tratam de modelos para séries temporais não-gaussianas que não estão sob os MLDG, dentre os quais pode-se citar o de Smith (1979), Smith (1981), Cox (1981), Smith & Miller (1986), Kaufmann (1987), Kitagawa (1987), Harvey & Fernandes (1989), Shephard & Pitt (1997), Jorgensen et al. (1999) e Durbin & Koopman (2000).

O problema com essas classes de modelos é sua tratabilidade analítica que é facilmente perdida, mesmo para componentes muito simples. Assim, a verossimilhança preditiva, que é fundamental para o processo de inferência, pode apenas ser obtida de forma aproximada. Portanto, a NGSSM proposta por Santos et al. (2010) tem como principal vantagem em relação aos trabalhos citados acima a tratabilidade analítica, onde as equações de evolução e a função de verossimilhança preditiva são exatas.

Part II

Artigos Científicos

Chapter 5

Modelling Volatility Using State Space Models with Heavy Tailed Distributions

Frank M. de Pinho[a], **Glaura C. Franco**[b], **Ralph S. Silva**[c]
[a]*IBMEC, Belo Horizonte, Brasil*
[b]*Universidade Federal de Minas Gerais, Belo Horizonte, Brasil*
[c]*Universidade Federal do Rio de Janeiro, Belo Horizonte, Brasil*

Abstract

This article deals with a non-Gaussian state space model (NGSSM), which is a generalization of the results in Smith & Miller (1986). The NGSSM is attractive because the likelihood can be analytically computed, thus avoiding the use of highly demanding computational algorithms such as the particle filter in order to make inference on the parameters. The paper focuses on stochastic volatility models in the NGSSM, where the observation equation is modelled with a heavy tailed distribution such as Log-normal, Log-gamma and Fréchet. Parameter estimation can be accomplished either using classical or Bayesian procedures and a simulation study shows that both methods lead to satisfactory results. In a real data application, the proposed stochastic volatility models in the NGSSM are compared with the autoregressive conditionally heteroscedastic and stochastic volatility models using South and North American stock price indexes.

Keyword: Bayesian and Classical Inference, Heavy Tailed Distributions, Non-Gaussian State Space Model, Stochastic Volatility, Stock price index.

5.1 Introduction

The global financial crisis has generated a significant instability in the prices of financial assets and particularly in the stock market. For this reason, a major concern among economists, fund managers and investment researchers is how long this crisis will impact the variability of asset prices. For this reason, researches focusing on the study and modeling of volatility has been intensified in the last few years.

Relying on the fact that the unconditional distribution of daily returns has fatter tails than the normal distribution, the usual time series models that assume normality and homoscedasticity are not appropriate to model volatility. Thus, more adequate procedures, especially the ones presenting conditional variance evolving on time, have been proposed. The most known approaches are the ones concerning conditional heteroscedastic models, such as ARCH (Engle, 1982), GARCH (Bollerslev, 1986), EGARCH (Nelson, 1991), TGARCH (Zakoian, 1994) and multivariate GARCH (Bauwens et al., 2006).

Taylor (1986) proposed the first stochastic volatility model, where the volatility is a stochastic function of the past volatility. Several studies on this approach have been developed, such as Melino & Turnbull (1990), Taylor (1994), Harvey et al. (1994), Jacquier et al. (1994), Eraker et al. (2003) and Raggi & Bordignon (2006).

Recently, a non Gaussian state space model was proposed by Santos et al. (2010). This procedure is a generalization of a result of Smith & Miller (1986), who proposed an exponential observation model with an exact evolution equation for the state. The work of Santos et al. (2010) allows for analytical computation of the marginal likelihood, which increases the applicability of the model and enables its use with a wide class of distributions for observational time series. Additionally, this model allows the relaxation of the normality and heteroscedasticity assumptions.

According to Tsay (2005), one of the main characteristics of volatility is that it evolves over time in a continuous way and it always varies within a fixed range. This means that volatility is often stationary. Due to the structure used in the model proposed by Santos et al. (2010), the only stochastic component is the level of the series, and it is built in a way similar to the local level model of Harvey (1989). Thus, the model is highly recommended to be applied to stationary series. Any other component, such as seasonality or structural breaks should be inserted as covariates.

There are some recent contributions in the literature that employ the state space approach to handle nonlinear and non Gaussian time series. Some examples are the works of Shephard (1994), extended by Deschamps (2011) for Bayesian estimation, that uses a local scale procedure for modeling volatility. Ferrante & Vidoni (1998) and Vidoni (1999) introduce non-linear and non Gaussian state space models with analytic updating recursions for filtering and prediction.

Thus, the purpose of this work is to present new models in the non-Gaussian state space family that can be used to model volatility. Among them, there is the class of heavy tailed distributions, much employed in the volatility literature, as in the works of Anderson (2001) and Chib et al. (2002). The models introduced here comprise the Log-normal, Log-gamma, Fréchet, Lévy and the Generalized Error Distribution (GED). In addition, the Pareto and Weibull models, already considered in Santos et al. (2010), are also presented.

Monte Carlo results for Bayesian and classical methods of inference in the estimation of the non-Gaussian state space model are performed for the distributions cited above. Additionally, the NGSSM addressed here is used to model the most known stock exchange indexes in North and South America and the fits are compared to the classical generalized autoregressive conditional heteroscedasticity (see GARCH; Bollerslev, 1986) models.

The paper is organized as follows. Section 6.2 defines the NGSSM and presents the inference procedures. Section 5.3 shows how to write the heavy tailed distributions cited above in the NGSSM form. Section 6.4 shows the results of the Monte Carlo simulation studies and Section 5.5 presents an application of heavy tailed models in the NGSSM to estimate the volatility of several stock exchange indexes. Section 6.5 concludes the work.

5.2 A non-Gaussian state space model

Santos et al. (2010) define a new family of non-Gaussian state space models, which is a generalization of the works of Smith & Miller (1986) and Harvey & Fernandes (1989). Let $\{y_t\}_{t=1}^n$ be a time series with probability function given by

$$p\left(y_t\middle|\,\mu_t,\boldsymbol{\varphi}\right)=\begin{cases} q\left(y_t,\boldsymbol{\varphi}\right)\mu_t^{r(y_t,\boldsymbol{\varphi})}\exp\left\{-\mu_t s\left(y_t,\boldsymbol{\varphi}\right)\right\},y_t\in H\left(\boldsymbol{\varphi}\right)\subset\mathbb{R} \\ 0,\ \text{otherwise,}\end{cases} \tag{5.1}$$

where n is the sample size, $\boldsymbol{\varphi}$ is a p-dimensional parameter vector, $\boldsymbol{\varphi}=(\varphi_1,\ldots,\varphi_p)'$, and functions $q\left(y_t,\boldsymbol{\varphi}\right)$, $r\left(y_t,\boldsymbol{\varphi}\right)$, $s\left(y_t,\boldsymbol{\varphi}\right)$ and $H\left(\boldsymbol{\varphi}\right)$ are such that $p\left(y_t\middle|\,\mu_t,\boldsymbol{\varphi}\right)\geqslant 0$ and the Lebesgue-Stieltjes integral $\int p\left(y_t\middle|\,\mu_t,\boldsymbol{\varphi}\right)dy_t=1$. If $r\left(y_t,\boldsymbol{\varphi}\right)=r\left(\boldsymbol{\varphi}\right)$, $s\left(y_t,\boldsymbol{\varphi}\right)=s\left(\boldsymbol{\varphi}\right)$ and $H\left(\boldsymbol{\varphi}\right)$ is a constant function (it does not depend on $\boldsymbol{\varphi}$), the distribution family becomes a special case of the exponential family.

The NGSSM considers $\{y_t\}_{t=1}^n$ following the distribution in equation 5.1 with the state given by

$$\mu_t=\lambda_t g\left(\boldsymbol{x}_t,\boldsymbol{\beta}\right),\quad\text{for}\quad t=1,\ldots,n,$$

where g is the link function, $\boldsymbol{x}_t$ is a vector of covariates and $\boldsymbol{\beta}$ (one of the components of $\boldsymbol{\varphi}$) is the regression coefficient vector. The dynamic level λ_t is given by the evolution equation $\lambda_t = \omega^{-1}\lambda_{t-1}\varsigma_t$, with the *prior* specification $\lambda_0 | \, Y_0 \sim \text{Gamma}\,(a_0; \, b_0)$. In this case, $\varsigma_t \sim \text{Beta}\,(\omega a_{t-1}, \, (1-\omega)\, a_{t-1})$, that is

$$\omega \, \frac{\lambda_t}{\lambda_{t-1}} \Bigg| \, \lambda_{t-1}, \boldsymbol{Y_{t-1}} \sim \text{Beta}\,(\omega a_{t-1}, \, (1-\omega)\, a_{t-1})\,, \quad \text{for} \quad t = 1, \ldots, n, \tag{5.2}$$

where $\boldsymbol{Y_{t-1}} = \{y_{t-1}, \ldots, y_1\}$ for $t \geqslant 1$, $0 < \omega < 1$ and Y_0 is the initial information. Parameter ω has the function of increasing multiplicatively the variance over time.

Taking the logarithm of the evolution equation, λ_t, it can be seen that it is the random walk equation used for the local level model (Harvey, 1989), that is

$$\ln(\lambda_t) = \ln(\lambda_{t-1}) + \xi_t,$$

where $\xi_t = \ln(\varsigma_t / \omega) \in \mathbb{R}$.

Theorem 1 in Santos et al. (2010) presents the equations for the exact evolution of the dynamic level and the predictive density function for the NGSSM, which are as follows.

1. The *prior* distribution $\lambda_t | \, \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi} \sim \text{Gamma}\left(a_{t|t-1}; \, b_{t|t-1}\right)$, where

$$a_{t|t-1} = \omega a_{t-1} \quad \text{and} \quad b_{t|t-1} = \omega b_{t-1}.$$

2. The *prior* distribution $\mu_t | \, \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi} \sim \text{Gamma}\left(c_{t|t-1}; \, d_{t|t-1}\right)$, where

$$c_{t|t-1} = \omega a_{t-1} \quad \text{and} \quad d_{t|t-1} = \omega b_{t-1} \left[g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right)\right]^{-1}.$$

   They are easily obtained from equation 5.1 and the scale property of the Gamma distribution.

3. The *posterior* distribution $\lambda_t = \mu_t \left[g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right)\right]^{-1} \Big| \boldsymbol{Y_t}, \boldsymbol{\varphi} \sim \text{Gamma}\left(a_t;\, b_t\right)$ where

$$a_t = a_{t|t-1} + r\left(y_t, \boldsymbol{\varphi}\right) \quad \text{and} \quad b_t = b_{t|t-1} + s\left(y_t, \boldsymbol{\varphi}\right) g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right).$$

4. The *posterior* distribution $\mu_t | \boldsymbol{Y_t}, \boldsymbol{\varphi} \sim \text{Gamma}\left(c_t;\, d_t\right)$, where

$$c_t = c_{t|t-1} + r\left(y_t, \boldsymbol{\varphi}\right) \quad \text{and} \quad d_t = d_{t|t-1} + s\left(y_t, \boldsymbol{\varphi}\right).$$

5. The predictive density function is given by

$$p\left(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi}\right) = \frac{\Gamma\left(r\left(y_t, \boldsymbol{\varphi}\right) + c_{t|t-1}\right) q\left(y_t, \boldsymbol{\varphi}\right) d_{t|t-1}^{c_{t|t-1}} I_{\left(y_t \in H(\boldsymbol{\varphi})\right)}}{\Gamma\left(c_{t|t-1}\right) \left[s\left(y_t, \boldsymbol{\varphi}\right) + d_{t|t-1}\right]^{r(y_t, \boldsymbol{\varphi}) + c_{t|t-1}}}. \tag{5.3}$$

5.2.1 Inference procedure

Parameter inference in the NGSSM can be performed either using classical or Bayesian procedures. Both are based on the likelihood function

$$L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right) = \prod_{t=1}^{n} p\left(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi}\right),$$

where $p\left(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi}\right)$ is given in equation 6.4.

- **Classical inference**

Classical inference for the parameters of the NGSSM is performed through maximum likelihood estimation. The log-likelihood function is calculated as

$$\begin{aligned}\ell\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right) = &\sum_{t=1}^{n} \ln \Gamma\left(r\left(y_t, \boldsymbol{\varphi}\right) + c_{t|t-1}\right) + \sum_{t=1}^{n} \ln\left(q\left(y_t, \boldsymbol{\varphi}\right)\right) - \sum_{t=1}^{n} \ln \Gamma\left(c_{t|t-1}\right) \\ &+ \sum_{t=1}^{n} c_{t|t-1} \ln\left(b_{t|t-1}\right) - \sum_{t=1}^{n} \left[r\left(y_t, \boldsymbol{\varphi}\right) + c_{t|t-1}\right] \ln\left[s\left(y_t, \boldsymbol{\varphi}\right) + d_{t|t-1}\right],\end{aligned}$$

where $a_0 > 0$ and $b_0 > 0$ (see Santos et al., 2010). Thus, the maximum likelihood estimator (MLE) for $\boldsymbol{\varphi}$ is given by

$$\hat{\varphi}_{ML} = \arg\max_{\boldsymbol{\varphi}} \ell\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right).$$

Due to the fact that $\ell\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right)$ is a nonlinear function of $\boldsymbol{\varphi}$, numerical procedures such as the BFGS algorithm proposed by Broyden (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970), should be used.

The asymptotic confidence interval for $\boldsymbol{\varphi}$ is built based on a numerical approximation for the Fisher information matrix $I_n(\boldsymbol{\varphi})$, using $I_n(\boldsymbol{\varphi}) \cong -G(\boldsymbol{\varphi})$, where $-G(\boldsymbol{\varphi})$ is the matrix of second derivatives of the log-likelihood function with respect to the parameters.

Let φ_i, $i = 1, \ldots, p$, be any component of $\boldsymbol{\varphi}$. Then, an asymptotic confidence interval of $100(1-\kappa)\%$ for φ_i is given by

$$\hat{\varphi}_i \pm z_{\kappa/2}\sqrt{\widehat{Var}(\hat{\varphi}_i)},$$

where $z_{\kappa/2}$ is the $\kappa/2$ percentile of the standard normal distribution and $\widehat{Var}(\hat{\varphi}_i)$ is obtained from the diagonal elements of the Fisher information matrix.

- **Bayesian inference**

The *posterior* distribution $\pi\left(\boldsymbol{\varphi} | \boldsymbol{Y_n}\right)$ of the parameter vector $\boldsymbol{\varphi}$ is given by

$$\pi\left(\boldsymbol{\varphi} | \boldsymbol{Y_n}\right) = \frac{L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right)\pi\left(\boldsymbol{\varphi}\right)}{\displaystyle\int L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right)\pi\left(\boldsymbol{\varphi}\right) d\boldsymbol{\varphi}},$$

where $L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right)$ is the likelihood function and $\pi\left(\boldsymbol{\varphi}\right)$ is the *prior* distribution for $\boldsymbol{\varphi}$. In this paper a proper and non informative Uniform distribution with respect to Bayes-Laplace is used. It is given by $\pi\left(\boldsymbol{\varphi}\right) = c$ for all possible values of $\boldsymbol{\varphi}$ in a determined range and 0 otherwise. The Bayesian estimates of the posterior mean (BE-Mean), the posterior median (BE-Median) and the credibility interval are obtained from a sample of the posterior distribution. The adaptive random walk Metropolis (ARWM) algorithm proposed by Roberts & Rosenthal (2009) (see also Haario et al., 2001) has been used to sample from the posterior distribution.

The ARWM works as follows. Suppose that given some initial $\boldsymbol{\varphi}_0$ from $\pi(\boldsymbol{\varphi}|\boldsymbol{Y}_n)$, the $j-1$ iterates $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{j-1}$ have been generated. The j^{th} iterate $\boldsymbol{\varphi}_j$ is generated from the proposal density $\eta_j(\boldsymbol{\varphi}|\boldsymbol{\psi})$ which may also depend on some other value of $\boldsymbol{\varphi}$ which is called $\boldsymbol{\psi}$. Let $\boldsymbol{\varphi}_j^p$ be the proposed value of $\boldsymbol{\varphi}_j$ generated from $\eta_j(\boldsymbol{\varphi}|\boldsymbol{\varphi}_{j-1})$. Then $\boldsymbol{\varphi}_j = \boldsymbol{\varphi}_j^p$ is taken with probability

$$\alpha(\boldsymbol{\varphi}_j^p, \boldsymbol{\varphi}_{j-1}) = \min\left\{1, \frac{\pi(\boldsymbol{\varphi}_j^p|\boldsymbol{Y}_j)}{\pi(\boldsymbol{\varphi}_{j-1})} \frac{\eta_j(\boldsymbol{\varphi}_{j-1}|\boldsymbol{\varphi}_j^p)}{\eta_j(\boldsymbol{\varphi}_j^p|\boldsymbol{\varphi}_{j-1})}\right\}, \tag{5.4}$$

and $\boldsymbol{\varphi}_j = \boldsymbol{\varphi}_{j-1}$ otherwise. In adaptive sampling the parameters of $\eta_j(\boldsymbol{\varphi}|\boldsymbol{\psi})$ are estimated from the iterates $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{j-2}$. Under appropriate regularity conditions the sequence of iterates $\boldsymbol{\varphi}_j, j \geqslant 1$, converges to draws from the target distribution $\pi(\boldsymbol{\varphi}|\boldsymbol{Y}_n)$. The proposal distribution in the ARWM algorithm used in this paper is given by a mixture of two normal distributions with mean components given by $\boldsymbol{\varphi}_{j-1}$. The first component has a small weight and a fixed covariance matrix while the second component has more weight, say 0.95, and a covariance matrix that is updated as iteration goes. For more details about the ARWM see Roberts & Rosenthal (2009) and Haario et al. (2001).

Credibility intervals for φ_i, $i = 1,...,p$ are built as follows. Given a value $0 < \kappa < 1$, the interval $[c_1,c_2]$ satisfying

$$\int_{c_1}^{c_2} \pi(\varphi_i \mid \boldsymbol{Y}_n)\, d\varphi_i = 1 - \kappa$$

is a credibility interval for φ_i with level $100(1-\kappa)\%$.

- **Model selection**

The adequacy of the model should be checked after fitting a model to a set of data. There are many methods of diagnosis suggested in the literature, and some of them are described below.

Harvey & Fernandes (1989) suggested a diagnosis method based on the standardized residuals, also known as Pearson residuals, which are defined as:

$$r_t^p = \frac{y_t - E\left(y_t \left| \mathbf{Y_{t-1}}, \boldsymbol{\varphi} \right.\right)}{\sqrt{Var\left(y_t \left| \mathbf{Y_{t-1}}, \boldsymbol{\varphi} \right.\right)}}.$$

The authors propose the following residual analysis:

1. Examine the plot of residuals vs. time and residuals vs. an estimate of the level component.

2. Verify if the sample variance of the standardized residuals is close 1. A value greater than 1 indicates overdispersion.

Another alternative is to use the deviance residuals (McCulagh & Nelder, 1989), which are given by:

$$r_t^d = \left\{ 2ln \left[\frac{p\left(y_t \left| y_t, \boldsymbol{\varphi} \right.\right)}{p\left(y_t \left| \hat{\phi}_t, \boldsymbol{\varphi} \right.\right)} \right] \right\}^{\frac{1}{2}},$$

where $\hat{\phi}_t = E\left(y_t \left| \mathbf{Y_{t-1}}, \boldsymbol{\varphi} \right.\right)$.

When two or more models present reasonable fits to the dta, it is necessary to choose one of them. According to Harvey (1989) the AIC and BIC criteria proposed, respectively, by Akaike (1974) and Schwarz (1978), are suitable procedures. They are defined by:

$$AIC = -2l\left(\hat{\boldsymbol{\varphi}}\right) + 2k$$

and

$$BIC = -2l(\hat{\varphi}) + 2k\ln(n),$$

where $l(\cdot)$ is the log-likelihood function, k the number of parameters and n the number of observations.

Hurvich & Tsai (1993) have proposed a correction in the AIC, called here AICc. Burnham & Anderson (2002) strongly recommend using AICc, rather than AIC, if n is small or k is large. The AICc criterion is defined by:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}.$$

5.3 Heavy tailed distributions in the NGSSM

In this section, some of the most used heavy tailed distributions, such as the Log-normal, Log-gamma, Fréchet, Lévy, Generalized Skew Normal (Skew GED), Pareto and Weibull, are discussed and they are proved to belong to the NGSSM.

The main characteristic of this kind of distribution is that it presents heavier tails than the normal distribution. The formal definition, found in Asmussen (2003), is as follows. A distribution function, F, of a random variable X belongs to the class of heavy right tail if $\lim_{x\to\infty} e^{\lambda x}[1-F(x)] = \infty$, for all $\lambda > 0$. This is equivalent to state that the moment generating function, $M_X(s)$, of F is infinite for all $s > 0$.

Teugels (1975), Embrechts et al. (1997) and Goldie & Klüppelberg (1998), among others, present a wide discussion about heavy tailed distribution properties and applications. Neyman & Scott (1971) and Green (1976) showed that there is a close relationship between the heavy tailed distribution family and the absolute or relative distribution outliers prone. That is, probability distributions that are contained in the heavy tailed distribution family are more propense to generate outliers.

5.3.1 Log-normal model

If a time series $\{y_t\}_{t=1}^{n}$ is generated from a Log-normal distribution with location parameter $\delta_t = \delta$, shape parameter $\gamma_t = \gamma$, unknown and invariant in time, and precision parameter σ_t^{-2}, restricted to $\sigma_t^{-2} = \mu_t > 0$ and $\gamma < y_t$, then

$$p\left(y_t | \mu_t, \boldsymbol{\varphi}\right) = \frac{\mu_t^{\frac{1}{2}}}{\left(y_t - \gamma\right)\sqrt{2\pi}} \exp\left\{-\mu_t \frac{\left[\ln\left(y_t - \gamma\right) - \delta\right]^2}{2}\right\} I_{(\gamma < y_t < \infty)},$$

where $\mu_t = \lambda_t g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi} = \left(\omega, \boldsymbol{\beta}, \delta, \gamma\right)'$.

The Log-normal model can be written in the NGSSM form as

$$q\left(y_t, \boldsymbol{\varphi}\right) = \left[\left(y_t - \gamma\right)\sqrt{2\pi}\right]^{-1}, \qquad r\left(y_t, \boldsymbol{\varphi}\right) = \frac{1}{2} \quad \text{and}$$

$$s\left(y_t, \boldsymbol{\varphi}\right) = \frac{\left[\ln\left(y_t - \gamma\right) - \delta\right]^2}{2}.$$

Thus the likelihood function $L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right) = \prod_{t=1}^{n} \left\{ \frac{\Gamma\left(\frac{1}{2} + c_{t|t-1}\right)\left[\left(y_t - \gamma\right)\sqrt{2\pi}\right]^{-1} d_{t|t-1}^{c_{t|t-1}} I_{(\gamma < y_t < \infty)}}{\Gamma\left(c_{t|t-1}\right)\left(d_{t|t-1} + \left[\ln\left(y_t - \gamma\right) - \delta\right]^2/2\right)^{\frac{1}{2} + c_{t|t-1}}} \right\}.$$

5.3.2 Log-gamma model

The Log-gamma distribution was presented by Consul & Jain (1971). If a time series $\{y_t\}_{t=1}^{n}$ is generated from a Log-gamma distribution with shape parameter $\alpha_t = \alpha$, unknown and invariant in time, and scale parameter $\alpha\mu_t$, restricted to $\alpha > 0$ and $\alpha\mu_t > 0$, then

$$p\left(y_t | \mu_t, \boldsymbol{\varphi}\right) = \frac{\left(\alpha\mu_t\right)^{\alpha}\left[\ln\left(y_t\right)\right]^{\alpha - 1}}{\Gamma\left(\alpha\right) y_t^{\alpha\mu_t + 1}} I_{(1 < y_t < \infty)},$$

where $\mu_t = \lambda_t g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi} = \left(\omega, \boldsymbol{\beta}, \alpha\right)'$.

The Log-gamma model can be written in the NGSSM form as

$$q\left(y_t,\boldsymbol{\varphi}\right)=\alpha^{\alpha}\left[\ln\left(y_t\right)\right]^{\alpha-1}\left[\Gamma\left(\alpha\right)y_t\right]^{-1},\quad r\left(y_t,\boldsymbol{\varphi}\right)=\alpha\quad\text{and}$$

$$s\left(y_t,\boldsymbol{\varphi}\right)=\alpha\ln\left(y_t\right).$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)=\prod_{t=1}^{n}\left\{\frac{\Gamma\left(\alpha+c_{t|t-1}\right)\alpha^{\alpha}\left[\ln\left(y_t\right)\right]^{\alpha-1}\left[\Gamma\left(\alpha\right)y_t\right]^{-1}d_{t|t-1}^{c_{t|t-1}}I_{(1<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(\alpha\ln\left(y_t\right)+d_{t|t-1}\right)^{\alpha+c_{t|t-1}}}\right\}.$$

5.3.3 Fréchet model

If a time series $\{y_t\}_{t=1}^{n}$ is generated from a Maximum Fréchet distribution with shape parameter $\alpha_t=\alpha$, location parameter $\gamma_t=\gamma$, unknown and invariant in time, and scale parameter μ_t^{α}, restricted to $\gamma<y_t$, $\alpha>0$ and $\mu_t^{\alpha}>0$, then

$$p\left(y_t|\,\mu_t,\boldsymbol{\varphi}\right)=\alpha\mu_t^{-1}\left(\frac{\mu_t}{y_t-\gamma}\right)^{\alpha+1}\exp\left\{-\left(\frac{\mu_t}{y_t-\gamma}\right)^{\alpha+1}\right\}I_{(\gamma<y_t<\infty)},$$

where $\mu_t^{\alpha}=\lambda_t g\left(\boldsymbol{x}_t,\boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi}=\left(\omega,\boldsymbol{\beta},\alpha,\gamma\right)'$.

The Maximum Fréchet model can be written in the NGSSM form as

$$q\left(y_t,\boldsymbol{\varphi}\right)=\alpha\left(y_t-\gamma\right)^{-\alpha-1},\quad r\left(y_t,\boldsymbol{\varphi}\right)=1\quad\text{and}\quad s\left(y_t,\boldsymbol{\varphi}\right)=\left(y_t-\gamma\right)^{-\alpha}.$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)=\prod_{t=1}^{n}\left\{\frac{\Gamma\left(1+c_{t|t-1}\right)\alpha\left(y_t-\gamma\right)^{-\alpha-1}\left(d_{t|t-1}\right)^{c_{t|t-1}}I_{(\gamma<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(\left(y_t-\gamma\right)^{-\alpha}+d_{t|t-1}\right)^{1+c_{t|t-1}}}\right\}.$$

The Minimum Fréchet model can be also easily written in the NGSSM form, just changing $(y_t-\gamma)$ for $(\gamma-y_t)$ and using the restriction $\gamma>y_t$ instead of $\gamma<y_t$.

5.3.4 Lévy model

If a time series $\{y_t\}_{t=1}^n$ is generated from a Lévy distribution with location parameter $\gamma_t = \gamma$, unknown and invariant in time, and precision parameter μ_t, restricted to $\mu_t > 0$ and $y_t > \gamma$, then

$$p\left(y_t \middle| \mu_t, \boldsymbol{\varphi}\right) = \frac{\mu_t^{\frac{1}{2}}}{\sqrt{2\pi\left(y_t - \gamma\right)^3}} \exp\left\{-\mu_t \left[2\left(y_t - \gamma\right)\right]^{-1}\right\} I_{(\gamma < y_t < \infty)},$$

where $\mu_t = \lambda_t g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi} = \left(\omega, \boldsymbol{\beta}, \gamma\right)'$.

The Lévy model can be written in the NGSSM form as

$$q\left(y_t, \boldsymbol{\varphi}\right) = \left[2\pi\left(y_t - \gamma\right)\right]^{-\frac{3}{2}}, \quad r\left(y_t, \boldsymbol{\varphi}\right) = \frac{1}{2} \quad \text{and} \quad s\left(y_t, \boldsymbol{\varphi}\right) = \left[2\left(y_t - \gamma\right)\right]^{-1}.$$

Thus the likelihood function $L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right) = \prod_{t=1}^{n} \left\{ \frac{\Gamma\left(\frac{1}{2} + c_{t|t-1}\right)\left[2\pi\left(y_t - \gamma\right)\right]^{-\frac{3}{2}} \left(d_{t|t-1}\right)^{c_{t|t-1}} I_{(\gamma < y_t < \infty)}}{\Gamma\left(c_{t|t-1}\right)\left(\left[2\left(y_t - \gamma\right)\right]^{-1} + d_{t|t-1}\right)^{\frac{1}{2} + c_{t|t-1}}} \right\}.$$

5.3.5 Skew GED model

The Generalized Skew Normal Distribution (Skew GED) is also known as the Skew Exponential Power Distribution. If a time series $\{y_t\}_{t=1}^n$ is generated from a Skew GED distribution with location parameter $\delta_t = \delta$, shape parameter $\alpha_t = \alpha$ and asymmetry parameter $\kappa_t = \kappa$, all of them unknown and invariant in time, and precision parameter μ_t, restricted to $\alpha > 0$, $\kappa > 0$ and $\mu_t > 0$, then

$$p\left(y_t \middle| \mu_t, \boldsymbol{\varphi}\right) = \frac{\kappa \mu_t^{\frac{1}{\alpha}}}{\Gamma\left(\alpha^{-1}\right)\left(1 + \kappa^2\right)} \exp\left\{-\mu_t \left\{\left[\kappa^{\alpha} z_t^{+}\right]^{\alpha} + \left[\kappa^{-\alpha} z_t^{-}\right]^{\alpha}\right\}\right\} I_{(-\infty < y_t < \infty)},$$

where $z_t = y_t - \delta$, $\mu_t = \lambda_t g(\boldsymbol{x}_t, \boldsymbol{\beta})$ and $\boldsymbol{\varphi} = (\omega, \boldsymbol{\beta}, \delta, \alpha, \kappa)'$,

$$u^+ = \begin{cases} u, \text{if } u \geqslant 0 \\ 0, \text{if } u < 0 \end{cases} \quad \text{and} \quad u^- = \begin{cases} -u, if\ u \leqslant 0 \\ 0, if\ u > 0 \end{cases}.$$

The Skew GED includes the Skew Normal distribution ($\alpha = 2$, $\kappa \neq 1$), the Normal distribution ($\alpha = 2$, $\kappa = 1$), the Skew Laplace distribution ($\alpha = 1$, $\kappa \neq 1$), the Laplace distribution ($\alpha = 1$, $\kappa = 1$) and the Uniform distribution ($\alpha \to \infty$).

The Skew GED model can be written in the NGSSM form as

$$q(y_t, \boldsymbol{\varphi}) = \frac{\kappa}{\Gamma(\alpha^{-1})(1+\kappa^2)}, \quad r(y_t, \boldsymbol{\varphi}) = \frac{1}{\alpha} \quad \text{and}$$

$$s(y_t, \boldsymbol{\varphi}) = \left[\kappa^{\alpha} z_t^+\right]^{\alpha} + \left[\kappa^{-\alpha} z_t^-\right]^{\alpha},$$

where $z_t = y_t - \delta$.

Thus the likelihood function $L(\boldsymbol{\varphi}; \boldsymbol{Y_n})$ is given by

$$L(\boldsymbol{\varphi}; \boldsymbol{Y_n}) = \prod_{t=1}^{n} \left\{ \frac{\Gamma\left(1/\alpha + c_{t|t-1}\right) \kappa \left[\Gamma\left(\alpha^{-1}\right)\left(1+\kappa^2\right)\right]^{-1} d_{t|t-1}^{c_{t|t-1}} I_{-\infty < y_t < \infty}}{\Gamma\left(c_{t|t-1}\right)\left(\left[\kappa^{\alpha} z_t^+\right]^{\alpha} + \left[\kappa^{-\alpha} z_t^-\right]^{\alpha} + d_{t|t-1}\right)^{1/\alpha + c_{t|t-1}}} \right\}.$$

For details about Skew GED random number generator see Ayebo & Kozubowski (2003).

5.3.6 Pareto model

If a time series $\{y_t\}_{t=1}^{n}$ is generated from a Pareto distribution with scale parameter μ_t, restricted to $y_t > 1$, then

$$p(y_t | \mu_t, \boldsymbol{\varphi}) = \mu_t y_t^{-\mu_t - 1} I_{(1 < y_t < \infty)},$$

where $\mu_t = \lambda_t g(\boldsymbol{x}_t, \boldsymbol{\beta})$ and $\boldsymbol{\varphi} = (\omega, \boldsymbol{\beta})'$.

The Pareto model can be written in the NGSSM form as

$$q\left(y_t,\boldsymbol{\varphi}\right)=y_t^{-1},\quad r\left(y_t,\boldsymbol{\varphi}\right)=1\quad\text{and}\quad s\left(y_t,\boldsymbol{\varphi}\right)=\ln\left(y_t\right).$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)=\prod_{t=1}^{n}\left\{\frac{\Gamma\left(1+c_{t|t-1}\right)y_t^{-1}d_{t|t-1}^{c_{t|t-1}}I_{(1<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(\ln\left(y_t\right)+d_{t|t-1}\right)^{1+c_{t|t-1}}}\right\}.$$

5.3.7 Weibull model

If a time series $\{y_t\}_{t=1}^n$ is generated from a Weibull distribution with location parameter $\upsilon_t=\upsilon$, unknown and invariant in time, and scale parameter μ_t, restricted to $\upsilon>0$, $\mu_t>0$ and $y_t>0$, then

$$p\left(y_t|\,\mu_t,\boldsymbol{\varphi}\right)=\upsilon\mu_t y_t^{\upsilon-1}\exp\left\{-\mu_t y_t^{\upsilon}\right\}I_{(0<y_t<\infty)},$$

where $\mu_t=\lambda_t g\left(\boldsymbol{x}_t,\boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi}=\left(\omega,\boldsymbol{\beta},\upsilon\right)'$.

