# Modeling Synthetic Aperture Radar Image Data

by

Donald Matthew Pianto

A thesis presented to the Universidade Federal de Pernambuco
in partial fulfillment of the requirements for the degree of
Doutor em MATEMÁTICA COMPUTACIONAL

Recife, February, 2008

**Pianto, Donald Matthew**
**Modeling synthetic aperture radar image data / Donald Matthew Pianto. – Recife : O Autor, 2008.**
**xviii, 108 p. : il., fig., tab.**

**Tese (doutorado) - Universidade Federal de Pernambuco. CCEN. Matemática Computational, 2008.**

**Inclui bibliografia e apêndices.**

**1. Processamento de imagens. 2. Distribuição GA0 . 3. Verossimilhança monótona. 4. Bootstrap. 5. Radar de abertura sintética. I. Título.**

**621.36 CDD (22.ed.) MEI2008-11**

UFPE 60 ANOS

*UNIVERSIDADE FEDERAL DE PERNAMBUCO*
Centro de Ciências Exatas e da Natureza
**Doutorado em Matemática Computacional**

**PARECER DA BANCA EXAMINADORA DE DEFESA DE TESE DE DOUTORADO**

# *DONALD MATTHEW PIANTO*

## *"MODELING SYNTHETIC APERTURE RADAR IMAGE DATA"*

A Banca Examinadora composta pelos Professores Francisco Cribari Neto (Presidente), Silvio de Barros Melo, ambos da Universidade Federal de Pernambuco, Alejandro César Frery Orgambide, da Universidade Federal de Alagoas, Borko Darko Stosic, da Universidade Federal Rural de Pernambuco, Maria da Conceição Sampaio de Sousa, da Universidade de Brasília, considera o candidato:

(✓) APROVADO COM DISTINÇÃO ( ) APROVADO ( ) REPROVADO

Secretaria do Programa de Doutorado em Matemática Computacional do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, aos 14 dias do mês de fevereiro de 2008.

PROF. FRANCISCO CRIBARI NETO
PRESIDENTE

PROF. SILVIO DE BARROS MELO
1° EXAMINADOR

PROF. ALEJANDRO CÉSAR FRERY ORGAMBIDE
2ª EXAMINADOR

PROF. BORKO DARKO STOSIC
3° EXAMINADOR

PROFª MARIA DA CONCEIÇÃO SAMPAIO DE SOUSA
4° EXAMINADOR

Av. Prof. Luis Barros Freire, s/nº, Cidade Universitária - CEP: 50670-901 - Recife - PE
FONE/FAX: (81) 2126-8412 - E-mail: ppgmc@ufpe.br - Home Page: www.ppgmc.ufpe.br

*to Maria Eduarda, Alma and Don*

# Acknowledgements

More than just dedicating this thesis to my wife and parents, I wish to thank them explicitly for all their love and support throughout my lifetime, not only during the period when this thesis was written. I'm so lucky to have you all.

I'd also like to thank Suely and José Galib for their confidence and support through the years. I don't believe that better in-laws exist.

Thanks to my adviser, Francisco Cribari-Neto, for pointing me in the right direction and not letting me stray when the results didn't come very quickly. I know that it must have been stressful to have me as an advisee, so thank you for your patience. Thank you also for your close reading of earlier versions of this text, the other members of my committee will never know how much easier you made it to read my thesis.

Thanks to Alejandro Frery for sharing data with me and for also sharing so much of his time, experience, and intuition. All the parts of this thesis are richer because of his participation, but the section on the analysis of real data would have been impossible without him. I must admit that I was somewhat disappointed by the less than perfect results for my bootstrap estimator in the Monte Carlo simulations, but Alejandro encouraged me by telling me that a decision about the usefulness of the estimators could only be made after examining real data. Thank you.

Thanks to Maria da Conceição Sampaio de Sousa for suggesting the Doctoral program in Computational Mathematics to me when I was deciding what to study and for her support and encouragement during the whole period.

Thanks to Sílvio Melo for his careful reading of my project defense which caught an error which had gotten past both me and Lauro Lins.

Thanks to Borko Darko Stošić for his participation in this committee and his suggestions on how to get back to work when I was stalled.

Thanks to Lauro Didier Lins for his participation in some of the early work presented in this thesis as well as the template for this thesis format. I really enjoyed working and studying with you.

Thanks to all the professors and students of the Doctorate in Computational Mathematics for the friendships and intellectual stimulus they provided. A special thanks to Klaus Vasconcellos, Andrei Toom, Jalila, Líliam, Alex, Calitéia, Alexandre, Mirele, Graça, Glória, Artur, Themis and all the rest.

Thanks to Valéria Bittencourt for all her help throughout the program and for making everything she is involved in run smoothly.

Thanks to Angela Farias for her help in getting this thesis submitted.

Thanks to my friends in Recife who made living here so much fun: Tiago, Juliana, André, Roberta, Isabela and Bruno César.

Thanks to Leontina for bringing her uplifting spirit to our home three times a week.

Finally my thanks to CAPES for financial support during my time in this program.

# Resumo

Nessa tese estudamos a estimação por máxima verossimilhança (MV) do parâmetro de aspereza da distribuição $\mathscr{G}_A^0$ de imagens com *speckle* (Frery *et al.*, 1997). Descobrimos que, satisfeita uma certa condição dos momentos amostrais, a função de verossimilhança é monótona e as estimativas MV são infinitas, implicando uma região plana. Implementamos quatro estimadores de correção de viés em uma tentativa de obter estimativas MV finitas. Três dos estimadores são obtidos da literatura sobre verossimilhança monótona (Firth, 1993; Jeffreys, 1946) e um, baseado em reamostragem, é proposto pelo autor. Fazemos experimentos numéricos de Monte Carlo para comparar os quatro estimadores e encontramos que não existe um favorito claro, a menos quando um parâmetro (dado a priori da estimação) toma um valor específico. Também aplicamos os estimadores a dados reais de radar de abertura sintética. O resultado desta análise mostra que os estimadores precisam ser comparados com base em suas habilidades de classificar regiões corretamente como ásperas, planas, ou intermediárias e não pelos seus vieses e erros quadráticos médios.

**Palavras-chave**: *speckle*, distribuição $\mathscr{G}_A^0$, radar de abertura sintética (SAR), imagens coerentes, máxima verossimilhança, *bootstrap*, reamostragem, verossimilhança monótona.

# Abstract

In this thesis we study maximum likelihood estimation (MLE) of the roughness parameter of the $\mathscr{G}_A^0$ distribution for speckled imagery (Frery *et al.*, 1997). We discover that when a certain criteria is satisfied by the sample moments, the likelihood function is monotone and MLE estimates are infinite, implying an extremely homogeneous region. We implement four bias correcting estimators in an attempt to obtain finite MLE estimates. Three of the estimators are taken from the literature on monotone likelihood (Firth, 1993; Jeffreys, 1946) and one, based on resampling, is proposed by the author. We perform Monte Carlo experiments to compare the four estimators and find that there is no clear favorite, except when one of the parameters (which is given before estimation) takes on a specific value. We also apply the estimators to real data obtained from synthetic aperture radar (SAR). It becomes clear from this analysis that the estimators need to be compared based on their ability to classify regions correctly as rough, smooth, or intermediate and not on their biases and mean squared errors.

**Keywords**: speckle, $\mathscr{G}_A^0$ distribution, synthetic aperture radar (SAR), coherent imaging, maximum likelihood, bootstrap, resampling, monotone likelihood.

# Contents

# List of Figures

# List of Tables

CHAPTER 1

# Introduction

## 1.1 What is speckle and why study it?

In non-coherent imaging, one observes energy reflected from targets. This energy may come in the form of, for example, incoherent visible light waves (photography). Images obtained with coherent illumination (sonar, laser imaging, ultrasound-B, and synthetic aperture radar—SAR) provide higher resolution of the target through the use of the phase information of the reflected signal, but are also subject to speckle noise. The general statistical model that we will use to study images in this work is the multiplicative model (Frery *et al.*, 1997). In the multiplicative model, the amplitude of the reflected signal in any pixel (picture element corresponding to a given region on the ground) results from the multiplication of the terrain backscatter and speckle noise, which are considered independent.

Dark areas in an image represent low backscatter and bright areas high backscatter. The terrain backscatter in any pixel depends on the roughness of the target, the angle of the incoming signal and the water content of the target, among other factors. Figure 1.1 provides some examples of different backscatter intensities. A smooth region, such as a lake or crop fields, will most likely have low backscatter, since the reflected radiation continues away from the sensor. This is not the case if the sensor is perpendicular to the smooth surface as in the mountain example. A forest will have intermediate backscatter since its moisture aids signal reflection and its rough surface will reflect light in many directions. A city will likely have high backscatter because of its significant heterogeneity and strong signal reflection.

Even when the sensor is pointed at a particular point on the ground, it may be receiving signals that were originally directed at other points. Since the original radiation was coherent, these signals may interfere to form a bright spot in a region with low terrain backscatter, or a

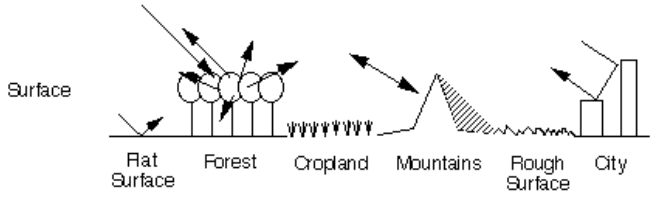


**Figure 1.1** Representation of the backscatter from targets of varying roughness. Figure taken from the European Space Agency (2008).

dark spot on a region of high terrain backscatter. When this happens, we say that the image is contaminated with speckle noise.

In Figure 1.2 laser light has been reflected off a wall and photographed by a digital camera. The "salt and pepper" contamination is called speckle and comes from the interference phenomenon discussed above. Astronomers have exploited their understanding of speckle to enhance the resolution of their telescopes (Fried, 1966; Labeyrie, 1970). In this work, we study the properties of a statistical model of speckled imagery (the $\mathcal{G}_A^0$ model) which has been used to classify targets as heterogeneous, intermediate, or homogeneous.

The images we will study come from synthetic aperture radar (SAR). They can be taken from satellites or planes. The aperture is synthetic because a relatively small antenna (10 meters, perhaps) can image the same area various times as it moves overhead, simulating a much larger antenna (see Figure 1.3). The large effective antenna size provides higher resolution than that available from the original antenna. Uses of SAR include (European Space Agency, 2008): sea ice monitoring; cartography; surface deformation detection; glacier monitoring; crop production forecasting; forest cover mapping; ocean wave spectra; urban planning; coastal surveillance (erosion); monitoring disasters such as forest fires, floods, volcanic eruptions, and oil spills.

Imagine a pilot with synthetic aperture radar (SAR) in his plane flying over an unknown area either covered by clouds or at night. If the pilot has engine or fuel trouble and needs to land his plane, it would be of utmost importance to land in the smoothest (most homogeneous) area possible. Visual examination of Figure 1.4 would not necessarily yield an answer of where to land. The models and estimators discussed in this thesis can be used to give a systematic

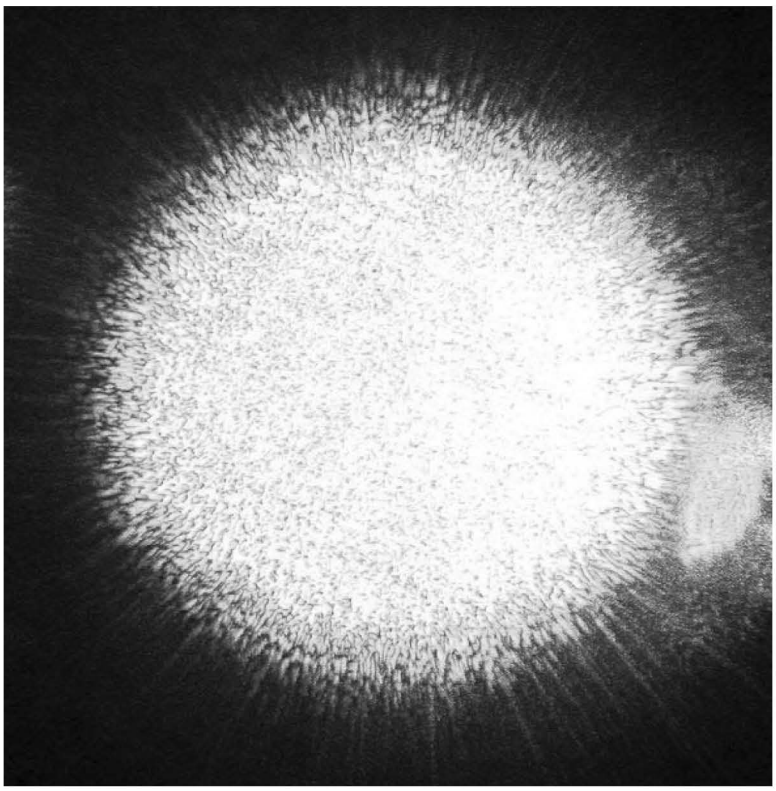

**Figure 1.2** Photograph of laser speckle taken with a digital camera.

answer to this distressed pilot.

## 1.2 What has been done and what we did

### 1.2.1 What has been done

The $\mathscr{G}_A^0$ model, which is the focus of this work, was introduced in Frery *et al.* (1997). The authors assume the multiplicative model and discuss the models for the speckle and amplitude backscatter used in the literature. They introduce the three parameter square root of the generalized inverse Gaussian distribution as a statistical model for the terrain backscatter. They show that for extremely heterogeneous targets, the terrain backscatter follows the two parameter square root of an inverse gamma distribution. The amplitude distribution of the total return is then shown follow the $\mathscr{G}_A^0$ law.

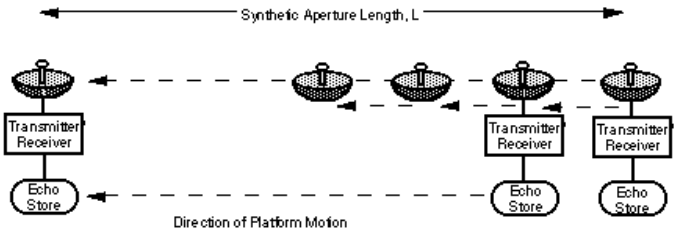


**Figure 1.3** Schematic of how synthetic aperture radar (SAR) functions. Figure taken from the European Space Agency (2008).

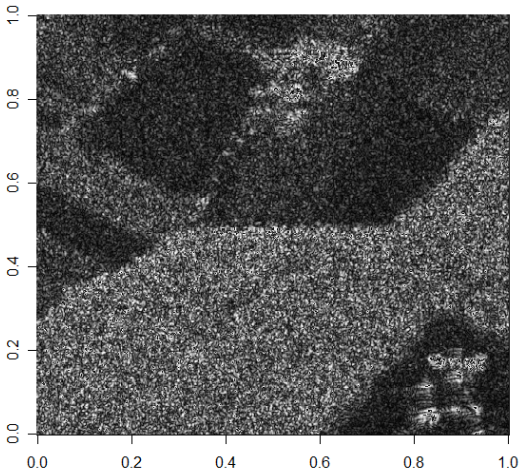


**Figure 1.4** Synthetic aperture radar (SAR) image. The amplitude data goes from black (0 amplitude) to white (maximum amplitude).

Since its inception, attempts to estimate the parameters of the $\mathcal{G}_A^0$ distribution using maximum likelihood have suffered from convergence problems and large biases in small samples. Cribari-Neto *et al.* (2002), Frery *et al.* (2004), Vasconcellos *et al.* (2005) and Silva *et al.* (2007) among others, tried to resolve the problem through bias correction, improved maximization algorithms and adjusted profile likelihood methods.

In Cribari-Neto *et al.* (2002) the authors attempt to correct for the finite sample bias of the maximum likelihood estimator of the $\mathcal{G}_A^0$ parameters by applying an adaptation of the better bootstrap bias correction of Efron (1990). Whereas Efron requires a closed form solution for the parameter estimates, Cribari-Neto *et al.* apply the correction to the estimating equations

which are then solved numerically. The bias of the resulting estimates is quite small when compared to simple maximum likelihood estimation. However, the authors exclude samples for which the numerical solution does not converge from their analysis.

In their analysis, Frery *et al.* (2004) include samples where the numerical solution does not converge. They suggest the use of a maximization algorithm which alternates maximization over one parameter with maximization over the other parameter of the $\mathcal{G}_A^0$ distribution. They discover that this new algorithm always converges while maximizing the $\mathcal{G}_A^0$ likelihood.

Vasconcellos *et al.* (2005) propose an analytic bias correction for the maximum likelihood estimation of the parameters of the $\mathcal{G}_I^0$ distribution. This law describes the distribution of radiation intensity rather than amplitude, but has the same parameters as the $\mathcal{G}_A^0$ distribution. They calculate the second order bias correction and implement it to obtain corrected estimators. The corrected estimators have better bias and mean squared error properties than the uncorrected maximum likelihood estimators. The correction requires that the uncorrected solution be finite, because the correction is a function of the original solution.