The Weibull model can be written in the NGSSM form as

$$q\left(y_t,\boldsymbol{\varphi}\right)=\upsilon y_t^{\upsilon-1},\quad r\left(y_t,\boldsymbol{\varphi}\right)=1\quad\text{and}\quad s\left(y_t,\boldsymbol{\varphi}\right)=y_t^{\upsilon}.$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)=\prod_{t=1}^{n}\left\{\frac{\Gamma\left(1+c_{t|t-1}\right)\upsilon y_t^{\upsilon-1}d_{t|t-1}^{c_{t|t-1}}I_{(0<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(y_t^{\upsilon}+d_{t|t-1}\right)^{1+c_{t|t-1}}}\right\}.$$

5.4 Monte Carlo study

In this section the performance of the Log-normal, Log-gamma, Fréchet, Lévy, Skew GED, Pareto and Weibull models is evaluated through a Monte Carlo experiment, using the maximum likelihood estimator (MLE) and the Bayesian estimators (BE-Mean and BE-Median). Asymptotic confidence interval and credibility interval for the parameter vector are also presented and they are compared with respect to the coverage rate, for a fixed level of 95%.

The number of Monte Carlo replications was set equal to 1,000 for time series of size $n = \{100;\, 200;\, 500\}$, generated under the *prior* specification $\lambda_0 |\, Y_0 \sim \text{Gamma}\,(100.0;\, 1.0)$, with a covariate $x_t = \sin\,(2\pi t/12)$, $t = 1, \ldots, n$.

For all distributions $\beta = 1.0$ and $\omega = (0.90, 0.95)$ but only results for $\omega = 0.90$ are presented here, as they were very similar to the case $\omega = 0.95$.

Specific parameters were set as follows: Log-normal ($\delta = 5.0$), Log-gamma ($\alpha = 5.0$), Fréchet ($\alpha = 5.0$), Skew GED ($\delta = 5.0$, $\alpha = 1.5$, $\kappa = 1.0$) and Weibull ($\upsilon = 5.0$). For the Log-normal, Fréchet and Lévy models the parameter γ was fixed at 0.0. For the Skew GED model the parameter α was fixed at 1.5, thus, there is a distribution with a tail heavier than the Skew Normal ($\alpha = 2.0$) and lighter than the Skew Laplace (both are particular cases of the Skew GED).

To calculate the maximum likelihood estimator, the BFGS algorithm assumed, as initial state condition $\lambda_0 |\, Y_0 \sim \text{Gamma}\,(0.01;\, 0.01)$, $\omega_0 = 0.50$ and $\beta_0 = \delta_0 = \alpha_0 = \upsilon_0 = \kappa_0 = 0.01$.

For the Bayesian estimation using the ARWM algorithm, chains of size 20,000 were generated with burn in of 5,000. The Uniform $(-5{,}000; 5{,}000)$ and Uniform $(0; 10{,}000)$ are used as the *prior* distribution for the parameters that are defined in $\Re$ and $\Re^+$, respectively. More details about the initial conditions in the ARWM algorithm and the Bayesian approach are available from the authors upon request.

All codes for NGSSM were developed by the authors in OX Metrics.

5.4.1 Empirical distribution of the estimators

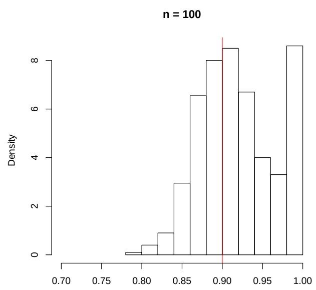
In this subsection, the empirical distribution of the MLE and Bayesian estimators for the parameters of the heavy tailed distribution in the NGSSM is investigated for time series of sizes $n = 100, 200, 500$. As the empirical distribution of the estimators for ω, β and the third parameter (δ for Log-normal and Skew GED, α for Log-gamma and Fréchet and υ for Weibull) is very similar for all models studied, only the results for the Log-normal model are presented here.

Figure 5.1 shows the empirical distribution based on 1,000 replications of the MLE, BE-Mean and BE-Median estimates for parameter ω. Series of small size shows an asymmetric behavior, always overestimating ω. It can be noted that the mode for the MLE is equal to 1.0. For larger series, the empirical distribution appears symmetric around the real value of the parameter. As expected, the variance decreases as the sample sizes increase.

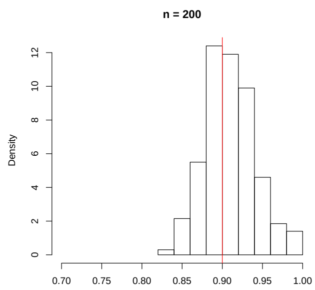
Figures 5.2 and 5.3 present the empirical distribution of the estimates of parameters β and δ, respectively, for the Log-normal model. The histograms are symmetric around the real value of the parameter for all sample sizes. For parameter δ, the MLE presents larger variability than the Bayesian estimators (this behavior only occurs in the Log-normal and Skew GED models). It can also be observed, as expected, that the variance of the estimates decreases with the increase of the sample size.

5.4.2 Point and interval estimation

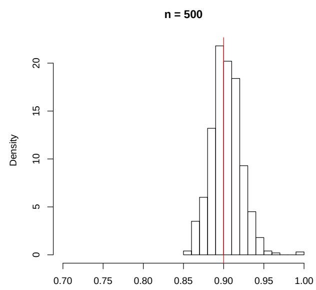
In this section, point and interval estimation for parameters of the models described in Section 3 are presented. Tables 1 to 7 show, respectively, the results for the Log-normal, Log-gamma, Fréchet, Lévy, Skew GED, Pareto and Weibull models. The average of 1,000 Monte Carlo replications of the MLE, BE-Mean and BE-Median, along with the mean square error (MSE), are presented. The tables also show the lower and upper limits and coverage rates (Cov Rate) of the asymptotic confidence intervals (Conf Int)



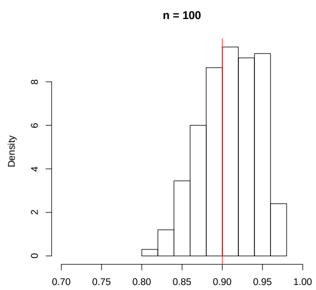
(a) MLE of ω



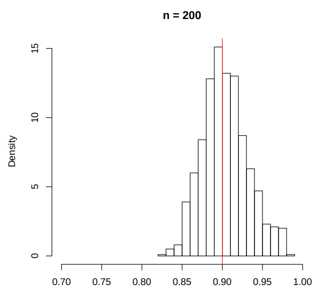
(b) MLE of ω



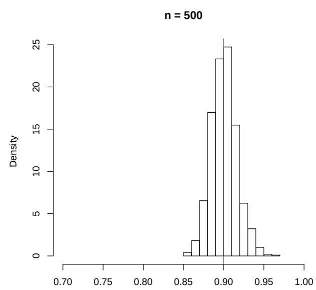
(c) MLE of ω



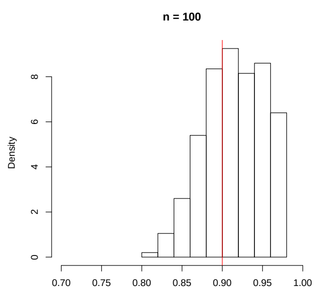
(d) BE-Mean of ω



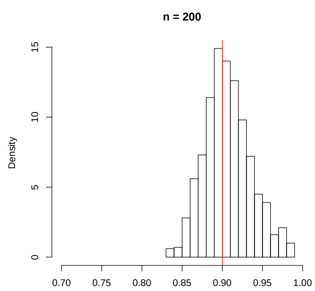
(e) BE-Mean of ω



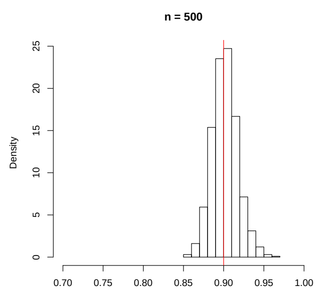
(f) BE-Mean of ω



(g) BE-Median of ω

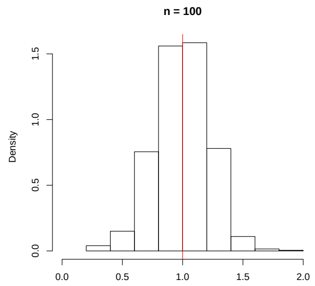


(h) BE-Median of ω

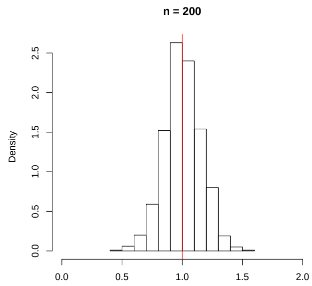


(i) BE-Median of ω

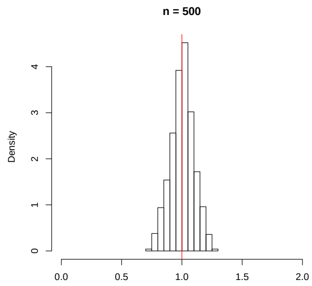
Figure 5.1: Histograms of the estimates (MLE, BE-Mean and BE-Median) of ω for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500.



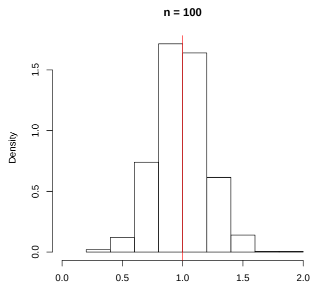
(a) MLE of β



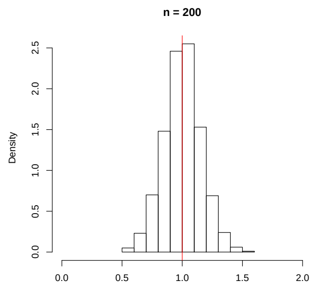
(b) MLE of β



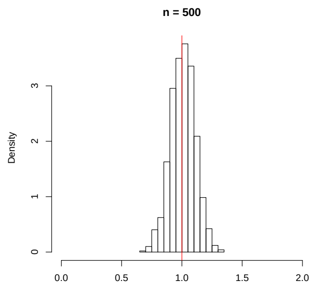
(c) MLE of β



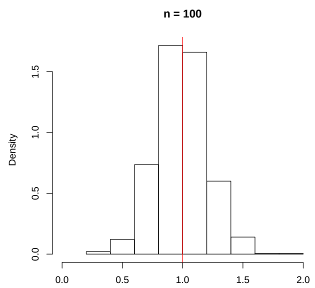
(d) BE-Mean of β



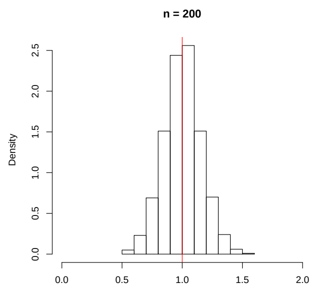
(e) BE-Mean of β



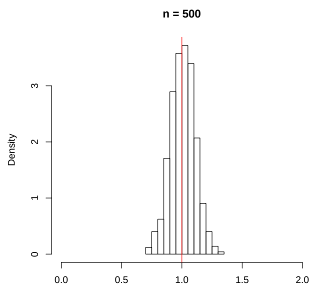
(f) BE-Mean of β



(g) BE-Median of β

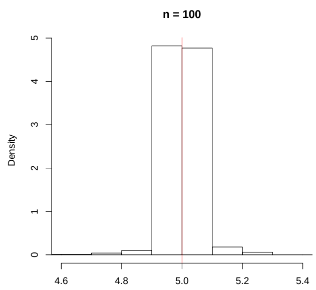


(h) BE-Median of β

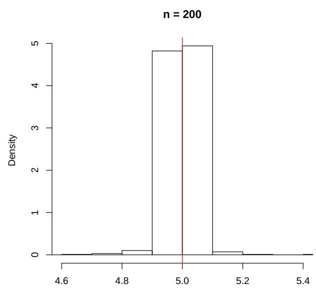


(i) BE-Median of β

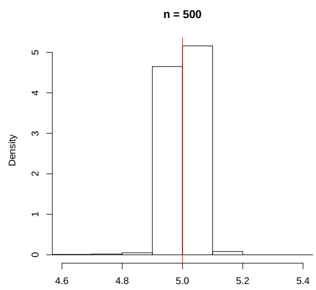
Figure 5.2: Histograms of the estimates (MLE, BE-Mean and BE-Median) of β for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500.



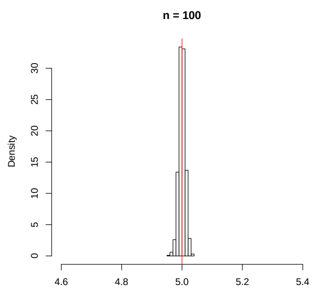
(a) MLE of δ



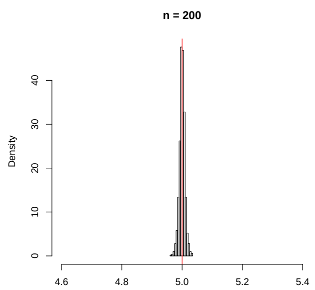
(b) MLE of δ



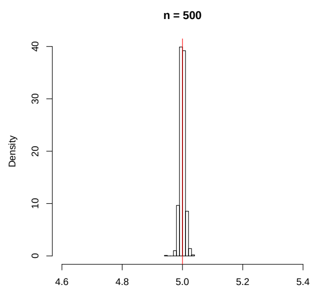
(c) MLE of δ



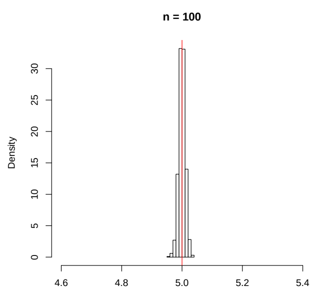
(d) BE-Mean of δ



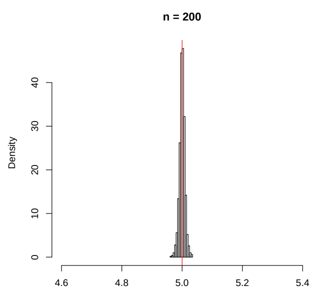
(e) BE-Mean of δ



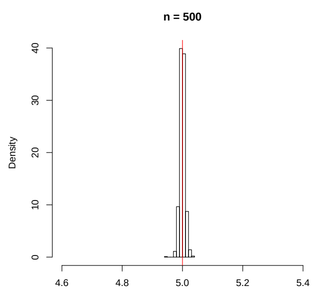
(f) BE-Mean of δ



(g) BE-Median of δ



(h) BE-Median of δ



(i) BE-Median of δ

Figure 5.3: Histograms of the estimates (MLE, BE-Mean and BE-Median) of δ for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500.

and of the confidence credibility intervals (Cred Int). Parameter γ for Log-normal, Féchet and Lévy and parameter α for the Skew GED distributions were kept fixed in the estimation stage.

The patterns are very similar for the parameter estimation in all models and therefore the conclusions will be jointly summarized for all cases. It can be observed that the estimation procedures seem consistent, as the MSE decreases as the sample sizes increase.

Table 5.1: Monte Carlo study for the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$).

n	φ	**MLE (MSE)**	**BE Mean (MSE)**	**BE Median (MSE)**	**Conf Int Cov Rate**	**Cred Int Cov Rate**
	ω	0.9206	0.9090	0.9149	[0.7407 ; 0.9644]	[0.8121 ; 0.9728]
		(0.0028)	(0.0013)	(0.0016)	0.916	0.983
100	β	0.9955	0.9915	0.9922	[0.5619 ; 1.4291]	[0.5575 ; 1.4223]
		(0.0507)	(0.0436)	(0.0436)	0.948	0.962
	δ	5.0006	5.0001	5.0001	[4.9441 ; 5.0570]	[4.9792 ; 5.0209]
		(0.0024)	(0.0001)	(0.0001)	0.932	0.951
	ω	0.9098	0.9039	0.9067	[0.8325 ; 0.9484]	[0.8429 ; 0.9490]
		(0.0011)	(0.0008)	(0.0009)	0.958	0.944
200	β	1.0032	1.0029	1.0030	[0.7031 ; 1.3033]	[0.7011 ; 1.3045]
		(0.0239)	(0.0246)	(0.0247)	0.944	0.940
	δ	4.9980	5.0002	5.0002	[4.9489 ; 5.0471]	[4.9832 ; 5.0171]
		(0.0020)	(0.0001)	(0.0001)	0.946	0.951
	ω	0.9038	0.9006	0.9018	[0.8659 ; 0.9311]	[0.8651 ; 0.9296]
		(0.0003)	(0.0003)	(0.0003)	0.949	0.953
500	β	1.0021	1.0076	1.0074	[0.8136 ; 1.1906]	[0.8183 ; 1.1968]
		(0.0090)	(0.0102)	(0.0102)	0.951	0.937
	δ	4.9996	4.9999	4.9999	[4.9586 ; 5.0406]	[4.9847 ; 5.0151]
		(0.0025)	(0.0001)	(0.0001)	0.944	0.948

Concerning parameter ω (the first line in all tables and all sample sizes), the MLE seems to consistently overestimate the true value, presenting larger bias and MSE than the Bayesian estimators, for small sample sizes. With respect to the Bayesian estimators, there is not much difference between BE-Mean and BE-Median and they are quite close to the true value of ω even for small samples. Concerning the intervals, it is interesting to note that, for all series of size $n = 100$, the coverage rate of the asymptotic confidence intervals is below the nominal rate and the coverage rate of the credibility

Table 5.2: Monte Carlo study for the Log-gamma model with ($\omega = 0.90$; $\beta = 1.0$; $\alpha = 5.0$).

n	φ	**MLE (MSE)**	**BE Mean (MSE)**	**BE Median (MSE)**	**Conf Int Cov Rate**	**Cred Int Cov Rate**
	ω	0.9245	0.8844	0.8935	[0.7673 ; 0.9687]	[0.7506 ; 0.9669]
		(0.0044)	(0.0026)	(0.0026)	0.794	0.960
100	β	0.9977	0.9983	0.9984	[0.8705 ; 1.1249]	[0.8695 ; 1.1273]
		(0.0043)	(0.0041)	(0.0041)	0.949	0.954
	α	5.1396	5.3720	5.3265	[3.6782 ; 6.6009]	[3.9632 ; 7.0443]
		(0.6493)	(0.7823)	(0.7375)	0.936	0.941
	ω	0.9128	0.8921	0.8964	[0.8286 ; 0.9536]	[0.8110 ; 0.9487]
		(0.0020)	(0.0012)	(0.0012)	0.869	0.952
200	β	0.9987	0.9975	0.9975	[0.9084 ; 1.0890]	[0.9066 ; 1.0883]
		(0.0021)	(0.0023)	(0.0023)	0.943	0.947
	α	5.0630	5.1783	5.1577	[4.0494 ; 6.0765]	[4.1986 ; 6.2794]
		(0.3097)	(0.3310)	(0.3213)	0.937	0.939
	ω	0.9026	0.8970	0.8987	[0.8559 ; 0.9343]	[0.8523 ; 0.9320]
		(0.0004)	(0.0004)	(0.0004)	0.952	0.952
500	β	0.9995	1.0000	1.0000	[0.9425 ; 1.0565]	[0.9430 ; 1.0570]
		(0.0008)	(0.0008)	(0.0008)	0.948	0.953
	α	5.0292	5.0667	5.0591	[4.3923 ; 5.6661]	[4.4519 ; 5.7283]
		(0.1085)	(0.1151)	(0.1139)	0.949	0.938

Table 5.3: Monte Carlo study for the Fréchet model with ($\omega = 0.90$; $\beta = 1.0$; $\alpha = 5.0$).

n	φ	**MLE (MSE)**	**BE Mean (MSE)**	**BE Median (MSE)**	**Conf Int Cov Rate**	**Cred Int Cov Rate**
	ω	0.9204	0.9021	0.9096	[0.7391 ; 0.9681]	[0.7880 ; 0.9740]
		(0.0029)	(0.0016)	(0.0018)	0.920	0.983
100	β	1.0093	1.0157	1.0145	[0.6752 ; 1.3433]	[0.6834 ; 1.3544]
		(0.0312)	(0.0288)	(0.0287)	0.938	0.957
	α	5.0368	5.1230	5.1143	[4.2355 ; 5.8381]	[4.3475 ; 5.9506]
		(0.1741)	(0.1719)	(0.1698)	0.940	0.944
	ω	0.9102	0.8988	0.9024	[0.8199 ; 0.9519]	[0.8263 ; 0.9509]
		(0.0012)	(0.0010)	(0.0010)	0.954	0.955
200	β	1.0046	1.0141	1.0134	[0.9518 ; 1.2407]	[0.7776 ; 1.2543]
		(0.0137)	(0.0161)	(0.0161)	0.956	0.935
	α	5.0106	5.0677	5.0631	[4.4404 ; 5.5808]	[4.5087 ; 5.6565]
		(0.0865)	(0.0892)	(0.0889)	0.956	0.946
	ω	0.9028	0.9002	0.9017	[0.8589 ; 0.9331]	[0.8592 ; 0.9328]
		(0.0004)	(0.0004)	(0.0004)	0.945	0.941
500	β	1.0004	1.0046	1.0044	[0.8514 ; 1.1494]	[0.8559 ; 1.1543]
		(0.0057)	(0.0059)	(0.0059)	0.949	0.949
	α	5.0062	5.0212	5.0190	[4.6437 ; 5.3688]	[4.6653 ; 5.3879]
		(0.0336)	(0.0352)	(0.0354)	0.957	0.947

Table 5.4: Monte Carlo study for the Lévy model with ($\omega = 0.90$; $\beta = 1.0$).

n	φ	MLE (MSE)	BE Mean (MSE)	BE Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
100	ω	0.9188	0.9115	0.9174	[0.7438 ; 0.9638]	[0.8155 ; 0.9740]
		(0.0026)	(0.0014)	(0.0017)	0.925	0.987
	β	0.9917	0.9897	0.9900	[0.5671 ; 1.4164]	[0.5607 ; 1.4176]
		(0.0496)	(0.0480)	(0.0480)	0.949	0.954
200	ω	0.9090	0.9040	0.9068	[0.8299 ; 0.9482]	[0.8481 ; 0.9364]
		(0.0010)	(0.0007)	(0.0008)	0.959	0.953
	β	0.9961	0.9454	0.9455	[0.6966 ; 1.2956]	[0.9508 ; 1.2283]
		(0.0238)	(0.0218)	(0.0218)	0.938	0.963
500	ω	0.9035	0.9015	0.9027	[0.8658 ; 0.9308]	[0.8658 ; 0.9306]
		(0.0003)	(0.0003)	(0.0003)	0.950	0.948
	β	0.9989	0.9938	0.9938	[0.8102 ; 1.1875]	[0.8049 ; 1.1827]
		(0.0100)	(0.0089)	(0.0089)	0.944	0.962

Table 5.5: Monte Carlo study for the Skew GED model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$; $\kappa = 1.0$).

n	φ	MLE (MSE)	BE Mean (MSE)	BE Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
	ω	0.9330	0.9051	0.9075	[0.7359 ; 0.9728]	[0.8321 ; 0.9631]
		(0.0031)	(0.0012)	(0.0015)	0.913	0.975
	β	1.0113	1.0043	1.0051	[0.6468 ; 1.3758]	[0.8554 ; 1.1494]
100		(0.0344)	(0.0057)	(0.0062)	0.945	0.969
	δ	5.0000	4.9998	4.9998	[4.9897 ; 5.0103]	[4.9981 ; 5.0016]
		(0.00003)	(0.00000)	(0.00000)	0.931	0.946
	κ	1.0058	1.0206	1.0226	[0.8152 ; 1.1963]	[0.9618 ; 1.0474]
		(0.0100)	(0.0035)	(0.0044)	0.945	0.944
	ω	0.9131	0.9045	0.9057	[0.8284 ; 0.9516]	[0.8527 ; 0.9539]
		(0.0011)	(0.0006)	(0.0009)	0.962	0.982
	β	1.0063	1.0037	0.0039	[0.7491 ; 1.2636]	[0.9151 ; 1.0933]
200		(0.0190)	(0.0038)	(0.0043)	0.934	0.949
	δ	4.9998	4.9999	4.9999	[4.9918 ; 5.0079]	[4.9988 ; 5.0013]
		(0.00002)	(0.00000)	(0.00000)	0.945	0.947
	κ	0.9986	1.0119	0.0124	[0.8755 ; 1.1217]	[0.9860 ; 1.0377]
		(0.0041)	(0.0012)	(0.0014)	0.943	0.938
	ω	0.9039	0.9011	0.9014	[0.8650 ; 0.9319]	[0.8773 ; 0.9235]
		(0.0003)	(0.0003)	(0.0004)	0.9440	0.958
	β	0.9989	1.0028	1.0027	[0.8374 ; 1.1605]	[0.9755 ; 1.0406]
500		(0.0067)	(0.0010)	(0.0011)	0.9560	0.968
	δ	5.0000	5.0001	5.0001	[4.9938 ; 5.0061]	[4.9990 ; 5.0012]
		(0.00001)	(0.00000)	(0.00000)	0.9320	0.941
	κ	1.0015	1.0108	1.0112	[0.9327 ; 1.0703]	[0.9941 ; 1.0255]
		(0.0014)	(0.0004)	(0.0004)	0.9440	0.939

Table 5.6: Monte Carlo study for the Pareto model with ($\omega = 0.90$; $\beta = 1.0$).

n	φ	MLE (MSE)	BE Mean (MSE)	BE Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
100	ω	0.9183	0.9048	0.9115	[0.7351 ; 0.9655]	[0.8004 ; 0.9721]
		(0.0026)	(0.0014)	(0.0017)	0.937	0.991
	β	0.9990	0.9941	0.9943	[0.7065 ; 1.2915]	[0.6967 ; 1.2899]
		(0.0227)	(0.0221)	(0.0221)	0.952	0.959
200	ω	0.9079	0.9016	0.9049	[0.8239 ; 0.9486]	[0.8346 ; 0.9500]
		(0.0011)	(0.0008)	(0.0009)	0.964	0.961
	β	0.9961	0.9995	0.9996	[0.7893 ; 1.2028]	[0.7914 ; 1.2073]
		(0.0110)	(0.0108)	(0.0108)	0.950	0.958
500	ω	0.9043	0.8996	0.9009	[0.8640 ; 0.9329]	[0.8609 ; 0.9307]
		(0.0003)	(0.0003)	(0.0003)	0.952	0.959
	β	1.0014	1.0013	1.0013	[0.8713 ; 1.1315]	[0.8709 ; 1.1318]
		(0.0043)	(0.0046)	(0.0046)	0.955	0.942

Table 5.7: Monte Carlo study for the Weibull model with ($\omega = 0.90$; $\beta = 1.0$; $\upsilon = 5.0$).

n	φ	MLE (MSE)	BE Mean (MSE)	BE Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
	ω	0.9233	0.8969	0.9041	[0.7409 ; 0.9684]	[0.7823 ; 0.9711]
		(0.0034)	(0.0017)	(0.0019)	0.892	0.972
100	β	1.0018	1.0294	1.0282	[0.6689 ; 1.3347]	[0.6943 ; 1.3711]
		(0.0284)	(0.0318)	(0.0317)	0.953	0.942
	υ	5.0204	5.1499	5.1412	[4.2224 ; 5.8183]	[4.3678 ; 5.9844]
		(0.1706)	(0.1939)	(0.1913)	0.949	0.944
	ω	0.9083	0.9008	0.9045	[0.8163 ; 0.9504]	[0.8285 ; 0.9521]
		(0.0012)	(0.0010)	(0.0010)	0.961	0.951
200	β	0.9979	1.0054	1.0049	[0.7620 ; 1.2338]	[0.7697 ; 1.2444]
		(0.0142)	(0.0149)	(0.0149)	0.952	0.949
	υ	5.0100	5.0490	5.0444	[4.4404 ; 5.5795]	[4.4940 ; 5.6320]
		(0.0872)	(0.0839)	(0.0835)	0.944	0.952
	ω	0.9035	0.8991	0.9005	[0.8599 ; 0.9337]	[0.8581 ; 0.9317]
		(0.0004)	(0.0003)	(0.0003)	0.939	0.960
500	β	1.0020	1.0058	1.0054	[0.8531 ; 1.1509]	[0.8574 ; 1.1557]
		(0.0056)	(0.0061)	(0.0061)	0.949	0.946
	υ	5.0133	5.0244	5.0222	[4.6503 ; 5.3764]	[4.6696 ; 5.3921]
		(0.0352)	(0.0389)	(0.0389)	0.951	0.935

intervals is above the nominal rate. For larger sample sizes, the coverage rates of both intervals are close to the 95% level, except the confidence interval for the Log-gamma model with $n = 200$.

Estimates of parameter β (the second parameter in all tables and all sample sizes) do not differ for the MLE and Bayesian estimators and are very close to the real value $\beta = 1.0$ for all models. The Log-normal and Lévy models present the largest MSE values for all sample sizes, while the Log-gamma possesses the smallest ones. Therefore, the limits of the asymptotic confidence and credibility intervals are larger for the Log-normal and Lévy models. The Fréchet, Skew GED, Pareto and Weibull models show the same pattern for the MSE, which are smaller than the values in the Log-normal but larger than the ones in the Log-gamma models. Nevertheless, the coverage rates are all very close to the 95% fixed level, for all models and all sample sizes.

The third parameter, which depends on the distribution employed, was set equal to 5.0 for all cases, except in the Pareto and Lévy models, where there is no extra parameter. For the Log-normal model, the behaviour is the same for all methods and the estimates are very close to 5.0, with very small MSE. The intervals show coverage rates very close to 95% and small width. For the Log-gamma model, the MLE presents a better performance compared to the Bayesian estimators, with smaller MSE. The coverage rates of the intervals are below the 95% nominal level and the widths are the largest ones. The Fréchet and Weibull models present a very similar behaviour, with the same magnitude for the estimates. In this case, the MLE is again the procedure with the best performance (smaller bias and MSE).

Concerning the fourth parameter in the Skew GED model, the MSE is larger for the MLE compared to the Bayesian estimators for all sample sizes, although its bias is smaller for sample sizes 100 and 500. The coverage rates are close to the 95% fixed level for all sample sizes.

5.5 Application to South and North American stock exchange indexes

Heavy tailed models in the NGSSM were fitted to the volatility of the following stock exchange indexes: S&P 500 and NASDAQ (USA), INMEX (Mexico), IBOVESPA (Brazil), MERVAL (Argentina) and IPSA (Chile) comprising the period 02/01/2007 to 05/16/2011. Considering only work days, each series possesses 1101, 1101, 1098, 1078, 1074 and 1092 observations, respectively. The models were adjusted with the own series with an one-day delay as a covariate and the exponential link function.

With the purpose of comparing the models in the NGSSM with some known procedures in the literature, GARCH models proposed by Bollerslev (1986) were also fitted to the series. The GARCH models are defined as follows.

$$y_t = \sigma_t \epsilon_t, t = 1, \cdots, n, \tag{5.5}$$

$$\sigma_t^2 = \theta_0 + \sum_{i=1}^{p} \theta_i \varepsilon_{t-1}^2 + \sum_{j=1}^{q} \phi_j \sigma_{t-j}^2 \tag{5.6}$$

where $\theta_0 > 0$, $\theta_i \geq 0$, $\phi_j \geq 0$ and $\sum_{k=1}^{r} (\theta_k + \phi_k) < 1$ with $i = 1, \ldots, p$, $j = 1, \ldots, q$ and $r = max\,(p,q)$. The following distributions were assumed for ϵ_t: Gaussian, Skew Gaussian, t-Student, Skew t-Student, GED and Skew GED. All models were estimated using the square of the log-return of the stock exchange indexes.

According to the results of the simulation study in Section 4, for large sample sizes the MLE and Bayesian estimators are very similar. Thus, for the comparison with GARCH models (Table 5.8), only the results of the MLE are presented.

The programs developed in Ox Metrics by the authors are used to estimate the NGSSM. For GARCH models, the fGARCH package in software R, which uses Quasi-Maximum Likelihood Estimation (QMLE) is employed to estimate the parameters. For more details see Bollerslev & Wooldridge (1992).

For the Log-normal, Fréchet and Lévy models the parameter γ was fixed at 0.0 and, consequently, not estimated. For the Log-gamma and Pareto models there is a constraint that the series should have values greater than 1.0. Thus, for these models a constant value 1.0 was added to the observations of all series.

Figure 5.4 presents the indexes and the log-returns of the six series. It can be observed, in all cases, an increase in the volatility around observations 400 and 500, which corresponds to the second semester of 2008, period of the Global Financial Crisis in 2008.

Model comparison was performed using the AICc, BIC and log-likelihood (LN LIKE) criteria (see Table 5.8). According to the three criteria, the Weibull model is the best one within the NGSSM models and the GARCH (1,1) with Skew t-Student errors is the best one in the GARCH family. Comparing the two approaches (NGSSM and GARCH) it is worth to note that, except for the Lévy model, all other models in the NGSSM family present better results than the GARCH models, with the Weibull model being the best one, followed closely by the Log-gamma model. The fit of the Weibull model was assessed by the Pearson residual for all series and it was not observed any evidence of inadequacy.

Table 5.9 presents the MLE, BE-Mean and BE-Median for parameters of the Weibull model fitted to the volatility series of all indexes. In addition, 95% asymptotic confidence and credibility intervals are also built. It is verified that all parameters are significant to the 5% level.

It is interesting to note that the parameter estimates are relatively close for all models, except for IPSA. Values of ω are between 0.93 and 0.94 for the USA, Mexico, Brazil and Argentina indexes and around 0.91 for Chile. This indicates a smaller impact of the crisis in the variance of the level of this series, as can be visualized in Figure 5.4.