In Silva *et al.* (2007), the authors maximize adjusted profile likelihoods to obtain improved estimates of the $\mathcal{G}_A^0$ distribution's roughness parameter. They implement corrections proposed by Cox & Reid (1987; 1989), Barndorff-Nielsen (1983), Fraser & Reid (1995) and Fraser *et al.* (1999). For point estimation, the Cox & Reid adjustment is preferred whereas, for signalized likelihood ratio tests, the Barndorff-Nielsen adjustment based on the results of Fraser & Reid and Fraser *et al.* is preferred. The adjustment was only applied to samples for which unadjusted maximum likelihood converged, since that solution was taken as a starting point for maximization of the adjusted likelihood.

### 1.2.2 What we did

In this thesis we closely study the algorithm suggested in Frery *et al.* (2004) and discover that, although convergence is obtained, the algorithm returns a value of the log-likelihood less than that returned by other algorithms which do not converge. These larger values of the log-likelihood are generally obtained for large (in magnitude) values of the parameters.

This discovery leads us to study the log-likelihood for large absolute values of the parameters of the $\mathcal{G}_A^0$ distribution. In doing so we find so called "monotone likelihood" (Bryson & Johnson, 1981; Loughin, 1998; Heinze & Schemper, 2001; Sartori, 2006). This is a situation where the log-likelihood achieves its maximum for infinite parameter values. One solution to this problem encountered in the literature is that of Firth (1993), which leads to alterations of the log-likelihood. We also discover a simple quantity (a ratio of sample moments) which determines if a given sample suffers from monotone likelihood or not.

Rather than only implementing the solutions available in the literature, we also proposed a new bootstrap estimator based on the bias corrections in Cribari-Neto *et al.* (2002). We find conditions which are necessary and sufficient for its existence in finite samples and prove its consistency. In fact, the estimator converges to the maximum likelihood estimator (MLE) as the sample size goes to infinity. We finally implement all the solutions discussed and perform an extensive Monte Carlo analysis and analyze real data with the new estimators. The proposed bootstrap estimator and the estimator proposed by Firth have large variances in the Monte Carlo experiment. However, when analyzing real data the important question is whether the region is heterogeneous, homogeneous or intermediate. Hence, a large negative estimate of the roughness parameter (which would cause a large mean squared error) simply implies a smooth region in real data. Future work should evaluate the estimators based on classification rather than conventional bias and mean squared error statistics (Mejail *et al.*, 2003).

## 1.3 Organization of the thesis

In Chapter 2 we discuss the $\mathcal{G}_A^0$ model of speckled imagery that we study in this work. Then we discuss maximum likelihood estimation of the parameters of the $\mathcal{G}_A^0$ model, with an emphasis on the algorithms used to numerically maximize the likelihood. In Section 2.3 we expand the log-likelihood for large values of the parameters and discover "monotone likelihood". In Section 2.4 we discuss a solution to monotone likelihood which was suggested by Firth (1993). Section 2.5 suggests a bootstrap based estimator to handle the monotone likelihood problem.

Chapter 3 presents the results of our Monte Carlo analysis of the various estimators presented in the previous chapter. Chapter 4 applies all the estimators to real data in several practical settings. Chapter 5 concludes with suggestions for future work.

CHAPTER 2

# The $\mathscr{G}_A^0$ Distribution and Likelihood Inference

## 2.1 The $\mathscr{G}_A^0$ model

This section follows the discussion in Frery *et al.* (2004). To deal with inference in speckled imagery, the first step is to develop a statistical model for the speckle noise. This noise can be assumed neither Gaussian nor additive. The $\mathscr{G}$ family of distributions, a type of multiplicative model proposed and assessed in Frery *et al.* (1997) and Mejail *et al.* (2001), is regarded as the universal model for speckled data. This model describes the statistical behavior of the amplitude or intensity of reflected light contaminated with speckle noise. When studying amplitude data for regions with heterogeneous target classes, $\mathscr{G}_A^0$ is the $\mathscr{G}$ family member of interest.

The $\mathscr{G}_A^0$ distribution has three parameters $(\alpha, \gamma, n)$. Its density is given by

$$f_Z(z) = \frac{2n^n\Gamma(n-\alpha)}{\gamma^\alpha\Gamma(n)\Gamma(-\alpha)} \frac{z^{2n-1}}{(\gamma+nz^2)^{n-\alpha}} \mathbb{1}_{\mathbb{R}_+}(z), \tag{2.1}$$

where $\mathbb{1}_{\mathbb{R}_+}(z)$ is the indicator function which equals one if $z \in \mathbb{R}_+$ and zero otherwise. The parameter space is given by $\alpha < 0$, $\gamma > 0$ and $n \geq 1$. When the first parameter, $\alpha$ (the roughness parameter), is close to zero (around $-1.1$, for example), the data correspond to highly heterogeneous areas (such as urban areas). In this case, the distribution has thick tails as can be seen in Figure 2.1. Values of $\alpha$ around $-5$ correspond to intermediate areas such as forests. The tails are thinner for this distribution. Values of $\alpha$ less than about $-15$ correspond to extremely homogeneous areas, such as crop fields. The density in this case is quite compressed with little weight in the tails. Even so, comparison with a Normal$(1,1)$ shows that the tails decay linearly whereas the Normal decays quadratically.

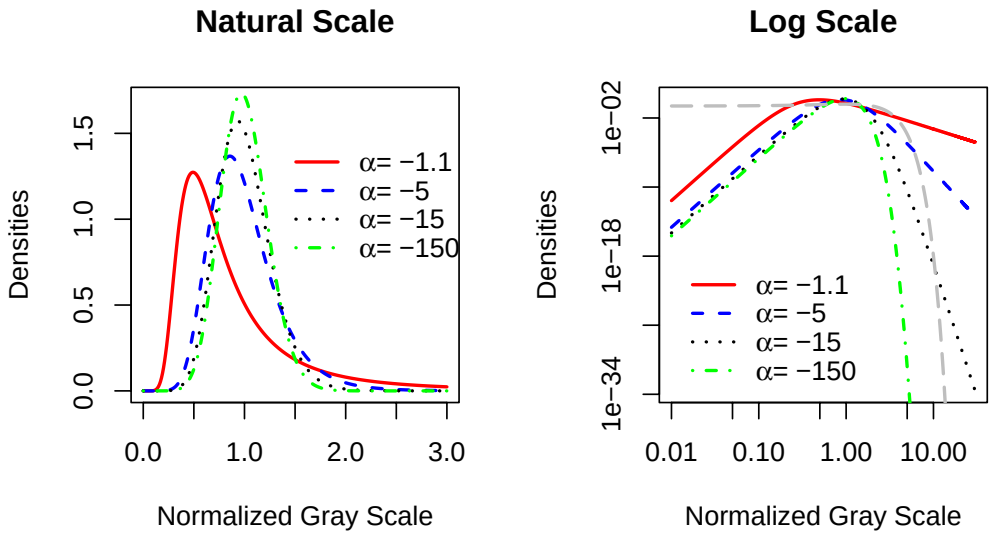


**Figure 2.1** Regular and log densities of the $\mathcal{G}_A^0(\alpha,\gamma,n=5)$ distribution for $\alpha \in \{-1.1,-5,-15,-150\}$ and $\gamma$ as in equation (2.7). (solid red, dashed blue, dotted black, and dash-dot green lines, respectively). Note that the expected value for all the above densities has been normalized to 1. The gray line in the log scale graph corresponds to a Normal$(1,1)$.

The parameter $\gamma$ is a scale factor which we will use for normalization. If $Z' \sim \mathcal{G}_A^0(\alpha,1,n)$ then $\sqrt{\gamma}Z' \sim \mathcal{G}_A^0(\alpha,\gamma,n)$. The parameter $n$ corresponds to the number of "looks" which formed the image.

The moments of $Z \sim \mathcal{G}_A^0(\alpha,\gamma,n)$ are given by

$$\mathbb{E}[Z^r] = \left(\frac{\gamma}{n}\right)^{r/2} \frac{\Gamma(-\alpha-r/2)\Gamma(n+r/2)}{\Gamma(-\alpha)\Gamma(n)} \tag{2.2}$$

with $n \geq 1$ and $r < -2\alpha$. When $r \geq -2\alpha > 0$ the $r$-th order moment is infinite.

The $\mathcal{G}_A^0$ cumulative distribution function is

$$F_Z(z) = \Upsilon_{2n,-2\alpha}\left(\frac{-\alpha z^2}{\gamma}\right),$$

where $\Upsilon_{2n,-2\alpha}$ is the cumulative distribution function of Snedecor's $\mathcal{F}$ law with $2n$ and $-2\alpha$ degrees of freedom. There are other forms of $F_Z$ but this one is useful because the functions $\Upsilon_{\cdot,\cdot}$ and $\Upsilon_{\cdot,\cdot}^{-1}$ are available in most statistical software packages. For instance, to generate samples

of a $\mathscr{G}_A^0(\alpha,\gamma,n)$ we simply use the inverse method

$$Z = (-\gamma \Upsilon^{-1}_{2n,-2\alpha}(U)/\alpha)^{1/2},$$

where $U$ is uniformly distributed on $(0,1)$.

We also generated the $\mathscr{G}_A^0$ realizations using the ratio of two gamma distributed random variables: $Z = \sqrt{XY}$, where $Y \sim \Gamma(n,n)$ and $X^{-1} \sim \Gamma(-\alpha,\gamma)$. The cumulative distribution functions of the outcomes generated using the $\Gamma$ law and the $\mathscr{F}$ law were then compared. They showed no systematic difference. Any differences in the tails got smaller as more and more realizations were used. Hence we used only the $\mathscr{F}$ law to generate our $\mathscr{G}_A^0$ variates.

Based on Equation (2.1) one can easily derive the following reduced log-likelihood function for an independent and identically distributed (iid) sample of size $N$:

$$\ell(\alpha,\gamma;z_1,\ldots,z_N,n) = N\ln\left(\frac{\Gamma(n-\alpha)}{\gamma^\alpha\Gamma(-\alpha)}\right) - (n-\alpha)\sum_{i=1}^{N}\ln(\gamma+nz_i^2). \tag{2.3}$$

Note that in $\ell$ we take the parameter $n$ as given, since this is generally the case in practical situations where $\mathscr{G}_A^0$ parameters are estimated.

Maximum likelihood estimates (MLEs) of the parameters are obtained by maximizing $\ell$. This can be done by setting the first derivative of $\ell$ with respect to each parameter to zero, yielding the following equations:

$$\begin{array}{rcl} 0 = \frac{\partial\ell}{\partial\alpha} & = & N\left(\psi(-\alpha)-\psi(n-\alpha)\right)+\sum_{i=1}^{N}\log(1+nz_i^2/\gamma) \\ 0 = \frac{\partial\ell}{\partial\gamma} & = & N\frac{(-\alpha)}{\gamma}-\frac{(n-\alpha)}{\gamma}\sum_{i=1}^{N}(1+nz_i^2/\gamma)^{-1}, \end{array} \tag{2.4}$$

where $\psi$ represents the digamma function, which is the derivative of the logarithm of the gamma function. Asymptotic inference about the parameters can be performed by considering the Fisher information matrix, $i$, which is equal to the expected value of the matrix of

second derivatives of $\ell$:

$$i(\alpha,\gamma) = \begin{pmatrix} N\left(\psi^{(1)}(-\alpha) - \psi^{(1)}(n-\alpha)\right) & \frac{N}{\gamma}\frac{n}{(n-\alpha)} \\ \frac{N}{\gamma}\frac{n}{(n-\alpha)} & \frac{N}{\gamma^2}\frac{(-\alpha)(n)}{(n-\alpha+1)} \end{pmatrix}, \tag{2.5}$$

where $\psi^{(1)}(\cdot)$ is the derivative of $\psi(\cdot)$. As $N \to \infty$, $i$ becomes the inverse of the asymptotic covariance matrix of the MLE and can be used to calculate confidence intervals for the parameters.

Because Equation (2.4) does not have a closed solution, it is necessary to use non-linear maximization routines to obtain our parameter estimates. These routines are the subject of the next section.

## 2.2 Maximum likelihood estimation of the $\mathcal{G}_A^0$ parameters

In this section we borrow freely from Lins & Pianto (2004). The literature reports that techniques for obtaining sensible estimates for the $\mathcal{G}_A^0$ parameters require large samples, *i.e.* hundreds or even thousands of observations (Mejail *et al.*, 2000). In Frery *et al.* (2004) the authors propose an alternative non-linear optimization method to maximize the $\mathcal{G}_A^0$ log-likelihood function which does a better job than traditional optimization methods (*e.g.* Downhill Simplex, Newton-Raphson, BFGS) when dealing with small samples. We discuss this method and two others in what follows.

### 2.2.1 Non-linear optimization

We wish to find the value of the vector of parameters, $p$, which maximizes a real valued function $f(p)$. In general, if the system of equations given by the first order conditions is non-linear, no closed form solution for the parameters exists. This is the case for the function $\ell$ in Equation (2.3), therefore, one must adopt an iterative strategy to find the solution. Three iterative

algorithms to find such solutions are discussed below.

#### 2.2.1.1 Algorithms

First, some notation. Let $p(t)$, an $n$-vector, be the value of our $n$ parameters at iteration $t$; let $g(t)$ be the gradient of our function at point $p(t)$; let $H(t)$ be the Hessian (matrix of second partial derivatives) at point $p(t)$.

The most commonly used iterative algorithm for non-linear function maximization is the Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) (Press *et al.*, 1992). In the BFGS algorithm a negative definite approximation, $W(t)$, to the Hessian is accumulated throughout the iteration scheme. After each iteration the next direction, $\Delta(t)$, to move the parameter estimate, $p(t)$, is calculated as in the Newton-Raphson method with $W(t)$ in place of $H(t)$: $\Delta(t) = -W^{-1}(t)g(t)$. In Frery *et al.* (2004) the authors report that the BFGS algorithm frequently failed to converge while estimating the parameters ($\alpha$ and $\gamma$) of Equation (2.3). We confirm this result below. The failure of BFGS is attributed to the extremely flat nature of the likelihood function which can cause the Hessian to be nearly singular and therefore void of useful information.

To resolve this problem Frery *et al.* (2004) suggested an iterative algorithm which does not rely on any information about the Hessian matrix, nor does it use the gradient information to determine the next direction in which to change the parameters. This algorithm (FCS) alternates between the maximization of the likelihood function with respect to each of the parameters. For example, in the case of Equation (2.3) one first maximizes the likelihood function over $\alpha$ and then over $\gamma$. The authors report that this algorithm *always* managed to converge, even in the situations where BFGS failed. This result is also confirmed below.

The last algorithm we wish to consider is that of steepest descent (SD). When choosing the next direction over which to maximize the function, SD uses the direction of the gradient, $\Delta(t) = g(t)$. As pointed out in Press *et al.* (1992), this means that the next maximization direction will always be perpendicular to the direction just searched. In the two-dimensional case, the FCS and SD algorithms should be the same except for the initial direction to search

and numerical roundoff. The latter will cause subsequent search directions to not necessarily be perpendicular in the SD algorithm. It is this expected equality between the methods that motivated our analysis of the SD algorithm.

#### 2.2.1.2 Convergence criteria

Up to this point we have only spoken of the next direction to choose in our search for a function's maximum. A critical part of any iterative procedure is the determination of whether or not the algorithm has converged. Let $\delta(t) = p(t) - p(t-1)$; let $\varepsilon_1$ and $\varepsilon_2$ be constants (*e.g.* $\varepsilon_1 = 10^{-4}$ and $\varepsilon_2 = 5 \times 10^{-3}$). Based on these definitions, `Ox` (Doornik, 2001) defines the following convergence criteria:

$$\begin{aligned}
\text{Strong Convergence} &= \begin{cases} 1, & \text{if } |q_i g_i| < \varepsilon_1 \text{ and } \left|\frac{\delta_i}{q_i}\right| < 10\varepsilon_1 \text{ for } i = 1,2,\ldots,n, \\ 0, & \text{otherwise,} \end{cases} \\
\text{Weak Convergence} &= \begin{cases} 1, & \text{if } |q_i g_i| < \varepsilon_2 \text{ for } i = 1,2,\ldots,n, \\ 0, & \text{otherwise,} \end{cases} \\
\text{where } q_i &= \begin{cases} 1, & \text{if } p_i = 0, \\ |p_i| + \frac{1}{\sqrt{n}}, & \text{if } p_i \neq 0\,. \end{cases}
\end{aligned} \tag{2.6}$$

One can see immediately that for $\varepsilon_1 \leq \varepsilon_2$ strong convergence implies weak convergence.

Convergence criteria are not universal. Hence, before comparing the results from any two given algorithms one should compare, and if possible unify, their convergence criteria. For instance, in Frery *et al.* (2004) the stopping criteria for the the FCS alternated optimization algorithm is given by

$$\left|\frac{\delta_1}{p_1}\right| + \left|\frac{\delta_2}{p_2}\right| < \varepsilon,$$

where the maximization is only being performed over two parameters ($p_1 = \alpha$ and $p_2 = \gamma$) and a typical value of $\varepsilon$ is $10^{-4}$. One cannot say if this criteria is more or less stringent than those of `Ox`. However, since `Ox` is known for its numerical soundness we consider only the `Ox`-like stopping criteria in Equation (2.6) for all algorithms studied.

All algorithms were implemented in the `Ox` programming language, version 3.1 (Doornik, 2001). All algorithms were allowed a maximum of 10000 iterations with tests for strong convergence performed after each iteration.[1] If after 10000 iterations strong convergence was not achieved, weak convergence was tested. In order to test the possibility of weak convergence for BFGS it was necessary to adapt the code provided in the `maximize` package. `Ox`'s `MaxNewton` procedure tries to limit the number of (possibly costly) function evaluations by performing an inexact line search during each step. Therefore we did not implement the SD algorithm by simply setting our function's Hessian to $-I$ and calling the `MaxNewton` routine provided in the `maximize` package. Rather we edited the code such that at each step the algorithm would find the "exact" minimum along the search direction. All calculations were performed on a Pentium IV computer running Windows XP.

### 2.2.2 Numerical results

We divided our small Monte Carlo analysis into 80 smaller Monte Carlo experiments, each indexed by a triple $(n,\alpha,N)$, where $n \in \{1,2,3,8\}$, $\alpha \in \{-15,-5,-3,-1\}$ and $N \in \{9,25,49,81,121\}$.[2] Given a triple, $(n,\alpha,N)$, the Monte Carlo experiment was to repeat the following procedure 100 times[3]:

1. Generate $N$ observations $z_1, z_2, \ldots, z_N$ from the distribution $\mathcal{G}_A^0(\alpha,\gamma^*,n)$, where

$$\gamma^* = n\left(\frac{\Gamma(n)\Gamma(-\alpha)}{\Gamma(n+1/2)\Gamma(-\alpha-1/2)}\right)^2 \tag{2.7}$$

   yields unit mean.

2. Maximize $\ell(\alpha,\gamma;z_1,\ldots,z_N,n)$ (Equation (2.3)) using FCS, SD and BFGS. Then, for each of these methods, record: if there was weak convergence or strong convergence; the $\hat{\alpha}$

[1] This is most likely an excessive number of iterations and leads to wasteful use of computing resources. Most likely 1000 or even 100 iterations would be sufficient to identify the problems we are studying.

[2] This was also done in Frery *et al.* (2004).

[3] Whereas the number of Monte Carlo replications is too small to perform any meaningful statistical inference, the results in this section are intended to discover possible systematic problems in $\mathcal{G}_A^0$ maximization. Such problems are in fact discovered and explored more thoroughly in the following sections.

and $\hat{\gamma}$ estimates (point declared as the one that maximizes $\ell$); the "maximum" value $\ell(\hat{\alpha},\hat{\gamma};z_1,\ldots,z_N,n)$; and the time spent on the maximization procedure.

A subset of the results of our experiment can be found in Tables 2.1 and 2.2. In both tables the triple $(n,\alpha,N)$ indexes the experiment and $\gamma^*$ is calculated based on this triple as in Equation (2.7). For Table 2.1, columns *FW*, *SW*, and *BW* show the percentage of the Monte Carlo repetitions for which the FCS, SD and BFGS methods (respectively) converged weakly. Columns *FS*, *SS* and *BS* show the percentage of the Monte Carlo repetitions where strong convergence of the methods FCS, SD and BFGS (respectively) was obtained. Columns *FB*, *SB*, and *BB* show the percentage of the repetitions for which weak or strong convergence was achieved *and* the log-likelihood evaluated at the estimated parameter was greater than or equal to the value of the other two methods for FCS, SD and BFGS, respectively. In Table 2.2, columns $b(\hat{\alpha}_{\text{FCS}})$, $b(\hat{\alpha}_{\text{SD}})$, and $b(\hat{\alpha}_{\text{BFGS}})$ give the estimated bias of the MLE estimate of $\alpha$ using FCS, SD and BFGS, respectively. Columns $\hat{\sigma}_{\text{FCS}}$, $\hat{\sigma}_{\text{SD}}$, and $\hat{\sigma}_{\text{BFGS}}$ give the standard deviation of the $\alpha$ estimates as obtained using FCS, SD and BFGS, respectively. Note that the results for the estimation of $\gamma$ are not reported because in practical situations only the parameter $\alpha$ is of interest.

Analyzing Table 2.1 we can see that the FCS method always converges weakly and that SD had weak convergence in almost all cases. On the other hand, the BFGS method had problems converging for small values of $\alpha$, $N$, and $n$. However, when it did converge it almost always provided the best estimate of the function maximum. For the experiments in which BFGS failed frequently, the table shows that SD converged strongly and provided the best estimate of the function maximum much more frequently than FCS. For example, in the experiment $(n,\alpha,N)=(1,-5,9)$ SD and FCS converged strongly 66 and 38 times and provided the largest function estimate 85 and 33 times, respectively. Hence SD appears to be a better choice for function maximization than FCS if BFGS fails.

Table 2.1 shows that, as $\alpha$, $N$, and $n$ increase, all optimization methods tend to give the same result. In other words, maximization becomes easier (numerically more stable).

**Table 2.1** Weak convergence, strong convergence, and best log-likelihood value for the 100 repetitions of a Monte Carlo experiment using FDS, SD and BFGS.

| $n$ | $\alpha$ | $\gamma^*$ | $N$ | *FW* | *FS* | *FB* | *SW* | *SS* | *SB* | *BW* | *BS* | *BB* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -15 | 18.15 | 9 | 100 | 36 | 29 | 99 | 57 | 70 | 42 | 37 | 42 |
| 1 | -15 | 18.15 | 49 | 100 | 55 | 44 | 100 | 63 | 58 | 60 | 55 | 60 |
| 1 | -15 | 18.15 | 121 | 100 | 78 | 62 | 100 | 80 | 65 | 87 | 73 | 86 |
| 1 | -5 | 5.42 | 9 | 100 | 38 | 33 | 100 | 66 | 85 | 40 | 38 | 40 |
| 1 | -5 | 5.42 | 49 | 100 | 82 | 71 | 100 | 83 | 75 | 82 | 79 | 82 |
| 1 | -5 | 5.42 | 121 | 100 | 95 | 84 | 100 | 95 | 83 | 97 | 94 | 97 |
| 1 | -3 | 2.88 | 9 | 100 | 52 | 42 | 99 | 73 | 69 | 54 | 50 | 54 |
| 1 | -3 | 2.88 | 49 | 100 | 94 | 83 | 100 | 94 | 84 | 95 | 93 | 95 |
| 1 | -3 | 2.88 | 121 | 100 | 99 | 97 | 100 | 99 | 97 | 99 | 99 | 99 |
| 1 | -1 | 0.41 | 9 | 100 | 80 | 74 | 100 | 97 | 87 | 82 | 80 | 81 |
| 1 | -1 | 0.41 | 49 | 100 | 100 | 96 | 100 | 100 | 98 | 100 | 100 | 100 |
| 1 | -1 | 0.41 | 121 | 100 | 100 | 99 | 100 | 100 | 98 | 100 | 100 | 100 |
| 2 | -15 | 16.13 | 9 | 100 | 37 | 33 | 100 | 52 | 76 | 40 | 38 | 40 |
| 2 | -15 | 16.13 | 49 | 100 | 73 | 67 | 100 | 71 | 71 | 77 | 73 | 76 |
| 2 | -15 | 16.13 | 121 | 100 | 84 | 73 | 100 | 81 | 72 | 88 | 85 | 88 |
| 2 | -5 | 4.82 | 9 | 100 | 57 | 52 | 100 | 66 | 86 | 59 | 57 | 59 |
| 2 | -5 | 4.82 | 49 | 100 | 97 | 94 | 100 | 95 | 92 | 98 | 97 | 98 |
| 2 | -5 | 4.82 | 121 | 100 | 99 | 96 | 100 | 99 | 95 | 100 | 99 | 100 |
| 2 | -3 | 2.56 | 9 | 100 | 69 | 64 | 99 | 77 | 84 | 69 | 69 | 69 |
| 2 | -3 | 2.56 | 49 | 100 | 100 | 97 | 100 | 100 | 97 | 100 | 100 | 100 |
| 2 | -3 | 2.56 | 121 | 100 | 100 | 99 | 100 | 100 | 99 | 100 | 100 | 100 |
| 2 | -1 | 0.36 | 9 | 100 | 93 | 89 | 99 | 98 | 91 | 93 | 93 | 93 |
| 2 | -1 | 0.36 | 49 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2 | -1 | 0.36 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | -15 | 15.49 | 9 | 100 | 41 | 38 | 100 | 41 | 67 | 48 | 41 | 48 |
| 3 | -15 | 15.49 | 49 | 100 | 79 | 72 | 100 | 73 | 72 | 82 | 79 | 82 |
| 3 | -15 | 15.49 | 121 | 100 | 84 | 74 | 100 | 84 | 74 | 91 | 90 | 91 |
| 3 | -5 | 4.63 | 9 | 100 | 70 | 67 | 98 | 71 | 78 | 72 | 72 | 72 |
| 3 | -5 | 4.63 | 49 | 100 | 99 | 99 | 100 | 99 | 99 | 99 | 99 | 99 |
| 3 | -5 | 4.63 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | -3 | 2.46 | 9 | 100 | 80 | 78 | 99 | 83 | 84 | 83 | 80 | 83 |
| 3 | -3 | 2.46 | 49 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | -3 | 2.46 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | -1 | 0.35 | 9 | 100 | 96 | 94 | 99 | 98 | 94 | 97 | 97 | 97 |
| 3 | -1 | 0.35 | 49 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | -1 | 0.35 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | -15 | 14.7 | 9 | 100 | 65 | 63 | 100 | 63 | 71 | 70 | 69 | 70 |
| 8 | -15 | 14.7 | 49 | 100 | 93 | 93 | 100 | 91 | 92 | 98 | 97 | 98 |
| 8 | -15 | 14.7 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | -5 | 4.39 | 9 | 100 | 86 | 86 | 100 | 85 | 88 | 88 | 88 | 88 |
| 8 | -5 | 4.39 | 49 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | -5 | 4.39 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | -3 | 2.34 | 9 | 100 | 96 | 96 | 100 | 96 | 96 | 97 | 97 | 97 |
| 8 | -3 | 2.34 | 49 | 100 | 100 | 100 | 100 | 100 | 99 | 100 | 100 | 100 |
| 8 | -3 | 2.34 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | -1 | 0.33 | 9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | -1 | 0.33 | 49 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | -1 | 0.33 | 121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

**Table 2.2** Estimated bias and standard deviation of $\hat{\alpha}$ for the 100 repetitions of a Monte Carlo Experiment using FDS, SD and BFGS.

| $n$ | $\alpha$ | $\gamma^*$ | $N$ | $b(\hat{\alpha}_{\text{FCS}})$ | $\hat{\sigma}_{\text{FCS}}$ | $b(\hat{\alpha}_{\text{SD}})$ | $\hat{\sigma}_{\text{SD}}$ | $b(\hat{\alpha}_{\text{BFGS}})$ | $\hat{\sigma}_{\text{BFGS}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -15 | 18.15 | 9 | -266.56 | 222.75 | -1.05e5 | 1.04e6 | -3.79e6 | 6.16e6 |
| 1 | -15 | 18.15 | 49 | -137.86 | 158.95 | -236.67 | 311.61 | -1.25e6 | 1.56e6 |
| 1 | -15 | 18.15 | 121 | -62.33 | 109.72 | -90.51 | 196.88 | -5.07e5 | 9.88e5 |
| 1 | -5 | 5.42 | 9 | -256.84 | 217.17 | -735.85 | 763.52 | -2.93e6 | 3.82e6 |
| 1 | -5 | 5.42 | 49 | -65.35 | 127.11 | -105.98 | 240.46 | -6.84e5 | 1.48e6 |
| 1 | -5 | 5.42 | 121 | -16.60 | 55.53 | -18.84 | 68.60 | -1.14e5 | 5.09e5 |
| 1 | -3 | 2.88 | 9 | -195.12 | 210.49 | -479.99 | 648.16 | -2.14e6 | 3.16e6 |
| 1 | -3 | 2.88 | 49 | -21.58 | 73.85 | -35.12 | 146.65 | -2.47e5 | 1.28e6 |
| 1 | -3 | 2.88 | 121 | -4.24 | 36.29 | -5.98 | 53.36 | -3.24e4 | 3.24e5 |
| 1 | -1 | 0.41 | 9 | -82.16 | 167.85 | -241.48 | 589.97 | -8.52e5 | 1.89e6 |
| 1 | -1 | 0.41 | 49 | -0.17 | 0.55 | -0.18 | 0.55 | -0.17 | 0.55 |
| 1 | -1 | 0.41 | 121 | -0.05 | 0.23 | -0.05 | 0.23 | -0.05 | 0.23 |
| 2 | -15 | 16.13 | 9 | -327.40 | 273.57 | -964.30 | 1299.23 | -4.29e6 | 9.37e6 |
| 2 | -15 | 16.13 | 49 | -103.80 | 173.70 | -137.74 | 262.29 | -1.11e6 | 2.11e6 |
| 2 | -15 | 16.13 | 121 | -49.88 | 113.06 | -48.01 | 120.59 | -3.77e5 | 1.07e6 |
| 2 | -5 | 4.82 | 9 | -246.58 | 293.54 | -644.87 | 1099.78 | -2.43e6 | 3.36e6 |
| 2 | -5 | 4.82 | 49 | -11.36 | 49.26 | -10.39 | 45.53 | -7.76e4 | 4.99e5 |
| 2 | -5 | 4.82 | 121 | -4.04 | 25.67 | -3.33 | 18.25 | -2.13e4 | 2.13e5 |
| 2 | -3 | 2.56 | 9 | -159.61 | 242.90 | -421.20 | 863.30 | -1.73e6 | 3.34e6 |
| 2 | -3 | 2.56 | 49 | -1.78 | 5.71 | -1.80 | 5.72 | -1.87 | 6.37 |
| 2 | -3 | 2.56 | 121 | -0.46 | 1.06 | -0.47 | 1.07 | -0.46 | 1.07 |
| 2 | -1 | 0.36 | 9 | -32.45 | 121.20 | -164.03 | 917.68 | -3.03e5 | 1.34e6 |
| 2 | -1 | 0.36 | 49 | -0.05 | 0.35 | -0.06 | 0.35 | -0.05 | 0.35 |
| 2 | -1 | 0.36 | 121 | -0.04 | 0.16 | -0.04 | 0.16 | -0.04 | 0.16 |
| 3 | -15 | 15.49 | 9 | -366.43 | 336.20 | -712.68 | 1211.79 | -4.65e6 | 5.65e6 |
| 3 | -15 | 15.49 | 49 | -88.35 | 162.84 | -83.66 | 169.71 | -8.67e5 | 1.93e6 |
| 3 | -15 | 15.49 | 121 | -49.35 | 114.15 | -42.62 | 102.92 | -2.73e5 | 8.77e5 |
| 3 | -5 | 4.63 | 9 | -177.16 | 281.25 | -344.84 | 910.78 | -1.73e6 | 3.36e6 |
| 3 | -5 | 4.63 | 49 | -5.65 | 25.24 | -4.90 | 18.24 | -1.41e4 | 1.41e5 |
| 3 | -5 | 4.63 | 121 | -0.65 | 4.15 | -0.67 | 4.15 | -0.68 | 4.36 |
| 3 | -3 | 2.46 | 9 | -104.04 | 220.51 | -144.84 | 372.02 | -1.03e6 | 2.55e6 |
| 3 | -3 | 2.46 | 49 | -0.72 | 2.11 | -0.73 | 2.13 | -0.73 | 2.12 |
| 3 | -3 | 2.46 | 121 | -0.11 | 0.69 | -0.12 | 0.69 | -0.11 | 0.69 |
| 3 | -1 | 0.35 | 9 | -15.46 | 83.55 | -1265.62 | 12553.65 | -7.15e4 | 5.82e5 |
| 3 | -1 | 0.35 | 49 | -0.04 | 0.22 | -0.04 | 0.22 | -0.04 | 0.22 |
| 3 | -1 | 0.35 | 121 | -0.03 | 0.16 | -0.03 | 0.16 | -0.03 | 0.16 |
| 8 | -15 | 14.70 | 9 | -282.88 | 417.33 | -318.09 | 572.95 | -2.98e6 | 5.11e6 |
| 8 | -15 | 14.70 | 49 | -31.17 | 108.79 | -23.93 | 78.06 | -1.62e5 | 8.90e5 |
| 8 | -15 | 14.70 | 121 | -3.27 | 8.86 | -3.30 | 8.84 | -3.32 | 8.97 |
| 8 | -5 | 4.39 | 9 | -109.30 | 289.32 | -106.06 | 303.45 | -1.65e6 | 6.43e6 |
| 8 | -5 | 4.39 | 49 | -0.64 | 2.00 | -0.64 | 2.00 | -0.64 | 2.00 |
| 8 | -5 | 4.39 | 121 | -0.03 | 0.92 | -0.04 | 0.92 | -0.03 | 0.92 |
| 8 | -3 | 2.34 | 9 | -21.89 | 109.61 | -18.43 | 95.35 | -1.73e5 | 1.07e6 |
| 8 | -3 | 2.34 | 49 | -0.24 | 0.91 | -0.24 | 0.91 | -0.24 | 0.91 |
| 8 | -3 | 2.34 | 121 | -0.05 | 0.46 | -0.05 | 0.46 | -0.05 | 0.46 |
| 8 | -1 | 0.33 | 9 | -0.55 | 0.99 | -0.55 | 1.00 | -0.55 | 1.00 |
| 8 | -1 | 0.33 | 49 | -0.07 | 0.24 | -0.07 | 0.24 | -0.07 | 0.24 |
| 8 | -1 | 0.33 | 121 | -0.02 | 0.12 | -0.02 | 0.12 | -0.02 | 0.12 |

In Table 2.2 we can see that the FCS method, for small values of $\alpha$, $N$, and $n$, tends to give the least biased and smallest standard deviation estimate for the parameter $\alpha$. The biases for SC and BFGS are astoundingly large for these values of the experimental parameters and continue to remain large even when only $N$ is small. As the experimental parameters increase FCS, SD and BFGS provide similar results. Hence, FCS appears to be a better choice if one desires more accurate parameter estimates, especially for small sample sizes.