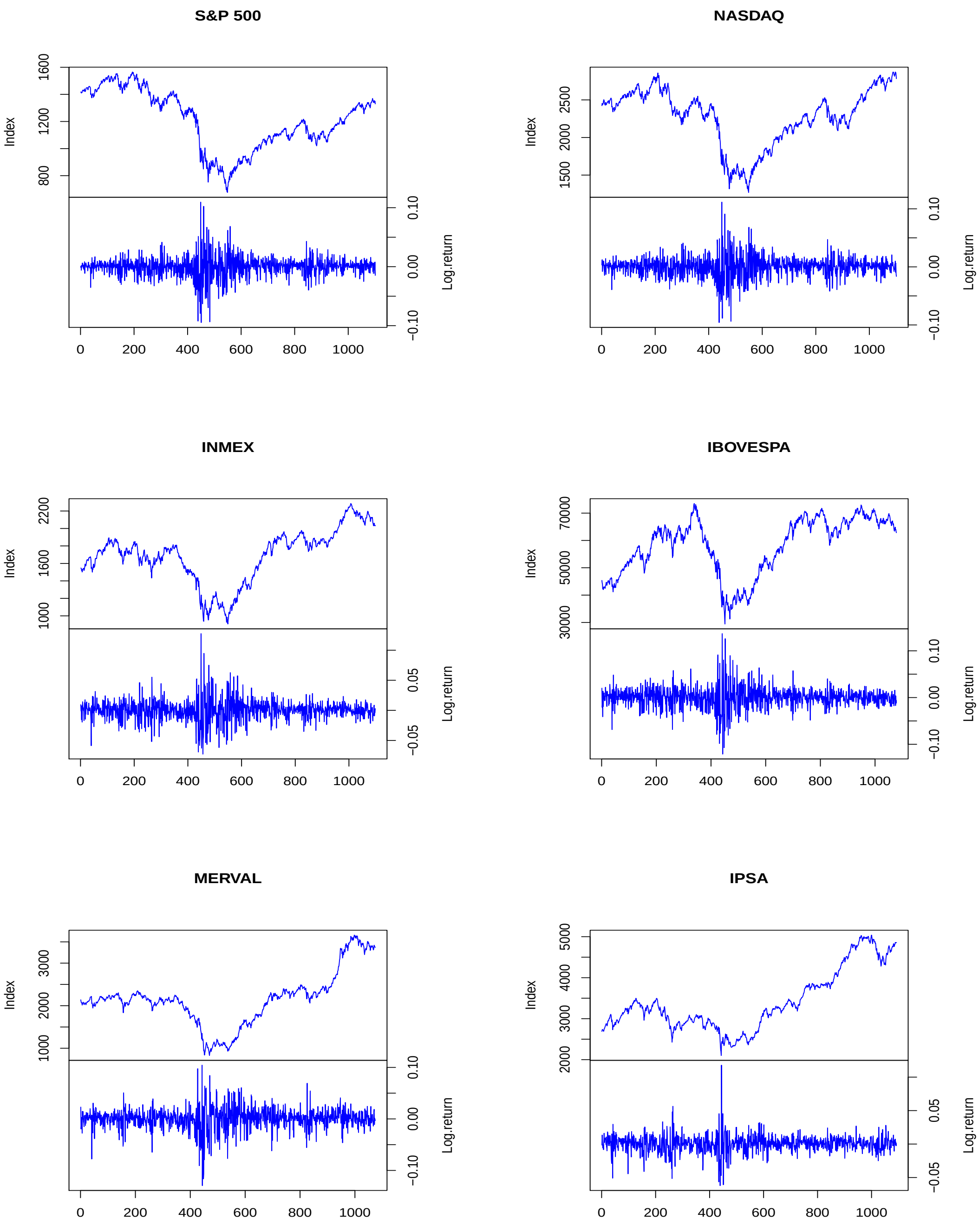


Figure 5.4: The index and the log-return of S&P 500, NASDAQ, INMEX, IBOVESPA, MERVAL and IPSA, in the period from 02/01/2007 to 05/16/2011.

Table 5.8: Fitted models for the North and South American stock indexes.

SERIES	NGSSM	AICc	BIC	LN LIKE	GARCH(1,1)	AICc	BIC	LN LIKE
	LOGNORMAL	-15.86	-15.85	8733.54	SKEW NORMAL	-14.08	-14.06	7753.78
	LOGGAMA	-16.16	-16.15	8900.48	NORMAL	-13.38	-13.36	7368.90
	FRÉCHET	-15.53	-15.52	8553.58	SKEW t-STUDENT	-15.17	-15.15	8352.86
S&P 500	LÉVY	-15.01	-15.00	8265.76	t-STUDENT	-14.43	-14.41	7946.50
	SKEW GED	-15.43	-15.41	8498.60	SKEW GED	-14.78	-14.76	8141.00
	PARETO	-15.58	-15.58	8581.54	GED	-14.38	-14.36	7920.16
	WEIBULL	**-16.22**	**-16.21**	**8933.75**				
	LOGNORMAL	-15.46	-15.45	8514.08	SKEW NORMAL	-13.84	-13.82	7622.17
	LOGGAMA	-15.78	-15.76	8688.91	NORMAL	-13.15	-13.14	7245.32
	FRÉCHET	-15.12	-15.11	8326.41	SKEW t-STUDENT	-14.85	-14.83	8176.67
NASDAQ	LÉVY	-14.64	-14.63	8058.82	t-STUDENT	-14.10	-14.09	7767.86
	SKEW GED	-15.10	-15.09	8318.60	SKEW GED	-14.37	-14.35	7913.89
	PARETO	-15.24	-15.23	8391.66	GED	-13.46	-13.44	7411.39
	WEIBULL	**-15.81**	**-15.80**	**8706.66**				
	LOGNORMAL	-15.32	-15.31	8413.60	SKEW NORMAL	-13.68	-13.66	7512.16
	LOGGAMA	-15.69	-15.67	8614.37	NORMAL	-12.91	-12.89	7089.70
	FRÉCHET	-14.94	-14.92	8203.37	SKEW t-STUDENT	-14.89	-14.87	8176.84
INMEX	LÉVY	-14.42	-14.41	7918.67	t-STUDENT	-14.15	-14.13	7773.55
	SKEW GED	-15.09	-15.07	8289.84	SKEW GED	-15.08	-15.06	8282.14
	PARETO	-15.24	-15.23	8368.69	GED	-13.97	-13.95	7672.04
	WEIBULL	**-15.71**	**-15.69**	**8626.82**				
	LOGNORMAL	-14.44	-14.43	7786.31	SKEW NORMAL	-12.83	-12.82	6921.41
	LOGGAMA	-14.73	-14.72	7944.69	NORMAL	-12.03	-12.01	6489.30
	FRÉCHET	-14.01	-14.00	7554.40	SKEW t-STUDENT	-13.98	-13.96	7537.56
IBOVESPA	LÉVY	-13.57	-13.56	7317.18	t-STUDENT	-13.20	-13.18	7118.99
	SKEW GED	-14.21	-14.19	7664.89	SKEW GED	-14.19	-14.17	7651.67
	PARETO	-14.30	-14.29	7710.37	GED	-12.96	-12.94	6988.77
	WEIBULL	**-14.75**	**-14.74**	**7952.81**				
	LOGNORMAL	-14.73	-14.71	7910.70	SKEW NORMAL	-12.78	-12.76	6868.08
	LOGGAMA	-15.02	-15.00	8068.03	NORMAL	-11.92	-11.90	6403.55
	FRÉCHET	-14.29	-14.28	7677.31	SKEW t-STUDENT	-14.15	-14.14	7604.65
MERVAL	LÉVY	-13.69	-13.68	7354.16	t-STUDENT	-13.35	-13.34	7174.86
	SKEW GED	-14.34	-14.33	7706.75	SKEW GED	-13.82	-13.80	7426.51
	PARETO	-14.46	-14.45	7766.39	GED	-13.44	-13.42	7220.48
	WEIBULL	**-15.04**	**-15.03**	**8079.62**				
	LOGNORMAL	-16.44	-16.43	8981.02	SKEW NORMAL	-14.76	-14.74	8060.62
	LOGGAMA	-16.70	-16.69	9121.19	NORMAL	-14.07	-14.05	7685.27
	FRÉCHET	-16.05	-16.03	8765.52	SKEW t-STUDENT	-16.06	-16.04	8774.13
IPSA	LÉVY	-15.62	-15.61	8531.17	t-STUDENT	-15.24	-15.22	8322.33
	SKEW GED	-16.13	-16.11	8808.75	SKEW GED	-16.28	-16.27	8895.41
	PARETO	-16.32	-16.31	8911.04	GED	-15.22	-15.21	8316.26
	WEIBULL	**-16.73**	**-16.71**	**9135.45**				

Obs.: In bold are the models with the smallest AICc and BIC and the largest log-likelihood (LN LIKE) for each series.

Table 5.9: Parameter estimates of the Weibull models for the volatility of the indexes.

NGSSM	φ	MLE	BE Mean	BE Median	Conf Int	Cred Int
	ω	0.9333	0.9308	0.9316	[0.9083 ; 0.9517]	[0.9080 ; 0.9506]
S&P 500	β	4.5686	4.4084	4.3641	[0.6582 ; 8.4789]	[0.6776 ; 8.1940]
	υ	0.5618	0.5631	0.5631	[0.5350 ; 0.5885]	[0.5363 ; 0.5897]
	ω	0.9423	0.9401	0.9407	[0.9184 ; 0.9594]	[0.9181 ; 0.9579]
NASDAQ	β	5.4782	5.4542	5.4979	[1.8609 ; 9.0955]	[1.8986 ; 8.8856]
	υ	0.5750	0.5760	0.5762	[0.5472 ; 0.6028]	[0.5479 ; 0.6039]
	ω	0.9305	0.9284	0.9290	[0.9031 ; 0.9504]	[0.9037 ; 0.9501]
INMEX	β	3.9082	3.8991	3.9007	[0.2876 ; 7.5289]	[0.2880 ; 7.4180]
	υ	0.5989	0.5996	0.5996	[0.5696 ; 0.6281]	[0.5703 ; 0.6281]
	ω	0.9410	0.9386	0.9391	[0.9158 ; 0.9588]	[0.9141 ; 0.9582]
IBOVESPA	β	5.3486	5.2530	5.2128	[2.4470 ; 8.2502]	[2.3158 ; 8.2850]
	υ	0.6039	0.6047	0.6042	[0.5741 ; 0.6337]	[0.5767 ; 0.6351]
	ω	0.9349	0.9322	0.9329	[0.9043 ; 0.9560]	[0.9047 ; 0.9554]
MERVAL	β	4.0468	4.0067	3.9604	[1.0735 ; 7.0201]	[1.1250 ; 7.0988]
	υ	0.5537	0.5547	0.5545	[0.5258 ; 0.5816]	[0.5276 ; 0.5833]
	ω	0.9145	0.9126	0.9128	[0.8858 ; 0.9363]	[0.8878 ; 0.9359]
IPSA	β	10.1068	9.9962	9.9389	[5.7859 ; 14.4278]	[5.9120 ; 14.3082]
	υ	0.6135	0.6139	0.6138	[0.5833 ; 0.6438]	[0.5848 ; 0.6443]

5.6 Conclusion

Due to the recent instability in the global economic scenario, a great variety of procedures to model volatility are being proposed in the econometric literature. In order to accommodate the main characteristics of this kind of series, the models need to, necessarily, incorporate heteroscedasticity and nonnormality assumptions.

Thus, the main objective of this work was to present some particular models in a non-Gaussian state space family (NGSSM), proposed by Santos et al. (2010), whose distribution function is contained in the family of heavy tailed distributions, such as the Log-normal, Log-gamma, Fréchet, Lévy, GED, Pareto and Weibull. The NGSSM, when combined with heavy tailed distributions, can produce better results than the classical methodologies often employed in econometric studies, such as the GARCH like families.

The superiority of the method addressed here was confirmed through the fit of the methodology to the main return indexes of North and South America, when compared to different GARCH models. The paper also presents the results of a Monte Carlo study comparing classical and Bayesian estimation for some heavy tailed distributions in the NGSSM. In general, the estimation procedures show very satisfactory results.

Future research encompasses the improvement of the maximum likelihood method to properly estimate ω for small samples and hypothesis test for the parameters.

Acknowledgements

The authors wish to acknowledge CAPES, CNPq and FAPEMIG for financial support.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716-723.

Asmussen, S., 2003. Applied Probability and Queues. Springer, Berlin.

Anderson, J., 2001. On the normal inverse Gaussian stochastic volatility model. Journal of Business and Economic Statistics, 19, 44-54.

Ayebo, A., Kozubowski, T.J., 2003. An asymmetric generalization of Gaussian and Laplace laws. Journal of Probability and Statistical Science, 1, 187-210.

Bauwens, L., Laurent, S., Rombouts, J.V.K., 2006. Multivariate GARCH models: A survey. Journal of Applied Econometrics, 21, 79-109.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307-327.

Bollerslev, T., Wooldridge J.M., 1992. Quasi-Maximum likelihood estimation and inference in dynamic models with time-varying covariance. Econometric Reviews 11, 143-172.

Broyden, C.G., 1970. The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics & Its Applications, 6, 76-90.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for sthocastic volatility models. Journal of Econometrics, 108, 281-316.

Consul, P.C., Jain, G.C., 1971. On the log-gamma distribution and its properties. Statistical Papers, 12(2), 100-106.

Deschamps, P.K., 2011. Bayesian estimation of an extended local scale stochastic volatility model. Journal of Econometrics, 162, 369-382.

Embrechts, P., Klüppelberg, C., Milosch, T., 1997. Modelling Extremal Events.

Springer, New York.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica, 50, 987-1007.

Eraker, B., Johanners, M., Polson, N.G., 2003. The impact of jumps in returns and volatility. Journal of Finance, 53, 1269-1330.

Ferrante, M., Vidoni, P., 1998. Finite dimensional filters for nonlinear stochastic difference equations with multiplicative noises. Stochastic Processes and Their Applications, 77, 69-81.

Fletcher, R., 1970. A new approach to variable metric algorithms. The Computer Journal, 13(3), 317-322.

Goldfard, D., 1970. A family of variable metric updates derived by variational means. Mathematics of Computation, 24(109), 23-26.

Goldie, C.M., Klüppelberg, C., 1998. Subexponential Distributions. A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhauser Boston, Cambridge, 435-459.

Green, R.F., 1976. Outlier-prone and outlier-resistant distributions. Journal of the American Statistical Association, 71(354), 502-505.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli, 7(2), 223-242.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Harvey, A.C., Fernandes, C., 1989. Time series models for count or qualitative observations. Journal of Business & Economic Statistics, 7(4), 407-417.

Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies, 61, 247-264.

Hurvich, C.M., Tsai, C.L., 1993. A corrected Akaike information criterion for vector autoregressive model selection. Journal of Time Series Analysis, 14, 271-279.

Jacquier, E., Polson, N.G., Rossi, P., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal of Businees & Economic Statistics, 12, 371-417.

McCulagh, P., Nelder. J.A., 1989. Generalized Linear Models. Chapman and Hall, London.

Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. Journal of Econometrics, 45, 239-265.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347-370.

Neyman, J., Scott, E.T., 1971. Outliers Proneness of Phenomena and Related Distributions, Optimizing Methods in Statistics. Academic Press, New York, 413-430.

Raggi, D., Bordignon, S., 2006. Comparing stochastic volatility models through Monte Carlo simulations. Computational Statistics and Data Analysis, 50, 1678-1699.

Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. Journal of Computational & Graphical Statistics, 18(2), 349-367.

Santos, T.R., Franco, G.C., Gamerman, D., 2010. Gamma family of dynamic models. Technical Report, 234, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro. http://www.dme.im.ufrj.br/arquivos/publicacoes/arquivo234.pdf

Schwarz, G.E., 1978. Estimating the dimension of a model. Annals of Statistics, 6(2), 461-464.

Shanno, D.F., 1970. Conditioning of quase-Newton methods for function minimization. Mathematics of Computation, 24(111), 647-656.

Shephard, N., 1994. Local scale model: state space alternative to integrated GARCH processes. Journal of Econometrics, 60, 181-202.

Smith, R.L., Miller, J.E., 1986. A non-Gaussian state space model and application to prediction of records. Journal of the Royal Statistical Society, Series B, 48(1), 79-88.

Sugiura, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Communication in Statistics, A7, 13-26.

Taylor, S.J., 1986. Modeling Financial Time Series. John Wiley & Sons.

Taylor, S.J., 1994. Modeling stochastic volatility: A review and comparative study. Mathematical Finance, 4, 183-204.

Teugels, J.L., 1975. The class of subexponential distributions. The Annals of Probability, 3(6), 1000-1011.

Tsay, R.S., 2005. Analysis of Financial Time Series. John Wiley & Sons, New Jersey.

Vidoni, P., 1999. Exponential family state space models based on conjugate latent process. Journal of Royal Statistical Society B., 61, 213-221.

West, M., Harrison, P.J., Migon, H.S., 1985. Dynamic generalized linear models and Bayesian forecasting (with discussion). Journal of the American Statistical Association, 81, 741-750.

Zakoian, J.M., 1994. Threshold heteroscedastic models. Journal of Economic Dynamics & Control, 18, 931-955.

Chapter 6

Penalized Likelihood for a Non Gaussian State Space Model Considering Heavy Tailed Distributions

Frank M. de Pinho[a], **Glaura C. Franco**[b]
[a]*IBMEC, Belo Horizonte, Brasil*
[b]*Universidade Federal de Minas Gerais, Belo Horizonte, Brasil*

Abstract

Santos et al. (2010) have proposed a non Gaussian model in the state space framework which accomodates a wide range of distributions. Although inference procedures for this new family work satisfactorily well, one of its parameters, ω, which impacts the variability of the model, is generally overestimated, regardless the estimation method used. This paper proposes a penalized likelihood function to reduce empirically the bias of the maximum likelihood estimator of parameter ω. Monte Carlo simulation studies are performed to measure the reduction of bias and mean square error of the obtained estimators.

Keyword: Monotone Likelihood, Maximum Likelihood Estimator, Heavy Tailed Distributions, BFGS, SQP, FSQP.

6.1 Introduction

Santos et al. (2010) have proposed a non Gaussian state space model (NGSSM), which is a generalization of the results of Smith & Miller (1986). This procedure comprises a dynamic model with exact evolution equation to any time series with exponential distribution, as well as transformations one by one of the series, allowing the analytical integration of the state and the achievement of the predictive likelihood.

Pinho et al. (2012) have studied some other distributions (all of them heavy tailed) that are special cases of the NGSSM, including the Log-normal, Log-gamma, Fréchet, Lévy, and the Skew Generalized Error Distribution (SGED). Pinho et al. (2012) also presented Monte Carlo experiments comparing Bayesian and classical methods of inference in the estimation of the NGSSM. The study was performed for time series of size larger than 100, however, it is quoted in the work that for series of smaller sizes there are problems in the estimation of parameter ω.

In this work the reasons and solutions to this problem are explored. It will be noted that parameter ω (known as the discount factor) presents, most of the times estimates close to the limit of the parameter space for this parameter. Thus, the goal of this work is to propose a penalty function for the likelihood, with the aim of correcting the bias of this estimator.

The paper is organized as follows. Section 6.2 defines the NGSSM. Section 6.3 shows the proposed penalized function for the maximum likelihood function and presents the inference procedures. Section 6.4 shows the results of the Monte Carlo studies to evaluate the penalized maximum likelihood estimator and Section 6.5 concludes the work.

6.2 A non-Gaussian state space model

Let $\{y_t\}_{t=1}^{n}$ be a time series. Santos et al. (2010) define a new family of non-Gaussian state space models (NGSSM), with exact marginal likelihood, if the probability (density) function of $\{y_t\}_{t=1}^{n}$ can be written in the form:

$$p(y_t|\mu_t,\boldsymbol{\varphi}) = q(y_t,\boldsymbol{\varphi})\mu_t^{r(y_t,\boldsymbol{\varphi})} \exp\left(-\mu_t s(y_t,\boldsymbol{\varphi})\right), \text{ for } y_t \in H(\boldsymbol{\varphi}) \subset \Re \qquad (6.1)$$

and $p(y_t|\mu_t,\boldsymbol{\varphi}) = 0$, otherwise. Functions $q(\cdot)$, $r(\cdot)$, $s(\cdot)$ and $H(\cdot)$ are such that $p(y_t|\mu_t,\boldsymbol{\varphi}) \geq 0$ and therefore $\mu_t > 0$, for all $t > 0$. It is also assumed that $\boldsymbol{\varphi}$ varies in the p-dimensional parameter space Φ.

A link function g relates the predictor to the parameter μ_t through the relation $\mu_t = \lambda_t g(x_t, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ are the regression coefficients of the covariate vector $\boldsymbol{x_t}$ and λ_t is the latent state variable.

The dynamic level λ_t is initialized with prior distribution $\lambda_0|Y_0 \sim Gamma(a_0,b_0)$ and evolves according to $\lambda_{t+1} = \omega^{-1}\lambda_t\varsigma_{t+1}$, where $\varsigma_{t+1}|\boldsymbol{Y}_t \sim Beta\left(\omega a_t,(1-\omega)a_t\right)$, $0 < \omega \leq 1$, $t = 1, 2, ...$, $\boldsymbol{Y}_t = \{Y_0, y_1, \ldots, y_t\}$ and Y_0 represents previously available information.

The prior and updated equations of the dynamic level are given, respectivelly, by (see Theorem 1 in Santos et al. (2010))

$$\mu_t|\,\mathbf{Y_{t-1}},\boldsymbol{\varphi} \sim \text{Gamma}\left(c_{t|t-1};\, d_{t|t-1}\right), \qquad (6.2)$$

where $c_{t|t-1} = \omega a_{t-1}$ and $d_{t|t-1} = \omega b_{t-1}\left[g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right)\right]^{-1}$, and

$$\mu_t|\,\mathbf{Y_t},\boldsymbol{\varphi} \sim \text{Gamma}\left(c_t;\, d_t\right), \qquad (6.3)$$

where $c_t = c_{t|t-1} + r\left(y_t, \boldsymbol{\varphi}\right)$ and $d_t = d_{t|t-1} + s\left(y_t, \boldsymbol{\varphi}\right)$.

The exact predictive density function is given by

$$p\left(y_t\right|\mathbf{Y_{t-1}},\boldsymbol{\varphi}) = \frac{\Gamma\left(r\left(y_t,\boldsymbol{\varphi}\right)+c_{t|t-1}\right) q\left(y_t,\boldsymbol{\varphi}\right) d_{t|t-1}^{c_{t|t-1}} I_{(y_t\in H(\boldsymbol{\varphi}))}}{\Gamma\left(c_{t|t-1}\right)\left[s\left(y_t,\boldsymbol{\varphi}\right)+d_{t|t-1}\right]^{r(y_t,\boldsymbol{\varphi})+c_{t|t-1}}}. \tag{6.4}$$

In Table 6.1 it can be seen special cases presented by Santos et al. (2010) and Pinho et al. (2012):

Table 6.1: Distributions in the NGSSM

Model	$\boldsymbol{\varphi}$	$q\left(y_t,\boldsymbol{\varphi}\right)$	$r\left(y_t,\boldsymbol{\varphi}\right)$	$s\left(y_t,\boldsymbol{\varphi}\right)$	$H\left(\boldsymbol{\varphi}\right)$
Log-normal†	$(\omega,\boldsymbol{\beta},\gamma,\delta)$	$\left[(y_t-\gamma)\sqrt{2\pi}\right]^{-1}$	$\frac{1}{2}$	$\frac{[\ln(y_t-\gamma)-\delta]^2}{2}$	(γ,∞)
Log-gamma†	$(\omega,\boldsymbol{\beta},\alpha)$	$\frac{\alpha^{\alpha}[ln(y_t)]^{\alpha-1}}{[\Gamma(\alpha)y_t]}$	α	$\alpha\ln\left(y_t\right)$	$(1,\infty)$
Fréchet†	$(\omega,\boldsymbol{\beta},\gamma,\alpha)$	$\alpha\left(y_t-\gamma\right)^{-\alpha-1}$	1	$(y_t-\gamma)^{-\alpha}$	(γ,∞)
Lévy†	$(\omega,\boldsymbol{\beta},\gamma)$	$\left[2\pi\left(y_t-\gamma\right)\right]^{-\frac{3}{2}}$	$\frac{1}{2}$	$\left[2\left(y_t-\gamma\right)\right]^{-1}$	(γ,∞)
Skew GED†	$(\omega,\boldsymbol{\beta},\kappa,\alpha,\delta)$	$\frac{\kappa}{\Gamma\left(\alpha^{-1}\right)\left(1+\kappa^2\right)}$	$\frac{1}{\alpha}$	$\left[\frac{(y_t-\delta)^+}{k^{-\alpha}}\right]^{\alpha}+\left[\frac{(y_t-\delta)^-}{k^{\alpha}}\right]^{\alpha}$	$(-\infty,\infty)$
Pareto†	$(\omega,\boldsymbol{\beta})$	y_t^{-1}	1	$\ln\left(y_t\right)$	$(1,\infty)$
Weibull†	$(\omega,\boldsymbol{\beta},\upsilon)$	$\upsilon y_t^{\upsilon-1}$	1	y_t^{υ}	$(0,\infty)$
Poisson	$(\omega,\boldsymbol{\beta})$	$(y_t!)^{-1}$	y_t	1	$\{0,1,\ldots\}$ †
Borel-Tanner	$(\omega,\boldsymbol{\beta},\gamma)$	$\frac{\gamma}{(y_t-\gamma)!}y_t^{y_t-\gamma-1}$	$y_t-\gamma$	y_t	$\{\gamma,\gamma+1,\ldots\}$
Gamma	$(\omega,\boldsymbol{\beta},\alpha)$	$\frac{\alpha^{\alpha}y_t^{\alpha-1}}{\Gamma(\alpha)}$	α	αy_t	$(0,\infty)$
Normal	$(\omega,\boldsymbol{\beta},\gamma)$	$[2\pi]^{-\frac{1}{2}}$	$\frac{1}{2}$	$\frac{(y_t-\gamma)^{-2}}{2}$	$(-\infty,\infty)$
Laplace	$(\omega,\boldsymbol{\beta},\gamma)$	$\frac{1}{\sqrt{2}}$	1	$\sqrt{2}\left\|y_t-\gamma\right\|$	$(-\infty,\infty)$
Inverse Gaussian	$(\omega,\boldsymbol{\beta},\gamma)$	$\frac{1}{\sqrt{2\pi y_t^3}}$	$\frac{1}{2}$	$\frac{(y_t-\gamma)^{-2}}{2y_t\gamma^2}$	$(0,\infty)$
Rayleigh	$(\omega,\boldsymbol{\beta},\gamma)$	y_t	1	$\frac{1}{2}\left(y_t-\gamma\right)^{-2}$	$(0,\infty)$
Generalized Gamma	$(\omega,\boldsymbol{\beta},\alpha,\upsilon)$	$\frac{\upsilon y_t^{\alpha-1}}{\Gamma\left(\frac{\alpha}{\upsilon}\right)}$	1	y_t^{υ}	$(0,\infty)$

Heavy tailed distributions.

In this paper, only the heavy tailed distributions are studied. It is important to note that the parameter vector $\boldsymbol{\varphi}$ of all models contains the parameters ω and $\boldsymbol{\beta}$. Parameter ω plays an important role in the NGSSM as it has the function of increasing multiplicatively the variance over time.

6.3 Penalized likelihood function for the NGSSM

Many papers in the literature deal with the problem of monotonicity of the likelihood function and, by consequence, the bias in the obtained estimates. In this direction it can be mentioned, among others, Cordeiro & McCullach (1991) that proposed bias correction to the estimator of the parameters of the generalized linear models (GLM); Firth

(1993) that proposed a penalized function (Jeffreys prior) for the likelihood function of the GLM to reduce the bias of parameters; Loughin (1998) that showed by Monte Carlo simulation that the likelihood function is monotone for the Cox regression and proposed a bootstrap approach to solve the problem of the classical estimation; Heinse & Schemper (2001) that also proposed a penalized function for the likelihood function in the Cox regression; Hahn & Newey (2004) and Bester & Hansen (2009) that proposed corrections for the maximum likelihood estimators of the nonlinear panel models.

The problem of monotonicity can be the case which arises in the maximum likelihood estimation of parameter ω in the NGSSM, for small samples. To investigate this assumption a broad study, including different heavy tailed distributions and maximization methods is performed. Besides, a penalty function for the likelihood function is proposed, in order to refine the estimation procedure of parameter ω.

6.3.1 Maximum Likelihood Estimator (MLE)

Classical inference for the parameters of the NGSSM can be performed through maximum likelihood estimation. The likelihood function is defined by $L_1(\boldsymbol{\varphi}; \boldsymbol{Y_n}) = \prod_{t=1}^{n} p(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi})$, where $p(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi})$ is given in equation 6.4. Then, the log-likelihood function is calculated as

$$\begin{aligned}\ell_1(\boldsymbol{\varphi}; \boldsymbol{Y_n}) &= \ln \prod_{t=1}^{n} p(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi}) \\ &= \sum_{t=1}^{n} \ln \Gamma\left(r(y_t, \boldsymbol{\varphi}) + c_{t|t-1}\right) + \sum_{t=1}^{n} \ln\left(q(y_t, \boldsymbol{\varphi})\right) - \sum_{t=1}^{n} \ln \Gamma\left(c_{t|t-1}\right) \\ &+ \sum_{t=1}^{n} c_{t|t-1} \ln\left(b_{t|t-1}\right) - \sum_{t=1}^{n} \left(r(y_t, \boldsymbol{\varphi}) + c_{t|t-1}\right) \ln\left(s(y_t, \boldsymbol{\varphi}) + d_{t|t-1}\right),\end{aligned}$$

Thus, the maximum likelihood estimator (MLE) for $\boldsymbol{\varphi}$ is given by

$$\hat{\boldsymbol{\varphi}}_{ML} = \arg\max_{\boldsymbol{\varphi}} \ell_1(\boldsymbol{\varphi}; \boldsymbol{Y_n}).$$

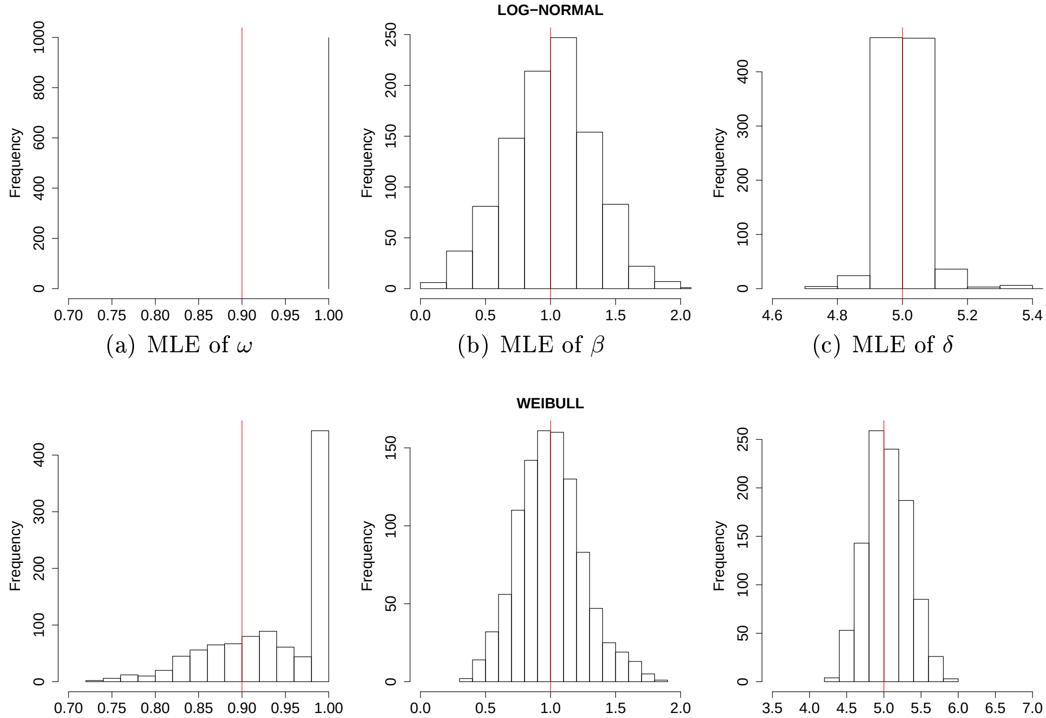
Due to the fact that $\ell_1(\boldsymbol{\varphi}; \boldsymbol{Y_n})$ is a nonlinear function of $\boldsymbol{\varphi}$, numerical procedures should be used. Santos et al. (2010) and Pinho et al. (2012) used the BFGS algorithm proposed by Broyden (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970).

Figure 6.1 presents 1000 Monte Carlo estimates for the MLE of $\boldsymbol{\varphi}$ in the NGSSM, using BFGS, for time series generated from the Log-Normal and Weibull models with size 50. It seems that this parameter is always overestimated and, in some cases, such as the Log-normal model, presents a mode in 1.00, which is the upper limit of the parameter space of ω. The results show that the adopted method presents problems only in the estimation of parameter ω. The behavior of the MLE for the Log Gamma, Pareto, Fréchet and Skew GED models (omitted here) is similar to the results presented by the Weibull model.

By the other hand, the estimation method adopted presents fewer problems as the size of the time series increases. For example, in Figure 6.2 the behavior of the estimates is very satisfactory for time series of size 200, for the same models, Log-normal and Weibull. Thus, this work has the aim of investigating this problem and to propose a solution.

The BFGS method does not impose any restriction on parameter ω. Nevertheless, this parameter should belong to the interval (0,1). Therefore the maximum likelihood estimate should be obtained through the transformation of a function f such that $f : \Re \rightarrow (0,1)$. Thus the first step is to evaluate the performance of the MLE by using other methods of maximization that allow the imposition of constraints on parameters. To this purpose, in this work it will also be used the Sequential Quadratic Programming (SQP) proposed by Nocedal & Wright (1999) and Feasible Sequential Quadratic Programming (FSQP) proposed by Lawrence & Tits (2001).

Table 6.2 presents 1000 Monte Carlo simulations for the percentage of times that the estimate of parameter ω is equal to 1.00, which is the limit of the parameter space, using BFGS, SQP and FSQP algorithms for the heavy tailed models. The real values



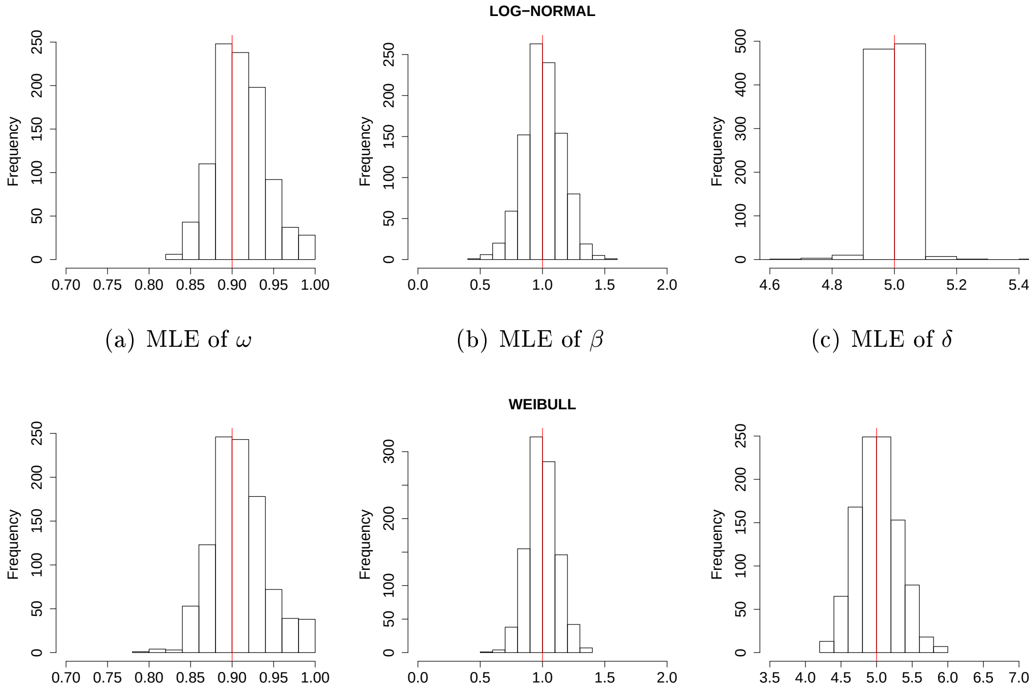
(a) MLE of ω (b) MLE of β (c) MLE of δ

(d) MLE of ω (e) MLE of β (f) MLE of υ

Figure 6.1: Histograms of 1000 estimates of the MLE, using BFGS, for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) and from the Weibull model with ($\omega = 0.90$; $\beta = 1.0$; $\upsilon = 5.0$), with size 50.

of parameter ω are 0.85, 0.90 and 0.95, for time series of size 50 and 100. It may be noted that for $n = 50$ the FSQP method presented the best performance for the Log-normal, Log-gamma, Weibull, Fréchet, Lévy, and Skew GED models, while the BFGS was better for the Pareto model. As emphasized above, it was expected that the BFGS maximization method presented worse results than FSQP and SQP because it is the only one that does not impose restrictions on the parameters.

These results are important because, whatever maximization method used, the MLE keeps presenting problems in the estimation of parameter ω. Therefore, these results justify the proposal of a penalty function for the likelihood function in order to reduce the bias in the estimation of parameter ω.



(a) MLE of ω (b) MLE of β (c) MLE of δ

(d) MLE of ω (e) MLE of β (f) MLE of υ

Figure 6.2: Histograms of 1000 estimates of the MLE, using BFGS, for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) and from the Weibull model with ($\omega = 0.90$; $\beta = 1.0$; $\upsilon = 5.0$), with size 200.

6.3.2 Penalized Maximum Likelihood Estimator

Before showing the proposed penalty function to correct the problems identified in Section 6.3.1, it is important to present the process of constructing this penalty function. After a thorough analysis of the results obtained by intensive Monte Carlo study it was noticed that:

- Except for parameter ω, the MLE for the other parameters showed good results, even for series of size 50 (see Figure 6.1);

- The maximum likelihood procedure presented some problems to estimate the real value of parameter ω when the sample size decreases. On the other hand, for

Table 6.2: Percentage of times that the maximum likelihood estimates of parameter ω is 1.00 in 1000 Monte Carlo simulations using BFGS, SQP and FSQP algorithms.

Model	ω	BFGS		SQP		FSQP	
		n=50	n=100	n=50	n=100	n=50	n=100
	0.85	1.000	0.054	0.315	0.046	0.314	0.046
LOG-NORMAL	**0.90**	1.000	0.187	0.516	0.168	0.514	0.167
	0.95	1.000	0.494	0.673	0.466	0.673	0.465
	0.85	0.435	0.183	0.273	0.071	0.273	0.071
LOG-GAMMA	**0.90**	0.630	0.317	0.392	0.145	0.392	0.145
	0.95	0.767	0.610	0.522	0.345	0.522	0.345
	0.85	0.281	0.054	0.290	0.062	0.289	0.062
PARETO	**0.90**	0.450	0.146	0.460	0.162	0.458	0.161
	0.95	0.612	0.414	0.618	0.425	0.616	0.425
	0.85	0.325	0.083	0.299	0.092	0.299	0.092
WEIBULL	**0.90**	0.534	0.208	0.460	0.205	0.460	0.205
	0.95	0.674	0.514	0.600	0.433	0.600	0.433
	0.85	0.327	0.073	0.285	0.071	0.285	0.071
FRÉCHET	**0.90**	0.550	0.176	0.483	0.157	0.483	0.157
	0.95	0.749	0.506	0.661	0.420	0.661	0.420
	0.85	0.360	0.048	0.314	0.046	0.311	0.045
LÉVY	**0.90**	0.568	0.179	0.512	0.180	0.511	0.181
	0.95	0.720	0.481	0.683	0.488	0.683	0.486
	0.85	0.301	0.059	0.304	0.055	0.304	0.052
SKEW GED	**0.90**	0.477	0.184	0.483	0.180	0.484	0.179
	0.95	0.659	0.450	0.659	0.438	0.658	0.438

large sample sizes the results are very good;

- The MLE showed the worst performance in the neighborhood of 1.00 (upper limit of the parameter space);
- Fixing the other parameters, the likelihood function increases when parameter ω increases, but this growth is very soft depending on the series.

Based on these observations, the penalty function should be such that it respects the following assumptions:

A1 The penalty function should be a function of parameter ω to influence the maximum point of the likelihood function;

A2 The penalty function should be set between 0 and 1 to have the same limits of the parameter space of ω;

A3 The penalty function should be a function of the size of the time series such that it influences the maximum point of the likelihood function only for time series of small size;

A4 The penalty function should not be a function of the other parameters of the model so that it does not influence their maximum likelihood estimates;

A5 The penalty function should have an inverse relationship to parameter ω close to 1.00. That is, the function must decrease near 1.00.

In view of the five assumptions **A1** $-$ **A5** above, the proposed penalty function, which has the aim of reducing the bias of the maximum likelihood estimator is defined as

$$v(\omega, n_1, n_2) = \frac{\Gamma(n_1 + n_2)}{\Gamma(n_1)\Gamma(n_2)} \omega^{n_1 - 1} (1 - \omega)^{n_2 - 1}, \tag{6.5}$$

where, $n_1 = \left\{ \frac{n+1}{n}, \left(\frac{n+1}{n}\right)^{\frac{1}{2}}, \left(\frac{n+1}{n}\right)^{\frac{1}{3}} \right\}$ and $n_2 = \left\{ \frac{n+1}{n}, \left(\frac{n+1}{n}\right)^{\frac{1}{2}}, \left(\frac{n+1}{n}\right)^{\frac{1}{3}} \right\}$, and n is the time series size.

It can be noted that the penalized function $v(\omega, n_1, n_2)$ is a function only of parameter ω and the time series size. Then, this function will affects directly the partial derivative of the likelihood function with respect to ω. Therefore it directly affect only the MLE of parameter ω.

Classical inference for the parameters of the NGSSM can also be performed through penalized maximum likelihood estimation. The log-penalized likelihood function is established in **Theorem 1**.

Theorem 1 Let $\{y_t\}_{t=1}^{n}$ be a time series with predictive distribution given in equation 6.4. If $v(\omega, n_1, n_2)$ is the penalty function described in equation 6.5, then the resulting log-penalized likelihood function is given by

$$
\begin{aligned}
\ell_2(\boldsymbol{\varphi}; \boldsymbol{Y_n}) = & \sum_{t=1}^{n} \ln \Gamma\left(r(y_t, \boldsymbol{\varphi}) + c_{t|t-1}\right) + \sum_{t=1}^{n} \ln\left(q(y_t, \boldsymbol{\varphi})\right) - \sum_{t=1}^{n} \ln \Gamma\left(c_{t|t-1}\right) \\
& + \sum_{t=1}^{n} c_{t|t-1} \ln\left(b_{t|t-1}\right) - \sum_{t=1}^{n} \left(r(y_t, \boldsymbol{\varphi}) + c_{t|t-1}\right) \ln\left(s(y_t, \boldsymbol{\varphi}) + d_{t|t-1}\right) \\
& + \sum_{t=1}^{n} \ln\left(\Gamma(n_1 + n_2)\right) - \sum_{t=1}^{n} \ln\left(\Gamma(n_1)\right) + \sum_{t=1}^{n} (n_1 - 1) \ln(\omega) \\
& + \sum_{t=1}^{n} (n_2 - 1) \ln(1 - \omega).
\end{aligned}
$$

Proof The proof is readily attained by multiplying the likelihood function, $L_1(\boldsymbol{\varphi}; \boldsymbol{Y_n})$ by the penalty function, $v(\omega, n_1, n_2)$.

Thus, the penalized maximum likelihood estimator (PMLE) for $\boldsymbol{\varphi}$ is given by

$$
\hat{\boldsymbol{\varphi}}_{PMLE} = \arg\max_{\boldsymbol{\varphi}} \ell_2(\boldsymbol{\varphi}; \boldsymbol{Y_n}).
$$

It should be noted that $\ell_2(\boldsymbol{\varphi}; \boldsymbol{Y_n})$ is also a nonlinear function of $\boldsymbol{\varphi}$, then the BFGS, SQP and FSQP algorithms of maximization should be used.

Table 6.3 shows nine different combinations of n_1 and n_2, where n_1 and n_2 are defined in equation 6.5 and n is the size of the time series. By consequence, nine penalty functions are obtained.

Table 6.3: Values of n_1 and n_2 for the penalized function $v(\omega, n_1, n_2)$.

PMLE	I	II	III	IV	V	VI	VII	VIII	IX
n_1	$\left(\frac{n+1}{n}\right)^{\frac{1}{3}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{3}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{3}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{2}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{2}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{2}}$	$\frac{n+1}{n}$	$\frac{n+1}{n}$	$\frac{n+1}{n}$
n_2	$\left(\frac{n+1}{n}\right)^{\frac{1}{3}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{2}}$	$\frac{n+1}{n}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{3}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{2}}$	$\frac{n+1}{n}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{3}}$	$\left(\frac{n+1}{n}\right)^{\frac{1}{2}}$	$\frac{n+1}{n}$

In Figure 6.3 it can be observed the behavior of some penalization functions (I, IV and VII) for time series of size 50, 100, 200 and 500. It is easy to see that function $v(\omega, n_1, n_2)$ is defined in the interval (0,1) and it is a decreasing function when the values of ω approach 1.00. Therefore, it will influence the maximum likelihood estimates of

ω as desired. It can also be observed that $v(\omega, n_1, n_2)$ is a function of the time series size, and for large n the function approaches a uniform function. Therefore, when n is large it will not influence the maximum likelihood estimates of ω, as desired.

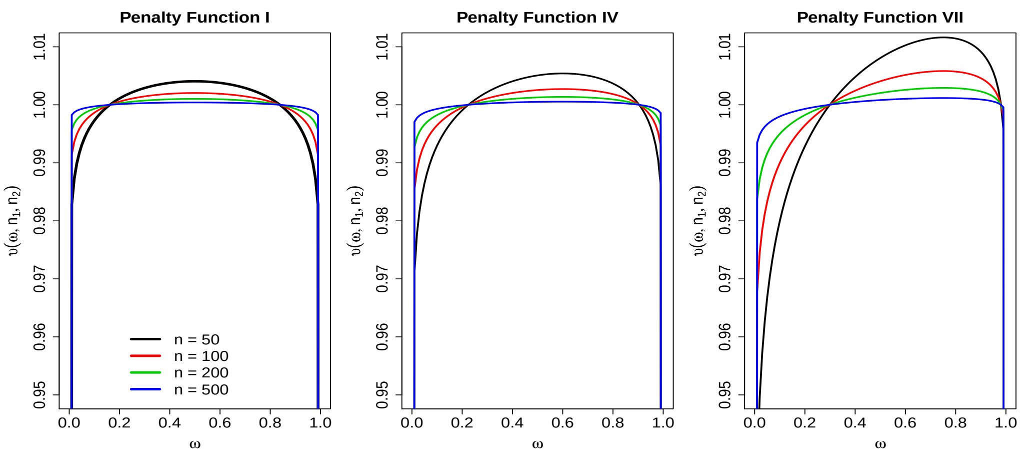


Figure 6.3: Penalty functions I (at left), IV (at center) and VII (at right) proposed to time series of size 50, 100, 200 and 500.

6.4 Monte Carlo study

In this section the performance of the penalized function in the MLE of the distributions presented in Table 6.2 is evaluated. To this purpose a broad Monte Carlo study was conducted with the nine penalized maximum likelihood estimators (PMLE) defined in Table 6.3.

All codes for NGSSM were developed by the authors in Ox Metrics.

The number of Monte Carlo replications was set equal to 1,000 for time series of size $n = \{50, 100\}$, generated with a covariate $x_t = \sin(2\pi t/12)$, $t = 1, \ldots, n$. For all distributions $\omega = (0.85, 0.90, 0.95)$ and the coefficient of the covariate is $\beta = 1.0$.

Specific parameters were set as follows: Log-normal ($\delta = 5.0$), Log-gamma ($\alpha = 5.0$), Fréchet ($\alpha = 5.0$), Skew GED ($\delta = 5.0$, $\alpha = 1.5$, $\kappa = 1.0$) and Weibull ($\upsilon = 5.0$). For the Log-normal, Fréchet and Lévy models the parameter γ was fixed at 0.0. For the

Skew GED model the parameter α was fixed at 1.5, thus, there is a distribution with a tail heavier than the Skew Normal ($\alpha = 2.0$) and lighter than the Skew Laplace (both are particular cases of the Skew GED).

To calculate the maximum likelihood estimator, the BFGS, SQP and FSQP assumed, as initial state condition $\lambda_0 | Y_0 \sim \text{Gamma}\,(0.01; 0.01)$, $\omega_0 = 0.50$ and $\beta_0 = \delta_0 = \alpha_0 = \upsilon_0 = \kappa_0 = 0.01$.

The estimates of MLE and PMLE by FSQP and SQP are nearly equal, then in this work only the results of MLE and PMLE estimates by SQP and BFGS will be presented.

Figures 6.4 and 6.5 present the reduction of bias and mean square error (MSE), in percentage, of the penalized function with respect to the MLE, for the BFGS and SQP methods, respectively. It is easy to see that all of the penalized estimators are able to reduce significantly the bias and MSE compared to the MLE for $\omega = (0.85, 0.90)$ in all models and time series sizes 50 and 100. For $\omega = 0.95$, only the PMLE I, IV and VII were able to reduce the bias and MSE. Thus, the next results are presented considering only these three functions.

Figures 6.6 and 6.7 present the boxplot of the MLE, PMLE I, PMLE IV and PMLE VII when parameter $\omega = 0.95$. It is easy to see that for all models the penalized estimators show results significantly better than the MLE, regardless the method of maximization used.

It is also interesting to note that the behavior of the MLE for the BFGS and SQP are different in the Log-normal and Log-gamma models. However, the behavior of the penalized estimators is robust with respect to the maximization algorithm used.

Tables 6.4, 6.5, 6.6 and 6.7 present, for time series of sizes 50 and 100, the bias and MSE for 1000 Monte Carlo estimates of MLE and nine different PMLE of ω by BFGS and SQP according to the default values of n_1 and n_2 showed in Table 6.3.

Except for a very few cases (showed in bold in the tables) the PMLE was able to substantially reduce the bias and MSE of the estimates of ω. The only case in which the PMLE was not able to improve the estimates was the Log-gamma with SQP and $\omega = 0.95$.

Tables 6.8, 6.9, 6.10 and 6.11 show, for time series of size 50 and 100, the estimates and MSE of parameter vector $\boldsymbol{\varphi}$ in the NGSSM, for 1000 Monte Carlo replication using MLE and PMLE I, IV and VII by BFGS and SQP. It is worth noting that the penalized functions can improve the estimates of ω without affecting the other parameters of $\boldsymbol{\varphi}$.

In Table 6.12 it is possible to analyze the asymptotic confidence intervals of the parameter vector $\boldsymbol{\varphi}$ obtained by the MLE and three different penalized estimators (PMLE I, PMLE IV and PMLE VII) for time series of size 50. It is easy to see that the coverage rates of the asymptotic confidence intervals for parameter ω obtained by the penalized estimators are better than the obtained by MLE, as they are closer to the nominal coverage level of 0.95. Therefore, the penalty function also improved the interval estimates of parameter ω. However, despite the improvement and except for the Log-gamma model that already had a coverage rate close to 0.95, all other coverage rates remain above the nominal rate.

It is necessary to highlight some unsatisfactory results regarding the confidence intervals. First, the coverage rates of the parameter ω for the Lévy model are very close to 1.00. Second, the coverage rates of parameters δ and κ for the Log-normal model are far below the nominal coverage rate of 0.95.

An alternative refinement of the confidence intervals can be achieved by bootstrap methods and the various types of bootstrap intervals.

6.5 Conclusion

This paper proposes methods of refining point estimation of parameter ω in the NGSSM for time series of small sizes, using a penalized likelihood function.

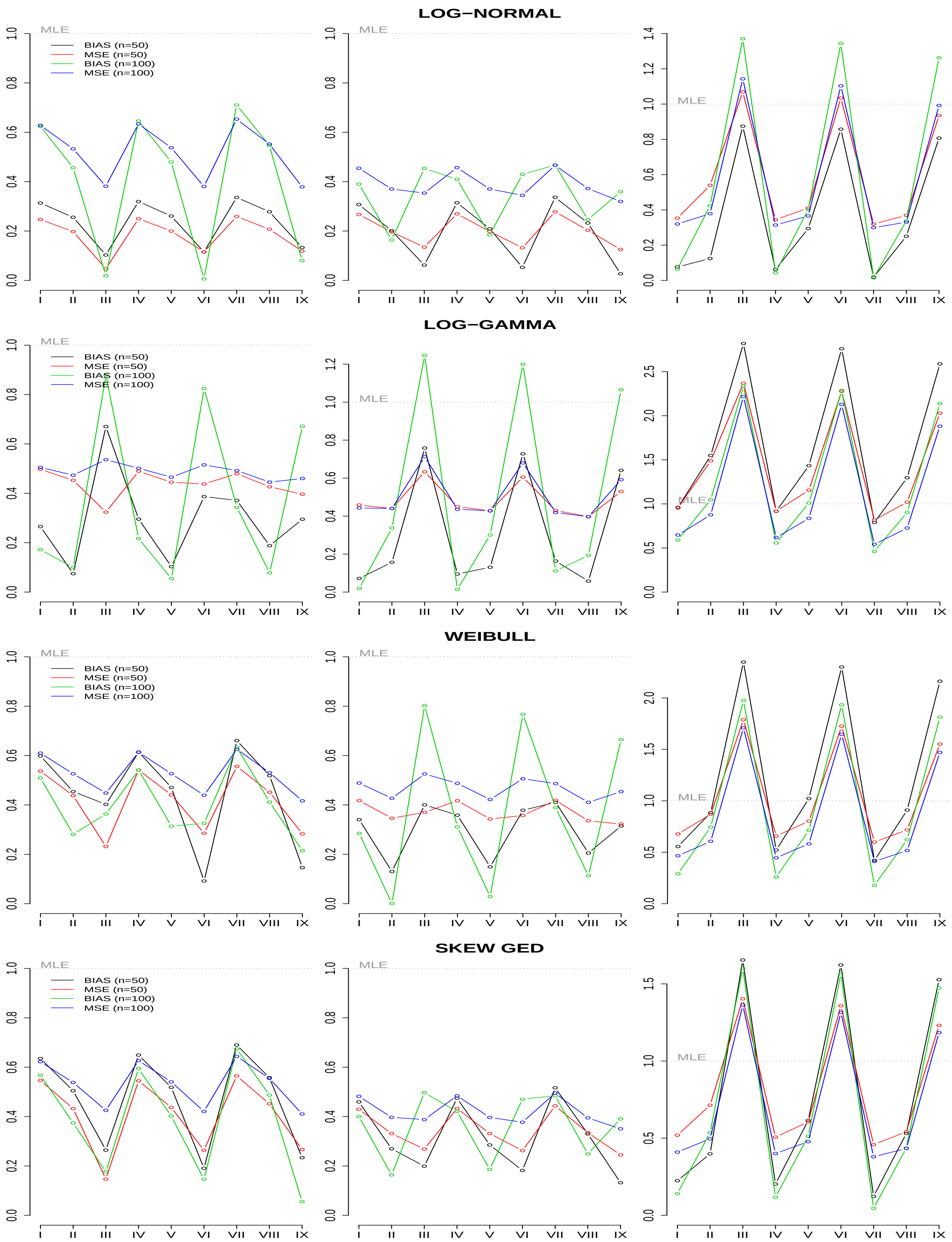


Figure 6.4: Percentage of bias and MSE of PMLE over the MLE, by BFGS, for the Log-normal, Log-gamma, Weibull and Skew GED models for $\omega = 0.85$ (at left), $\omega = 0.90$ (at center) and $\omega = 0.95$ (at right).

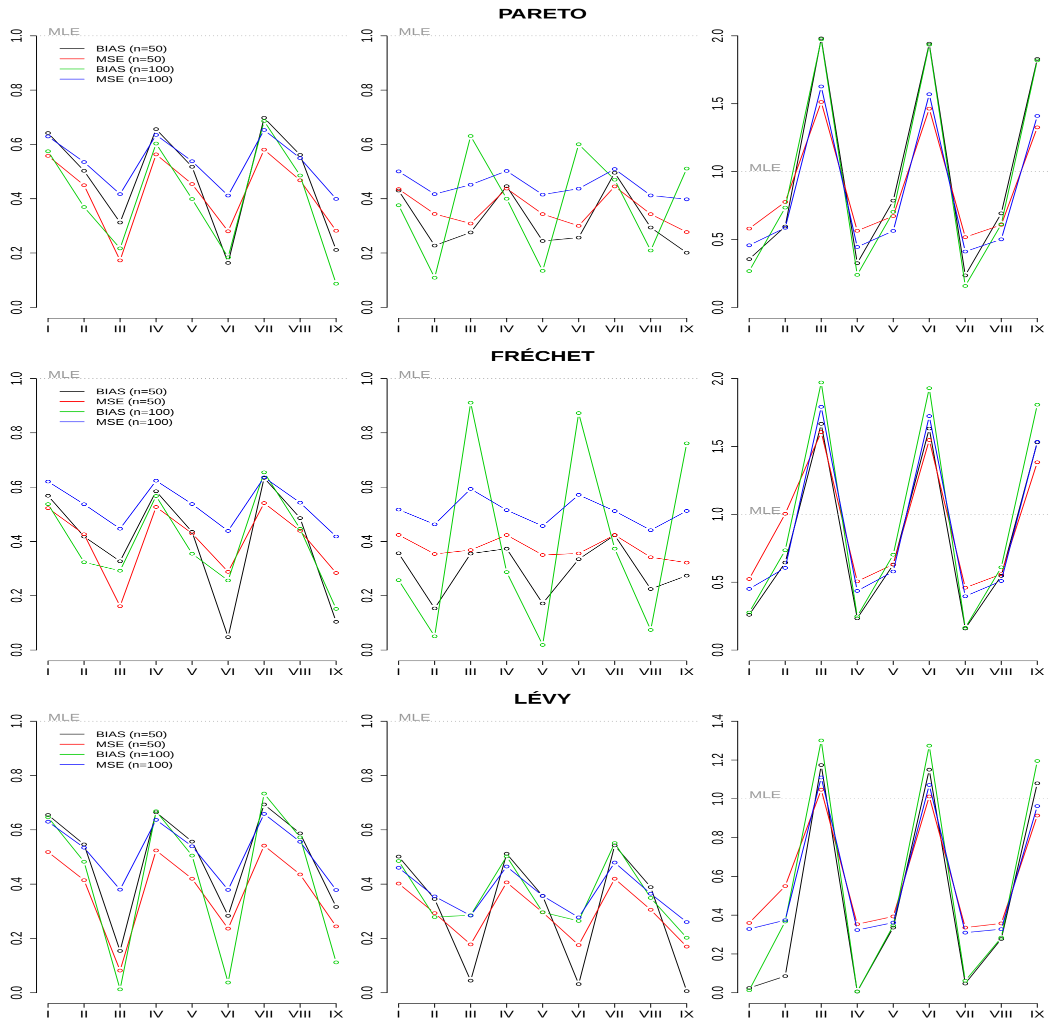


Figure 6.5: Percentage of bias and MSE of PMLE over the MLE, by BFGS, for the Pareto, Fréchet and Lévy models for $\omega = 0.85$ (at left), $\omega = 0.90$ (at center) and $\omega = 0.95$ (at right).

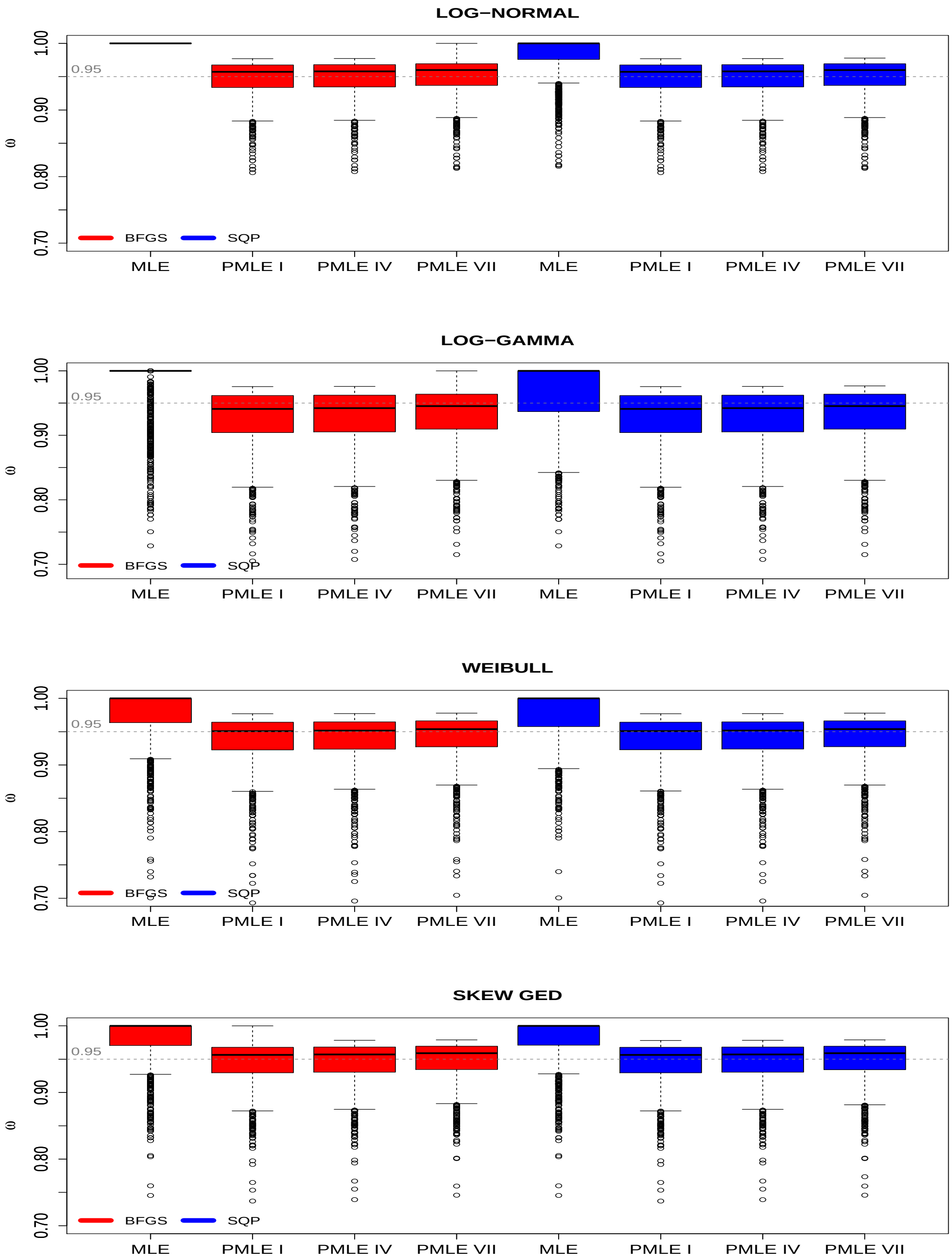


Figure 6.6: Boxplot of the 1000 estimates (MLE, PMLE I, PMLE IV and PMLE VII) for $\omega = 0.95$, by BFGS and SQP, for time series of size 50 and for Log-normal, Pareto, Weibull and Skew GED models.

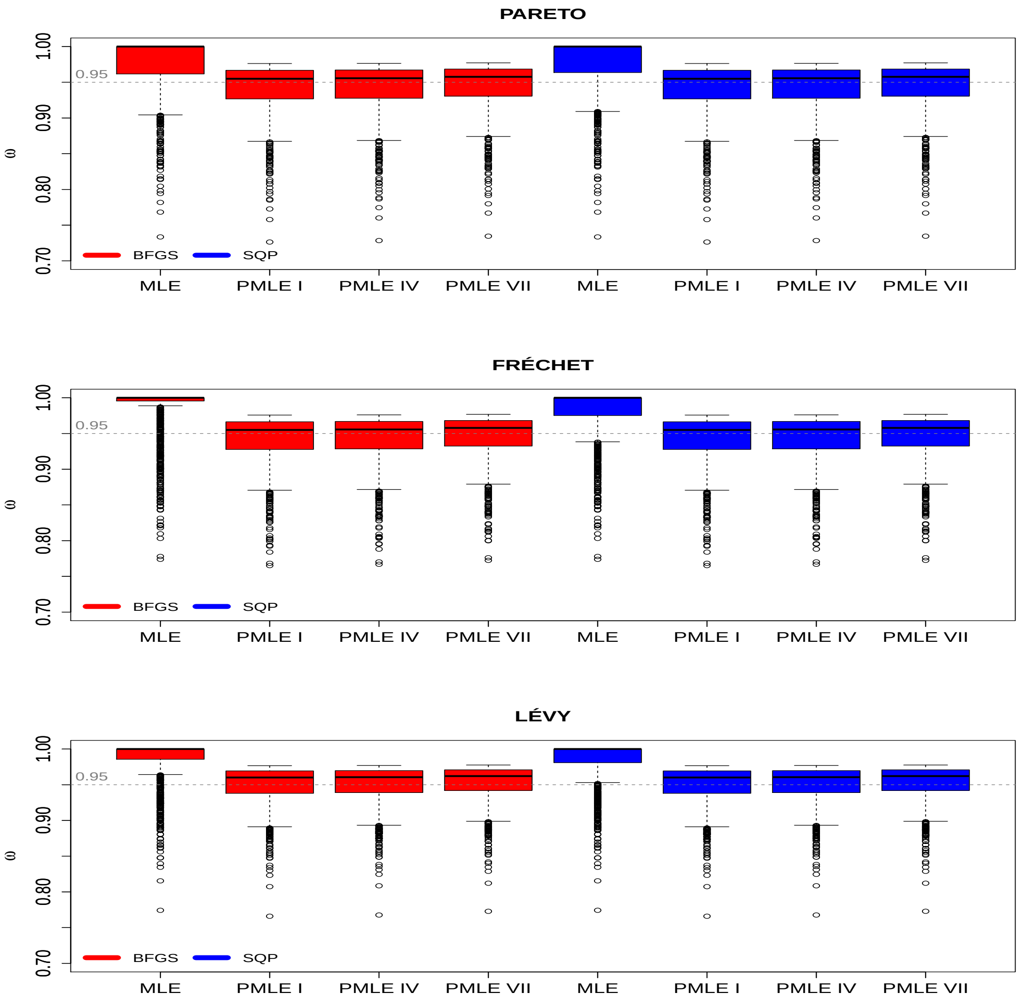


Figure 6.7: Boxplot of the 1000 estimates (MLE, PMLE I, PMLE IV and PMLE VII) for $\omega = 0.95$, by BFGS and SQP, for time series of size 50 and for Log-normal, Pareto, Weibull and Skew GED models.

Table 6.4: Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Log-normal and Log-gamma models).