The fact that, when BFGS does not converge, FCS provides the least biased estimates while SD provides a higher function value may imply that the likelihood is extremely flat, but increasing, as $-\alpha$ and $\gamma$ increase. This would explain the astoundingly large parameter estimates provided by BFGS before failing to converge as it searches for the maximum at larger and larger parameter values. The fact that FCS is less biased may only be a consequence of the fact that the MLE estimate is extremely biased and that FCS is losing precision earlier than the other algorithms. To explore this possibility we expand the reduced log likelihood function for large values of the parameters in the next section.

## 2.3 The behavior of the reduced log likelihood for large values of the parameters

In this section we expand the log-likelihood, Equation (2.3), for large values of the parameters $-\alpha$ and $\gamma$, with $n$ and $N$ fixed:

$$\begin{aligned} \ell &= \ell(\alpha,\gamma;z_1,\ldots,z_N,n) = [\ln(\Gamma(n-\alpha)) - \ln(\Gamma(-\alpha))] + \\ &\quad + \left[(-\alpha)\ln(\gamma) - \frac{n-\alpha}{N}\sum_{i=1}^{N}\ln(\gamma+nz_i^2)\right]. \end{aligned} \tag{2.8}$$

To do so, we use the following familiar expansions of the factorial (Sterling's approximation):

$$\lim_{x\to\infty}\ln(\Gamma(x)) = \ln(2\pi)/2 + (x-1/2)\ln x - x + c_1/x + \mathcal{O}(x^{-3}),$$

where $c_1 = 0.08\overline{3}$, and the logarithm:

$$\lim_{x\to\infty} \ln(1+1/x) = 1/x - 1/(2x^2) + \mathscr{O}(x^{-3}).$$

We do not use the expansion of the logarithm for the $\ln(x)$ term in the expansion of the factorial. The logarithm expansion is only used when a term of the form $\ln(1+1/x)$ with $x$ large is encountered. Inserting these expansions into the first term in brackets in Equation (2.8) for $-\alpha$ and $\gamma$ large, we obtain to $\mathscr{O}(\min(-\alpha,\gamma)^{-2})$:

$$\begin{aligned} & [\ln(2\pi)/2 + (n-\alpha-1/2)\ln(n-\alpha) - (n-\alpha) + c_1/(n-\alpha) \\ - \; & \{\ln(2\pi)/2 + (-\alpha-1/2)\ln(-\alpha) - (-\alpha) + c_1/(-\alpha)\}\,] \\ = \; & \left[ n\ln(-\alpha) + \frac{1}{2}\left(\frac{n}{-\alpha}\right)(n-1)\right] \end{aligned} \tag{2.9}$$

and we obtain the following for the second term in brackets:

$$\begin{aligned} & \left[ (-\alpha)\ln(\gamma) - \frac{(n-\alpha)}{N}\sum_{i=1}^{N}\left\{\ln(\gamma) + nz_i^2/\gamma - (nz_i^2/\gamma)^2/2\right\}\right] \\ = \; & \left[ -n\ln(\gamma) - \frac{n^2}{\gamma}\overline{z^2} - n\left(\frac{-\alpha}{\gamma}\right)\overline{z^2} + \left(\frac{n^2}{2}\right)\left(\frac{-\alpha}{\gamma^2}\right)\overline{z^4}\right], \end{aligned} \tag{2.10}$$

where $\overline{z^2}$ and $\overline{z^4}$ represent the sample second and fourth moments. Adding the last lines of equations (2.9) and (2.10) we see that the reduced log-likelihood becomes (to $\mathscr{O}(\min(-\alpha,\gamma)^{-2})$):

$$\ell = n\ln\left(\frac{-\alpha}{\gamma}\right) - \left(\frac{-\alpha}{\gamma}\right)n\overline{z^2} + \frac{n(n-1)}{2}\left(\frac{1}{-\alpha}\right) + n^2\left\{\left(\frac{-\alpha}{\gamma}\right)\frac{\overline{z^4}}{2} - \overline{z^2}\right\}\left(\frac{1}{\gamma}\right). \tag{2.11}$$

The first two terms in Equation (2.11) dominate the log-likelihood, being of $\mathscr{O}(1)$ if $-\alpha$ and $\gamma$ go to infinity at the same rate, whereas the last two terms are of $\mathscr{O}(1/\min(-\alpha,\gamma))$.

If one considers only the first two terms of equation (2.11) the resulting likelihood is maximized for $\gamma/(-\alpha) = \overline{z^2}$. In fact, this is exactly the ratio which is observed when the BFGS MLE estimates do not converge in our Monte Carlo experiment.

Now we consider the last two terms of Equation (2.11). We replace all occurrences of

$\gamma/(-\alpha)$ by $\overline{z^2}$. The term of interest is

$$\frac{n(n-1)}{2}\left(\frac{1}{-\alpha}\right)+n^2\left\{\left(\frac{1}{\overline{z^2}}\right)\frac{\overline{z^4}}{2}-\overline{z^2}\right\}\left(\frac{1}{\gamma}\right). \tag{2.12}$$

If Equation (2.12) is negative, then increasing $-\alpha$ and $\gamma$ while maintaining their ratio at $\overline{z^2}$ will increase the likelihood. This is because the order one terms remain at their maximum and the negative term goes to zero, increasing the log-likelihood. In this case, the MLE estimates are $(\hat{\alpha},\hat{\gamma})_{\text{MLE}}=(-\infty,\infty)$ with $\hat{\gamma}/(-\hat{\alpha})=\overline{z^2}$. When this occurs one has a "monotone likelihood" (Bryson & Johnson, 1981).

In our case a little algebra reveals that Equation (2.12) is negative when

$$\left(\frac{\overline{z^4}}{\left(\overline{z^2}\right)^2}\right)<\frac{n+1}{n}. \tag{2.13}$$

Once again, this agrees with our Monte Carlo results. When the above inequality is satisfied the BFGS MLE estimates diverge and when it is not satisfied they converge to finite values.

When Frery *et al.* (1997) proposed the $\mathscr{G}_A^0$ model they also discussed the relationship to other models of speckled data. In particular, they show in their Equation (7) that when $-\alpha,\gamma\to\infty$ with $-\alpha/\gamma=\beta_2$ the $\mathscr{G}_A^0$ distribution converges in distribution to a $\Gamma^{1/2}(n,n\beta_2)$ which is a model for homogeneous data. Hence divergence of the estimated parameters implies a smooth target.

### 2.3.1 Solutions for monotone likelihood in the literature

The literature on monotone likelihood touches logistic regression (Kolassa, 1997), the Cox proportional hazard survival model (Bryson & Johnson, 1981; Heinze & Schemper, 2001; Loughin, 1998), generalized linear models and Weibull models (Clarkson & Jennrich, 1991), as well as skewed normal and skewed $t$ model estimation (Sartori, 2006).

In Clarkson & Jennrich (1991) and Kolassa (1997) the interest is in a regression setting

where a divergent parameter can make sense conceptually. For example, if belonging to a certain group always indicates success in a logistic regression model, or if members of different groups have no overlap in survival times in a Cox proportional hazard survival model. In such cases Clarkson & Jennrich (1991) implement an "extended maximum likelihood" estimator, which allows some of the parameters to be infinite while still allowing inference to be performed on the other parameters. In Kolassa (1997) a similar attitude is taken and saddlepoint expansions are made to perform inference on the finite parameters.

In Loughin (1998) no explicit suggestions are made for improving inference in the face of monotone likelihood. Finally, Heinze & Schemper (2001) and Sartori (2006) implement a bias correction suggested by Firth (1993). Since the other solutions leave the estimates of the infinite parameters infinite, we follow Heinze & Schemper (2001) and Sartori (2006) and use Firth's (1993) bias correction.

## 2.4 Firth's method of bias correction for monotone likelihood

Generally speaking, MLE is biased in finite samples. This bias, $b(\theta) = \mathbb{E}(\hat{\theta}) - \theta$, depends on the true value of the parameter, $\theta$, and can be expanded asymptotically as

$$b(\theta) = b_1(\theta)/N + b_2(\theta)/N^2 + \cdots, \tag{2.14}$$

where $N$ represents the size of the sample used to estimate the parameter.

Other than Firth's method, there exist many specific methods to reduce the bias in MLE estimates (see, for example Schaefer, 1983; Cordeiro & McCullagh, 1991; Cordeiro & Cribari-Neto, 1998; Leung & Wang, 1998). Here, we discuss two general bias reduction approaches in MLE situations.

One approach, pioneered by Quenouille (1957) and more recently revived by Efron & Tibshirani (1993), involves re-estimating the parameters of interest for resamples of the original data to obtain the distribution of the estimator. This approach is computationally intensive, but

does not require analytical calculations. In our case, it cannot be implemented in its standard form since the original parameter estimate which is to be bias corrected is not necessarily finite. Even if it is finite, there is no assurance that it will remain finite in the re-sampled data. This is the main argument of Loughin (1998), where he cautions about the use of bootstrap techniques for bias reduction in Cox models where monotone likelihood may occur.

The second method is based on explicitly calculating the term $b_1(\theta)$ and then plugging the estimated parameter into this bias term, yielding the following bias corrected estimate:

$$\hat{\theta}_{BC} = \hat{\theta} - b_1(\hat{\theta})/N. \tag{2.15}$$

As above, this method fails when $\hat{\theta}$ is infinite.

### 2.4.1 Firth's idea

Firth's (1993) motivation for dealing with this problem was in the context of generalized linear models, where the parameter estimates are biased away from zero and in extreme cases can be infinite. His idea is the following: since the estimates may not exist, we should change the estimating equations to correct for the bias before estimating.

The idea behind his bias correction is the following. For all ML problems, the expected value of the score function, $U_\theta = \partial\ell/\partial\theta$, evaluated at the true parameter value, $\theta$, is zero:

$$\mathbb{E}[(U_\theta)_r] = \int_{-\infty}^{+\infty} \frac{\partial f(x)}{\partial \theta_r} \frac{1}{f(x)} f(x)dx = \frac{\partial}{\partial \theta_r} \int_{-\infty}^{+\infty} f(x)dx = 0, \;\; \forall r \in 1 \ldots k, \tag{2.16}$$

when $\theta$ is a $k-$component vector. However, the score is generally not linear in $\theta$, hence when one calculates the ML estimate by equating the value of the score to zero a bias is induced. In Figure 2.2 one can see that the required shift in the score function is $U_\theta b(\theta) = -i(\theta)b(\theta)$, where $i(\theta)$ is Fisher's information.

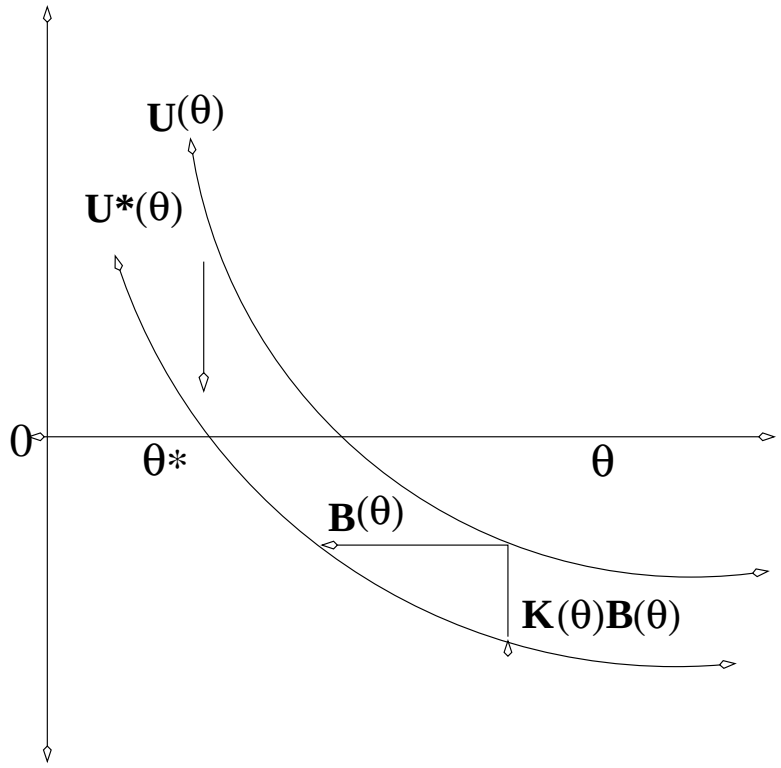


**Figure 2.2** Figure which illustrates Firth's heuristic argument for bias correction by shifting of the score function. In this figure $K(\theta)$ represents the information (the slope of the score) and $B(\theta)$ the bias. $\hat{\theta}$ is located where the original score function, $U(\theta)$, crosses zero.

### 2.4.2 Derivation of the general correction

Firth (1993) formalizes the above heuristic argument to calculate the necessary shift in the score. Here, following Firth (see also McCullagh, 1987), some notation is necessary to specify the log-likelihood derivatives and their cumulants:

$$U_r(\theta) = \partial \ell / \partial \theta^r, \quad U_{rs} = \partial^2 \ell / \partial \theta^r \partial \theta^s \tag{2.17}$$

where $\theta = (\theta^1, \ldots, \theta^p)$ is the parameter vector. The joint cumulants are

$$\kappa_{r,s} = N^{-1}\mathbb{E}[U_r U_s], \quad \kappa_{r,s,t} = N^{-1}\mathbb{E}[U_r U_s U_t], \quad \kappa_{r,st} = N^{-1}\mathbb{E}[U_r U_{st}]. \tag{2.18}$$

Throughout the calculations we will freely use the following results:

$$\kappa_{rs} + \kappa_{r,s} = 0, \quad \kappa_{rst} + \kappa_{r,st} + \kappa_{s,rt} + \kappa_{t,rs} + \kappa_{r,s,t} = 0. \tag{2.19}$$

Firth then considers a general alteration to the score (specified component by component):

$$U_r^*(\theta) = U_r(\theta) + A_r(\theta). \tag{2.20}$$

Performing an expansion around the true parameter, he obtains

$$\mathbb{E}[A] = -i(\theta)b_1(\theta)/N + \mathscr{O}(N^{-1/2}). \tag{2.21}$$

Two obvious choices which respect Equation (2.21) are $A^{(E)} = -i(\theta)b_1(\theta)/N$, which uses $i(\theta) = \kappa_{r,s}$ (the expected information), and $A^{(O)} = -I(\theta)b_1(\theta)/N$, which uses $I(\theta) = -U_{rs}$ (the observed information).

### 2.4.3 The Jeffreys invariant prior

Firth notes that for an exponential family model in canonical form the observed information does not depend on the data and hence $A^{(E)} = A^{(O)}$. In this case he shows that $A_r(\theta) = -\kappa^{u,v}\kappa_{ruv}/2$, where $\kappa^{u,v}$ represents the inverse of the expected Fisher information matrix. Hence,

$$A_r(\theta) = \frac{1}{2}\mathrm{tr}\left\{ i^{-1}(\theta)\left(\frac{\partial i(\theta)}{\partial \theta^r}\right)\right\} = \frac{\partial}{\partial \theta^r}\left\{\frac{1}{2}\log\det|i(\theta)|\right\} \tag{2.22}$$

and the bias correction corresponds to finding the mode of the posterior distribution after using the Jeffreys (1946) invariant prior (that is, maximizing $L^* = L|i(\theta)|^{1/2}$).

The $\mathscr{G}_A^0$ model is not an exponential family model (much less so in canonical form). The observed information matrix depends on the data:

$$I(\alpha,\gamma) = \begin{pmatrix} N\left[\psi^{(1)}(-\alpha) - \psi^{(1)}(n-\alpha)\right] & \frac{N}{\gamma} - \sum_{i=1}^N (\gamma + nz_i^2)^{-1} \\ \frac{N}{\gamma} - \sum_{i=1}^N (\gamma + nz_i^2)^{-1} & \frac{N(-\alpha)}{\gamma^2} - (n-\alpha)\sum_{i=1}^N (\gamma + nz_i^2)^{-2} \end{pmatrix}. \tag{2.23}$$

Even so, we wish to implement the Jeffreys invariant prior using both the observed and expected information matrices since this prior penalizes parameter regions where the information is small.