Model	ω	MLE BIAS (MSE)	I BIAS (MSE)	II BIAS (MSE)	III BIAS (MSE)	IV BIAS (MSE)	V BIAS (MSE)	VI BIAS (MSE)	VII BIAS (MSE)	VIII BIAS (MSE)	IX BIAS (MSE)
n=50	**0.85**	15.000	4.699	3.836	-1.544	4.794	3.910	1.729	5.042	4.172	2.001
		(2.250)	(0.555)	(0.445)	(0.106)	(0.562)	(0.450)	(0.260)	(0.582)	(0.466)	(0.267)
LN	**0.90**	10.000	3.073	2.014	-0.621	3.147	2.093	-0.531	3.365	2.324	-0.268
BFGS		(1.000)	(0.267)	(0.196)	(0.135)	(0.270)	(0.198)	(0.132)	(0.278)	(0.203)	(0.125)
	0.95	5.000	-0.381	-0.621	-4.373	-0.311	-1.471	-4.287	-0.093	-1.251	-4.034
		(0.250)	(0.088)	(0.135)	**(0.267)**	(0.086)	(0.103)	**(0.259)**	(0.080)	(0.092)	(0.234)
	0.85	7.132	4.711	3.825	1.637	4.794	3.910	1.730	5.036	4.171	2.001
		(1.024)	(0.556)	(0.444)	(0.258)	(0.562)	(0.450)	(0.260)	(0.581)	(0.466)	(0.267)
LN	**0.90**	6.202	3.073	2.014	-0.621	3.147	2.093	-0.531	3.365	2.324	-0.268
SQP		(0.649)	(0.267)	(0.196)	(0.135)	(0.270)	(0.198)	(0.132)	(0.278)	(0.203)	(0.125)
	0.95	3.162	-0.381	-1.544	**-4.373**	-0.311	-1.471	**-4.287**	-0.107	-1.255	**-4.034**
		(0.217)	(0.088)	(0.106)	**(0.267)**	(0.086)	(0.103)	**(0.259)**	(0.081)	(0.092)	**(0.234)**
n=100	**0.85**	3.216	2.006	1.469	-0.061	2.077	1.542	0.020	2.284	1.759	0.259
		(0.379)	(0.238)	(0.202)	(0.144)	(0.240)	(0.204)	(0.144)	(0.247)	(0.209)	(0.143)
LN	**0.90**	3.250	1.267	0.533	-1.475	1.331	0.601	-1.397	1.518	0.801	-1.166
BFGS		(0.321)	(0.146)	(0.119)	(0.114)	(0.147)	(0.119)	(0.111)	(0.150)	(0.120)	(0.103)
	0.95	2.717	-0.173	-1.151	**-3.725**	-0.119	-1.091	**-3.650**	0.039	-0.918	**-3.429**
		(0.159)	(0.051)	(0.060)	**(0.182)**	(0.050)	(0.058)	**(0.176)**	(0.048)	(0.053)	(0.158)
	0.85	3.138	2.496	2.209	1.387	2.530	2.244	1.424	2.633	2.348	1.536
		(0.370)	(0.283)	(0.257)	(0.200)	(0.284)	(0.258)	(0.201)	(0.288)	(0.262)	(0.203)
LN	**0.90**	3.151	2.012	1.589	0.453	2.042	1.621	0.488	2.133	1.715	0.591
SQP		(0.310)	(0.193)	(0.165)	(0.119)	(0.193)	(0.166)	(0.119)	(0.196)	(0.168)	(0.119)
	0.95	2.662	0.910	0.306	-1.225	0.935	0.332	-1.194	1.010	0.410	-1.104
		(0.157)	(0.067)	(0.055)	(0.063)	(0.067)	(0.055)	(0.062)	(0.067)	(0.054)	(0.058)
n=50	**0.85**	5.799	1.538	0.432	-3.886	1.713	0.598	-2.242	2.153	1.090	-1.708
		(1.278)	(0.635)	(0.579)	(0.414)	(0.626)	(0.569)	(0.560)	(0.613)	(0.545)	(0.507)
LG	**0.90**	5.354	0.384	-0.838	-4.059	0.508	-0.699	-3.897	0.874	-0.309	-3.431
BFGS		(0.808)	(0.371)	(0.356)	(0.512)	(0.363)	(0.346)	(0.489)	(0.347)	(0.321)	(0.428)
	0.95	2.622	-2.516	**-4.059**	**-7.391**	-2.402	**-3.760**	**-7.235**	-2.070	**-3.400**	**-6.786**
		(0.344)	(0.328)	**(0.512)**	**(0.814)**	(0.315)	**(0.397)**	**(0.783)**	(0.281)	**(0.351)**	**(0.697)**
	0.85	4.506	1.538	0.432	-2.427	1.713	0.598	-2.242	2.153	1.090	-1.708
		(1.040)	(0.635)	(0.579)	(0.580)	(0.626)	(0.569)	(0.560)	(0.613)	(0.545)	(0.507)
LG	**0.90**	3.933	0.381	-0.838	**-4.059**	0.508	-0.699	-3.897	0.890	-0.309	-3.431
SQP		(0.647)	(0.370)	(0.356)	(0.512)	(0.363)	(0.346)	(0.489)	(0.345)	(0.321)	(0.428)
	0.95	1.292	**-2.516**	**-3.886**	**-7.391**	**-2.402**	**-3.760**	**-7.235**	**-2.073**	**-3.400**	**-6.786**
		(0.341)	(0.328)	**(0.414)**	**(0.814)**	(0.315)	**(0.397)**	**(0.783)**	(0.281)	**(0.351)**	**(0.697)**
n=100	**0.85**	3.101	0.532	-0.309	-2.721	0.672	-0.168	-2.557	1.064	0.241	-2.083
		(0.637)	(0.322)	(0.302)	(0.342)	(0.319)	(0.296)	(0.328)	(0.313)	(0.284)	(0.293)
LG	**0.90**	3.053	-0.060	-1.030	**-3.803**	0.043	-0.916	**-3.662**	0.339	-0.589	**-3.253**
BFGS		(0.441)	(0.195)	(0.194)	(0.314)	(0.192)	(0.189)	(0.300)	(0.185)	(0.175)	(0.261)
	0.95	2.460	-1.451	**-2.577**	**-5.748**	-1.370	**-2.486**	**-5.621**	-1.136	-2.223	**-5.259**
		(0.202)	(0.131)	(0.177)	**(0.448)**	(0.125)	(0.169)	**(0.430)**	(0.109)	(0.147)	**(0.380)**
	0.85	2.296	1.304	0.847	-0.465	1.369	0.914	-0.392	1.560	1.111	-0.178
		(0.482)	(0.369)	(0.344)	(0.307)	(0.369)	(0.343)	(0.304)	(0.368)	(0.340)	(0.296)
LG	**0.90**	2.191	0.857	0.336	-1.164	0.905	0.386	-1.101	1.046	0.533	-0.929
SQP		(0.332)	(0.225)	(0.208)	(0.200)	(0.224)	(0.207)	(0.197)	(0.222)	(0.203)	(0.189)
	0.95	1.588	-0.290	-0.940	-2.685	-0.254	-0.901	-2.638	-0.126	-0.785	-2.500
		(0.159)	(0.117)	(0.123)	(0.186)	(0.115)	(0.120)	(0.182)	(0.105)	(0.113)	(0.170)

Obs.: Bias and MSE are multiplied by $\times 10^2$ and in bold are the cases which the PMLE do not decrease the bias or MSE.

A comparison of methods for maximization, which includes BFGS, SQP and FSQP is performed to verify if the problem of estimating parameter ω is related to the maximization method used.

The results showed that the penalty function improves significantly the estimates of parameter ω. In particular, the estimators PMLE I, PMLE IV and PMLE VII showed lower bias and MSE for all models and time series of size 50 and 100 and $\omega = (0.85, 0.90, 0.95)$.

Table 6.5: Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Pareto and Weibull models).

Model	ω	MLE BIAS (MSE)	I BIAS (MSE)	II BIAS (MSE)	III BIAS (MSE)	IV BIAS (MSE)	V BIAS (MSE)	VI BIAS (MSE)	VII BIAS (MSE)	VIII BIAS (MSE)	IX BIAS (MSE)
n=50	**0.85**	6.608	4.241	3.322	-2.064	4.336	3.420	1.080	4.612	3.709	1.396
		(0.975)	(0.544)	(0.438)	(0.168)	(0.549)	(0.442)	(0.273)	(0.566)	(0.456)	(0.275)
P	**0.90**	5.445	2.341	1.238	-1.501	2.429	1.330	-1.398	2.699	1.601	-1.095
BFGS		(0.609)	(0.265)	(0.209)	(0.188)	(0.266)	(0.209)	(0.183)	(0.271)	(0.209)	(0.169)
	0.95	2.523	-0.895	-1.501	**-4.998**	-0.820	-1.983	**-4.900**	-0.593	-1.746	**-4.615**
		(0.242)	(0.140)	(0.188)	**(0.366)**	(0.136)	(0.163)	**(0.354)**	(0.125)	(0.147)	**(0.321)**
	0.85	6.722	4.241	3.321	0.972	4.336	3.420	1.080	4.611	3.708	1.396
		(0.994)	(0.544)	(0.438)	(0.273)	(0.549)	(0.442)	(0.273)	(0.566)	(0.456)	(0.275)
P	**0.90**	5.557	2.341	1.238	-1.502	2.429	1.330	-1.398	2.699	1.600	-1.095
SQP		(0.617)	(0.265)	(0.209)	(0.188)	(0.266)	(0.209)	(0.183)	(0.271)	(0.209)	(0.169)
	0.95	2.591	-0.895	-2.063	**-4.998**	-0.820	-1.983	**-4.901**	-0.593	-1.746	**-4.616**
		(0.240)	(0.140)	(0.168)	**(0.366)**	(0.136)	(0.163)	**(0.354)**	(0.125)	(0.147)	**(0.321)**
n=100	**0.85**	2.982	1.714	1.101	-0.647	1.798	1.189	-0.548	2.046	1.448	-0.259
		(0.367)	(0.231)	(0.196)	(0.153)	(0.233)	(0.197)	(0.151)	(0.240)	(0.202)	(0.146)
P	**0.90**	3.014	1.133	0.328	-1.902	1.205	0.405	-1.810	1.417	0.630	-1.540
BFGS		(0.303)	(0.152)	(0.126)	(0.137)	(0.152)	(0.126)	(0.132)	(0.154)	(0.125)	(0.120)
	0.95	2.191	-0.584	-1.610	**-4.326**	-0.523	-1.541	**-4.239**	-0.344	-1.341	**-3.986**
		(0.154)	(0.070)	(0.090)	**(0.250)**	(0.068)	(0.086)	**(0.241)**	(0.063)	(0.077)	**(0.217)**
	0.85	3.071	2.280	1.941	1.003	2.321	1.983	1.048	2.444	2.108	1.181
		(0.384)	(0.279)	(0.251)	(0.195)	(0.280)	(0.252)	(0.195)	(0.284)	(0.255)	(0.197)
P	**0.90**	3.139	1.935	1.481	0.238	1.969	1.517	0.278	2.070	1.621	0.394
SQP		(0.317)	(0.198)	(0.171)	(0.127)	(0.199)	(0.171)	(0.126)	(0.201)	(0.173)	(0.125)
	0.95	2.271	0.515	-0.094	-1.694	0.542	-0.065	-1.659	0.630	0.022	-1.555
		(0.154)	(0.078)	(0.071)	(0.094)	(0.077)	(0.070)	(0.092)	(0.076)	(0.068)	(0.087)
n=50	**0.85**	6.556	3.923	2.977	-2.637	4.026	3.085	0.601	4.330	3.399	0.958
		(1.005)	(0.540)	(0.440)	(0.233)	(0.544)	(0.443)	(0.287)	(0.559)	(0.453)	(0.284)
W	**0.90**	5.493	1.871	0.715	-2.199	1.969	0.820	-2.079	2.256	1.127	-1.729
BFGS		(0.656)	(0.274)	(0.227)	(0.243)	(0.274)	(0.225)	(0.234)	(0.274)	(0.220)	(0.211)
	0.95	2.485	-1.381	-2.199	**-5.836**	-1.295	**-2.541**	**-5.718**	-1.046	-2.261	**-5.373**
		(0.279)	(0.189)	(0.243)	**(0.499)**	(0.183)	(0.224)	**(0.482)**	(0.167)	(0.201)	**(0.433)**
	0.85	6.546	3.923	2.977	0.478	4.026	3.085	0.601	4.330	3.399	0.958
		(0.992)	(0.540)	(0.440)	(0.288)	(0.544)	(0.443)	(0.287)	(0.559)	(0.453)	(0.284)
W	**0.90**	5.350	1.871	0.715	-2.199	1.990	0.820	-2.079	2.276	1.127	-1.729
SQP		(0.623)	(0.274)	(0.227)	(0.243)	(0.270)	(0.225)	(0.234)	(0.271)	(0.220)	(0.211)
	0.95	2.366	-1.359	**-2.637**	**-5.836**	-1.274	**-2.541**	**-5.718**	-1.026	-2.243	**-5.373**
		(0.258)	(0.185)	(0.233)	**(0.499)**	(0.179)	(0.224)	**(0.482)**	(0.164)	(0.196)	**(0.433)**
n=100	**0.85**	3.166	1.612	0.890	-1.150	1.713	0.995	-1.030	2.006	1.303	-0.681
		(0.463)	(0.283)	(0.244)	(0.208)	(0.284)	(0.244)	(0.203)	(0.290)	(0.246)	(0.193)
W	**0.90**	3.014	0.858	0.001	-2.419	0.938	0.088	-2.313	1.172	0.340	-2.003
BFGS		(0.340)	(0.166)	(0.145)	(0.179)	(0.166)	(0.144)	(0.172)	(0.165)	(0.140)	(0.154)
	0.95	2.384	-0.693	-1.774	**-4.709**	-0.618	-1.700	**-4.610**	-0.424	-1.483	**-4.323**
		(0.172)	(0.080)	(0.104)	**(0.294)**	(0.077)	(0.100)	**(0.283)**	(0.071)	(0.089)	**(0.253)**
	0.85	2.366	-1.359	-2.637	-5.836	-1.274	-2.541	-5.718	-1.026	-2.243	-5.373
		(0.258)	(0.185)	(0.233)	(0.499)	(0.179)	(0.224)	(0.482)	(0.164)	(0.196)	(0.433)
W	**0.90**	3.186	1.704	1.223	-0.099	1.747	1.262	-0.054	1.861	1.379	0.077
SQP		(0.352)	(0.210)	(0.185)	(0.147)	(0.211)	(0.185)	(0.146)	(0.212)	(0.185)	(0.143)
	0.95	2.300	0.474	-0.167	-1.865	0.503	-0.135	-1.827	0.589	-0.042	-1.714
		(0.162)	(0.086)	(0.079)	(0.109)	(0.085)	(0.078)	(0.107)	(0.084)	(0.076)	(0.101)

Obs.: Bias and MSE are multiplied by $\times 10^2$ and in bold are the cases which the PMLE do not decrease the bias or MSE.

Some other important results were observed. First the MLE using BFGS presented worse results than SQP and FSQP in the estimation of ω. Second the penalized estimators are robust with respect to the maximization method used. Third the penalized estimators are also able to slightly improve the results of the asymptotic confidence interval for ω.

Future research includes further evaluation on the performance of the maximization methods (computational time, bias and MSE) for the parameters of NGSSM in large

Table 6.6: Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Fréchet and Lévy models).

Model	ω	MLE BIAS (MSE)	I BIAS (MSE)	II BIAS (MSE)	III BIAS (MSE)	IV BIAS (MSE)	V BIAS (MSE)	VI BIAS (MSE)	VII BIAS (MSE)	VIII BIAS (MSE)	IX BIAS (MSE)
n=50	**0.85**	6.370	3.619	2.661	-2.082	3.725	2.772	0.302	4.039	3.095	0.662
		(1.021)	(0.533)	(0.435)	(0.165)	(0.538)	(0.438)	(0.294)	(0.553)	(0.448)	(0.290)
F	**0.90**	5.747	2.048	0.881	-2.042	2.145	0.985	-1.923	2.430	1.292	-1.574
BFGS		(0.684)	(0.290)	(0.242)	(0.252)	(0.290)	(0.239)	(0.243)	(0.290)	(0.234)	(0.220)
	0.95	3.169	-0.822	-2.042	**-5.281**	-0.740	-1.993	**-5.170**	-0.501	-1.732	**-4.848**
		(0.251)	(0.131)	**(0.252)**	**(0.403)**	(0.127)	(0.158)	**(0.388)**	(0.115)	(0.140)	**(0.347)**
	0.85	6.313	3.619	2.662	0.180	3.725	2.772	0.302	4.039	3.095	0.662
		(0.995)	(0.534)	(0.435)	(0.296)	(0.538)	(0.438)	(0.294)	(0.553)	(0.448)	(0.290)
F	**0.90**	5.671	2.048	0.881	-2.041	2.145	0.985	-1.922	2.430	1.292	-1.574
SQP		(0.655)	(0.290)	(0.242)	(0.252)	(0.290)	(0.239)	(0.243)	(0.290)	(0.234)	(0.220)
	0.95	2.918	-0.822	-2.082	**-5.281**	-0.740	-1.993	**-5.170**	-0.501	-1.732	**-4.848**
		(0.239)	(0.131)	(0.165)	**(0.403)**	(0.127)	(0.158)	**(0.388)**	(0.115)	(0.140)	**(0.347)**
n=100	**0.85**	3.264	1.754	1.056	-0.953	1.852	1.158	-0.836	2.137	1.457	-0.495
		(0.452)	(0.280)	(0.243)	(0.202)	(0.282)	(0.243)	(0.198)	(0.287)	(0.245)	(0.189)
F	**0.90**	2.768	0.712	-0.140	-2.521	0.794	-0.052	-2.415	1.033	0.205	-2.107
BFGS		(0.322)	(0.166)	(0.149)	(0.191)	(0.166)	(0.147)	(0.184)	(0.165)	(0.142)	(0.165)
	0.95	2.404	-0.664	-1.767	**-4.738**	-0.596	-1.688	**-4.637**	-0.393	-1.465	**-4.344**
		(0.166)	(0.075)	(0.100)	**(0.298)**	(0.072)	(0.096)	**(0.286)**	(0.066)	(0.084)	**(0.255)**
	0.85	3.332	2.395	2.014	0.942	2.442	2.063	0.994	2.584	2.207	1.149
		(0.461)	(0.334)	(0.303)	(0.242)	(0.335)	(0.304)	(0.242)	(0.339)	(0.307)	(0.242)
F	**0.90**	2.824	1.571	1.075	-0.241	1.609	1.115	-0.196	1.723	1.235	-0.063
SQP		(0.323)	(0.208)	(0.183)	(0.151)	(0.208)	(0.183)	(0.150)	(0.209)	(0.183)	(0.147)
	0.95	2.326	0.502	-0.136	-1.862	0.531	-0.104	-1.822	0.616	-0.011	-1.703
		(0.155)	(0.080)	(0.074)	(0.106)	(0.080)	(0.073)	(0.103)	(0.078)	(0.071)	(0.097)
n=50	**0.85**	7.960	5.213	4.349	-1.226	5.291	4.431	2.255	5.520	4.672	2.516
		(1.112)	(0.576)	(0.461)	(0.091)	(0.583)	(0.467)	(0.263)	(0.603)	(0.485)	(0.272)
L	**0.90**	6.634	3.327	2.290	-0.295	3.396	2.364	-0.210	3.599	2.580	0.039
BFGS		(0.690)	(0.277)	(0.202)	(0.123)	(0.280)	(0.204)	(0.121)	(0.290)	(0.211)	(0.117)
	0.95	3.444	-0.084	-0.295	**-4.043**	-0.021	-1.157	**-3.961**	0.163	-0.955	**-3.719**
		(0.223)	(0.080)	(0.123)	**(0.234)**	(0.079)	(0.088)	**(0.226)**	(0.075)	(0.080)	(0.204)
	0.85	7.554	5.213	4.349	2.167	5.290	4.431	2.255	5.519	4.671	2.516
		(1.034)	(0.576)	(0.461)	(0.260)	(0.583)	(0.467)	(0.263)	(0.603)	(0.485)	(0.272)
L	**0.90**	6.304	3.327	2.290	-0.295	3.396	2.364	-0.211	3.599	2.580	0.039
SQP		(0.651)	(0.277)	(0.202)	(0.123)	(0.280)	(0.204)	(0.121)	(0.290)	(0.211)	(0.117)
	0.95	3.255	-0.084	-1.226	**-4.044**	-0.021	-1.157	**-3.961**	0.163	-0.955	**-3.720**
		(0.217)	(0.080)	(0.091)	**(0.234)**	(0.079)	(0.088)	**(0.226)**	(0.075)	(0.080)	(0.204)
n=100	**0.85**	3.169	2.049	1.529	0.039	2.119	1.601	0.119	2.325	1.814	0.354
		(0.342)	(0.216)	(0.183)	(0.130)	(0.218)	(0.185)	(0.130)	(0.226)	(0.190)	(0.130)
L	**0.90**	3.637	1.765	1.013	-1.038	1.826	1.079	-0.962	2.006	1.270	-0.736
BFGS		(0.325)	(0.150)	(0.115)	(0.092)	(0.151)	(0.116)	(0.090)	(0.156)	(0.119)	(0.085)
	0.95	2.777	-0.036	-1.020	**-3.612**	0.017	-0.961	**-3.538**	0.171	-0.789	**-3.319**
		(0.159)	(0.052)	(0.059)	**(0.176)**	(0.051)	(0.057)	**(0.170)**	(0.049)	(0.052)	(0.153)
	0.85	3.142	2.529	2.244	1.449	2.564	2.279	1.485	2.667	2.383	1.594
		(0.337)	(0.258)	(0.233)	(0.181)	(0.259)	(0.234)	(0.182)	(0.264)	(0.238)	(0.184)
L	**0.90**	3.633	2.529	2.098	0.937	2.558	2.128	0.970	2.645	2.218	1.069
SQP		(0.325)	(0.204)	(0.172)	(0.114)	(0.205)	(0.173)	(0.115)	(0.208)	(0.176)	(0.116)
	0.95	2.800	1.047	0.443	-1.094	1.071	0.469	-1.063	1.141	0.545	-0.974
		(0.160)	(0.070)	(0.058)	(0.062)	(0.070)	(0.057)	(0.061)	(0.070)	(0.057)	(0.058)

Obs.: Bias and MSE are multiplied by $\times 10^2$ and in bold are the cases which the PMLE do not decrease the bias or MSE.

series. This is interesting because in simulation studies and real applications showed in Pinho et al. (2012) and Santos et al. (2010) only the BFGS was employed.

Another suggestion for future research is the evaluation of bootstrap methods and different boostrap confidence intervals for obtaining intervals to ω that produce better results than the asymptotic confidence interval.

Table 6.7: Bias and MSE of MLE and 9 different PMLE for ω by BFGS and SQP, for time series of sizes 50 and 100 (Skew GED model).

Model	φ	MLE BIAS (MSE)	I BIAS (MSE)	II BIAS (MSE)	III BIAS (MSE)	IV BIAS (MSE)	V BIAS (MSE)	VI BIAS (MSE)	VII BIAS (MSE)	VIII BIAS (MSE)	IX BIAS (MSE)
n=50	**0.85**	6.999	4.444	3.533	-1.849	4.544	3.627	1.335	4.829	3.905	1.634
		(1.014)	(0.554)	(0.439)	(0.149)	(0.552)	(0.443)	(0.267)	(0.573)	(0.458)	(0.270)
SGED	**0.90**	5.761	2.646	1.552	-1.146	2.732	1.642	-1.047	2.977	1.900	-0.761
BFGS		(0.630)	(0.271)	(0.209)	(0.169)	(0.273)	(0.209)	(0.165)	(0.279)	(0.211)	(0.154)
	0.95	2.877	-0.645	-1.146	**-4.760**	-0.583	-1.770	**-4.667**	-0.353	-1.527	**-4.393**
		(0.237)	(0.123)	(0.169)	**(0.333)**	(0.120)	(0.144)	**(0.322)**	(0.109)	(0.128)	**(0.292)**
	0.85	7.051	4.439	3.533	1.230	4.531	3.627	1.335	4.812	3.905	1.634
		(1.023)	(0.546)	(0.439)	(0.267)	(0.552)	(0.443)	(0.267)	(0.571)	(0.458)	(0.270)
SGED	**0.90**	5.854	2.651	1.553	-1.146	2.731	1.645	-1.048	2.976	1.898	-0.761
SQP		(0.638)	(0.271)	(0.209)	(0.169)	(0.273)	(0.209)	(0.165)	(0.279)	(0.211)	(0.154)
	0.95	2.892	-0.655	-1.849	**-4.760**	-0.583	-1.770	**-4.667**	-0.371	-1.528	**-4.394**
		(0.236)	(0.123)	(0.149)	**(0.333)**	(0.120)	(0.144)	**(0.322)**	(0.112)	(0.128)	**(0.292)**
n=100	**0.85**	2.925	1.660	1.095	-0.517	1.739	1.176	-0.427	1.972	1.425	-0.162
		(0.375)	(0.233)	(0.202)	(0.159)	(0.235)	(0.203)	(0.158)	(0.241)	(0.208)	(0.154)
SGED	**0.90**	3.264	1.306	0.535	-1.620	1.374	0.606	-1.535	1.581	0.812	-1.275
BFGS		(0.333)	(0.160)	(0.132)	(0.129)	(0.161)	(0.132)	(0.126)	(0.165)	(0.132)	(0.117)
	0.95	2.505	-0.352	-1.342	**-4.009**	-0.295	-1.283	**-3.927**	-0.115	-1.091	**-3.690**
		(0.160)	(0.066)	(0.079)	**(0.218)**	(0.064)	(0.076)	**(0.210)**	(0.061)	(0.069)	**(0.189)**
	0.85	2.874	2.181	1.871	1.005	2.220	1.910	1.047	2.339	2.027	1.170
		(0.366)	(0.277)	(0.251)	(0.200)	(0.278)	(0.252)	(0.201)	(0.282)	(0.255)	(0.202)
SGED	**0.90**	3.222	2.085	1.645	0.445	2.117	1.678	0.482	2.215	1.776	0.591
SQP		(0.329)	(0.210)	(0.181)	(0.132)	(0.210)	(0.182)	(0.131)	(0.213)	(0.183)	(0.131)
	0.95	2.473	0.735	0.127	-1.426	0.761	0.155	-1.394	0.838	0.237	-1.297
		(0.159)	(0.077)	(0.068)	(0.082)	(0.077)	(0.068)	(0.081)	(0.077)	(0.066)	(0.077)

Obs.: Bias and MSE are multiplied by $\times 10^2$ and in bold are the cases which the PMLE do not decrease the bias or MSE.

Acknowledgements

The authors wish to acknowledge CAPES, CNPq and FAPEMIG for financial support.

References

Broyden, C.G., 1970. The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics & Its Applications, 6, 76-90.

Cordeiro, G.M., McCullagh, P., 1995. Bias Correction in Generalized Linear Models. Journal of the Royal Statistical Society , 53(3), 629-643.

Davis, W.W., 1977. Robust interval estimation of the innovation variance of an ARMA model. The Annals of Statistics, 5(4), 700-708.

Firth, D., 1993. Bias Reduction of Maximum Likelihood Estimates. Biometrika, 80(1), 27-38.

Fletcher, R., 1970. A new approach to variable metric algorithms. The Computer Journal, 13(3), 317-322.

Table 6.8: Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for $\boldsymbol{\varphi}$ by BFGS and SQP (Log-normal and Log-gamma models).

		MLE			PMLE - BFGS			PMLE - SQP		
Model	φ	**BFGS (MSE)**	**SQP (MSE)**	**FSQP (MSE)**	**I (MSE)**	**IV (MSE)**	**VII (MSE)**	**I (MSE)**	**IV (MSE)**	**VII (MSE)**
n=50	$\omega = 0.90$	1.0000	0.9620	0.9619	0.9307	0.9315	0.9337	0.9307	0.9315	0.9337
		(0.0100)	(0.0065)	(0.0065)	(0.0027)	(0.0027)	(0.0028)	(0.0027)	(0.0027)	(0.0028)
LN	$\beta = 1.0$	1.0042	1.0092	1.0092	1.0076	1.0076	1.0077	1.0076	1.0076	1.0077
		(0.1105)	(0.1039)	(0.1039)	(0.1028)	(0.1028)	(0.1028)	(0.1028)	(0.1028)	(0.1028)
	$\delta = 5.0$	5.0004	5.0007	5.0007	5.0007	5.0007	5.0007	5.0007	5.0007	5.0007
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
	$\omega = 0.95$	1.0000	0.9816	0.9816	0.9462	0.9469	0.9491	0.9462	0.9469	0.9489
		(0.0025)	(0.0022)	(0.0022)	(0.0009)	(0.0009)	(0.0008)	(0.0009)	(0.0009)	(0.0008)
LN	$\beta = 1.0$	1.0127	1.0067	1.0067	1.0027	1.0028	1.0031	1.0027	1.0028	1.0030
		(0.0947)	(0.0968)	(0.0968)	(0.0986)	(0.0986)	(0.0983)	(0.0986)	(0.0986)	(0.0983)
	$\delta = 5.0$	5.0003	5.0004	5.0004	5.0003	5.0003	5.0003	5.0003	5.0003	5.0003
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
n=100	$\omega = 0.90$	0.9325	0.9315	0.9314	0.9127	0.9133	0.9152	0.9201	0.9204	0.9213
		(0.0032)	(0.0031)	(0.0031)	(0.0015)	(0.0015)	(0.0015)	(0.0019)	(0.0019)	(0.0020)
LN	$\beta = 1.0$	1.0087	1.0086	1.0086	1.0069	1.0070	1.0071	1.0075	1.0075	1.0076
		(0.0493)	(0.0493)	(0.0493)	(0.0490)	(0.0490)	(0.0490)	(0.0491)	(0.0491)	(0.0491)
	$\delta = 5.0$	5.0001	5.0001	5.0001	5.0001	5.0001	5.0001	5.0001	5.0001	5.0001
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0369)	(0.0001)
	$\omega = 0.95$	0.9772	0.9766	0.9766	0.9483	0.9488	0.9504	0.9591	0.9593	0.9601
		(0.0016)	(0.0016)	(0.0016)	(0.0005)	(0.0005)	(0.0005)	(0.0007)	(0.0007)	(0.0007)
LN	$\beta = 1.0$	0.9986	0.9987	0.9988	0.9966	0.9966	0.9967	0.9972	0.9972	0.9973
		(0.0473)	(0.0473)	(0.0473)	(0.0470)	(0.0470)	(0.0470)	(0.0470)	(0.0470)	(0.0470)
	$\delta = 5.0$	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
n=50	$\omega = 0.90$	0.9535	0.9393	0.9393	0.9038	0.9051	0.9087	0.9038	0.9051	0.9089
		(0.0081)	(0.0065)	(0.0065)	(0.0037)	(0.0036)	(0.0035)	(0.0037)	(0.0036)	(0.0035)
LG	$\beta = 1.0$	0.9976	0.9980	0.9980	0.9980	0.9980	0.9980	0.9980	0.9980	0.9980
		(0.0087)	(0.0087)	(0.0087)	(0.0087)	(0.0087)	(0.0087)	(0.0087)	(0.0087)	(0.0087)
	$\alpha = 5.0$	5.3310	5.3997	5.3997	5.5129	5.5071	5.4903	5.5129	5.5071	5.4895
		(1.5648)	(1.6280)	(1.6280)	(1.7954)	(1.7826)	(1.7448)	(1.7957)	(1.7827)	(1.7446)
	$\omega = 0.95$	0.9762	0.9629	0.9629	0.9248	0.9260	0.9293	0.9248	0.9260	0.9293
		(0.0034)	(0.0034)	(0.0034)	(0.0033)	(0.0032)	(0.0028)	(0.0033)	(0.0032)	(0.0028)
LG	$\beta = 1.0$	0.9970	0.9973	0.9973	0.9974	0.9974	0.9974	0.9974	0.9974	0.9974
		(0.0083)	(0.0083)	(0.0083)	(0.0083)	(0.0083)	(0.0083)	(0.0083)	(0.0083)	(0.0083)
	$\alpha = 5.0$	5.3178	5.3798	5.3798	5.4886	5.4838	5.4701	5.4886	5.4838	5.4702
		(1.3789)	(1.4545)	(1.4545)	(1.6349)	(1.6229)	(1.5909)	(1.6349)	(1.6230)	(1.5912)
n=100	$\omega = 0.90$	0.9305	0.9219	0.9219	0.8994	0.9004	0.9034	0.9086	0.9090	0.9105
		(0.0044)	(0.0033)	(0.0033)	(0.0020)	(0.0019)	(0.0018)	(0.0023)	(0.0022)	(0.0022)
LG	$\beta = 1.0$	0.9970	0.9971	0.9971	0.9970	0.9970	0.9970	0.9970	0.9970	0.9970
		(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)
	$\alpha = 5.0$	5.1568	5.1996	5.1996	5.2856	5.2808	5.2668	5.2494	5.2472	5.2404
		(0.6727)	(0.6636)	(0.6636)	(0.7194)	(0.7142)	(0.6993)	(0.6925)	(0.6901)	(0.6829)
	$\omega = 0.95$	0.9746	0.9659	0.9659	0.9355	0.9363	0.9386	0.9471	0.9475	0.9487
		(0.0020)	(0.0016)	(0.0016)	(0.0013)	(0.0012)	(0.0011)	(0.0012)	(0.0012)	(0.0011)
LG	$\beta = 1.0$	0.9998	0.9996	0.9996	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996
		(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)
	$\alpha = 5.0$	5.1805	5.2215	5.2215	5.3178	5.3145	5.3052	5.2806	5.2791	5.2736
		(0.5828)	(0.5857)	(0.5857)	(0.6650)	(0.6611)	(0.6504)	(0.6312)	(0.6296)	(0.6247)

Goldfard, D., 1970. A family of variable metric updates derived by variational means. Mathematics of Computation, 24(109), 23-26.

Hahn, J., Newey, W.A., 2004. Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. Econometrica, 72(4), 1295-1319.

Bester, C.A., Hansen, C., 2009. A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects. Journal of Business and Economic Statistics, 27(2) 131-148.

Heinze, G., Schemper, M., 2001. A Solution to the Problem of Monotone Likelihood

Table 6.9: Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for φ by BFGS and SQP (Pareto and Weibull models).

Model	φ	MLE			PMLE - BFGS			PMLE - SQP		
		BFGS (MSE)	SQP (MSE)	FSQP (MSE)	I (MSE)	IV (MSE)	VII (MSE)	I (MSE)	IV (MSE)	VII (MSE)
n=50	$\omega = 0.90$	0.9545	0.9556	0.9553	0.9234	0.9243	0.9270	0.9234	0.9243	0.9270
P		(0.0061)	(0.0062)	(0.0061)	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0027)
	$\beta = 1.0$	0.9930	0.9926	0.9927	0.9927	0.9927	0.9926	0.9927	0.9927	0.9926
		(0.0461)	(0.0461)	(0.0460)	(0.0463)	(0.0463)	(0.0462)	(0.0463)	(0.0463)	(0.0462)
	$\omega = 0.95$	0.9752	0.9759	0.9757	0.9410	0.9418	0.9441	0.9410	0.9418	0.9441
P		(0.0024)	(0.0024)	(0.0024)	(0.0014)	(0.0014)	(0.0013)	(0.0014)	(0.0014)	(0.0013)
	$\beta = 1.0$	0.9953	0.9952	0.9952	0.9938	0.9938	0.9938	0.9938	0.9938	0.9938
		(0.0442)	(0.0441)	(0.0441)	(0.0444)	(0.0444)	(0.0444)	(0.0444)	(0.0444)	(0.0444)
n=100	$\omega = 0.90$	0.9301	0.9314	0.9313	0.9113	0.9121	0.9142	0.9193	0.9197	0.9207
P		(0.0030)	(0.0032)	(0.0032)	(0.0015)	(0.0015)	(0.0015)	(0.0020)	(0.0020)	(0.0020)
	$\beta = 1.0$	0.9963	0.9964	0.9964	0.9958	0.9959	0.9959	0.9961	0.9961	0.9961
		(0.0206)	(0.0206)	(0.0206)	(0.0206)	(0.0206)	(0.0206)	(0.0206)	(0.0206)	(0.0206)
	$\omega = 0.95$	0.9719	0.9727	0.9727	0.9442	0.9448	0.9466	0.9551	0.9554	0.9563
P		(0.0015)	(0.0015)	(0.0015)	(0.0007)	(0.0007)	(0.0006)	(0.0008)	(0.0008)	(0.0008)
	$\beta = 1.0$	0.9998	0.9998	0.9998	0.9987	0.9988	0.9988	0.9992	0.9992	0.9992
		(0.0205)	(0.0204)	(0.0204)	(0.0204)	(0.0204)	(0.0204)	(0.0204)	(0.0204)	(0.0204)
n=50	$\omega = 0.90$	0.9549	0.9535	0.9535	0.9187	0.9197	0.9226	0.9187	0.9199	0.9228
		(0.0066)	(0.0062)	(0.0062)	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0027)
W	$\beta = 1.0$	0.9977	0.9982	0.9982	1.0095	1.0091	1.0079	1.0095	1.0089	1.0077
		(0.0555)	(0.0559)	(0.0559)	(0.0572)	(0.0571)	(0.0569)	(0.0572)	(0.0569)	(0.0567)
	$\upsilon = 5.0$	5.0658	5.0705	5.0705	5.1284	5.1262	5.1198	5.1285	5.1257	5.1193
		(0.2552)	(0.2685)	(0.2685)	(0.2836)	(0.2824)	(0.2791)	(0.2836)	(0.2828)	(0.2794)
	$\omega = 0.95$	0.9749	0.9737	0.9737	0.9362	0.9371	0.9395	0.9364	0.9373	0.9397
		(0.0028)	(0.0026)	(0.0026)	(0.0019)	(0.0018)	(0.0017)	(0.0018)	(0.0018)	(0.0016)
W	$\beta = 1.0$	1.0070	1.0065	1.0065	1.0159	1.0157	1.0149	1.0158	1.0156	1.0148
		(0.0573)	(0.0573)	(0.0573)	(0.0598)	(0.0597)	(0.0595)	(0.0598)	(0.0597)	(0.0595)
	$\upsilon = 5.0$	5.1010	5.1040	5.1040	5.1619	5.1602	5.1553	5.1614	5.1597	5.1548
		(0.2432)	(0.2522)	(0.2522)	(0.2773)	(0.2763)	(0.2732)	(0.2779)	(0.2768)	(0.2737)
n=100	$\omega = 0.90$	0.9301	0.9319	0.9319	0.9086	0.9094	0.9117	0.9170	0.9175	0.9186
		(0.0034)	(0.0035)	(0.0035)	(0.0017)	(0.0017)	(0.0017)	(0.0021)	(0.0021)	(0.0021)
W	$\beta = 1.0$	1.0045	1.0035	1.0035	1.0130	1.0127	1.0115	1.0096	1.0094	1.0088
		(0.0275)	(0.0275)	(0.0275)	(0.0280)	(0.0280)	(0.0279)	(0.0278)	(0.0278)	(0.0277)
	$\upsilon = 5.0$	5.0359	5.0315	5.0315	5.0820	5.0799	5.0738	5.0634	5.0623	5.0592
		(0.1547)	(0.1562)	(0.1562)	(0.1611)	(0.1604)	(0.1584)	(0.1576)	(0.1572)	(0.1563)
	$\omega = 0.95$	0.9738	0.9730	0.9730	0.9431	0.9438	0.9458	0.9547	0.9550	0.9559
		(0.0017)	(0.0016)	(0.0016)	(0.0008)	(0.0008)	(0.0007)	(0.0009)	(0.0009)	(0.0008)
W	$\beta = 1.0$	1.0185	1.0189	1.0189	1.0272	1.0269	1.0262	1.0238	1.0237	1.0234
		(0.0261)	(0.0263)	(0.0263)	(0.0273)	(0.0273)	(0.0272)	(0.0269)	(0.0268)	(0.0268)
	$\upsilon = 5.0$	5.0703	5.0722	5.0722	5.1240	5.1224	5.1183	5.1035	5.1028	5.1010
		(0.1432)	(0.1463)	(0.1463)	(0.1607)	(0.1599)	(0.1582)	(0.1542)	(0.1539)	(0.1531)

in Cox Regression. Biometrics, 57, 114-119.

Lawrence, C.T., Tits, A.L., 2001. A computationally efficient feasible sequential quadratic programming algorithm. SIAM Journal of Optimization, 11(4), 1092-1118.

Loughin, T.M., 1998. On the Bootstrap and Monotone Likelihood in the Cox Proportional Hazards Regression Model. Lifetime Data Analysis, 4, 393-403.

Nocedal, J., Wright, J.W., 1999. Numerical Optimization. Springer Verlag, New York.

Pinho, F.M., Franco, G.C., Silva, R.S., 2012. Modelling Volatility Using State Space Models with Heavy Tailed Distributions. Working paper.

Santos, T.R., Franco, G.C., Gamerman, D., 2010. Gamma family of dynamic mod-

Table 6.10: Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for $\boldsymbol{\varphi}$ by BFGS and SQP (Fréchet and Lévy models).

			MLE			PMLE - BFGS			PMLE - SQP	
Model	$\boldsymbol{\varphi}$	**BFGS (MSE)**	**SQP (MSE)**	**FSQP (MSE)**	**I (MSE)**	**IV (MSE)**	**VII (MSE)**	**I (MSE)**	**IV (MSE)**	**VII (MSE)**
n=50	$\omega = 0.90$	0.9575	0.9567	0.9567	0.9205	0.9215	0.9243	0.9205	0.9215	0.9243
		(0.0068)	(0.0066)	(0.0066)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
F	$\beta = 1.0$	1.0047	1.0052	1.0052	1.0155	1.0151	1.0140	1.0155	1.0151	1.0140
		(0.0598)	(0.0602)	(0.0602)	(0.0619)	(0.0618)	(0.0616)	(0.0619)	(0.0618)	(0.0616)
	$\alpha = 5.0$	5.0571	5.0602	5.0602	5.1210	5.1189	5.1125	5.1210	5.1189	5.1126
		(0.2729)	(0.2856)	(0.2856)	(0.3078)	(0.3065)	(0.3029)	(0.3078)	(0.3065)	(0.3029)
	$\omega = 0.95$	0.9817	0.9792	0.9792	0.9418	0.9426	0.9450	0.9418	0.9426	0.9450
		(0.0025)	(0.0024)	(0.0024)	(0.0013)	(0.0013)	(0.0012)	(0.0013)	(0.0013)	(0.0012)
F	$\beta = 1.0$	1.0133	1.0140	1.0140	1.0226	1.0223	1.0216	1.0226	1.0223	1.0216
		(0.0528)	(0.0529)	(0.0529)	(0.0547)	(0.0546)	(0.0544)	(0.0547)	(0.0546)	(0.0544)
	$\alpha = 5.0$	5.0588	5.0652	5.0652	5.1187	5.1172	5.1127	5.1188	5.1172	5.1128
		(0.2246)	(0.2336)	(0.2336)	(0.2542)	(0.2533)	(0.2508)	(0.2542)	(0.2533)	(0.2508)
n=100	$\omega = 0.90$	0.9277	0.9282	0.9282	0.9071	0.9079	0.9103	0.9157	0.9161	0.9172
		(0.0032)	(0.0032)	(0.0032)	(0.0017)	(0.0017)	(0.0016)	(0.0021)	(0.0021)	(0.0021)
F	$\beta = 1.0$	0.9945	0.9942	0.9942	1.0022	1.0019	1.0008	0.9989	0.9987	0.9982
		(0.0264)	(0.0266)	(0.0266)	(0.0267)	(0.0267)	(0.0266)	(0.0266)	(0.0265)	(0.0265)
	$\alpha = 5.0$	5.0334	5.0320	5.0320	5.0778	5.0757	5.0695	5.0587	5.0577	5.0546
		(0.1536)	(0.1558)	(0.1558)	(0.1607)	(0.1600)	(0.1580)	(0.1574)	(0.1571)	(0.1562)
	$\omega = 0.95$	0.9740	0.9733	0.9733	0.9434	0.9440	0.9461	0.9550	0.9553	0.9562
		(0.0017)	(0.0016)	(0.0016)	(0.0007)	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0008)
F	$\beta = 1.0$	1.0068	1.0071	1.0071	1.0159	1.0157	1.0149	1.0123	1.0122	1.0119
		(0.0270)	(0.0272)	(0.0272)	(0.0280)	(0.0279)	(0.0278)	(0.0275)	(0.0275)	(0.0275)
	$\alpha = 5.0$	5.0727	5.0748	5.0748	5.1261	5.1246	5.1204	5.1057	5.1051	5.1033
		(0.1521)	(0.1578)	(0.1578)	(0.1701)	(0.1695)	(0.1679)	(0.1640)	(0.1637)	(0.0412)
n=50	$\omega = 0.90$	0.9663	0.9630	0.9629	0.9333	0.9340	0.9360	0.9333	0.9340	0.9360
L		(0.0069)	(0.0065)	(0.0065)	(0.0028)	(0.0028)	(0.0029)	(0.0028)	(0.0028)	(0.0029)
	$\beta = 1.0$	0.9842	0.9820	0.9821	0.9808	0.9808	0.9808	0.9808	0.9808	0.9808
		(0.0878)	(0.0883)	(0.0883)	(0.0890)	(0.0890)	(0.0889)	(0.0890)	(0.0890)	(0.0889)
	$\omega = 0.95$	0.9844	0.9826	0.9826	0.9492	0.9498	0.9516	0.9492	0.9498	0.9516
L		(0.0022)	(0.0022)	(0.0022)	(0.0008)	(0.0008)	(0.0007)	(0.0008)	(0.0008)	(0.0007)
	$\beta = 1.0$	1.0002	0.9990	0.9990	0.9964	0.9964	0.9965	0.9964	0.9964	0.9965
		(0.0856)	(0.0865)	(0.0865)	(0.0864)	(0.0864)	(0.0863)	(0.0864)	(0.0864)	(0.0863)
n=100	$\omega = 0.90$	0.9364	0.9363	0.9364	0.9176	0.9183	0.9201	0.9253	0.9256	0.9264
L		(0.0033)	(0.0033)	(0.0033)	(0.0015)	(0.0015)	(0.0016)	(0.0020)	(0.0020)	(0.0021)
	$\beta = 1.0$	0.9974	0.9968	0.9968	0.9963	0.9964	0.9965	0.9967	0.9967	0.9967
		(0.0451)	(0.0453)	(0.0453)	(0.0448)	(0.0448)	(0.0448)	(0.0449)	(0.0449)	(0.0449)
	$\omega = 0.95$	0.9778	0.9780	0.9779	0.9496	0.9502	0.9517	0.9605	0.9607	0.9614
L		(0.0016)	(0.0016)	(0.0016)	(0.0005)	(0.0005)	(0.0005)	(0.0007)	(0.0007)	(0.0007)
	$\beta = 1.0$	0.9951	0.9949	0.9949	0.9924	0.9924	0.9926	0.9932	0.9933	0.9933
		(0.0411)	(0.0412)	(0.0412)	(0.0412)	(0.0412)	(0.0412)	(0.0412)	(0.0412)	(0.0311)

els. Technical Report, 234, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro.

Shanno, D.F., 1970. Conditioning of quase-Newton methods for function minimization. Mathematics of Computation, 24(111), 647-656.

Smith, R.L., Miller, J.E., 1986. A non-Gaussian state space model and application to prediction of records. Journal of the Royal Statistical Society, Series B, 48(1), 79-88.

Tukey, J., 1958. Bias and confidence in not quite large samples (abstract). The Annals of Mathematical Statistics, 29(2), 614.

Table 6.11: Estimates and MSE of MLE by BFGS, SQP and FSQP and 3 different PMLE for φ by BFGS and SQP (Skew GED model).

		MLE			PMLE - BFGS			PMLE - SQP		
Model	φ	**BFGS (MSE)**	**SQP (MSE)**	**FSQP (MSE)**	**I (MSE)**	**IV (MSE)**	**VII (MSE)**	**I (MSE)**	**IV (MSE)**	**VII (MSE)**
n=50	$\omega = 0.90$	0.9576	0.9585	0.9588	0.9265	0.9273	0.9298	0.9265	0.9273	0.9298
		(0.0063)	(0.0064)	(0.0064)	(0.0027)	(0.0027)	(0.0028)	(0.0027)	(0.0027)	(0.0028)
	$\beta = 1.0$	0.9887	0.9893	0.9891	0.9880	0.9881	0.9879	0.9881	0.9881	0.9879
SGED		(0.0703)	(0.0703)	(0.0705)	(0.0701)	(0.0698)	(0.0700)	(0.0699)	(0.0699)	(0.0699)
	$\delta = 5.0$	4.9996	4.9996	4.9996	4.9996	4.9996	4.9995	4.9996	4.9996	4.9996
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	$\kappa = 1.0$	1.0074	1.0078	1.0077	1.0075	1.0073	1.0071	1.0081	1.0073	1.0071
		(0.0315)	(0.0308)	(0.0308)	(0.0306)	(0.0303)	(0.0309)	(0.0307)	(0.0305)	(0.0305)
	$\omega = 0.95$	0.9788	0.9789	0.9788	0.9436	0.9442	0.9465	0.9434	0.9442	0.9463
		(0.0024)	(0.0024)	(0.0024)	(0.0012)	(0.0012)	(0.0011)	(0.0012)	(0.0012)	(0.0011)
	$\beta = 1.0$	0.9944	0.9948	0.9947	0.9931	0.9933	0.9935	0.9932	0.9933	0.9931
SGED		(0.0737)	(0.0736)	(0.0736)	(0.0742)	(0.0742)	(0.0738)	(0.0743)	(0.0742)	(0.0743)
	$\delta = 5.0$	4.9999	4.9998	4.9998	4.9998	4.9998	4.9998	4.9998	4.9998	4.9998
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	$\kappa = 1.0$	1.0123	1.0091	1.0091	1.0097	1.0098	1.0097	1.0103	1.0098	1.0102
		(0.0331)	(0.0284)	(0.0284)	(0.0287)	(0.0287)	(0.0287)	(0.0292)	(0.0287)	(0.0287)
n=100	$\omega = 0.90$	0.9326	0.9322	0.9322	0.9131	0.9137	0.9158	0.9209	0.9212	0.9222
		(0.0033)	(0.0033)	(0.0033)	(0.0016)	(0.0016)	(0.0016)	(0.0021)	(0.0021)	(0.0021)
	$\beta = 1.0$	1.0135	1.0133	1.0133	1.0127	1.0127	1.0121	1.0129	1.0129	1.0130
SGED		(0.0371)	(0.0372)	(0.0372)	(0.0369)	(0.0369)	(0.0375)	(0.0369)	(0.0369)	(0.0369)
	$\delta = 5.0$	4.9997	4.9997	4.9997	4.9997	4.9997	4.9996	4.9997	4.9997	4.9997
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	$\kappa = 1.0$	1.0042	1.0043	1.0043	1.0043	1.0043	1.0035	1.0042	1.0042	1.0042
		(0.0104)	(0.0104)	(0.0104)	(0.0103)	(0.0103)	(0.0112)	(0.0103)	(0.0103)	(0.0103)
	$\omega = 0.95$	0.9751	0.9747	0.9747	0.9465	0.9471	0.9488	0.9573	0.9576	0.9584
		(0.0016)	(0.0016)	(0.0016)	(0.0007)	(0.0006)	(0.0006)	(0.0008)	(0.0008)	(0.0008)
	$\beta = 1.0$	1.0099	1.0100	1.0100	1.0093	1.0093	1.0078	1.0095	1.0096	1.0096
SGED		(0.0311)	(0.0311)	(0.0311)	(0.0312)	(0.0312)	(0.0327)	(0.0311)	(0.0311)	(0.1631)
	$\delta = 5.0$	4.9998	4.9998	4.9998	4.9998	4.9998	4.9995	4.9998	4.9998	4.9998
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
	$\kappa = 1.0$	1.0013	1.0014	1.0014	1.0013	1.0013	0.9994	1.0014	1.0014	1.0014
		(0.0113)	(0.0112)	(0.0112)	(0.0110)	(0.0110)	(0.0128)	(0.0111)	(0.0111)	(0.0111)

Table 6.12: 95% Asymptotic confidence interval of MLE by BFGS 3 differents PMLE using BFGS for time series of size 50.

Model	φ	MLE (Cov Rate)	PMLE I (Cov Rate)	PMLE VII (Cov Rate)	PMLE IV (Cov Rate)
LN	$\omega = 0.95$	[0.7876 ; 1.0000] 1.000	[0.5106 ; 0.9925] 0.986	[0.5067 ; 0.9929] 0.980	[0.5150 ; 0.9931] 0.985
	$\beta = 1.0$	[0.3963 ; 1.6291] 0.949	[0.3869 ; 1.6185] 0.949	[0.4098 ; 1.6329] 0.934	[0.3886 ; 1.6173] 0.947
	$\delta = 5.0$	[4.9743 ; 5.0263] 0.934	[4.9751 ; 5.0256] 0.928	[4.9750 ; 5.0250] 0.918	[4.9751 ; 5.0256] 0.927
LG	$\omega = 0.95$	[0.6882 ; 0.9943] 0.970	[0.4662 ; 0.9864] 0.941	[0.4749 ; 0.9875] 0.947	[0.4765 ; 0.9874] 0.946
	$\beta = 1.0$	[0.8169 ; 1.1772] 0.948	[0.8177 ; 1.1772] 0.948	[0.8207 ; 1.1794] 0.931	[0.8176 ; 1.1769] 0.947
	$\alpha = 5.0$	[3.2047 ; 7.4309] 0.960	[3.2889 ; 7.6882] 0.964	[3.2960 ; 7.6625] 0.965	[3.2873 ; 7.6476] 0.964
P	$\omega = 0.95$	[0.5983 ; 0.9956] 0.984	[0.5082 ; 0.9908] 0.969	[0.5015 ; 0.9913] 0.976	[0.5138 ; 0.9915] 0.974
	$\beta = 1.0$	[0.5788 ; 1.4118] 0.946	[0.5747 ; 1.4128] 0.952	[0.5818 ; 1.4204] 0.957	[0.5755 ; 1.4121] 0.951
W	$\omega = 0.95$	[0.6151 ; 0.9949] 0.981	[0.4875 ; 0.9896] 0.964	[0.4863 ; 0.9909] 0.970	[0.4875 ; 0.9896] 0.972
	$\beta = 1.0$	[0.5303 ; 1.4838] 0.944	[0.5451 ; 1.4868] 0.944	[0.5540 ; 1.4975] 0.953	[0.5451 ; 1.4868] 0.945
	$\upsilon = 5.0$	[4.0011 ; 6.2009] 0.963	[4.0759 ; 6.2480] 0.964	[4.0670 ; 6.2413] 0.958	[4.0759 ; 6.2480] 0.964
F	$\omega = 0.95$	[0.6728 ; 0.9967] 0.991	[0.4827 ; 0.9919] 0.978	[0.4851 ; 0.9913] 0.974	[0.4895 ; 0.9926] 0.983
	$\beta = 1.0$	[0.5463 ; 1.4804] 0.965	[0.5521 ; 1.4931] 0.968	[0.5346 ; 1.4700] 0.962	[0.5520 ; 1.4911] 0.968
	$\alpha = 5.0$	[3.9826 ; 6.1350] 0.975	[0.40414 ; 6.1961] 0.972	[4.0563 ; 6.2150] 0.956	[4.0392 ; 6.1863] 0.972
L	$\omega = 0.95$	[0.6471 ; 0.9974] 0.994	[0.5148 ; 0.9932] 0.990	[0.5144 ; 0.9935] 0.987	[0.5202 ; 0.9937] 0.991
	$\beta = 1.0$	[0.3977 ; 1.6026] 0.965	[0.3959 ; 1.5969] 0.965	[0.3835 ; 1.5861] 0.935	[0.3969 ; 1.5961] 0.965
SGED	$\omega = 0.95$	[0.6122 ; 0.9962] 0.986	[0.5045 ; 0.9915] 0.974	[0.4932 ; 0.9924] 0.976	[0.5099 ; 0.9922] 0.980
	$\beta = 1.0$	[0.4627 ; 1.5260] 0.952	[0.4607 ; 1.5254] 0.952	[0.4757 ; 1.5369] 0.947	[0.4621 ; 1.5249] 0.954
	$\delta = 5.0$	[4.9861 ; 5.0137] 0.856	[4.9862 ; 5.0133] 0.856	[4.9867 ; 5.0132] 0.864	[4.9862 ; 5.0134] 0.860
	$\kappa = 1.0$	[0.7211 ; 1.3035] 0.910	[0.7227 ; 1.2968] 0.91	[0.7288 ; 1.2988] 0.908	[0.7228 ; 1.2967] 0.908

Chapter 7

Bootstrapping Non Gaussian State Space Models

Frank M. de Pinho[a], **Glaura C. Franco**[b]
[a]*IBMEC, Belo Horizonte, Brasil*
[b]*Universidade Federal de Minas Gerais, Belo Horizonte, Brasil*

Abstract

This paper proposes some different bootstrap procedures for inference in a non Gaussian family of state space models (NGSSM), introduced by Santos et al. (2010). Confidence intervals for the parameters of the NGSSM can be built using the asymptotic normality assumption of the maximum likelihood estimators, but subjected to certain regularity conditions that may not be satisfied. Some previous studies have shown, empirically, that the coverage rate of the asymptotic confidence intervals are far from the true confidence level assumed, especially for small samples. Thus, this paper evaluates the performance of three bootstrap confidence intervals in three different bootstrap methods applied to the NGSSM. The results show that the bootstrap confidence interval with bias-correction using a parametric bootstrap is the procedure which shows the best performance.

Keyword: Heavy Tailed Distribution, Penalized Maximum Likelihood Estimator, Bootstrap Confidence Intervals, BFGS.

7.1 Introduction

Santos et al. (2010) have proposed a non Gaussian state space model (NGSSM) with exact marginal likelihood function, which is a generalization of the results of Smith & Miller (1986). In their paper they present a filtering method that allows the estimation of the dynamic parameter and also show methods of smoothing and forecasting.

Pinho et al. (2012) have proposed heavy tailed distributions as special cases of the NGSSM. They presented some Monte Carlo results comparing Bayesian and classical methods of inference in the estimation of the NGSSM for the heavy tailed distributions. The results of the point and interval estimators, whether classical or Bayesian, were very satisfactory when the size of the series is large (greater than 100).

However, Pinho & Franco (2012) showed that, for small series, the maximum likelihood estimator (MLE) provides unsatisfactory results in the estimation of one of the parameters of the NGSSM. This parameter, called ω, plays a very important role in the NGSSM because it has the function of increasing multiplicatively the variance over time. The parameter space of ω is $(0, 1)$ and for the Monte Carlo simulation study performed it was seen that, for small series, the estimate of ω is always close to 1.0, regardless the real value of this parameter. Then, to solve this problem Pinho & Franco (2012) proposed a penalized maximum likelihood estimator (PMLE) and demonstrate empirically that there is a significant improvement in the estimates of parameter ω.

Confidence intervals for the parameters of the NGSSM were also built in the work of Pinho & Franco (2012), using the asymptotic properties of the MLE. However, the results for parameter ω, even using the penalized function were unsatisfactory, because the coverage rates remained above the nominal level used in the Monte Carlo study. With the aim of improving the results for the confidence intervals, especially for small series, this paper proposes some bootstrap procedures in the NGSSM and also employs different bootstrap confidence intervals proposed by Efron & Tibshirani (1993).

The paper is organized as follows. Section 7.2 defines the NGSSM and shows the estimators used for point or interval estimation of parameters of the NGSSM. Section 7.3 shows the bootstrap scheme to construct a bootstrap series in the NGSSM and describes the bootstrap confidence intervals utilized. Section 7.4 shows the results of the Monte Carlo simulation studies to evaluate the behavior of the bootstrap confidence intervals proposed. Section 7.5 concludes the work.

7.2 A non-Gaussian state space model

Santos et al. (2010) define a new family of non-Gaussian state space models, which is a generalization of the works of Smith & Miller (1986) and Harvey & Fernandes (1989).

A time series $\{y_t\}_{t=1}^n$ is in this class of models if it satisfies the following assumptions:

A0 Its probability (density) function can be written in the form:

$$p(y_t|\mu_t,\boldsymbol{\varphi}) = q(y_t,\boldsymbol{\varphi})\mu_t^{r(y_t,\boldsymbol{\varphi})} \exp\left(-\mu_t s(y_t,\boldsymbol{\varphi})\right), \text{ for } y_t \in H(\boldsymbol{\varphi}) \subset \Re \tag{7.1}$$

and $p(y_t|\mu_t,\boldsymbol{\varphi}) = 0$, otherwise. Functions $q(\cdot)$, $r(\cdot)$, $s(\cdot)$ and $H(\cdot)$ are such that $p(y_t|\mu_t,\boldsymbol{\varphi}) \geq 0$ and therefore $\mu_t > 0$, for all $t > 0$. It is also assumed that $\boldsymbol{\varphi}$ varies in the p-dimensional parameter space Φ.

A1 If x_t is a covariate vector, the link function g relates the predictor to the parameter μ_t through the relation $\mu_t = \lambda_t g(x_t, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ are the regression coefficients (one of the components of $\boldsymbol{\varphi}$) and λ_t is the latent state variable related to the description of the dynamic level. If the predictor is linear, then $g(x_t, \boldsymbol{\beta}) = g(x_t^{'}\boldsymbol{\beta})$.

A2 The dynamic level λ_t evolves according to the system equation $\lambda_{t+1} = \omega^{-1}\lambda_t\varsigma_{t+1}$, where $\varsigma_{t+1}|\boldsymbol{Y}_t \sim Beta\,(\omega a_t,(1-\omega)a_t)$, $0 < \omega \leq 1$, $t = 1, 2, ...,$ that is, $\omega\frac{\lambda_{t+1}}{\lambda_t} \mid \lambda_t, \boldsymbol{Y}_t \sim Beta\,(\omega a_t, (1-\omega)a_t)$, $\boldsymbol{Y}_t = \{Y_0, y_1, \ldots, y_t\}$ and Y_0 represents previously available information.

A3 The dynamic level λ_t is initialized with prior distribution $\lambda_0|Y_0 \sim Gamma(a_0,b_0)$.

Theorem 1 in Santos et al. (2010) present the equations for the exact evolution of the dynamic level and the predictive density function for the NGSSM. They are presented below.

- *Prior* distribution $\mu_t|\, \boldsymbol{Y_{t-1}},\boldsymbol{\varphi} \sim \text{Gamma}\left(c_{t|t-1};\, d_{t|t-1}\right)$, where

$$c_{t|t-1} = \omega a_{t-1},$$

$$d_{t|t-1} = \omega b_{t-1} \left[g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right)\right]^{-1}.$$

- Online or updated distribution $\mu_t|\, \boldsymbol{Y_t},\boldsymbol{\varphi} \sim \text{Gamma}\left(c_t;\, d_t\right)$, where

$$c_t = c_{t|t-1} + r\left(y_t, \boldsymbol{\varphi}\right),$$

$$d_t = d_{t|t-1} + s\left(y_t, \boldsymbol{\varphi}\right).$$

- Predictive density function is given by

$$p\left(y_t|\, \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi}\right) = \frac{\Gamma\left(r\left(y_t, \boldsymbol{\varphi}\right) + c_{t|t-1}\right) q\left(y_t, \boldsymbol{\varphi}\right) d_{t|t-1}^{c_{t|t-1}} I_{(y_t \in H(\boldsymbol{\varphi}))}}{\Gamma\left(c_{t|t-1}\right) \left[s\left(y_t, \boldsymbol{\varphi}\right) + d_{t|t-1}\right]^{r(y_t,\boldsymbol{\varphi}) + c_{t|t-1}}}. \tag{7.2}$$

Santos et al. (2010) and Pinho et al. (2012) presents some special cases of the NGSSM as follow in the Table 7.1.

Classical estimation for the parameter vector $\boldsymbol{\varphi}$, which contains ω, β and specific parameters of the distribution used (see Table 7.1), is performed through maximum likelihood procedures. As already pointed out in the previous section, there are some convergence problems in the estimation of parameter ω for small series.

Thus, Pinho & Franco (2012) have proposed a penalty function to reduce the bias

Table 7.1: Cases of the NGSSM

Model	$\boldsymbol{\varphi}$	$q(y_t, \boldsymbol{\varphi})$	$r(y_t, \boldsymbol{\varphi})$	$s(y_t, \boldsymbol{\varphi})$	$H(\boldsymbol{\varphi})$
Log-normal†	$(\omega, \boldsymbol{\beta}, \gamma, \delta)$	$\left[(y_t - \gamma)\sqrt{2\pi}\right]^{-1}$	$\frac{1}{2}$	$\frac{[\ln(y_t-\gamma)-\delta]^2}{2}$	(γ,∞)
Log-gamma†	$(\omega, \boldsymbol{\beta}, \alpha)$	$\frac{\alpha^{\alpha}[ln(y_t)]^{\alpha-1}}{[\Gamma(\alpha)y_t]}$	α	$\alpha \ln(y_t)$	$(1,\infty)$
Fréchet†	$(\omega, \boldsymbol{\beta}, \gamma, \alpha)$	$\alpha(y_t - \gamma)^{-\alpha-1}$	1	$(y_t - \gamma)^{-\alpha}$	(γ,∞)
Lévy†	$(\omega, \boldsymbol{\beta}, \gamma)$	$[2\pi(y_t - \gamma)]^{-\frac{3}{2}}$	$\frac{1}{2}$	$[2(y_t - \gamma)]^{-1}$	(γ,∞)
Skew GED†	$(\omega, \boldsymbol{\beta}, \kappa, \alpha, \delta)$	$\frac{\kappa}{\Gamma\left(\alpha^{-1}\right)\left(1+\kappa^2\right)}$	$\frac{1}{\alpha}$	$\left[\frac{(y_t-\delta)^+}{k^{-\alpha}}\right]^{\alpha} + \left[\frac{(y_t-\delta)^-}{k^{\alpha}}\right]^{\alpha}$	$(-\infty,\infty)$
Pareto†	$(\omega, \boldsymbol{\beta})$	y_t^{-1}	1	$\ln(y_t)$	$(1,\infty)$
Weibull†	$(\omega, \boldsymbol{\beta}, \upsilon)$	$\upsilon y_t^{\upsilon-1}$	1	y_t^{υ}	$(0,\infty)$
Poisson	$(\omega, \boldsymbol{\beta})$	$(y_t!)^{-1}$	y_t	1	$\{0,1,\ldots\}$
Borel-Tanner	$(\omega, \boldsymbol{\beta}, \gamma)$	$\frac{\gamma}{(y_t-\gamma)!} y_t^{y_t-\gamma-1}$	$y_t - \gamma$	y_t	$\{\gamma,\gamma+1,\ldots\}$
Gamma	$(\omega, \boldsymbol{\beta}, \alpha)$	$\frac{\alpha^{\alpha} y_t^{\alpha-1}}{\Gamma(\alpha)}$	α	αy_t	$(0,\infty)$
Normal	$(\omega, \boldsymbol{\beta}, \gamma)$	$[2\pi]^{-\frac{1}{2}}$	$\frac{1}{2}$	$\frac{(y_t-\gamma)^{-2}}{2}$	$(-\infty,\infty)$
Laplace	$(\omega, \boldsymbol{\beta}, \gamma)$	$\frac{1}{\sqrt{2}}$	1	$\sqrt{2}\,\lvert y_t - \gamma\rvert$	$(-\infty,\infty)$
Inverse Gaussian	$(\omega, \boldsymbol{\beta}, \gamma)$	$\frac{1}{\sqrt{2\pi y_t^3}}$	$\frac{1}{2}$	$\frac{(y_t-\gamma)^{-2}}{2y_t\gamma^2}$	$(0,\infty)$
Rayleigh	$(\omega, \boldsymbol{\beta}, \gamma)$	y_t	1	$\frac{1}{2}(y_t - \gamma)^{-2}$	$(0,\infty)$
Generalized Gamma	$(\omega, \boldsymbol{\beta}, \alpha, \upsilon)$	$\frac{\upsilon y_t^{\alpha-1}}{\Gamma\left(\frac{\alpha}{\upsilon}\right)}$	1	y_t^{υ}	$(0,\infty)$

† In this paper, only the heavy tailed distributions are studied.

of the maximum likelihood estimator, which is given by:

$$v(\omega, n_1, n_2) = \frac{\Gamma(n_1 + n_2)}{\Gamma(n_1)\Gamma(n_2)}\omega^{n_1-1}(1-\omega)^{n_2-1}, \tag{7.3}$$

where, $n_1 = \left\{\frac{n+1}{n}, \left(\frac{n+1}{n}\right)^{\frac{1}{2}}, \left(\frac{n+1}{n}\right)^{\frac{1}{3}}\right\}$ and $n_2 = \left\{\frac{n+1}{n}, \left(\frac{n+1}{n}\right)^{\frac{1}{2}}, \left(\frac{n+1}{n}\right)^{\frac{1}{3}}\right\}$.

Pinho et al. (2012) proposed the penalized likelihood function as the multiplication of the likelihood function, $L_1(\boldsymbol{\varphi}; \boldsymbol{Y_n}) = \prod_{t=1}^{n} p(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi})$, and the penalty function $v(\omega, n_1, n_2)$. Thus

$$L_2(\boldsymbol{\varphi}; \boldsymbol{Y_n}) = \prod_{t=1}^{n} p(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi}) \times v(\omega, n_1, n_2), \tag{7.4}$$

where $p(y_t | \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi})$ is given in equation 7.2 and $v(\omega, n_1, n_2)$ is given in equation 7.3.