To perform the calculation using the expected information we use Equation (2.5) to obtain

$$i^{-1}(\alpha,\gamma)=\frac{1}{N}\frac{1}{|i_\alpha(\alpha)|}\begin{pmatrix} \frac{(n)(-\alpha)}{(n-\alpha+1)} & \frac{-\gamma n}{(n-\alpha)} \\ \frac{-\gamma n}{(n-\alpha)} & \gamma^2(\psi^{(1)}(-\alpha)-\psi^{(1)}(n-\alpha)), \end{pmatrix} \tag{2.24}$$

where

$$i_\alpha(\alpha)=\begin{pmatrix} \psi^{(1)}(-\alpha)-\psi^{(1)}(n-\alpha) & \frac{n}{n-\alpha} \\ \frac{n}{n-\alpha} & \frac{(n)(-\alpha)}{(n-\alpha+1)}. \end{pmatrix}$$

The derivatives of the information which we require are given by

$$\frac{\partial i(\alpha,\gamma)}{\partial\alpha}=N\begin{pmatrix} -\psi^{(2)}(-\alpha)+\psi^{(2)}(n-\alpha) & \frac{n}{\gamma}(n-\alpha)^{-2} \\ \frac{n}{\gamma}(n-\alpha)^{-2} & -\frac{n^2}{\gamma^2(n-\alpha+1)^2} \end{pmatrix} \tag{2.25}$$

and

$$\frac{\partial i(\alpha,\gamma)}{\partial\gamma}=N\begin{pmatrix} 0 & -\frac{n}{\gamma^2(n-\alpha)} \\ -\frac{n}{\gamma^2(n-\alpha)} & -\frac{2n(-\alpha)}{\gamma^3(n-\alpha+1)} \end{pmatrix}. \tag{2.26}$$

Using Equation (2.22) we can calculate $(1/2)\partial(\log|i(\alpha,\gamma)|)/\partial\gamma=-1/\gamma$ and

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial\alpha}\log|i(\alpha,\gamma)| &= \frac{1}{2|i_\alpha(\alpha)|}\left\{\frac{n(-\alpha)}{(n-\alpha+1)}\left(-\psi^{(2)}(-\alpha)+\psi^{(2)}(n-\alpha)\right)\right. \\ &\quad \left.-\frac{2n^2}{(n-\alpha)^3}-\frac{n(n+1)}{(n-\alpha+1)^2}\left(\psi^{(1)}(-\alpha)-\psi^{(1)}(n-\alpha)\right)\right\}. \end{aligned}\tag{2.27}$$

The above equations were implemented in `Ox` (Doornik, 2001) and tested by comparing the analytical and numeric derivatives based on adding half the log of the determinant of the information to the log-likelihood.

The calculation using the observed information is based on Equation (2.23). Following similar steps to those above for the expected information we arrive at the slightly more complicated

formulas:

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial\alpha}\log|I(\alpha,\gamma)| &= \frac{N^2}{2|I(\alpha,\gamma)|}\left\{\left(-\psi^{(2)}(-\alpha)+\psi^{(2)}(n-\alpha)\right)\times\right.\\ &\quad\left(\frac{-\alpha}{\gamma^2}-\frac{n-\alpha}{N}\sum(\gamma+nz_i^2)^{-2}\right)+\\ &\quad\left(\psi^{(1)}(-\alpha)-\psi^{(1)}(n-\alpha)\right)\times\\ &\quad\left.\left(\frac{1}{N}\sum(\gamma+nz_i^2)^{-2}-\frac{1}{\gamma^2}\right)\right\} \qquad (2.28)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial\gamma}\log|I(\alpha,\gamma)| &= \frac{N^2}{|I(\alpha,\gamma)|}\left\{\left(\frac{1}{N}\sum(\gamma+nz_i^2)^{-1}-\frac{1}{\gamma}\right)\times\right.\\ &\quad\left(\frac{1}{N}\sum(\gamma+nz_i^2)^{-2}-\frac{1}{\gamma^2}\right)+\\ &\quad\left(\psi^{(1)}(-\alpha)-\psi^{(1)}(n-\alpha)\right)\times\\ &\quad\left.\left(\frac{(n-\alpha)}{N}\sum(\gamma+nz_i^2)^{-3}-\frac{(-\alpha)}{\gamma^3}\right)\right\}. \qquad (2.29)\end{aligned}$$

Once again, these last two formulas were checked by comparing numerical and analytic derivatives in `Ox`.

### 2.4.4 The general correction

Finally, we implement the bias correction for the general case derived by Firth. The modifications of the score function based on the expected and observed information are given by

$$A_r^{(E)} = \kappa^{u,v}(\kappa_{r,u,v}+\kappa_{r,uv})/2 \qquad (2.30)$$

and

$$A_r^{(O)} = -U_{rs}\kappa^{s,t}\kappa^{u,v}(\kappa_{t,u,v}+\kappa_{t,uv})/(2N). \qquad (2.31)$$

Note that these formulas correspond to many terms because repeated indices are summed over. For example,

$$A_\alpha^{(E)} = \kappa^{\alpha,\alpha}(\kappa_{\alpha,\alpha,\alpha} + \kappa_{\alpha,\alpha\alpha})/2 + \kappa^{\gamma,\gamma}(\kappa_{\alpha,\gamma,\gamma} + \kappa_{\alpha,\gamma\gamma})/2 + \kappa^{\alpha,\gamma}(\kappa_{\alpha,\alpha,\gamma} + \kappa_{\alpha,\alpha\gamma}). \quad (2.32)$$

Exploiting the fact that $U_{\alpha\alpha}$ does not depend on the data and using the restrictions in Equation (2.19) the above equation reduces to

$$A_\alpha^{(E)} = \kappa^{\alpha,\alpha}(-\kappa_{\alpha\alpha\alpha})/2 + \kappa^{\gamma,\gamma}(-\kappa_{\alpha\gamma\gamma} - 2\kappa_{\gamma,\alpha\gamma})/2 + \kappa^{\alpha,\gamma}(-\kappa_{\alpha,\alpha\gamma}). \quad (2.33)$$

The corresponding equation for $\gamma$ is

$$A_\gamma^{(E)} = \kappa^{\alpha,\alpha}(-\kappa_{\alpha,\alpha\gamma}) + \kappa^{\gamma,\gamma}(-\kappa_{\gamma\gamma\gamma} - 2\kappa_{\gamma,\gamma\gamma})/2 + \kappa^{\alpha,\gamma}(-\kappa_{\alpha\gamma\gamma} - \kappa_{\alpha,\gamma\gamma} - \kappa_{\gamma,\alpha\gamma}). \quad (2.34)$$

In order to explicitly form these terms we need the following expectations for a random variable $Z \sim \mathscr{G}_A^0(\alpha,\gamma,n)$:

$$\mathbb{E}[(1+nZ^2/\gamma)^{-a}] = \frac{\Gamma(n-\alpha)}{\Gamma(n-\alpha+a)}\frac{\Gamma(-\alpha+a)}{\Gamma(-\alpha)}.$$

We also need

$$\mathbb{E}[\frac{\log(1+nZ^2/\gamma)}{(1+nZ^2/\gamma)^b}] = (\psi(n-\alpha+b) - \psi(-\alpha+b))\frac{\Gamma(n-\alpha)}{\Gamma(n-\alpha+b)}\frac{\Gamma(-\alpha+b)}{\Gamma(-\alpha)}.$$

We only calculated the correction term using the expected information since the observed information contains many more terms because of the double sum. Unfortunately, in this case, we do not have a separate calculation of the penalized likelihood function, which we can use as a check on our calculations and implementation.

## 2.5 A resampling solution for monotone likelihood

As commented at the beginning of section 2.4, Loughin (1998) discourages the use of bootstrap resampling methods for bias estimation and correction when monotone likelihood has a significant probability of occurring in the resamples. He argues that the results obtained are only valid conditional on the MLE estimates being finite and he demonstrates a severe underestimation of the bias when monotone likelihood occurs frequently.

However, Cribari-Neto *et al.* (2002) successfully implement just such bootstrap bias corrections. Their results remain conditional on the MLE estimates being finite, however, their bias correction is quite good (especially the "better" bootstrap bias estimate based on Efron (1990)).

### 2.5.1 Bootstrap bias estimates

In this section we briefly discuss simple non-parametric bootstrap bias estimates, Efron's (1990) "better" bias estimates, and Cribari-Neto *et al.*'s (2002) adaptation of Efron's estimator.

Let $X \overset{\text{iid}}{\sim} f_\theta(x)$ be a random sample of size $N$ from a distribution $F_\theta$. The non-parametric bootstrap approximates $F$ by $\hat{F}$, the empirical distribution function based on the data. One generates samples from $\hat{F}$ by sampling with replacement from the data. A non-parametric bootstrap bias estimate is formed by averaging the bootstrap parameter estimate, $\hat{\theta}_i$, over many samples from $\hat{F}$ and subtracting the original estimate. A bootstrap bias corrected estimate (BBC) is then formed by subtracting this bias from the original estimate:

$$\hat{\theta}_{\text{BBC}} = \hat{\theta} - \left( \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}_i - \hat{\theta} \right). \tag{2.35}$$

Efron's (1990) suggestion requires a little notation. In each of the $B$ bootstrap resamples above, the sample can be described by the weight that each observation receives in the new empirical distribution function. For the original sample, each observation received weight $1/N$. This can be succinctly recorded in a vector $P^0 = 1/N(1,\ldots,1)$. For the $i$-th bootstrap sample

this vector becomes

$$P^i = \frac{1}{N}\left(\#\{X_1\}_i, \ldots, \#\{X_N\}_i\right),$$

where $\#\{X_j\}_i$ represents the number of times $X_j$ occurs in the $i$-th bootstrap sample. The vector is called the resampling vector. Efron's idea is based on the possibility of writing the parameter estimate as a closed function of the data using the vector $P^0$. For example, the estimate of the mean can be written $\bar{X} = T(P^0) = P^0 \cdot x$. An estimate of the second moment could be written $\bar{X^2} = T(P^0) = P^0 \cdot (x^2)$.

In order to accelerate the convergence of the bias estimate such that fewer bootstrap repetitions are required, Efron suggests that when calculating the estimate of the bias that one subtracts the parameter estimate resulting from using

$$P^* = \frac{1}{B}\sum_{i=1}^{B} P_i,$$

$T(P^*)$. The new better bootstrap bias correction (BBBC) estimate would then be

$$\hat{\theta}_{\text{BBBC}} = \hat{\theta} - \left(\frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i - T(P^*)\right). \tag{2.36}$$

It makes intuitive sense that this would yield faster convergence to an accurate bias estimate as the number of bootstrap samples ($B$) increases, since for small values of $B$ the obtained samples could be significantly different from the original sample. In this case the true bias and the difference between the bootstrap samples and the original sample would get confounded in the bias estimate. We note that as $B \to \infty$ for fixed $N$, $P^* \to P^0$.

As Cribari-Neto *et al.* (2002) note, the MLEs for the $\mathscr{G}_A^0$ model do not have closed forms. However, to implement the BBBC they take an approach similar to Firth's above. They write the estimating equations as a function of $P^0$ and then use the estimate obtained by replacing $P^0$ with $P^*$ in the estimating equations to correct the bias. They rewrite the log-likelihood in

Equation (2.3) as

$$(\hat{\alpha},\hat{\gamma}) = T(P^0) = \arg\max_{(\alpha,\gamma)} \left\{ \ln\left(\frac{\Gamma(n-\alpha)}{\gamma^{\alpha}\Gamma(-\alpha)}\right) - (n-\alpha)(P^0 \cdot \ln(\gamma + nz^2)) \right\}. \tag{2.37}$$

They then calculate $T(P^*)$ obtained by replacing $P^0$ by $P^*$ in Equation (2.37) and insert it into Equation (2.36) to generate their BBBC estimate.

Given that Cribari-Neto *et al.*'s bias corrected estimates were all conditional on the MLE estimate being finite, it makes sense for them to have used the BBBC estimate (and for it to have performed well) since many samples were being discarded and there was no guarantee that as $B \to \infty$ for fixed $N$, $P^* \to P^0$.

### 2.5.2 A bootstrap estimator

In this section we suggest an estimator based on bootstrap resampling which works even in the presence of monotone likelihood. We first define the estimator and then comment on its properties. In Equation (2.13), on page 21, we defined the criterion which if satisfied yields infinite MLE parameter estimates. Our suggestion is to perform a non-parametric bootstrap of the data, keeping only those bootstrap samples where the divergence criteria is not satisfied until a certain pre-determined number of non-diverging bootstrap samples are obtained. Then our estimate is given by $T(P^*)$ as calculated in Equation (2.37). Whereas this estimator requires much resampling and checking of the divergence criteria, it only requires one non-linear maximization.

Note that it would be a grave error to use the average of the parameter estimates from the resamples as our estimator. To see this consider the following heuristic argument. Given a sample, one hopes that the empirical distribution function calculated from that sample $\hat{F}$ is a reasonable approximation to the population distribution function $F$. By resampling with replacement from our data, we essentially draw samples from $\hat{F}$. We know that the true parameter value for the distribution $\hat{F}$ is $\hat{\theta}$, our MLE estimate. If the average of the parameter estimates from the resamples is larger than $\hat{\theta}$ one should believe that $\hat{\theta}$ is likely larger than the popula-

tion parameter value, $\theta$, by the same amount. Summarizing by looking at Equation (2.35), a bootstrap bias corrected parameter estimate, $\hat{\theta}_{\mathrm{BBC}}$, will be smaller than our original estimate if the average of the resample estimates is larger than our original estimate and therefore this average should not be used as an estimator.

For our proposed estimator to be useful it must be consistent and it must exist and be finite for the great majority of samples. The following are the possible problems that could occur with the estimator: (i) no bootstrap samples may exist which do not satisfy the divergence criteria, Equation (2.13); (ii) even if we have $B$ valid bootstrap samples, the pseudo-sample corresponding to $P^*$ may satisfy Equation (2.13); (iii) the estimator may not be consistent (converge in probability to the true parameter value). In the following paragraphs we treat each of the problems listed above.

In order to resolve problem (i) we must determine what value of $P^i$, the resampling vector, maximizes $r = \overline{z^4}/(\overline{z^2})^2$. If this maximum value satisfies the divergence criteria, then no bootstrap sample exists. The ratio of the fourth sample moment to the second sample moment squared is similar to a measure of kurtosis. The more weight the distribution has in the upper tail, the higher the value will be (we only mention the upper tail because $z > 0$ and the moments in question are not centered).

Define $x = \max(z_1, \ldots, z_N)$. The fourth sample moment can then be written as $\overline{z^4} = x^4(1 + \sum a_i^4)/N$, where $z_i = a_i x$, and the second sample moment as $\overline{z^2} = x^2(1 + \sum a_i^2)/N$. Define $a_{\min} = \min(a_1, \ldots, a_N)$. The values of $a_{\min} \leq a_i \leq 1$ which maximize $r$ correspond to choices of $a_i = a_{\min}$ or $a_i = 1$ (we choose only the minimum or maximum in our resample). Assuming that for $N$ observations we set $N_{\max}$ observations to the maximum and the rest to the minimum, $r$ can be written as

$$r = N \frac{N_{\max} + (N - N_{\max}) a_{\min}^4}{(N_{\max} + (N - N_{\max}) a_{\min}^2)^2}.$$

Rewriting in terms of $f_{\max} = N_{\max}/N$ and maximizing with respect to $f_{\max}$ we find that

$$f_{\max} = \frac{a_{\min}^2}{1 + a_{\min}^2}.$$

For a resample with a fraction $f_{\max}$ of observations in the maximum and the rest in the minimum,

$$r_{\max} = \left( \frac{1 + a_{\min}^2}{2a_{\min}} \right)^2 .$$

As $a_{\min}$ increases between zero and one, $r_{\max}$ decreases. It is of interest to calculate the largest value of $a_{\min}$ for which $r_{\max}$ violates the divergence criteria, given $N$ and $n$. We call this value $a_{\max}$ and hope that the notation can handle the abuse. As defined, $a_{\max}$ satisfies

$$a_{\max} = \frac{1}{\sqrt{n}} \left[ \sqrt{n+1} - 1 \right] .$$

In Table 2.3 we tabulate the probability of having a sample with $a_{\min} > a_{\max}$. The table shows that problem (i) gets worse for larger values of $-\alpha$; varies non-monotonically with $n$, and gets better with more observations, $N$. In the worst case presented in the table, $(\alpha, n, N) = (-15, 3, 9)$, the probability of generating a sample for which no valid bootstrap samples exist is 3.9%. However, for $N = 25$ the probability falls to much less ($1.7 \times 10^{-3}$%). This is impressive if one notes in Table 3.1 that the probability of monotone likelihood is 22% for $(\alpha, n, N) = (-15, 3, 49)$. In our Monte Carlo experiments using the bootstrap estimator, we use a smallest sample size of $N = 49$. In this case the worst case probability that our estimator will not exist is about $1 \times 10^{-8}$%.

We continue to problem (ii) mentioned above. The problem is the possibility that the pseudo-sample corresponding to $P^*$ will satisfy the divergence criteria, even though each bootstrap sample used to construct it did not. We demonstrate algebraically that this is not possible. Each bootstrap sample satisfies

$$(P^i \cdot z^4)/(P^i \cdot z^2)^2 > c$$

which we can write as $d_i \equiv a_i/b_i^2 > c$. We wish to determine if $P^* = \sum_{i=1}^{B} P^i/B$ satisfies

$$(P^* \cdot z^4)/(P^* \cdot z^2)^2 > c.$$

**Table 2.3** Probability of generating a sample of size $N$ from a $\mathcal{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ chosen according to equation (2.7), such that all possible bootstrap samples satisfy the divergence criterion, equation (2.13). Note that 0.0e+00 represents a number smaller than 1.0e−16.

| $\alpha$ | $n$ | $N=9$ | $N=25$ | $N=49$ | $N=81$ | $N=121$ |
|---|---|---|---|---|---|---|
| −15 | 1 | 22.00e−3 | 340.0000e−8 | 43.0000e−13 | 0.00e+00 | 0.00e+00 |
| −15 | 2 | 36.00e−3 | 1400.0000e−8 | 630.0000e−13 | 0.00e+00 | 0.00e+00 |
| −15 | 3 | 39.00e−3 | 1700.0000e−8 | 1200.0000e−13 | 0.00e+00 | 0.00e+00 |
| −15 | 8 | 26.00e−3 | 440.0000e−8 | 72.0000e−13 | 0.00e+00 | 0.00e+00 |
| −5 | 1 | 16.00e−3 | 110.0000e−8 | 4.3000e−13 | 0.00e+00 | 0.00e+00 |
| −5 | 2 | 19.00e−3 | 180.0000e−8 | 11.0000e−13 | 0.00e+00 | 0.00e+00 |
| −5 | 3 | 16.00e−3 | 100.0000e−8 | 3.5000e−13 | 0.00e+00 | 0.00e+00 |
| −5 | 8 | 4.30e−3 | 1.8000e−8 | 0.0011e−13 | 0.00e+00 | 0.00e+00 |
| −3 | 1 | 11.00e−3 | 36.0000e−8 | 0.4600e−13 | 0.00e+00 | 0.00e+00 |
| −3 | 2 | 10.00e−3 | 28.0000e−8 | 0.2700e−13 | 0.00e+00 | 0.00e+00 |
| −3 | 3 | 7.10e−3 | 8.9000e−8 | 0.0260e−13 | 0.00e+00 | 0.00e+00 |
| −3 | 8 | 1.20e−3 | 0.0360e−8 | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| −1 | 1 | 2.20e−3 | 0.2700e−8 | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| −1 | 2 | 0.95e−3 | 0.0210e−8 | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| −1 | 3 | 0.42e−3 | 0.0017e−8 | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| −1 | 8 | 0.03e−3 | 4.0e−15 | 0.00e+00 | 0.00e+00 | 0.00e+00 |

This can be written as

$$\left(\frac{1}{B}\sum_{i=1}^{B} a_i\right) \Big/ \left(\frac{1}{B}\sum_{i=1}^{B} b_i\right)^2.$$

Define $b_{\max} = \max_i(b_i)$ and divide the numerator and denominator by $b_{\max}^2$ to obtain

$$\left(\frac{1}{B}\sum_{i=1}^{B} d_i(b_i/b_{\max})^2\right) \Big/ \left(\frac{1}{B}\sum_{i=1}^{B}(b_i/b_{\max})\right)^2,$$

which, since each $d_i > c$, is strictly greater than

$$Bc\left(\sum_{i=1}^{B}(b_i/b_{\max})^2\right) \Big/ \left(\sum_{i=1}^{B}(b_i/b_{\max})\right)^2.$$

Define $e = (e_1,\ldots,e_B)$ with $e_i = b_i/b_{\max}$ and $1_B = (1,\ldots,1)$ as vectors of length $B$. We must

now show that $Bc(e \cdot e)/(e \cdot 1_B)^2 \geq c$, which can be rewritten as,

$$(e \cdot 1_B)^2/(e \cdot e) \leq B.$$

Define $\hat{e}$ as the unit norm vector in the $e$ direction and $\hat{1}_B$ as the unit norm vector in the $1_B$ direction. Divide the top and bottom of the left hand side of the last display by the squared norm of $e$ and both sides by the squared norm of $1_B$ (which is $B$) to obtain

$$(\hat{e} \cdot \hat{1}_B)^2/(\hat{e} \cdot \hat{e}) \leq 1.$$

The last inequality proves the result. It holds because $\hat{e} \cdot \hat{e} = 1$ and $\hat{e} \cdot \hat{1}_B \leq \|\hat{e}\| \cdot \|\hat{1}_B\| = 1$.

The third possible problem (that the estimator may not be consistent) will be considered in steps. First we show that the probability of violating the divergence criteria goes to zero as $N$ goes to infinity. We have

$$\text{plim}(r) = \frac{\mathbb{E}[Z^4]}{\mathbb{E}[Z^2]^2} = \frac{(-\alpha-1)(n+1)}{(-\alpha-2)(n)}, \tag{2.38}$$

where plim represents convergence in probability and the first equality holds because each sample moment converges in probability independently to the population moment and therefore the ratio of the sample moments converges to the ratio of the population moments. The second equality comes from the known population moments for $\mathscr{G}_A^0$ variables in Equation (2.2) on page 10. For $-\alpha > 2$ the right hand side of Equation (2.38) is always greater than $(n+1)/n$ and the divergence criteria is violated. For $-\alpha < 2$ the data are so heterogeneous and disperse that the fourth moment does not exist (is infinite) and the divergence criterion is violated. Therefore, because of Equation (2.38), the probability that an iid sample drawn from a $\mathscr{G}_A^0$ law satisfies the divergence criterion (Equation (2.13)) goes to zero as $N \to \infty$.

Second, since $\hat{F}$ converges to $F$ almost surely as $N \to \infty$, the probability of a bootstrap sample satisfying the divergence criterion also goes to zero. In this case we recover the result that for $N \to \infty$, and $B \to \infty$, $P^* \to P^0$ and our estimator is asymptotically equivalent to MLE which is consistent.

CHAPTER 3

# A Monte Carlo Study of the MLE, Firth, and Bootstrap Estimators

## 3.1 Numerical results using Firth's bias corrections

To evaluate the corrections described in Sections 2.4.3 and 2.4.4, we performed a Monte Carlo experiment with 32 sub-experiments where for each triple, $(n, \alpha, N)$ with $n \in \{1,2,3,8\}$, $\alpha \in \{-15,-5,-3,-1\}$, and $N \in \{49,121\}$, we generated 10000 samples with $N$ observations from a $\mathscr{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ chosen as in Equation (2.7). For each sample, parameter estimates for $\alpha$ and $\gamma$ were obtained using standard MLE, MLE with the Jeffreys invariant prior calculated with the expected and observed information matrix, and with Firth's score modification using the expected information (four estimation schemes). The first three estimations were performed using the BFGS algorithm (Press *et al.*, 1992) in Ox (Doornik, 2001). The last estimation was performed by using Ox's non-linear equation solver, since we did not have a closed form for the log-likelihood.

The results in Table 3.1 show that the percentage of samples which satisfy the divergence criteria is only slightly larger than the percentage of times the BFGS algorithm did not converge strongly while maximizing the unadjusted log-likelihood. Unexpectedly, there is a slight increase in the number of failures for simple MLE as $\alpha$ changes from $-3$ to $-1$, even though the fraction of diverging samples consistently decreases. This same behavior is encountered for all the other likelihoods as well (except perhaps when $n = 8$). Maybe we need to expand the likelihood for small values of the parameters to understand what is happening.

The use of the Jeffreys invariant prior with expected information has the largest effect on the failure to converge strongly. In nearly all cases BFGS converged strongly with this likelihood.

The exceptions correspond to the strange behavior commented upon in the previous paragraph. Firth's correction with the expected information had the worst strong convergence properties after uncorrected MLE. As this log-likelihood was maximized using the non-linear solver code, we tested the effect of using this solver on all the other log-likelihoods. The results were disastrous. Non-strong-convergence rates rose significantly for all log-likelihoods. The non-linear solver's success is strongly dependent on the initial starting point. The presented results were obtained by using the solution from the maximization of the Jeffreys invariant prior with expected information as the starting point for the non-linear solver. Strong convergence rates for the estimator with the Jeffreys invariant prior with observed information were quite poor, especially for small values of $n$.

It is of interest to consider the rates of weak convergence as well. These results are presented in Table 3.2. Comparison with the previous table reveals that the Jeffreys with expected and observed information either converge strongly or not at all. The MLE and Firth estimator with expected information have many situations in which weak, but not strong convergence is obtained. Jeffreys with observed information and unadjusted MLE have high rates of non-convergence.

In Tables 3.3 and 3.4 we encounter the estimated bias for each estimator of $\alpha$ calculated using all samples, and using only those samples for which the divergence criteria is not violated, respectively. The improvement for the unadjusted log-likelihood is enormous when we ignore the diverging samples. The Jeffreys prior with expected information is nearly unaffected when diverging samples are excluded, indicating the robustness of this estimator. Firth's estimator with expected information tends to improve its bias when the diverging samples are excluded, suggesting that the correction has not fully resolved the monotone likelihood problem. The Jeffreys prior with observed information is minimally affected by exclusion of the diverging samples. This is somewhat surprising since the correction depends on the data, but does not seem to be influenced by monotone likelihood. In terms of the absolute value of the bias (and not its change with the exclusion of diverging samples), Table 3.3 and Figure 3.1 do not suggest that any of the three modifications is consistently less biased. The biases for the Jeffreys prior based on expected information and Firth's correction seem to consistently have opposite signs

and similar magnitudes (except when $\alpha = -1$, which may be a result of the non-convergence of the Firth estimator for this parameter value).

Now we consider the mean squared error (MSE) of the estimators of $\alpha$ in Tables 3.5 and 3.6, once again, with and without the diverging samples, respectively. The inclusion of the diverging samples wreaks havoc on the MSE of the unadjusted MLE and on Firth's correction. The MSE of the Jeffreys prior with expected information is hardly affected, whereas that with observed information is minimally affected in the opposite direction from what one would expect. The MSE decreases, albeit by a small amount, when the diverging samples are included. By examining Figure 3.3 one can see that the MSE of Jeffreys with observed information is smaller than that with expected information when $\alpha = -15$, larger for values of $\alpha$ between $-15$ and $-1$ and equal for $\alpha = -1$ and $N = 121$. There is no clear best choice between these two estimators since one does not know ahead of time what parameter value one is estimating. However, for $n = 8$ (which is known information) the Jeffreys prior with expected information is clearly better than all other alternatives.

Tables 3.7 and 3.8 present similar results for the bias and MSE of the estimators of $\gamma$ calculated for all samples. Once again the Jeffreys prior with expected information demonstrates superior MSE when $n = 8$. The MSE for Firth's correction with expected information explodes for $\alpha = -1$ when $n < 8$.

In the absence of further evidence, the only clear "best" estimator is the Jeffreys prior with expected information when $n = 8$. As was clear from the tables and figures, in almost all cases as the sample size increases, the estimators improved their bias and MSE. In the next section we implement and examine the estimator based on resampling.

## 3.2 Numerical results for the bootstrap estimator

In this section we report the results of the same Monte Carlo experiment described in Section 3.1 for the bootstrap estimator with $10 * N$ bootstrap samples.

In Table 3.1 one can see that monotone likelihood has practically no effect on whether or not the bootstrap estimator converges strongly or not. The only noticeable trend is that which was commented in Section 3.1 when $\alpha = -1$. The bootstrap estimator seems to handle the divergence problem. Also, comparison with Table 3.2 shows that the bootstrap estimator either converges strongly or not at all.

Comparison of the last columns of Tables 3.3 and 3.4 shows that the bias of the bootstrap estimator is minimally affected by the inclusion of diverging samples. There is a slight downward shift in the bias (as is to be expected), however, this significantly reduces the magnitude of the bias when $\alpha = -15$ and increases it only slightly for other values of $\alpha$. Considering only Table 3.3 and Figure 3.1 one can see that the bootstrap estimator's bias is consistently among the smallest biases for all values of $\alpha$ (except when $n = 8$). Figure 3.2 shows a direct comparison between the Jeffreys prior with expected information and the bootstrap estimator. They are comparable, but the Jeffreys prior with expected information continues its dominance for $n = 8$, whereas the bootstrap estimator dominates when $\alpha = -15$.

Comparing the last column of Tables 3.5 and 3.6, one sees that the bootstrap estimator's MSE is inflated by the inclusion of diverging samples. This inflation is worst when $\alpha = -15$, causing the estimator to become imprecise. Even so, one can see in Figure 3.3, that for $n = 1$ it compares favorably to the Jeffreys prior with observed information and is no worse or better than the Jeffreys prior with expected information. For $n = 2$ it is generally between the priors. For larger values of $n$ its large MSE when $\alpha = -15$ make it an unlikely candidate for use. The only two estimators for which the form of the correction to be applied depends on the data are the Jeffreys prior with observed information and the bootstrap estimator. Both of these estimators show an increase of MSE as $N$ goes from 49 to 121 with $\alpha = -15$ and $n = 3$.

Concerning estimation of $\gamma$, Tables 3.7 and 3.8 show that the bias of the bootstrap estimator is comparable with that of the Jeffreys prior with observed information. Examination of the MSE shows a slight inefficiency relative to the prior with observed information when $\alpha = -15$ and an efficiency gain for all other values of $\alpha$. However, the Jeffreys prior with expected information continues dominant when $n = 8$.

**Table 3.1** Percentage of times that the sample of size $N$ drawn from a $\mathscr{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ as in (2.7): (Div.)— satisfied the divergence criteria; (MLE, Jeffreys with Expected Information, Firth with Expected Information, Jeffreys with Observed Information, Bootstrap with $10*N$ replications)— failed to converge strongly.

| Id | $n$ | $\alpha$ | $N$ | Div. | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | 0.462 | 0.455 | 0.000 | 0.142 | 0.089 | 0.000 |
| 2 | 1 | −15 | 121 | 0.327 | 0.323 | 0.000 | 0.204 | 0.052 | 0.000 |
| 3 | 1 | −5 | 49 | 0.219 | 0.217 | 0.001 | 0.044 | 0.142 | 0.001 |
| 4 | 1 | −5 | 121 | 0.063 | 0.061 | 0.000 | 0.030 | 0.077 | 0.000 |
| 5 | 1 | −3 | 49 | 0.088 | 0.089 | 0.004 | 0.014 | 0.152 | 0.002 |
| 6 | 1 | −3 | 121 | 0.006 | 0.007 | 0.000 | 0.003 | 0.040 | 0.001 |
| 7 | 1 | −1 | 49 | 0.000 | 0.019 | 0.021 | 0.014 | 0.034 | 0.021 |
| 8 | 1 | −1 | 121 | 0.000 | 0.018 | 0.016 | 0.013 | 0.018 | 0.012 |
| 9 | 2 | −15 | 49 | 0.333 | 0.330 | 0.000 | 0.180 | 0.045 | 0.000 |
| 10 | 2 | −15 | 121 | 0.174 | 0.171 | 0.000 | 0.144 | 0.021 | 0.000 |
| 11 | 2 | −5 | 49 | 0.065 | 0.065 | 0.000 | 0.026 | 0.073 | 0.000 |
| 12 | 2 | −5 | 121 | 0.004 | 0.004 | 0.000 | 0.003 | 0.030 | 0.000 |
| 13 | 2 | −3 | 49 | 0.011 | 0.011 | 0.000 | 0.003 | 0.044 | 0.001 |
| 14 | 2 | −3 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.003 | 0.000 |
| 15 | 2 | −1 | 49 | 0.000 | 0.018 | 0.020 | 0.015 | 0.019 | 0.022 |
| 16 | 2 | −1 | 121 | 0.000 | 0.016 | 0.015 | 0.010 | 0.016 | 0.015 |
| 17 | 3 | −15 | 49 | 0.239 | 0.238 | 0.000 | 0.155 | 0.019 | 0.000 |
| 18 | 3 | −15 | 121 | 0.086 | 0.085 | 0.000 | 0.083 | 0.014 | 0.000 |
| 19 | 3 | −5 | 49 | 0.020 | 0.020 | 0.000 | 0.010 | 0.032 | 0.000 |
| 20 | 3 | −5 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.006 | 0.000 |
| 21 | 3 | −3 | 49 | 0.001 | 0.001 | 0.000 | 0.001 | 0.010 | 0.000 |
| 22 | 3 | −3 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 23 | 3 | −1 | 49 | 0.000 | 0.012 | 0.010 | 0.007 | 0.008 | 0.013 |
| 24 | 3 | −1 | 121 | 0.000 | 0.007 | 0.005 | 0.003 | 0.006 | 0.005 |
| 25 | 8 | −15 | 49 | 0.036 | 0.036 | 0.000 | 0.032 | 0.000 | 0.000 |
| 26 | 8 | −15 | 121 | 0.001 | 0.001 | 0.000 | 0.002 | 0.000 | 0.000 |
| 27 | 8 | −5 | 49 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 28 | 8 | −5 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 29 | 8 | −3 | 49 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 30 | 8 | −3 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 31 | 8 | −1 | 49 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 32 | 8 | −1 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