Then, the penalized log-likelihood function is calculated as

$$\begin{aligned}\ell_2(\boldsymbol{\varphi};\boldsymbol{Y_n}) &= \sum_{t=1}^{n}\ln\Gamma\left(r(y_t,\boldsymbol{\varphi})+c_{t|t-1}\right)+\sum_{t=1}^{n}\ln\left(q(y_t,\boldsymbol{\varphi})\right)-\sum_{t=1}^{n}\ln\Gamma\left(c_{t|t-1}\right)\\ &+\sum_{t=1}^{n}c_{t|t-1}\ln\left(b_{t|t-1}\right)-\sum_{t=1}^{n}\left(r(y_t,\boldsymbol{\varphi})+c_{t|t-1}\right)\ln\left(s(y_t,\boldsymbol{\varphi})+d_{t|t-1}\right)\\ &+\sum_{t=1}^{n}\ln\left(\Gamma(n_1+n_2)\right)-\sum_{t=1}^{n}\ln\left(\Gamma(n_1)\right)+\sum_{t=1}^{n}(n_1-1)\ln(\omega)\\ &+\sum_{t=1}^{n}(n_2-1)\ln(1-\omega),\end{aligned}$$

Thus, the penalized maximum likelihood estimator (PMLE) for $\boldsymbol{\varphi}$ is given by

$$\hat{\boldsymbol{\varphi}}_{PMLE}=\arg\max_{\boldsymbol{\varphi}}\ell_2(\boldsymbol{\varphi};\boldsymbol{Y_n}).$$

$\ell_2(\boldsymbol{\varphi};\boldsymbol{Y_n})$ is a nonlinear function of $\boldsymbol{\varphi}$ and does not have an analytic form for the partial derivatives of the log-likelihood function and the penalized log-likelihood function, respectively, then numerical procedures should be used. In this paper the maximization method used is the BFGS algorithm proposed by Broyden (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970) because Pinho & Franco (2012) showed that the behavior of the penalized estimators is robust with respect to the maximization algorithm used.

Pinho et al. (2012) evaluated nine combinations of values of n_1 and n_2, for the penalty function. According to their results, in this paper it will be used the combination $n_1=\left(\frac{n+1}{n}\right)^{\frac{1}{2}}$ and $n_2=\left(\frac{n+1}{n}\right)^{\frac{1}{3}}$ as they presented the best results to reduce the bias and mean square error for all models. In Figure 7.1 it can be observed the behavior of the penalized function $v(\omega,n_1,n_2)$ for time series size 50 and 100, in the intervals $\omega=(0.00;1.00)$ (at left) and $\omega=(0.80;1.00)$ (at right).

The asymptotic confidence interval for $\boldsymbol{\varphi}$ is built based on a numerical approximation by BFGS for the Fisher information matrix $I_n(\boldsymbol{\varphi})$, using $I_n(\boldsymbol{\varphi})\cong -G(\boldsymbol{\varphi})$, where

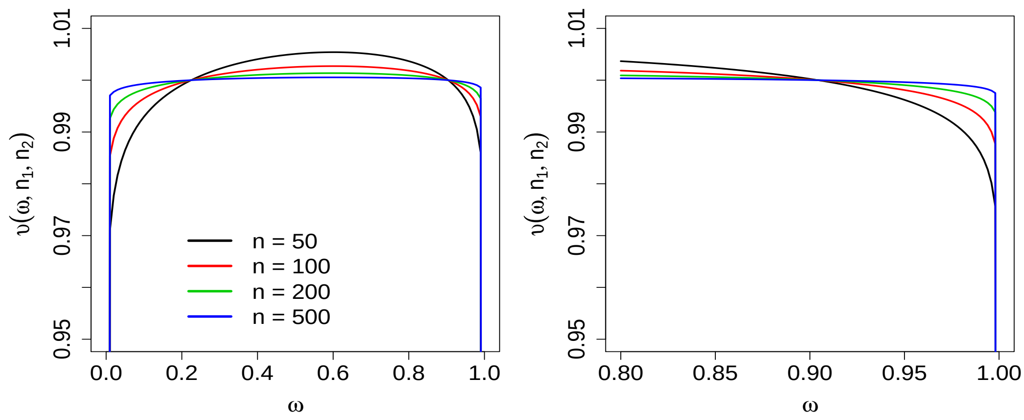


Figure 7.1: Penalty functions IV proposed to time series of size 50, 100, 200 and 500.

$-G(\boldsymbol{\varphi})$ is the matrix of second derivatives with respect to the parameters of the log-likelihood function $\ell_1(\boldsymbol{\varphi};\boldsymbol{Y_n}) = \ln L_1(\boldsymbol{\varphi};\boldsymbol{Y_n})$ or the log-penalized likelihood function $\ell_2(\boldsymbol{\varphi};\boldsymbol{Y_n}) = \ln L_2(\boldsymbol{\varphi};\boldsymbol{Y_n})$. As the computation of the derivatives is not an easy task, numerical derivatives are used (Franco et al., 2008).

Let φ_i, $i = 1, \ldots, p$, be any component of $\boldsymbol{\varphi}$. Then, an asymptotic confidence interval of $100(1-\phi)\%$ for φ_i is given by

$$\hat{\varphi}_i \pm z_{\phi/2}\sqrt{\widehat{Var}(\hat{\varphi}_i)},$$

where $z_{\phi/2}$ is the $\phi/2$ percentile of the standard normal distribution and $\widehat{Var}(\hat{\varphi}_i)$ is obtained from the diagonal elements of the Fisher information matrix.

7.3 Bootstrap methods

The jackknife proposed by Tukey (1958) and the bootstrap proposed by Efron (1979), under the condition of independent and identically distributed observations, have become well established as nonparametric estimators of the variance of a statistic.

Davis (1977), Freedman (1984) and Efron & Tibshirani (1986) extended this procedure to other measures of statistical accuracy such as bias and prediction error, and complicated data structures such as time series (ARMA models with inovations independent and identically distributed), censored data and regression models. Kunsch (1989) extended these proposals for the case where the observations form a general stationary sequence. Many other articles were published on bootstrap methods for ARMA family and its extensions, including: Thombs & Schucany (1990), McCullough (1994), Souza & Neto (1996), Buhlmann & Kunsch (1999), Pascual et al. (2000), Kim (2002), Franco & Reisen (2004) and Alonso et al. (2006).

In the context of the Gaussian state space model there is the pioneering work of Stoffer & Wall (1991), where the bootstrap is proposed as a method for assessing the precision of Gaussian maximum likelihood estimates of the parameters of linear state space models. After that, Stoffer & Wall (2002) and Stoffer & Wall (2004) discuss about a bootstrap approach to evaluate conditional forecast errors in ARMA models, using the state space form, and that a resampling procedure can provide insight into the validity of the model. Rodriguez & Ruiz (2009) proposed a bootstrap procedure for constructing prediction intervals in Gaussian state space models that does not need the backward representation of the model and is based on obtaining the intervals directly for the observations. Comparatively, the bootstrap procedure proposed by Stoffer & Wall (2002) is further complicated by the fact that the intervals are obtained for the prediction errors instead of the observations.

Franco & Souza (2002) and Franco et al. (2008) treat the problem of assessing the accuracy of hyperparameters for a specific Gaussian state space models (local level model, linear trend model and basic structural model). In these papers, a Monte Carlo study is used to compare the performance of parametric and nonparametric bootstrap in the calculation of standard deviations and confidence intervals for the hyperparameters.

Thus, in an attempt to obtain better confidence intervals for the parameters of the NGSSM, this work proposes three different bootstrap procedures, along with three bootstrap confidence intervals introduced by Efron & Tibshirani (1993)

7.3.1 Bootstrap schemes

In this paper, three bootstrap schemes are evaluated for the NGSSM.

- **Scheme 01 (parametric bootstrap)**

  Step 1: Obtain the maximum likelihood estimates $\hat{\boldsymbol{\varphi}}$ of the vector parameter $\boldsymbol{\varphi}$;

  Step 2: Generate B bootstrap series y_t^*, of size T, where $y_t^* \sim NGSSM\left(\mu_t, \hat{\boldsymbol{\varphi}}\right)$;

  Step 3: Obtain the bootstrap maximum likelihood estimatives $\boldsymbol{\varphi}^*$ of the vector parameter $\boldsymbol{\varphi}$.

  This bootstrap scheme was proposed by Efron & Tibshirani (1993).

- **Scheme 02 (bootstrap on standardized Pearson residual)**

  Step 1: Obtain the maximum likelihood estimates $\hat{c}_{t|t-1}$ and $\hat{d}_{t|t-1}$ of parameters $c_{t|t-1}$ and $d_{t|t-1}$ of the *prior* distribution of the dinamic parameter μ_t and the maximum likelihood estimates $\hat{\boldsymbol{\varphi}}$ of the vector parameter $\boldsymbol{\varphi}$;

  Step 2: Calculate $\hat{\mu} = \frac{\hat{c}_{t|t-1}}{\hat{d}_{t|t-1}}$, $\hat{y}_t = E\left(y_t|\, \hat{\mu}_t, \hat{\boldsymbol{\varphi}}\right)$, $Var\left(y_t|\, \hat{\mu}_t, \hat{\boldsymbol{\varphi}}\right)$ and $\hat{\varepsilon}_t = \frac{y_t - \hat{y}_t}{\sqrt{Var(y_t|\hat{\mu}_t, \hat{\boldsymbol{\varphi}})}}$ (standardized Pearson residual);

  Step 3: Resample $\hat{\varepsilon}_t$ and obtain B samples ε_t^* independent and identically distributed, of size T;

  Step 4: Obtain B bootstrap series y_t^*, of size T, by $y_t^* = \hat{y}_t + \varepsilon_t^* \sqrt{Var\left(y_t|\, \hat{\mu}_t, \hat{\boldsymbol{\varphi}}\right)}$;

  Step 5: Obtain the bootstrap maximum likelihood estimates $\boldsymbol{\varphi}^*$ of the vector parameter $\boldsymbol{\varphi}$.

  This bootstrap scheme was adapted to NGSSM from Davison & Hinkley (1997).

- **Scheme 03 (bootstrap on transformed standardized Pearson residual)**

  Step 1: Obtain the maximum likelihood estimates $\hat{c}_{t|t-1}$ and $\hat{d}_{t|t-1}$ of parameters $c_{t|t-1}$ and $d_{t|t-1}$ of the *prior* distribution of the dinamic parameter μ_t and obtain the estimates of the maximum likelihood $\hat{\boldsymbol{\varphi}}$ of the vector parameter $\boldsymbol{\varphi}$;

  Step 2: Calculate $\hat{\mu} = \frac{\hat{c}_{t|t-1}}{\hat{d}_{t|t-1}}$, $\hat{y}_t = E\left(y_t|\, \hat{\mu}_t, \hat{\boldsymbol{\varphi}}\right)$, $Var\left(y_t|\, \hat{\mu}_t, \hat{\boldsymbol{\varphi}}\right)$ and $\hat{\varepsilon}_t = \frac{h(y_t)-h(\hat{y}_t)}{\sqrt{Var(y_t|\hat{\mu}_t,\hat{\boldsymbol{\varphi}})\dot{h}^2(\hat{y}_t)}}$ (transformed standardized Pearson residual);

  Step 3: Resample $\hat{\varepsilon}_t$ and obtain B samples ε_t^* independent and identically distributed, of size T;

  Step 4: Obtain B bootstrap series y_t^*, of size T, by
  $y_t^* = h^{-1}\left[h\left(\hat{y}_t\right) + \varepsilon_t^* \sqrt{Var\left(y_t|\, \hat{\mu}_t, \hat{\boldsymbol{\varphi}}\right) \dot{h}^2\left(\hat{y}_t\right)}\right]$;

  Step 5: Obtain the maximum likelihood bootstrap estimates of $\boldsymbol{\varphi}^*$ of the vector parameter $\boldsymbol{\varphi}$.

This bootstrap scheme was adapted to NGSSM from Davison & Hinkley (1997).

7.3.2 Bootstrap confidence intervals

In this work three methods proposed by Efron & Tibshirani (1986) to construct bootstrap confidence intervals are employed. They are: Percentile interval (%Int), Bootstrap-t (Boot-t) and Bias-corrected (BC). For each one of the methods described below it is first necessary to generate B bootstrap series $y_t^{*1}, y_t^{*2}, \cdots, y_t^{*B}$ and calculate the bootstrap estimate of parameter φ, $\hat{\varphi}^*$. A short description of each method follows.

- **Percentile**

  The ϕ and $(1-\phi)$ percentiles of the bootstrap distribution of $\hat{\varphi}$ can be defined by

$$\left[\hat{\varphi}^{*(\phi)}; \hat{\varphi}^{*(1-\phi)}\right].$$

Thus, after estimating the values of φ for each of the B bootstrap series, take the $100\phi^{th}$ ordered value as the lower interval point and the $100\left(1-\phi\right)^{th}$ ordered value as the upper interval point.

- **Bootstrap-t**

After generating the bootstrap series, compute the statistic

$$Z^{*b} = \frac{\hat{\varphi}^{*b} - \hat{\varphi}}{\hat{se}^{*b}},$$

where $\hat{se}^{*b}$ is the estimated standard error of $\hat{\varphi}^{*}$ for the bootstrap series y_t^{*b}. After the generation of bootstrap series, a table of percentiles of the empirical distribution Z^{*b} is obtained. Thus, the bootstrap-t confidence interval is given by

$$\left[\hat{\varphi} - \hat{t}^{(1-\phi)}\hat{se}; \hat{\varphi} - \hat{t}^{(\phi)}\hat{se}\right],$$

where $\hat{t}^{(1-\phi)}$ and $\hat{t}^{(\phi)}$ are, respectively, the ϕ and $(1-\phi)$ percentile of the empirical distributiton of Z^{*b} and $\hat{se}$ is the standard error of $\hat{\varphi}$, which can be obtained though the bootstrap samples.

- **Bias-corrected**

The Bias-corrected interval is defined by

$$\left[\hat{\varphi}^{*(\phi_1)}; \hat{\varphi}^{*(\phi_2)}\right],$$

where $\phi_1 = \Phi\left(2z_0 + z^{(\phi)}\right)$ and $\phi_2 = \Phi\left(2z_0 + z^{(1-\phi)}\right)$. The function Φ is the cumulative distribution function of a standard normal $N\left(0;1\right)$ and $z^{(\phi)}$ its $100\phi^{th}$ percentile point. The value of $\hat{z}_0$ is calculated using the proportion of $\hat{\varphi}^{*b}$ in the bootstrap samples that are smaller than the $\hat{\varphi}$ in the original series. Then:

$$z_0 = \Phi^{-1}\left(\frac{\#\hat{\varphi}^{*b} < \hat{\varphi}}{B}\right).$$

7.4 Monte Carlo study

In this section the performance of the bootstrap methods and bootstrap confidence intervals for parameters of the NGSSM are evaluated through a Monte Carlo experiment using the maximum likelihood estimator (MLE) and the penalized maximum likelihood estimator (PMLE) as defined in Section 7.2. The asymptotic confidence interval and bootstrap confidence interval for the parameter vector are presented and they are compared with respect to the coverage rate, for a fixed level of 95% ($\phi = 0.05$). The NGSSM cases evaluated are the heavy tailed models. They are: Log-normal (LN), Log-gamma (LG), Fréchet (F), Lévy (L), Skew GED (SGED), Pareto (P) and Weibull (W) models.

To obtain the estimates of maximum likelihood or penalized maximum likelihood of the NGSSM parameters is used the BFGS algorithm.

To obtain the estimates of bootstrap interval by Scheme 03 (bootstrap on transformed standardized Pearson residual) is used $h(\bullet) = ln(\bullet)$.

The number of Monte Carlo and bootstrap replications was set equal to 1,000 for time series of size $n = \{50\}$, generated with a covariate $x_t = \sin(2\pi t/12)$, $t = 1, \ldots, n$. For all distributions $\omega = (0.85; 0.90; 0.95)$ and the coefficient of the covariate is $\beta = 1.0$.

Specific parameters were set as follows: Log-normal ($\delta = 5.0$), Log-gamma ($\alpha = 5.0$), Fréchet ($\alpha = 5.0$), Skew GED ($\delta = 5.0$, $\alpha = 1.5$, $\kappa = 1.0$) and Weibull ($\upsilon = 5.0$). For the Log-normal, Fréchet and Lévy models the parameter γ was fixed at 0.0. For the Skew GED model the parameter α was fixed at 1.5, thus, there is a distribution with a tail heavier than the Skew Normal ($\alpha = 2.0$) and lighter than the Skew Laplace (both are particular cases of the Skew GED).

To calculate the maximum likelihood estimator the BFGS assumed as initial state condition $\lambda_0 | Y_0 \sim \text{Gamma}\,(0.01; 0.01)$, $\omega_0 = 0.50$ and $\beta_0 = \delta_0 = \alpha_0 = \upsilon_0 = \kappa_0 = 0.01$.

All codes for NGSSM were developed by the authors in Ox Metrics.

Table 7.2 presents the interval estimates of the **MLE** for the vector parameter $\boldsymbol{\varphi}$ of the 1000 Monte Carlo, for time series of size 50. The intervals are the asymptotic confidence interval (Asym Int), the percentile bootstrap interval (% Int), the bootstrap-t interval (Boot-t) and bootstrap bias-corrected (BC) (three bootstrap intervals by **parametric bootstrap methods**). Except for the asymptotic confidence interval of the Log-Gamma model which has a coverage rate very close to the nominal level of 95%, the asymptotic confidence interval, for all other models had unsatisfactory results. More specifically, the results were unsatisfactory to all models for the parameter ω which presented a coverage rate far above the nominal level and for the Skew GED model, where the parameters δ and κ presented a coverage rate far below the nominal level. In general, the three bootstrap intervals show worse results than the asymptotic confidence interval when is used the MLE.

Table 7.3 presents four interval estimates of the **PMLE** for the vector parameter $\boldsymbol{\varphi}$ of the 1000 Monte Carlo, for time series of size 50. The intervals are the asymptotic confidence interval and three bootstrap intervals by **parametric bootstrap method**. It is easy to see that the BC interval present for all parameters of the Weibull and Fréchet models a coverage rate almost equal to the nominal rate (difference less than 0.007). The BC interval, for all parameters of the Log-Normal, Pareto, Lévy Skew GED models the difference between the coverage rate and the nominal rate is less than 0.015. The Boot-t interval for all parameters of the Féchet, Lévy and Skew GED models show also a difference between the coverage rate and the nominal level less than 0.015. The intervals can be observed in Figures 7.2 and 7.3.

Table 7.4 presents four interval estimates of the **PMLE** for the vector parameter $\boldsymbol{\varphi}$ of the 1000 Monte Carlo, for time series of size 50. The intervals are the asymptotic confidence interval and three bootstrap intervals by **standardized Pearson residual bootstrap method**. It is easy to see that only the BC interval, for all parameters of Log-normal model provides satisfactory results. For other models, at least for a parameter, the bootstrap intervals show a big difference between the coverage rate and the nominal rate.

The estimates of bootstrap on standardized Pearson residual and transformed standardized Pearson residual are nearly equal, then in this work the results of transformed standardized Pearson residual will be omitted.

7.5 Conclusion

This paper has employed bootstrap techniques to obtain the empirical distribution of the estimates of parameters of the non Gaussian State Space family proposed by Santos et al. (2010) and extended to heavy tailed distributions by Pinho et al. (2012) with the objective of refining the parameter interval estimates, for time series of small sizes.

It can be concluded that the best confidence interval was the bootstrap bias-corrected interval (BC) obtained by parametric bootstrap when the PMLE proposed by Pinho & Franco (2012) was used.

Therefore, it can also be concluded that the penalty function proposed by Pinho & Franco (2012), besides improving the point estimates of parameter vector φ also improves the interval estimates when it is reconciled with parametric bootstrap method and bootstrap bias-corrected interval.

Acknowledgements

The authors wish to acknowledge CAPES, CNPq and FAPEMIG for financial support.

Table 7.2: Parametric Bootstrap - bootstrap estimates, range and coverage rate by MLE.

		MLE - BFGS			
Model	φ	**Conf Int Range (Cov Rate)**	**% Int Range (Cov Rate)**	**Boot-t Range (Cov Rate)**	**BC Range (Cov Rate)**
	$\omega = 0.95$	[0.7797 ; 1.0000]	[1.0000 ; 1.0000]	[1.0000 ; 1.0000]	[1.0000 ; 1.0000]
		0.2203	0.0000	0.0000	0.0000
		(1.000)	(0.000)	(0.000)	(0.000)
LN	$\beta = 1.0$	[0.4301 ; 1.6377]	[0.2867 ; 1.5380]	[0.4085 ; 1.6598]	[0.4126 ; 1.6645]
		1.2076	1.2513	1.2513	1.2519
		(0.937)	(0.963)	(0.948)	(0.947)
	$\delta = 5.0$	[4.9747 ; 5.0260]	[4.9736 ; 5.0270]	[4.9736 ; 5.0270]	[4.9736 ; 5.0270]
		0.0513	0.0534	0.0534	0.0534
		(0.955)	(0.967)	(0.967)	(0.964)
	$\omega = 0.95$	[0.6963 ; 0.9940]	[0.4400 ; 0.6350]	[0.8851 ; 1.0383]	[0.4511 ; 0.6355]
		0.2977	0.1950	0.1532	0.1844
		(0.963)	(0.269)	(0.205)	(0.271)
LG	$\beta = 1.0$	[0.8147 ; 1.1786]	[0.1809 ; 0.2810]	[0.9571 ; 1.0396]	[0.2157 ; 0.3104]
		0.3639	0.1001	0.0825	0.0947
		(0.953)	(0.153)	(0.193)	(0.247)
	$\alpha = 5.0$	[3.2076 ; 7.5284]	[0.7653 ; 4.5321]	[4.5973 ; 6.9919]	[0.7883 ; 4.8424]
		4.3208	3.7668	2.3946	4.0541
		(0.957)	(0.267)	(0.205)	(0.271)
	$\omega = 0.95$	[0.5760 ; 0.9961]	[0.7576 ; 0.9785]	[0.8008 ; 1.0217]	[0.7885 ; 0.9785]
		0.4201	0.2209	0.2209	0.1900
P		(0.986)	(0.957)	(0.953)	(0.897)
	$\beta = 1.0$	[0.5665 ; 1.4021]	[0.4087 ; 1.2212]	[0.5785 ; 1.3910]	[0.5389 ; 1.3481]
		0.8356	0.8125	0.8125	0.8092
		(0.952)	(0.887)	(0.911)	(0.917)
	$\omega = 0.95$	[0.6468 ; 0.9960]	[0.7869 ; 0.9570]	[0.8375 ; 1.0076]	[0.8010 ; 0.9570]
		0.3492	0.1701	0.1701	0.1560
		(0.982)	(0.914)	(0.874)	(0.815)
W	$\beta = 1.0$	[0.5445 ; 1.4785]	[0.3645 ; 1.1722]	[0.6266 ; 1.4343]	[0.5133 ; 1.3595]
		0.9340	0.8077	0.8077	0.8462
		(0.950)	(0.855)	(0.850)	(0.861)
	$\upsilon = 5.0$	[3.9872 ; 6.1508]	[3.5604 ; 5.3369]	[4.2370 ; 6.0135]	[3.7283 ; 5.5874]
		2.1636	1.7765	1.7765	1.8591
		(0.975)	(0.865)	(0.887)	(0.892)
	$\omega = 0.95$	[0.6460 ; 0.9962]	[0.7851 ; 0.9540]	[0.8384 ; 1.0072]	[0.7953 ; 0.9540]
		0.3502	0.1689	0.1688	0.1587
		(0.989)	(0.908)	(0.863)	(0.836)
F	$\beta = 1.0$	[0.5351 ; 1.4663]	[0.3552 ; 1.1557]	[0.6187 ; 1.4193]	[0.5027 ; 1.3412]
		0.9312	0.8005	0.8006	0.8385
		(0.957)	(0.836)	(0.855)	(0.865)
	$\alpha = 5.0$	[3.9858 ; 6.1638]	[3.5382 ; 5.2995]	[4.2494 ; 6.0107]	[3.7020 ; 5.5466]
		2.1780	1.7613	1.7613	1.8446
		(0.963)	(0.862)	(0.871)	(0.877)
	$\omega = 0.95$	[0.6576 ; 0.9978]	[0.8477 ; 0.9895]	[0.8657 ; 1.0076]	[0.8701 ; 0.9895]
		0.3402	0.1418	0.1419	0.1194
		(0.994)	(0.979)	(0.933)	(0.759)
L	$\beta = 1.0$	[0.3838 ; 1.6011]	[0.2421 ; 1.4304]	[0.3981 ; 1.5864]	[0.3804 ; 1.5704]
		1.2173	1.1883	1.1883	1.1900
		(0.960)	(0.934)	(0.944)	(0.941)
	$\omega = 0.95$	[0.5961 ; 0.9964]	[0.8090 ; 0.9970]	[0.8269 ; 1.0149]	[0.8330 ; 0.9970]
		0.4003	0.1880	0.1880	0.1640
		(0.983)	(0.994)	(0.974)	(0.873)
SGED	$\beta = 1.0$	[0.4879 ; 1.5504]	[0.3301 ; 1.4284]	[0.4669 ; 1.5651]	[0.4657 ; 1.5607]
		1.0625	1.0983	1.0982	1.0950
		(0.952)	(0.961)	(0.953)	(0.950)
	$\delta = 5.0$	[4.9867 ; 5.0139]	[4.9528 ; 4.9878]	[4.9828 ; 5.0177]	[4.9528 ; 4.9877]
		0.0272	0.0350	0.0349	0.0349
		(0.858)	(0.954)	(0.950)	(0.952)
	$\kappa = 1.0$	[0.7264 ; 1.3178]	[0.7131 ; 1.4587]	[0.7023 ; 1.4479]	[0.7137 ; 1.4635]
		0.5914	0.7456	0.7456	0.7498
		(0.906)	(0.945)	(0.949)	(0.938)

Table 7.3: Parametric Bootstrap - bootstrap estimates, range and coverage rate by PMLE.

		PMLE IV - BFGS			
Model	φ	**Conf Int** **Range** **(Cov Rate)**	**% Int** **Range** **(Cov Rate)**	**Boot-t** **Range** **(Cov Rate)**	**BC** **Range** **(Cov Rate)**
	$\omega = 0.95$	[0.5067 ; 0.9929] 0.4862 (0.980)	[0.8191 ; 0.9741] 0.1550 (1.000)	[0.8346 ; 0.9896] 0.1550 (0.918)	[0.8484 ; 0.9744] 0.1260 (0.938)
LN	$\beta = 1.0$	[0.4098 ; 1.6329] 1.2231 (0.934)	[0.2528 ; 1.5351] 1.2823 (0.958)	[0.3789 ; 1.6611] 1.2822 (0.939)	[0.3843 ; 1.6665] 1.2822 (0.942)
	$\delta = 5.0$	[4.9750 ; 5.0250] 0.0500 (0.918)	[4.9725 ; 5.0276] 0.0551 (0.954)	[4.9724 ; 5.0275] 0.0551 (0.952)	[4.9724 ; 5.0276] 0.0552 (0.954)
	$\omega = 0.95$	[0.4749 ; 0.9875] 0.5126 (0.947)	[0.2901 ; 0.9643] 0.6742 (0.984)	[0.5455 ; 1.2188] 0.6733 (0.962)	[0.4985 ; 0.9702] 0.4717 (0.918)
LG	$\beta = 1.0$	[0.8207 ; 1.1794] 0.3587 (0.931)	[0.6786 ; 1.0549] 0.3763 (0.944)	[0.8132 ; 1.1901] 0.3769 (0.905)	[0.8125 ; 1.1685] 0.3560 (0.930)
	$\alpha = 5.0$	[3.2960 ; 7.6625] 4.3665 (0.965)	[2.7525 ; 14.1047] 11.3522 (0.998)	[2.1786 ; 13.3538] 11.1752 (0.962)	[2.8077 ; 14.7801] 11.9724 (1.000)
P	$\omega = 0.95$	[0.5015 ; 0.9913] 0.4898 (0.976)	[0.7501 ; 0.9723] 0.2222 (1.000)	[0.7814 ; 1.0036] 0.2222 (0.985)	[0.8119 ; 0.9735] 0.1616 (0.957)
	$\beta = 1.0$	[0.5818 ; 1.4204] 0.8386 (0.957)	[0.4354 ; 1.2914] 0.8560 (0.940)	[0.5731 ; 1.4291] 0.8560 (0.961)	[0.5752 ; 1.4305] 0.8553 (0.964)
	$\omega = 0.95$	[0.4863 ; 0.9909] 0.5046 (0.970)	[0.7750 ; 0.9723] 0.1973 (1.000)	[0.7935 ; 0.9909] 0.1974 (0.944)	[0.8122 ; 0.9727] 0.1605 (0.948)
W	$\beta = 1.0$	[0.5540 ; 1.4975] 0.9435 (0.953)	[0.4207 ; 1.3308] 0.9101 (0.961)	[0.5944 ; 1.5045] 0.9101 (0.931)	[0.5807 ; 1.5358] 0.9551 (0.950)
	$\upsilon = 5.0$	[4.0670 ; 6.2413] 2.1743 (0.958)	[4.0274 ; 6.1099] 2.0825 (0.962)	[4.1911 ; 6.2736] 2.0825 (0.946)	[4.1723 ; 6.3352] 2.1629 (0.951)
	$\omega = 0.95$	[0.4851 ; 0.9913] 0.5062 (0.974)	[0.7789 ; 0.9725] 0.1936 (0.999)	[0.7982 ; 0.9917] 0.1935 (0.962)	[0.8187 ; 0.9729] 0.1542 (0.951)
F	$\beta = 1.0$	[0.5346 ; 1.4700] 0.9354 (0.962)	[0.4037 ; 1.3111] 0.9074 (0.953)	[0.5709 ; 1.4784] 0.9075 (0.948)	[0.5595 ; 1.5100] 0.9505 (0.955)
	$\alpha = 5.0$	[4.0563 ; 6.2150] 2.1587 (0.956)	[4.0272 ; 6.0987] 2.0715 (0.960)	[4.1774 ; 6.2489] 2.0715 (0.946)	[4.1612 ; 6.3033] 2.1421 (0.956)
L	$\omega = 0.95$	[0.5144 ; 0.9935] 0.4791 (0.987)	[0.8284 ; 0.9742] 0.1458 (1.000)	[0.8430 ; 0.9888] 0.1458 (0.935)	[0.8555 ; 0.9745] 0.1190 (0.931)
	$\beta = 1.0$	[0.3835 ; 1.5861] 1.2026 (0.935)	[0.2310 ; 1.4571] 1.2261 (0.947)	[0.3710 ; 1.5971] 1.2261 (0.950)	[0.3750 ; 1.6024] 1.2274 (0.954)
	$\omega = 0.95$	[0.4932 ; 0.9924] 0.4992 (0.976)	[0.7889 ; 0.9742] 0.1853 (0.999)	[0.8105 ; 0.9958] 0.1853 (0.947)	[0.8303 ; 0.9763] 0.1460 (0.948)
SGED	$\beta = 1.0$	[0.4757 ; 1.5369] 1.0612 (0.947)	[0.3143 ; 1.4302] 1.1159 (0.957)	[0.4445 ; 1.5604] 1.1159 (0.950)	[0.4500 ; 1.5641] 1.1141 (0.956)
	$\delta = 5.0$	[4.9867 ; 5.0132] 0.0265 (0.864)	[4.9824 ; 5.0175] 0.0351 (0.974)	[4.9823 ; 5.0175] 0.0352 (0.962)	[4.9824 ; 5.0175] 0.0351 (0.965)
	$\kappa = 1.0$	[0.7288 ; 1.2988] 0.5700 (0.908)	[0.7166 ; 1.4367] 0.7201 (0.950)	[0.7013 ; 1.4208] 0.7195 (0.943)	[0.7174 ; 1.4426] 0.7252 (0.941)

Table 7.4: Bootstrap on standardized Pearson residual - bootstrap estimates, range and coverage rate by PMLE.