**Table 3.2** Percentage of times that the sample of size $N$ drawn from a $\mathscr{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ as in (2.7): (Div.)— satisfied the divergence criteria; (MLE, Jeffreys with Expected Information, Firth with Expected Information, Jeffreys with Observed Information, Bootstrap with $10*N$ replications)— failed to converge weakly.

| Id | $n$ | $\alpha$ | $N$ | Div. | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | 0.462 | 0.384 | 0.000 | 0.000 | 0.089 | 0.000 |
| 2 | 1 | −15 | 121 | 0.327 | 0.228 | 0.000 | 0.000 | 0.052 | 0.000 |
| 3 | 1 | −5 | 49 | 0.219 | 0.168 | 0.001 | 0.000 | 0.142 | 0.001 |
| 4 | 1 | −5 | 121 | 0.063 | 0.037 | 0.000 | 0.000 | 0.077 | 0.000 |
| 5 | 1 | −3 | 49 | 0.088 | 0.068 | 0.004 | 0.002 | 0.152 | 0.002 |
| 6 | 1 | −3 | 121 | 0.006 | 0.004 | 0.000 | 0.000 | 0.040 | 0.001 |
| 7 | 1 | −1 | 49 | 0.000 | 0.019 | 0.021 | 0.009 | 0.034 | 0.021 |
| 8 | 1 | −1 | 121 | 0.000 | 0.018 | 0.016 | 0.005 | 0.018 | 0.012 |
| 9 | 2 | −15 | 49 | 0.333 | 0.284 | 0.000 | 0.000 | 0.045 | 0.000 |
| 10 | 2 | −15 | 121 | 0.174 | 0.121 | 0.000 | 0.000 | 0.021 | 0.000 |
| 11 | 2 | −5 | 49 | 0.065 | 0.051 | 0.000 | 0.000 | 0.073 | 0.000 |
| 12 | 2 | −5 | 121 | 0.004 | 0.002 | 0.000 | 0.000 | 0.030 | 0.000 |
| 13 | 2 | −3 | 49 | 0.011 | 0.008 | 0.000 | 0.000 | 0.044 | 0.001 |
| 14 | 2 | −3 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.003 | 0.000 |
| 15 | 2 | −1 | 49 | 0.000 | 0.018 | 0.020 | 0.008 | 0.019 | 0.022 |
| 16 | 2 | −1 | 121 | 0.000 | 0.016 | 0.015 | 0.007 | 0.016 | 0.015 |
| 17 | 3 | −15 | 49 | 0.239 | 0.208 | 0.000 | 0.000 | 0.019 | 0.000 |
| 18 | 3 | −15 | 121 | 0.086 | 0.068 | 0.000 | 0.000 | 0.014 | 0.000 |
| 19 | 3 | −5 | 49 | 0.020 | 0.017 | 0.000 | 0.000 | 0.032 | 0.000 |
| 20 | 3 | −5 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.006 | 0.000 |
| 21 | 3 | −3 | 49 | 0.001 | 0.001 | 0.000 | 0.000 | 0.010 | 0.000 |
| 22 | 3 | −3 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 23 | 3 | −1 | 49 | 0.000 | 0.012 | 0.010 | 0.005 | 0.008 | 0.013 |
| 24 | 3 | −1 | 121 | 0.000 | 0.007 | 0.005 | 0.003 | 0.006 | 0.005 |
| 25 | 8 | −15 | 49 | 0.036 | 0.033 | 0.000 | 0.000 | 0.000 | 0.000 |
| 26 | 8 | −15 | 121 | 0.001 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| 27 | 8 | −5 | 49 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 28 | 8 | −5 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 29 | 8 | −3 | 49 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 30 | 8 | −3 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 31 | 8 | −1 | 49 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 32 | 8 | −1 | 121 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

**Table 3.3** Bias of $\hat{\alpha}$ for each of the estimation methods with 10000 samples of size $N$ drawn from a $\mathcal{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ as in (2.7).(MLE, Jeffreys with expected information, Firth with expected information, Jeffreys with observed information, bootstrap with $10*N$ replications).

| Id | $n$ | $\alpha$ | $N$ | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | −1.70e+06 | 1.24e+01 | −2.81e−03 | 8.77e+00 | 6.17e+00 |
| 2 | 1 | −15 | 121 | −8.70e+05 | 1.06e+01 | −1.05e+01 | 5.11e+00 | 2.68e+00 |
| 3 | 1 | −5 | 49 | −1.16e+06 | 2.79e+00 | −2.66e+00 | −7.24e−01 | −7.78e−01 |
| 4 | 1 | −5 | 121 | −1.17e+05 | 1.77e+00 | −2.46e+00 | −1.56e+00 | −1.25e+00 |
| 5 | 1 | −3 | 49 | −2.53e+05 | 1.15e+00 | −1.24e+00 | −1.85e+00 | −1.07e+00 |
| 6 | 1 | −3 | 121 | −1.77e+04 | 5.77e−01 | −4.80e−01 | −8.63e−01 | −7.98e−01 |
| 7 | 1 | −1 | 49 | −8.40e+02 | 6.59e−02 | −2.17e+00 | −3.22e−01 | −2.10e−01 |
| 8 | 1 | −1 | 121 | −8.78e−02 | 1.26e−02 | −2.45e+00 | 2.66e−03 | −9.15e−02 |
| 9 | 2 | −15 | 49 | −1.56e+06 | 1.06e+01 | −1.06e+01 | 4.79e+00 | 2.00e+00 |
| 10 | 2 | −15 | 121 | −4.75e+05 | 8.09e+00 | −1.23e+01 | −5.33e−01 | −6.31e−01 |
| 11 | 2 | −5 | 49 | −2.22e+05 | 1.85e+00 | −2.62e+00 | −1.77e+00 | −1.57e+00 |
| 12 | 2 | −5 | 121 | −7.19e+03 | 9.08e−01 | −7.70e−01 | −2.03e+00 | −1.09e+00 |
| 13 | 2 | −3 | 49 | −3.29e+04 | 6.39e−01 | −7.07e−01 | −1.11e+00 | −8.59e−01 |
| 14 | 2 | −3 | 121 | −3.83e+02 | 2.58e−01 | −1.24e−01 | −3.24e−01 | −3.38e−01 |
| 15 | 2 | −1 | 49 | −1.59e−01 | −5.58e−03 | −5.20e+00 | −1.05e−02 | −1.63e−01 |
| 16 | 2 | −1 | 121 | −6.67e−02 | −1.12e−02 | −3.50e+00 | −1.83e−02 | −7.66e−02 |
| 17 | 3 | −15 | 49 | −1.19e+06 | 9.16e+00 | −1.50e+01 | 1.37e+00 | −2.75e−01 |
| 18 | 3 | −15 | 121 | −2.38e+05 | 6.30e+00 | −1.06e+01 | −9.59e−01 | −2.79e+00 |
| 19 | 3 | −5 | 49 | −7.52e+04 | 1.28e+00 | −2.01e+00 | −2.28e+00 | −1.57e+00 |
| 20 | 3 | −5 | 121 | −1.10e+00 | 5.49e−01 | −3.12e−01 | −1.94e+00 | −8.51e−01 |
| 21 | 3 | −3 | 49 | −3.75e+03 | 4.07e−01 | −2.76e−01 | −7.61e−01 | −7.07e−01 |
| 22 | 3 | −3 | 121 | −2.42e−01 | 1.51e−01 | −7.00e−02 | −2.43e−01 | −2.45e−01 |
| 23 | 3 | −1 | 49 | −1.17e−01 | 6.18e−03 | −1.74e+00 | −7.14e−03 | −1.25e−01 |
| 24 | 3 | −1 | 121 | −4.55e−02 | 2.73e−03 | −1.39e+00 | −8.23e−03 | −2.82e−02 |
| 25 | 8 | −15 | 49 | −2.22e+05 | 4.81e+00 | −7.59e+00 | −3.11e+00 | −4.42e+00 |
| 26 | 8 | −15 | 121 | −2.99e+03 | 2.30e+00 | −1.83e+00 | −2.76e+00 | −3.11e+00 |
| 27 | 8 | −5 | 49 | −3.34e+01 | 4.34e−01 | −1.92e−01 | −2.73e+00 | −8.08e−01 |
| 28 | 8 | −5 | 121 | −3.04e−01 | 1.40e−01 | −8.84e−02 | −9.60e−01 | −2.58e−01 |
| 29 | 8 | −3 | 49 | −3.42e−01 | 1.18e−01 | −6.98e−02 | −2.29e−01 | −2.99e−01 |
| 30 | 8 | −3 | 121 | −1.12e−01 | 5.46e−02 | −1.35e−02 | −3.34e−02 | −1.54e−01 |
| 31 | 8 | −1 | 49 | −6.52e−02 | 9.82e−03 | −1.13e−01 | 7.52e−05 | −5.87e−02 |
| 32 | 8 | −1 | 121 | −2.59e−02 | 3.41e−03 | −3.61e−03 | 2.69e−03 | −2.72e−02 |

**Table 3.4** Same as Table 3.3 except only data for which the divergence criteria is not satisfied was used to calculate the bias of $\hat{\alpha}$.

| Id | $n$ | $\alpha$ | $N$ | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | −1.46e+02 | 1.28e+01 | 1.04e+01 | 9.08e+00 | 9.46e+00 |
| 2 | 1 | −15 | 121 | −2.44e+02 | 1.12e+01 | 6.85e+00 | 6.54e+00 | 6.48e+00 |
| 3 | 1 | −5 | 49 | −9.84e+01 | 3.01e+00 | 1.47e+00 | −6.01e−01 | 6.28e−01 |
| 4 | 1 | −5 | 121 | −5.18e+01 | 1.91e+00 | −4.37e−02 | −1.21e+00 | −6.29e−01 |
| 5 | 1 | −3 | 49 | −4.12e+01 | 1.24e+00 | −4.16e−02 | −1.76e+00 | −5.03e−01 |
| 6 | 1 | −3 | 121 | −1.63e+01 | 5.94e−01 | −2.71e−01 | −8.16e−01 | −6.74e−01 |
| 7 | 1 | −1 | 49 | −2.53e−01 | 6.67e−02 | −2.16e+00 | −3.21e−01 | −2.05e−01 |
| 8 | 1 | −1 | 121 | −8.78e−02 | 1.26e−02 | −2.45e+00 | 2.66e−03 | −9.15e−02 |
| 9 | 2 | −15 | 49 | −2.09e+02 | 1.12e+01 | 7.20e+00 | 6.35e+00 | 6.26e+00 |
| 10 | 2 | −15 | 121 | −1.45e+02 | 8.72e+00 | 1.55e+00 | 1.02e+00 | 2.20e+00 |
| 11 | 2 | −5 | 49 | −6.15e+01 | 2.01e+00 | 9.60e−02 | −1.38e+00 | −7.56e−01 |
| 12 | 2 | −5 | 121 | −3.34e+00 | 9.27e−01 | −5.25e−01 | −1.98e+00 | −1.04e+00 |
| 13 | 2 | −3 | 49 | −2.36e+00 | 6.69e−01 | −2.14e−01 | −1.03e+00 | −7.29e−01 |
| 14 | 2 | −3 | 121 | −3.67e−01 | 2.60e−01 | −1.06e−01 | −3.18e−01 | −3.38e−01 |
| 15 | 2 | −1 | 49 | −1.59e−01 | −5.58e−03 | −5.20e+00 | −1.05e−02 | −1.63e−01 |
| 16 | 2 | −1 | 121 | −6.67e−02 | −1.12e−02 | −3.50e+00 | −1.83e−02 | −7.66e−02 |
| 17 | 3 | −15 | 49 | −1.43e+02 | 9.87e+00 | 4.23e+00 | 3.13e+00 | 3.66e+00 |
| 18 | 3 | −15 | 121 | −1.31e+02 | 6.81e+00 | −1.39e+00 | 1.44e−01 | −5.46e−01 |
| 19 | 3 | −5 | 49 | −6.02e+00 | 1.36e+00 | −4.19e−01 | −2.06e+00 | −1.23e+00 |
| 20 | 3 | −5 | 121 | −1.10e+00 | 5.49e−01 | −3.12e−01 | −1.94e+00 | −8.51e−01 |
| 21 | 3 | −3 | 49 | −1.05e+00 | 4.13e−01 | −1.95e−01 | −7.44e−01 | −6.82e−01 |
| 22 | 3 | −3 | 121 | −2.42e−01 | 1.51e−01 | −7.00e−02 | −2.43e−01 | −2.45e−01 |
| 23 | 3 | −1 | 49 | −1.17e−01 | 6.18e−03 | −1.74e+00 | −7.14e−03 | −1.25e−01 |
| 24 | 3 | −1 | 121 | −4.55e−02 | 2.73e−03 | −1.39e+00 | −8.23e−03 | −2.82e−02 |
| 25 | 8 | −15 | 49 | −2.53e+01 | 5.17e+00 | −2.02e+00 | −2.54e+00 | −2.97e+00 |
| 26 | 8 | −15 | 121 | −6.32e+00 | 2.32e+00 | −1.65e+00 | −2.74e+00 | −2.99e+00 |
| 27 | 8 | −5 | 49 | −8.12e−01 | 4.36e−01 | −1.80e−01 | −2.73e+00 | −8.01e−01 |
| 28 | 8 | −5 | 121 | −3.04e−01 | 1.40e−01 | −8.84e−02 | −9.60e−01 | −2.58e−01 |
| 29 | 8 | −3 | 49 | −3.42e−01 | 1.18e−01 | −6.98e−02 | −2.29e−01 | −2.99e−01 |
| 30 | 8 | −3 | 121 | −1.12e−01 | 5.46e−02 | −1.35e−02 | −3.34e−02 | −1.54e−01 |
| 31 | 8 | −1 | 49 | −6.52e−02 | 9.82e−03 | −1.13e−01 | 7.52e−05 | −5.87e−02 |
| 32 | 8 | −1 | 121 | −2.59e−02 | 3.41e−03 | −3.61e−03 | 2.69e−03 | −2.72e−02 |

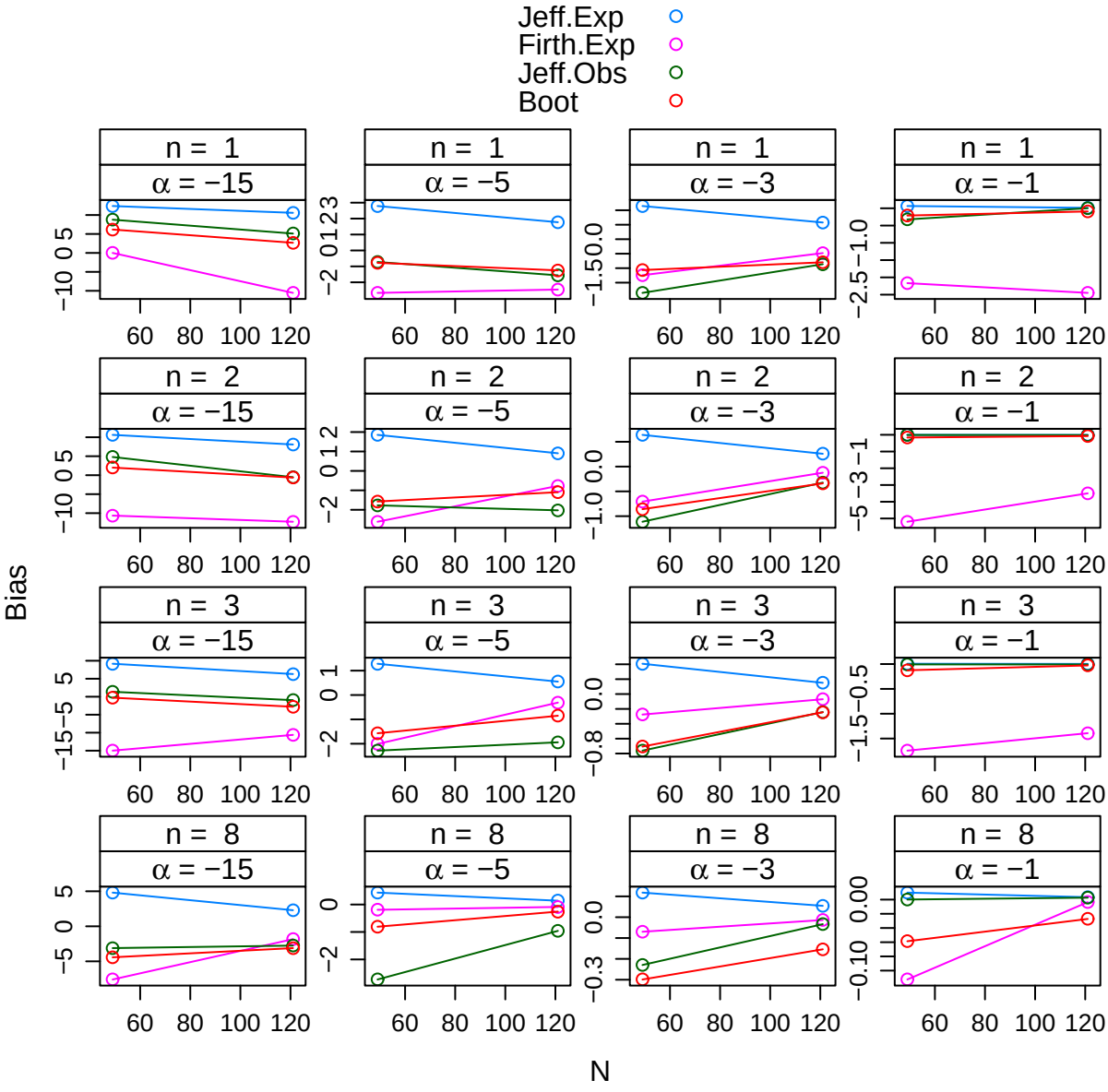


**Figure 3.1** Bias of the Jeffreys with expected and observed information, Firth with expected information, and bootstrap estimators of $\alpha$.