		PMLE IV - BFGS		
Model	φ	**% Int Range (Cov Rate)**	**Boot-t Range (Cov Rate)**	**BC Range (Cov Rate)**
	$\omega = 0.95$	[0.7591 ; 0.9707]	[0.8082 ; 1.0200]	[0.8376 ; 0.9743]
		0.2116	0.2118	0.1367
		(0.988)	(0.988)	(0.948)
LN	$\beta = 1.0$	[0.2345 ; 1.6222]	[0.3158 ; 1.7041]	[0.3216 ; 1.7113]
		1.3877	1.3883	1.3897
		(0.975)	(0.969)	(0.966)
	$\delta = 5.0$	[4.9186 ; 5.0270]	[4.9235 ; 5.0320]	[4.9291 ; 5.0276]
		0.1084	0.1085	0.0985
		(0.947)	(0.944)	(0.952)
	$\omega = 0.95$	[0.3088 ; 0.9588]	[0.5505 ; 1.2004]	[0.5277 ; 0.9688]
		0.6500	0.6499	0.4411
		(0.874)	(0.994)	(0.923)
LG	$\beta = 1.0$	[0.6689 ; 1.0667]	[0.8037 ; 1.2017]	[0.8042 ; 1.1804]
		0.3978	0.3980	0.3762
		(0.776)	(0.971)	(0.960)
	$\alpha = 5.0$	[2.5685 ; 11.5139]	[2.9405 ; 11.8595]	[2.8145 ; 15.0984]
		8.9454	8.9190	12.2839
		(0.996)	(0.987)	(0.997)
	$\omega = 0.95$	[0.5000 ; 0.9623]	[0.8022 ; 1.2652]	[0.5480 ; 0.9739]
		0.4623	0.4630	0.4259
P		(0.905)	(1.000)	(0.982)
	$\beta = 1.0$	[0.0100 ; 1.1925]	[0.6631 ; 1.8287]	[0.0100 ; 1.4746]
		1.1825	1.1656	1.4646
		(0.843)	(0.988)	(0.991)
	$\omega = 0.95$	[0.7722 ; 0.9706]	[0.8016 ; 1.0000]	[0.8248 ; 0.9723]
		0.1984	0.1984	0.1475
		(0.994)	(0.958)	(0.954)
W	$\beta = 1.0$	[0.3649 ; 1.2556]	[0.6018 ; 1.4925]	[0.5792 ; 1.5237]
		0.8907	0.8907	0.9445
		(0.905)	(0.914)	(0.943)
	$\upsilon = 5.0$	[3.6009 ; 5.6566]	[4.2067 ; 6.2624]	[4.0754 ; 6.3744]
		2.0557	2.0557	2.2990
		(0.914)	(0.953)	(0.981)
	$\omega = 0.95$	[0.6749 ; 0.9641]	[0.7456 ; 1.0349]	[0.8069 ; 0.9718]
		0.2892	0.2893	0.1649
		(0.910)	(0.991)	(0.948)
F	$\beta = 1.0$	[0.6869 ; 1.7226]	[0.5502 ; 1.5859]	[0.5943 ; 1.5619]
		1.0357	1.0357	0.9676
		(0.997)	(0.960)	(0.961)
	$\alpha = 5.0$	[2.0252 ; 3.4561]	[4.5140 ; 5.9449]	[3.9615 ; 4.2323]
		1.4309	1.4309	0.2708
		(0.002)	(0.794)	(0.082)
	$\omega = 0.95$	[0.5315 ; 0.9707]	[0.7323 ; 1.1715]	[0.6077 ; 0.9752]
		0.4392	0.4392	0.3675
		(0.985)	(1.000)	(0.960)
L	$\beta = 1.0$	[0.0100 ; 1.2523]	[0.5630 ; 1.8865]	[0.0913 ; 1.7789]
		1.2423	1.3235	1.6876
		(0.870)	(0.844)	(0.979)
	$\omega = 0.95$	[0.7378 ; 0.9713]	[0.7906 ; 1.0239]	[0.8232 ; 0.9773]
		0.2335	0.2333	0.1541
		(0.989)	(0.996)	(0.971)
SGED	$\beta = 1.0$	[0.2941 ; 1.4829]	[0.4013 ; 1.5896]	[0.4079 ; 1.5942]
		1.1888	1.1883	1.1863
		(0.967)	(0.969)	(0.961)
	$\delta = 5.0$	[4.9843 ; 5.0153]	[4.9843 ; 5.0154]	[4.9838 ; 5.0158]
		0.0310	0.0311	0.0320
		(0.904)	(0.931)	(0.920)
	$\kappa = 1.0$	[0.7071 ; 1.4371]	[0.6890 ; 1.4186]	[0.7036 ; 1.4439]
		0.7300	0.7296	0.7403
		(0.930)	(0.947)	(0.943)

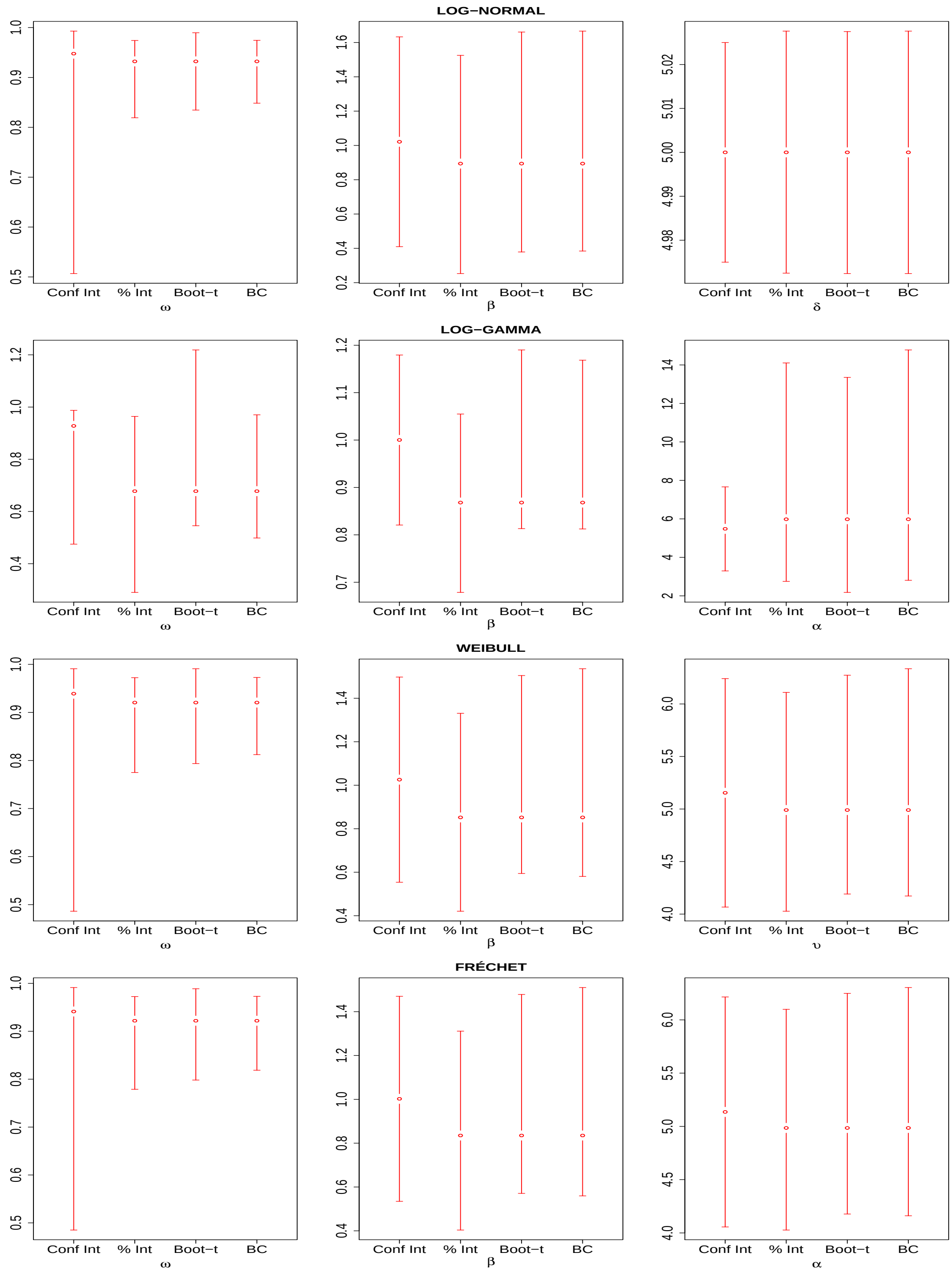


Figure 7.2: Parametric Bootstrap - Asymptotic confidence interval and bootstrap confidence interval by PMLE for the estimates of vector parameter φ of the Log-normal, Log-gamma, Weibull and Fréchet models.

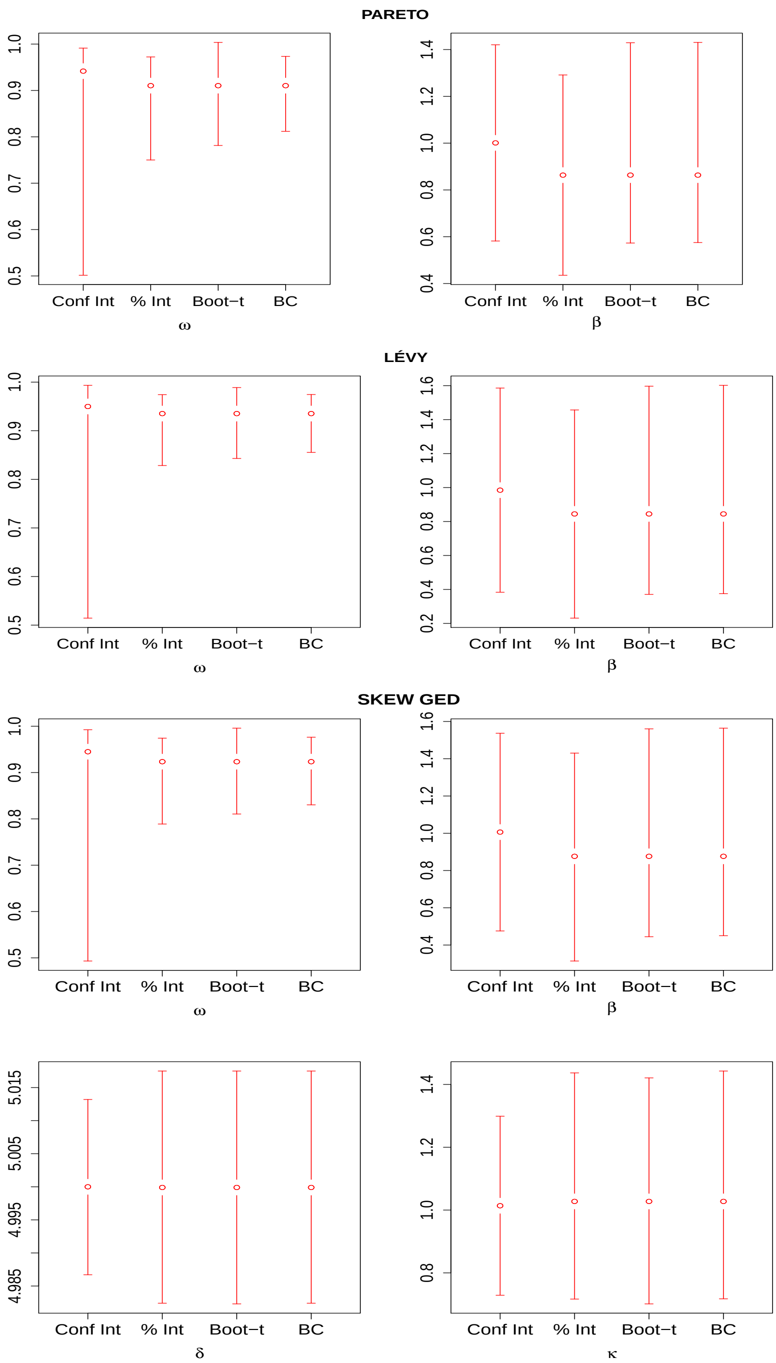


Figure 7.3: Parametric Bootstrap - Asymptotic confidence interval and bootstrap confidence inverval by PMLE for the estimates of vector parameter φ of the Pareto, Lévy and Skew GED models.

References

Alonso, A.M., Daniel, P., Romo, J., 2006. Introducing model uncertainty by moving blocks bootstrap. Statistical Papers 47, 167-179.

Broyden, C.G., 1970. The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics & Its Applications, 6, 76-90.

Buhlmann, P., Kunsch, H.R., 1999. Block length selection in the bootstrap for time series. Computational Statistics & Data Analysis, 31, 295-310.

Davis, W.W., 1977. Robust interval estimation of the innovation variance of an ARMA model. The Annals of Statistics, 5(4), 700-708.

Davison, A.C, Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge University Press.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1), 1-26.

Efron, B., Tibshirani, R.J., 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy (with discussion). Statistical Science, 1(1), 54-77.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Chapman & Hall, New York.

Fletcher, R., 1970. A new approach to variable metric algorithms. The Computer Journal, 13(3), 317-322.

Franco, G.C., Reisen, V.A., 2004. Bootstrap Techniques in Semiparametric Estimation Methods for ARFIMA Models: A Comparison Study. Computational Statistics, 19, 243-259.

Franco, G.C., Souza, R.C., 2002. A Comparison of Methods for Bootstrapping in the Local Level Model. Journal of Forecasting, 21, 27-38.

Franco, G.C., Santos, T.R., Ribeiro, J.A., Cruz, F.R., 2008. Confidence intervals for hyperparameters in structural models. Communications in Statistcs: Simulation and Computation, 37 (3), 486-497.

Freedman, D.A., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models. The Annals of Statistics, 12(3), 827-842.

Goldfard, D., 1970. A family of variable metric updates derived by variational means. Mathematics of Computation, 24(109), 23-26.

Hahn, J., Newey, W.A., 2004. Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. Econometrica, 72(4), 1295-1319.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Harvey, A.C., Fernandes, C., 1989. Time Series Models for Count or Qualitative Observations. Journal of Business & Economic Statistics, 7(4), 407-417.

Kim, J.H., 2002. Bootstrap Prediction Intervals for Autoregressive Models of Unknown or Infinite Lag Order. Journal of Forecasting, 21, 265-280.

Kunsch, H.R., 1989. The Jackknife and the bootstrap for general stationary observations. The Annals of Statistics, 17(3), 1217-1241.

Lawrence, C.T., Tits, A.L., 2001. A computationally efficient feasible sequential quadratic programming algorithm. SIAM Journal of Optimization, 11(4), 1092-1118.

Loughin, T.M., 1998. On the Bootstrap and Monotone Likelihood in the Cox Proportional Hazards Regression Model. Lifetime Data Analysis, 4, 393-403.

McCullough, B.D., 1994. Bootstrapping forecast intervals: An application to $AR(p)$ models. Journal of Forecasting, 13(1), 51-66.

Nocedal, J., Wright, J.W., 1999. Numerical Optimization. Springer Verlag, New York.

Pascual, L., Romo, J., Ruiz, E., 2000. Bootstrap predictive inference for ARIMA processes. Journal of Time Series Analysis, 25(4), 449-465.

Pinho, F.M., Franco, G.C., Silva, R.S., 2012. Modelling Volatility Using State Space

Models with Heavy Tailed Distributions. Working paper.

Pinho, F.M., Franco, G.C., 2012. Penalized Likelihood for a Non Gaussian State Space Model Considering Heavy Tailed Distributions. Working paper.

Politis, D.M., Romano, J.P., 1994. The Stationary Bootstrap. Journal of the American Statistical Association, 89(428), 1303-1313.

Rodriguez, A., Ruiz, E., 2009. Bootstrap prediction intervals in state space models. Journal of Time Series Analysis, 30(2), 167-178.

Santos, T.R., Franco, G.C., Gamerman, D., 2010. Gamma family of dynamic models. Technical Report, 234, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro.

Souza, R.C., Neto, A.C., 1996. A Bootstrap Simulation Study in $ARMA(p,q)$ Structures. Journal of Forecasting, 15(4), 343-353.

Shanno, D.F., 1970. Conditioning of quase-Newton methods for function minimization. Mathematics of Computation, 24(111), 647-656.

Smith, R.L., Miller, J.E., 1986. A non-Gaussian state space model and application to prediction of records. Journal of the Royal Statistical Society, Series B, 48(1), 79-88.

Stoffer, D.S., Wall, K.D., 1991. Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter. Journal of the American Statistical Association, 86(416), 1024-1033.

Stoffer, D.S., Wall, K.D., 2002. A state space approach to bootstrapping conditional forecasts in ARMA models. Journal of Time Series Analysis, 23(6), 733-751.

Stoffer, D.S., Wall, K.D., 2004. Resampling in State Space Models. Chapter 9 of State Space and Unobserved Component Models: Theory and Applications. A. Harvey, S.J. Koopman and N. Shephard (edictors). Cambridge University Press.

Thombs, L.A., Schucany, W.R., 1990. Bootstrap Prediction Intervals for Autoregression. Journal of the American Statistical Association, 85(410), 486-492.

Tukey, J., 1958. Bias and confidence in not quite large samples (abstract). The

Annals of Mathematical Statistics, 29(2), 614.

Chapter 8

Considerações Finais

Este trabalho teve como objetivo geral ampliar o conhecimento sobre os NGSSM quanto às distribuições nela contidas, quanto aos métodos de estimação dos parâmetros e quanto a sua aplicabilidade a conjuntos de dados reais. Pode-se elencar novos conhecimentos produzidos a partir deste trabalho:

- Demonstrou-se que outras cinco distribuições de caudas pesadas estão contidas na NGSSM, além as propostas por Santos et al. (2010). São elas a Log-normal, Log-gama, Fréchet, Lévy e Skew GED.

- Observou-se, empiricamente, que o estimador de máxima verossimilhança e os estimadores bayesianos (média e mediana a posteriori), para os parâmetros da NGSSM, são assintoticamente não viesados e consistentes.

- Observou-se, empiricamente, que o estimador de máxima verossimilhança sobrestima o parâmetro ω e, por consequência, subestima a variabilidade de séries temporais pequenas. Estes resultados provocaram a necessidade da proposição de estimadores pontuais clássicos mais adequados.

- Propôs-se estimadores de máxima verossimilhança penalizados, para os parâmetros da NGSSM, a fim de mitigar o viés apresentado pelo estimador de máxima verossimilhança para séries temporais pequenas.

- Observou-se, empiricamente, que o estimador de máxima verossimilhança penalizado, para os parâmetros da NGSSM, proposto neste trabalho apresenta viés significativamente menor que o estimador de máxima verossimilhança.

- Demonstrou-se, por meio de Simulação Monte Carlo, que o intervalo de confiança assintótico e o intervalo de credibilidade apresentaram taxas de cobertura muito próximas às taxas nominais utilizadas no estudo empírico para séries temporais maiores que $n = 100$. Em contrapartida, os resultados do intervalo de confiança assintótico apresentaram taxas de cobertura distantes das taxas nominais utilizadas para séries temporais com $n = 50$. Estes resultados provocaram a necessidade da proposição de estimadores intervalares (considerando a inferência clássica) mais adequados.

- Propôs-se métodos bootstrap adaptados à NGSSM para a construção de intervalos de confiança bootstrap para os parâmetros da NGSSM.

- Observou-se, empiricamente, que os intervalos de confiança bootstrap com correção de viés obtido a partir do bootstrap paramétrico (método adaptado à NGSSM) apresentam taxas de cobertura muito próximas da taxa nominal utilizadas no estudo empírico.

- Demonstrou-se que para as séries $S\&P500$, $NASDAQ$, $IBOVESPA$, $INMEX$, $MERVAL$, $IPSA$, para o período de 02/01/2007 to 05/16/2011, que os modelos de cauda pesada da NGSSM apresentam melhores ajustes que os modelos da família GARCH, considerando-se os critérios AICc, BIC e log-verossimilhança.

- Demonstrou-se que para as séries $S\&P500$, $NASDAQ$, $IBOVESPA$, $INMEX$, $MERVAL$, $IPSA$, para o período de 02/01/2007 to 05/16/2011, dentre os modelos de cauda pesada da NGSSM, o modelo Weibull apresentou melhores ajustes, considerando-se os critérios AICc, BIC e log-verossimilhança.

A despeito de todas conclusões obtidas neste trabalho que propicia um maior conhecimento e uma melhor compreensão sobre a NGSSM, pode-se afirmar que há um vasto campo de pesquisa sobre esta nova família de modelos proposta por Santos et al. (2010). Pode-se elencar possíveis trabalhos futuros sobre a NGSSM.

- Desenvolver um pacote em R e/ou OxMetrics para facilitar o acesso de pesquisadores a esta nova família de modelos.
- Obter novas distribuições de probabilidade que são casos particulares da NGSSM.
- Extender a NGSSM por meio da substituição dos parâmetros estáticos dos modelos que estão contido no vetor de parâmetros φ em parâmetros dinâmicos.
- Extender a NGSSM para o caso multivariado.
- Avaliar mistura de modelos com a NGSSM, como por exemplo AR-NGSSM, MA-NGSSM, ARMA-NGSSM, ARMAX-NGSSM, dentre outros.
- Estimar e avaliar a qualidade dos ajustes dos modelos da NGSSM e comparar com outras famílias de modelos utilizados na literatura contemporânea para séries de commodities, outras séries financeiras e de outras outras áreas do conhecimento, tais como climatologia, confiabilidade, neurociência, dentre outras.
- Estimar e avaliar a qualidade dos ajustes dos modelos da NGSSM e comparar com outras famílias de modelos utilizados na literatura contemporânea para volatilidade realizada de séries financeiras.
- Explorar esta nova família de modelos dentro da teoria de gerenciamento de risco de ativos/portfólios de investimentos.

Referências Bibliográficas

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716-723.

Alonso, A.M., Daniel, P., Romo, J., 2006. Introducing model uncertainty by moving blocks bootstrap. Statistical Papers 47, 167-179.

Asmussen, S., 2000. Ruin Prbabilities. World Sicientic, Singapura.

Asmussen, S., 2003. Applied Probability and Queues. Springer, Berlin.

Anderson, J., 2001. On the normal inverse Gaussian stochastic volatility model. Journal of Business and Economic Statistics, 19, 44-54.

Ayebo, A., Kozubowski, T.J., 2003. An asymmetric generalization of Gaussian and Laplace laws. Journal of Probability and Statistical Science, 1, 187-210.

Bauwens, L., Laurent, S., Rombouts, J.V.K., 2006. Multivariate GARCH models: A survey. Journal of Applied Econometrics, 21, 79-109.

Bester, C.A., Hansen, C., 2009. A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects. Journal of Business and Economic Statistics, 27(2) 131-148.

Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. Regular Variation. Cambridge University Press, Cambridge.

Beirlant, J., Goegebeur, Y., Segers J., Teugels, J., 2004. Statistics Extremes: Theory and Applications. John Wiley & Sons.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307-327.

Bollerslev, T., Wooldridge J.M., 1992. Quasi-Maximum likelihood estimation and inference in dynamic models with time-varying covariance. Econometric Reviews 11, 143-172.

Broyden, C.G., 1970. The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics & Its Applications, 6, 76-90.

Buhlmann, P., Kunsch, H.R., 1999. Block length selection in the bootstrap for time series. Computational Statistics & Data Analysis, 31, 295-310.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag.

Casella, G., Berger, R.L., 2002. Statistical Inference. Thonson Learning, Buxbury.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for sthocastic volatility models. Journal of Econometrics, 108, 281-316.

Chiogna, M., Gaetan, C., 2002. Dynamic generalized linear models with application to environmental epidemiology. Applied Statistics, 51, 453-468.

Chover, J., Ney, P., Wainger, S., 1972. Functions of probability measures. Journal of Analysis Mathematical, 26, 255-302.

Chystiakov, V.P., 1964. A theorem on sums of independent positive random variables and its aplications to banching random processes. Theory of Probability Applied, 9, 640-448.

Commandeur, J.J.F., Koopman, S.J., 2007. An Introduction to State Space Time Series Analysis. Oxford University Press, Oxford.

Consul, P.C., Jain, G.C., 1971. On the log-gamma distribution and its properties. Statistical Papers, 12(2), 100-106.

Cordeiro, G.M., McCullagh, P., 1995. Bias Correction in Generalized Linear Models. Journal of the Royal Statistical Society , 53(3), 629-643.

Cox, D.R., 1981. Statistical analysis of time series: some recent developments. Scandanavian Journal of Statistics, 8, 93-115.

Davis, W.W., 1977. Robust interval estimation of the innovation variance of an ARMA model. The Annals of Statistics, 5(4), 700-708.

Davison, A.C, Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge University Press.

Deschamps, P.K., 2011. Bayesian estimation of an extended local scale stochastic volatility model. Journal of Econometrics, 162, 369-382.

Durbin, J., Koopman, S.J., 2000. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). Journal of the Royal Statistical Society, series B, 62, 3-56.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1), 1-26.

Efron, B., Tibshirani, R.J., 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy (with discussion). Statistical Science, 1(1), 54-77.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Chapman & Hall, New York.

Embrechts, P., Godie, C.M., 1980. On clousure and factorization theorems for subexponential and related distributions. Journal of Austral Mathematical Society, series A, 243-256.

Embrechts, P., Klüppelberg, C., Milosch, T., 1997. Modelling Extremal Events. Springer, New York.

Embrechts, P., Omey, E., 1984. A property of longtailed distributions. Journal of Applied Probability, 21, 80-87.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica, 50, 987-1007.

Eraker, B., Johanners, M., Polson, N.G., 2003. The impact of jumps in returns and volatility. Journal of Finance, 53, 1269-1330.

Fahrmeir, L., 1987. Regression models for nonstationary categorical time series. Journal of Time Series Analysis, 8, 147-160.

Ferrante, M., Vidoni, P., 1998. Finite dimensional filters for nonlinear stochastic difference equations with multiplicative noises. Stochastic Processes and Their Applications, 77, 69-81.

Firth, D., 1993. Bias Reduction of Maximum Likelihood Estimates. Biometrika, 80(1), 27-38.

Fletcher, R., 1970. A new approach to variable metric algorithms. The Computer Journal, 13(3), 317-322.

Franco, G.C., Gamerman, D., Santos, T.R., 2009. Modelos de espaço de estados: abordagens clássica e bayesiana. 13^a Escola de Séries Temporais e Econometria, São Carlos.

Franco, G.C., Reisen, V.A., 2004. Bootstrap Techniques in Semiparametric Estimation Methods for ARFIMA Models: A Comparison Study. Computational Statistics, 19, 243-259.

Franco, G.C., Souza, R.C., 2002. A Comparison of Methods for Bootstrapping in the Local Level Model. Journal of Forecasting, 21, 27-38.

Franco, G.C., Santos, T.R., Ribeiro, J.A., Cruz, F.R., 2008. Confidence intervals for hyperparameters in structural models. Communications in Statistcs: Simulation and Computation, 37 (3), 486-497.

Freedman, D.A., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models. The Annals of Statistics, 12(3), 827-842.

Fruhwirth-Schnatter, S., 1994. Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. Statistics and Computing, 4, 259-269.

Gamerman, D., West, M., 1987. An application of dynamic survival models in unemployment studies. The Statistician, 36, 269-274.

Gamerman, D., 1991. Dynamic Bayesian models for survival data. Applied Statistics, 40, 63-79.

Gamerman, D., 1998. Markov chain Monte Carlo for dynamic generalized linear models. Biometrika, 85, 215-227.

Goldfard, D., 1970. A family of variable metric updates derived by variational means. Mathematics of Computation, 24(109), 23-26.

Goldie, C.M., Klüppelberg, C., 1998. Subexponential Distributions. A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhauser Boston, Cambridge, 435-459.

Godolphin, E.J., Triantafyllopoulos, K., 2006. Decomposition of time series models in state-space form. Computational Statistics and Data Analysis, 50, 2232-2246.

Goldfard, D., 1970. A family of variable metric updates derived by variational means. Mathematics of Computation, 24(109), 23-26.

Green, R.F., 1974. A note on outlier-prone families of distributions. The annals of Statistics, 2(6), 1293-1295.

Green, R.F., 1976. Outlier-prone and outlier-resistant distributions. Journal of the American Statistical Association, 71(354), 502-505.

Grunwald, G.K., Raftery, A.E., Guttorp, P., 1993. Time series of Continuos proportions. Journal of the Royal Statistical Society , series B, 55(1), 103-116.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli, 7(2), 223-242.

Hahn, J., Newey, W.A., 2004. Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. Econometrica, 72(4), 1295-1319.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Harvey, A.C., Fernandes, C., 1989. Time series models for count or qualitative observations. Journal of Business & Economic Statistics, 7(4), 407-417.

Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies, 61, 247-264.

Hemming, K., Shaw, J.E.H., 2002. A parametric dynamic survival model applied to breast cancer survival times. Applied Statistics, 51, 421-435.

Heinze, G., Schemper, M., 2001. A Solution to the Problem of Monotone Likelihood in Cox Regression. Biometrics, 57, 114-119.

Holt, C.C., 1957. Forecasting Seasoals and Trends by Exponentially Weighted Moving Averages. Office of Naval (ONR 52), Carnegie Institute of Technology, Pittsburgh.

Hurvich, C.M., Tsai, C.L., 1993. A corrected Akaike information criterion for vector autoregressive model selection. Journal of Time Series Analysis, 14, 271-279.

Jacquier, E., Polson, N.G., Rossi, P., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal of Businees & Economic Statistics, 12, 371-417.

Jorgensen, B., Lundbye-Christensen, S., Song, P.X.K., Sun, L., 1999. A state space models for multivariate longitudinal count data. Biometrika, 86, 169-181.

Junior, J.D.O.S., 2007. Considerações sobre a relação entre distribuições de cauda pesada e conflitos de informação em inferência bayesiana. Dissertação de mestrado em Estatistica, UNICAMP.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME. Journal of Basic Engineering, 82, 35-45.

Kalman, R.E., Bucy, R.S., 1961. New results in filtering and prediction theory. Transactions of the ASME. Journal of Basic Engineering, 83, 95-108.

Kaufmann, H., 1987. Regression models for nonstationary categorical time series: asymptotic estimation theory. Annals of Statistics, 15, 79-98.

Kim, J.H., 2002. Bootstrap Prediction Intervals for Autoregressive Models of Unknown or Infinite Lag Order. Journal of Forecasting, 21, 265-280.

Kitagawa, G., 1987. Non-Gaussian state-space modelling of nonstationary time series. Journal of the American Statistical Association, 82, 1032-1063.

Klüppelberg, C., 1988. Subexponential distributions and integrated tails. Journal of Applied Probability, 25, 132-141.

Kunsch, H.R., 1989. The Jackknife and the bootstrap for general stationary observations. The Annals of Statistics, 17(3), 1217-1241.

Lawrence, C.T., Tits, A.L., 2001. A computationally efficient feasible sequential quadratic programming algorithm. SIAM Journal of Optimization, 11(4), 1092-1118.

Lindsey, J.K., Lambert, P., 1995. Dynamic generalized linear models and repeated measurements. Journal of Statistical Planning and Inference, 47, 129-139.

Loughin, T.M., 1998. On the Bootstrap and Monotone Likelihood in the Cox Proportional Hazards Regression Model. Lifetime Data Analysis, 4, 393-403.

McCulagh, P., Nelder. J.A., 1989. Generalized Linear Models. Chapman and Hall, London.

McCullough, B.D., 1994. Bootstrapping forecast intervals: An application to $AR(p)$ models. Journal of Forecasting, 13(1), 51-66.

Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. Journal of Econometrics, 45, 239-265.

Nelder. J.A., Wedderburn, R.W.M., 1972. Generalized linear models. Journal of the Royal Statistical Society, series A, 135, 370-384.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347-370.

Neyman, J., Scott, E.T., 1971. Outliers proneness of phenomena and related distributions, optimizing methods in statistics. Academic Press, New York, 413-430.

Nocedal, J., Wright, J.W., 1999. Numerical Optimization. Springer Verlag, New York.

Pascual, L., Romo, J., Ruiz, E., 2000. Bootstrap predictive inference for ARIMA processes. Journal of Time Series Analysis, 25(4), 449-465.

Pinho, F.M., Franco, G.C., Silva, R.S., 2012. Modelling Volatility Using State Space Models with Heavy Tailed Distributions. Working paper.

Pinho, F.M., Franco, G.C., 2012. Penalized Likelihood for a Non Gaussian State Space Model Considering Heavy Tailed Distributions. Working paper.

Politis, D.M., Romano, J.P., 1994. The Stationary Bootstrap. Journal of the American Statistical Association, 89(428), 1303-1313.

Raggi, D., Bordignon, S., 2006. Comparing stochastic volatility models through Monte Carlo simulations. Computational Statistics and Data Analysis, 50, 1678-1699.

Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. Journal of Computational & Graphical Statistics, 18(2), 349-367.

Rodriguez, A., Ruiz, E., 2009. Bootstrap prediction intervals in state space models. Journal of Time Series Analysis, 30(2), 167-178.

Santana, F.T., 2008. Distribuições Subexponenciais. VIII ERMAC - Encontro Regional de Matemática Aplicada e Computacional - Universidade Federal do Rio Grande do Norte.

Santos, T.R., 2009. Inferência sobre os hiperparâmetros dos modelos estruturais sob a perspectiva clássica e bayesiana. Dissertação de mestrado em Estatística - UFMG.

Santos, T.R., Franco, G.C., Gamerman, D., 2010. Gamma family of dynamic models. Technical Report, 234, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro.

Schwarz, G.E., 1978. Estimating the dimension of a model. Annals of Statistics, 6(2), 461-464.

Shanno, D.F., 1970. Conditioning of quase-Newton methods for function minimization. Mathematics of Computation, 24(111), 647-656.

Shephard, N., 1994. Local scale model: state space alternative to integrated GARCH processes. Journal of Econometrics, 60, 181-202.

Shephard, N., Pitt, M.K., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika, 84, 653-667.

Shiryaev, A.N., 1989. Probability. Springer, New York.

Smith, J.Q., 1979. A Generalization of the Bayesian Steady Forecasting Model. Journal of the Royal Statistical Society, series B, 41, 375-387.

Smith, J.Q., 1981. The Multiparameter Steady Model. Journal of the Royal Statistical Society, series B, 43, 256-260.

Smith, R.L., Miller, J.E., 1986. A non-Gaussian state space model and application to prediction of records. Journal of the Royal Statistical Society, Series B, 48(1), 79-88.

Souza, R.C., Neto, A.C., 1996. A Bootstrap Simulation Study in $ARMA(p, q)$ Structures. Journal of Forecasting, 15(4), 343-353.

Stoffer, D.S., Wall, K.D., 1991. Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter. Journal of the American Statistical Association, 86(416), 1024-1033.

Stoffer, D.S., Wall, K.D., 2002. A state space approach to bootstrapping conditional forecasts in ARMA models. Journal of Time Series Analysis, 23(6), 733-751.

Stoffer, D.S., Wall, K.D., 2004. Resampling in State Space Models. Chapter 9 of State Space and Unobserved Component Models: Theory and Applications. A. Harvey, S.J. Koopman and N. Shephard (edictors). Cambridge University Press.

Sugiura, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Communication in Statistics, A7, 13-26.

Taylor, S.J., 1986. Modeling Financial Time Series. John Wiley & Sons.

Taylor, S.J., 1994. Modeling stochastic volatility: A review and comparative study. Mathematical Finance, 4, 183-204.

Thombs, L.A., Schucany, W.R., 1990. Bootstrap Prediction Intervals for Autoregression. Journal of the American Statistical Association, 85(410), 486-492.

Teugels, J.L., 1975. The class of subexponential distributions. The Annals of Probability, 3(6), 1000-1011.

Tsay, R.S., 2005. Analysis of Financial Time Series. John Wiley & Sons, New Jersey.

Tukey, J., 1958. Bias and confidence in not quite large samples (abstract). The Annals of Mathematical Statistics, 29(2), 614.

Vidoni, P., 1999. Exponential family state space models based on conjugate latent process. Journal of Royal Statistical Society B., 61, 213-221.

Winters, P.R., 1960. Forecasting sales by exponentially weighted moving averages. Management Science, 6, 324-342.

Yakymiv, A.L., 1997. Some properties of subexponential distributions. Mathematical Notes, 62(1), 116-121.

West, M., Harrison, J., 1997. Bayesian forecasting and dynamic models. Springer, New York.

West, M., Harrison, P.J., Migon, H.S., 1985. Dynamic Generalized Linear Models and Bayesian Forecasting (with discussion). Journal of the American Statistical Association, 81, 741-750.

Zakoian, J.M., 1994. Threshold heteroscedastic models. Journal of Economic Dynamics & Control, 18, 931-955.