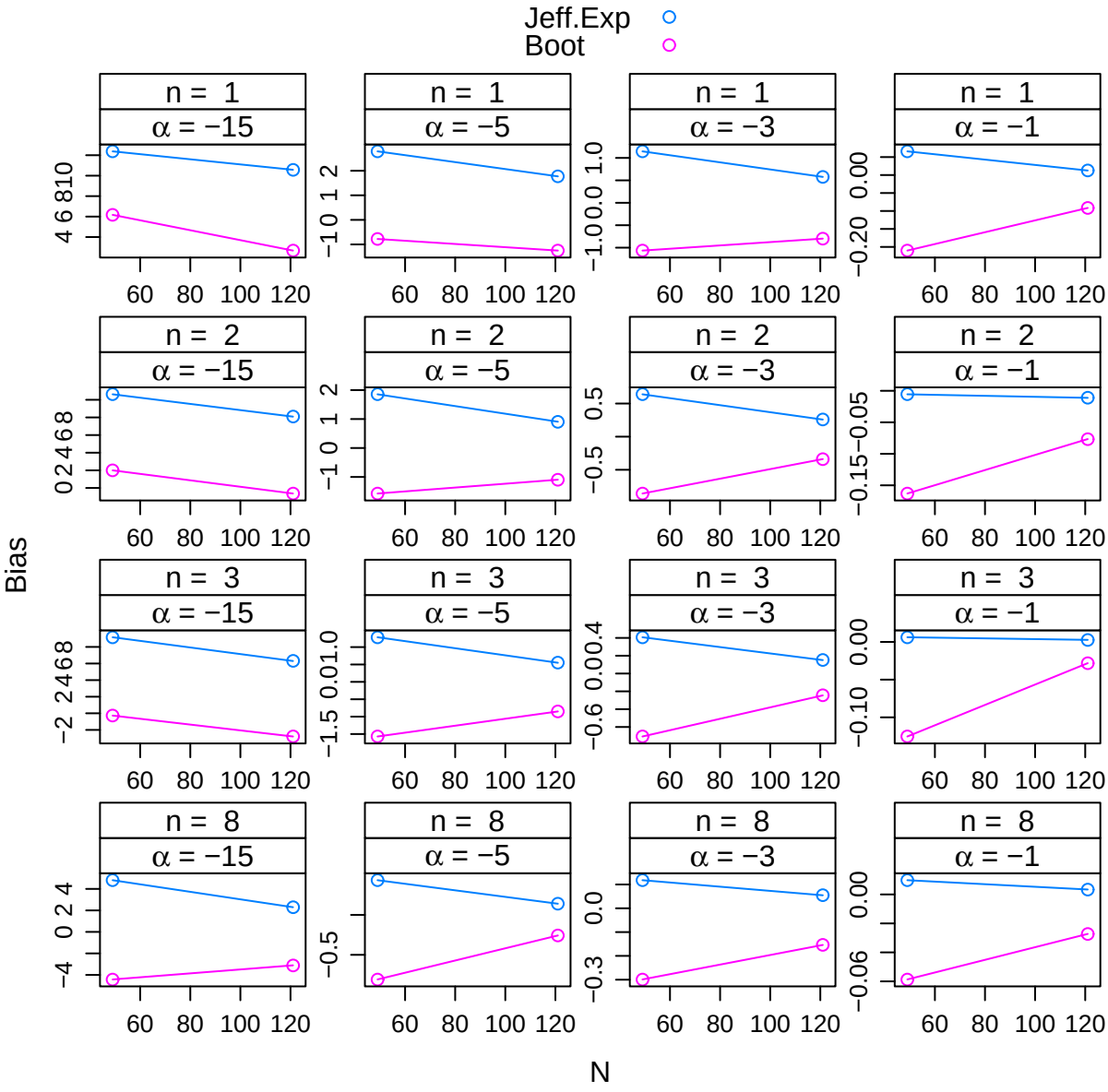


**Figure 3.2** Bias of the Jeffreys with expected information and bootstrap estimators of $\alpha$.

**Table 3.5** Mean squared error (MSE) of the estimators of $\alpha$ for each of the estimation methods for 10000 samples of size $N$ drawn from a $\mathscr{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ as in (2.7).(MLE, Jeffreys with expected information, Firth with expected information, Jeffreys with observed information, bootstrap with $10*N$ replications).

| Id | $n$ | $\alpha$ | $N$ | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | 7.73e+13 | 153.65 | 1245.68 | 90.92 | 90.12 |
| 2 | 1 | −15 | 121 | 4.33e+13 | 113.12 | 17424.90 | 40.21 | 73.61 |
| 3 | 1 | −5 | 49 | 2.10e+15 | 8.13 | 908.82 | 21.67 | 18.42 |
| 4 | 1 | −5 | 121 | 5.50e+11 | 3.93 | 148.31 | 20.33 | 14.36 |
| 5 | 1 | −3 | 49 | 3.89e+12 | 1.61 | 70.50 | 29.97 | 11.09 |
| 6 | 1 | −3 | 121 | 8.40e+11 | 0.76 | 19.22 | 11.78 | 5.60 |
| 7 | 1 | −1 | 49 | 1.65e+09 | 0.18 | 1309.51 | 9.67 | 0.46 |
| 8 | 1 | −1 | 121 | 1.44e−01 | 0.11 | 1202.05 | 0.12 | 0.14 |
| 9 | 2 | −15 | 49 | 8.59e+13 | 114.23 | 4600.91 | 39.06 | 88.75 |
| 10 | 2 | −15 | 121 | 5.32e+12 | 69.06 | 6181.81 | 26.90 | 72.75 |
| 11 | 2 | −5 | 49 | 1.48e+12 | 4.34 | 363.63 | 25.34 | 21.17 |
| 12 | 2 | −5 | 121 | 1.88e+10 | 2.10 | 28.62 | 23.87 | 11.21 |
| 13 | 2 | −3 | 49 | 1.91e+11 | 0.93 | 202.96 | 16.62 | 6.51 |
| 14 | 2 | −3 | 121 | 6.81e+08 | 0.50 | 2.14 | 3.49 | 1.27 |
| 15 | 2 | −1 | 49 | 4.04e−01 | 0.32 | 27001.69 | 0.27 | 0.41 |
| 16 | 2 | −1 | 121 | 1.41e−01 | 0.14 | 10163.64 | 0.14 | 0.16 |
| 17 | 3 | −15 | 49 | 1.62e+14 | 86.82 | 7960.03 | 24.66 | 95.47 |
| 18 | 3 | −15 | 121 | 1.05e+12 | 46.01 | 2686.39 | 42.64 | 111.16 |
| 19 | 3 | −5 | 49 | 9.65e+11 | 3.01 | 1451.32 | 26.92 | 20.12 |
| 20 | 3 | −5 | 121 | 5.44e+02 | 1.59 | 5.13 | 19.87 | 6.50 |
| 21 | 3 | −3 | 49 | 1.58e+10 | 0.80 | 9.21 | 9.44 | 4.45 |
| 22 | 3 | −3 | 121 | 7.90e−01 | 0.40 | 0.59 | 2.21 | 0.87 |
| 23 | 3 | −1 | 49 | 2.42e−01 | 0.17 | 501.76 | 0.20 | 0.28 |
| 24 | 3 | −1 | 121 | 7.31e−02 | 0.05 | 1070.88 | 0.09 | 0.04 |
| 25 | 8 | −15 | 49 | 1.51e+13 | 34.23 | 2535.83 | 72.26 | 175.53 |
| 26 | 8 | −15 | 121 | 2.93e+10 | 17.67 | 137.89 | 54.49 | 91.46 |
| 27 | 8 | −5 | 49 | 1.06e+07 | 1.97 | 5.11 | 32.31 | 6.68 |
| 28 | 8 | −5 | 121 | 1.38e+00 | 0.91 | 1.14 | 11.12 | 1.20 |
| 29 | 8 | −3 | 49 | 1.22e+00 | 0.59 | 0.83 | 3.23 | 1.06 |
| 30 | 8 | −3 | 121 | 2.93e−01 | 0.23 | 0.26 | 0.55 | 0.34 |
| 31 | 8 | −1 | 49 | 6.04e−02 | 0.06 | 62.05 | 0.06 | 0.06 |
| 32 | 8 | −1 | 121 | 1.85e−02 | 0.02 | 0.02 | 0.02 | 0.02 |

**Table 3.6** Same as Table 3.5 except only data for which the divergence criteria is not satisfied was used to calculate the MSE of $\hat{\alpha}$.

| Id | $n$ | $\alpha$ | $N$ | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | 1.94e+07 | 164.16 | 1021.55 | 106.30 | 93.15 |
| 2 | 1 | −15 | 121 | 3.36e+07 | 125.20 | 61.85 | 54.46 | 50.19 |
| 3 | 1 | −5 | 49 | 1.47e+07 | 9.20 | 13.64 | 26.79 | 3.71 |
| 4 | 1 | −5 | 121 | 6.51e+06 | 4.15 | 7.31 | 18.36 | 6.65 |
| 5 | 1 | −3 | 49 | 3.55e+06 | 1.74 | 35.18 | 31.92 | 3.14 |
| 6 | 1 | −3 | 121 | 2.18e+06 | 0.74 | 9.88 | 11.39 | 3.60 |
| 7 | 1 | −1 | 49 | 7.68e−01 | 0.18 | 1310.11 | 9.67 | 0.41 |
| 8 | 1 | −1 | 121 | 1.44e−01 | 0.11 | 1202.05 | 0.12 | 0.14 |
| 9 | 2 | −15 | 49 | 4.63e+07 | 126.96 | 65.04 | 54.02 | 49.34 |
| 10 | 2 | −15 | 121 | 2.15e+07 | 77.75 | 66.09 | 14.56 | 27.89 |
| 11 | 2 | −5 | 49 | 1.27e+07 | 4.60 | 7.19 | 22.95 | 8.45 |
| 12 | 2 | −5 | 121 | 8.16e+03 | 2.03 | 10.29 | 22.98 | 10.07 |
| 13 | 2 | −3 | 49 | 8.47e+02 | 0.88 | 6.88 | 15.89 | 4.36 |
| 14 | 2 | −3 | 121 | 1.75e+00 | 0.49 | 0.99 | 3.40 | 1.27 |
| 15 | 2 | −1 | 49 | 4.04e−01 | 0.32 | 27001.69 | 0.27 | 0.41 |
| 16 | 2 | −1 | 121 | 1.41e−01 | 0.14 | 10163.64 | 0.14 | 0.16 |
| 17 | 3 | −15 | 49 | 2.46e+07 | 98.55 | 50.36 | 21.69 | 33.94 |
| 18 | 3 | −15 | 121 | 3.08e+07 | 49.96 | 131.57 | 21.80 | 44.52 |
| 19 | 3 | −5 | 49 | 1.91e+04 | 2.92 | 10.44 | 23.70 | 12.12 |
| 20 | 3 | −5 | 121 | 5.44e+02 | 1.59 | 5.13 | 19.87 | 6.50 |
| 21 | 3 | −3 | 49 | 8.32e+01 | 0.77 | 2.65 | 9.14 | 3.81 |
| 22 | 3 | −3 | 121 | 7.90e−01 | 0.40 | 0.59 | 2.21 | 0.87 |
| 23 | 3 | −1 | 49 | 2.42e−01 | 0.17 | 501.76 | 0.20 | 0.28 |
| 24 | 3 | −1 | 121 | 7.31e−02 | 0.05 | 1070.88 | 0.09 | 0.04 |
| 25 | 8 | −15 | 49 | 1.91e+05 | 34.43 | 152.41 | 55.13 | 91.91 |
| 26 | 8 | −15 | 121 | 6.21e+04 | 17.37 | 102.14 | 54.20 | 83.52 |
| 27 | 8 | −5 | 49 | 6.69e+00 | 1.95 | 3.81 | 32.12 | 6.19 |
| 28 | 8 | −5 | 121 | 1.38e+00 | 0.91 | 1.14 | 11.12 | 1.20 |
| 29 | 8 | −3 | 49 | 1.22e+00 | 0.59 | 0.83 | 3.23 | 1.06 |
| 30 | 8 | −3 | 121 | 2.93e−01 | 0.23 | 0.26 | 0.55 | 0.34 |
| 31 | 8 | −1 | 49 | 6.04e−02 | 0.06 | 62.05 | 0.06 | 0.06 |
| 32 | 8 | −1 | 121 | 1.85e−02 | 0.02 | 0.02 | 0.02 | 0.02 |

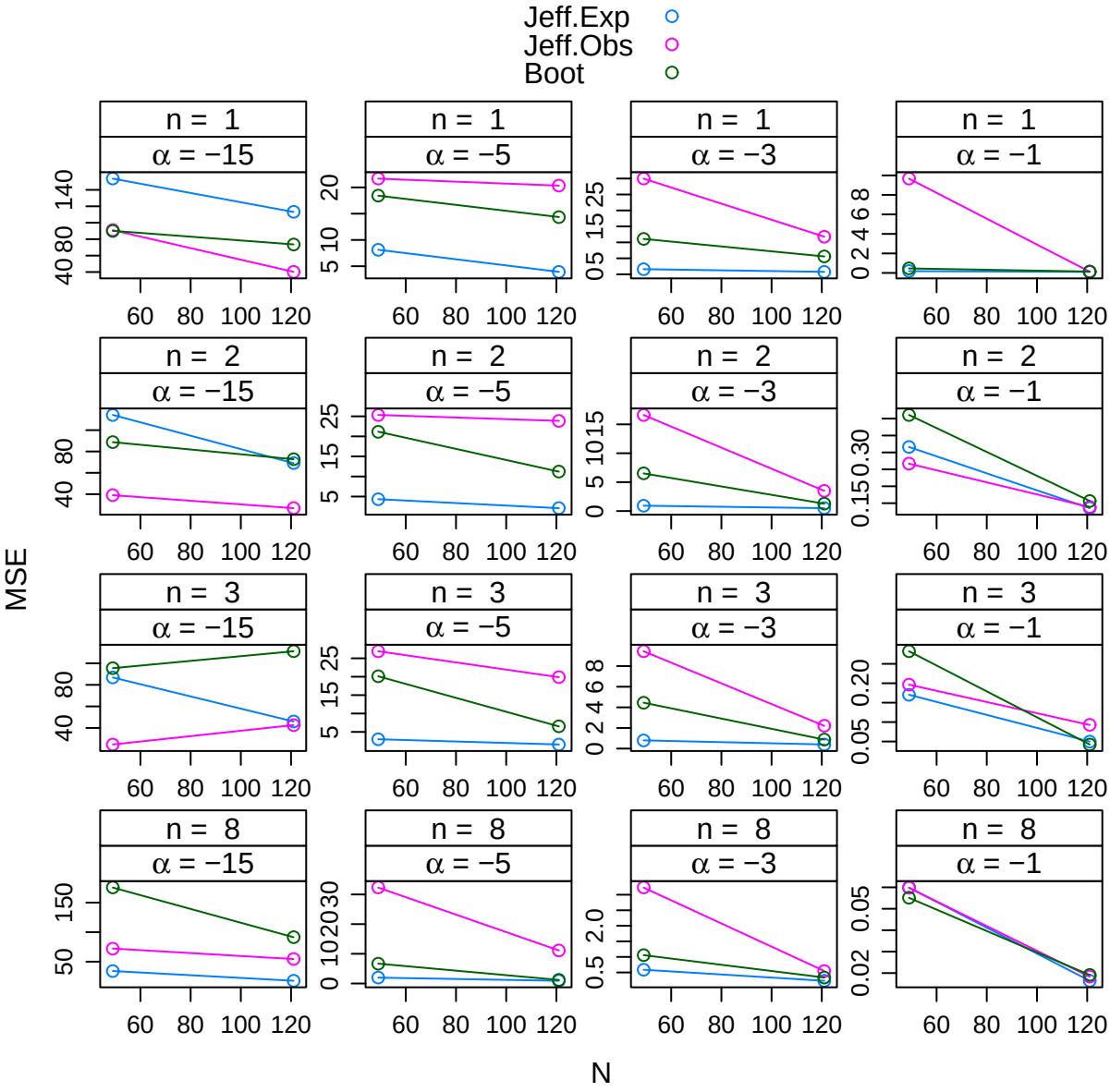


**Figure 3.3** MSE of the Jeffreys with expected and observed information and bootstrap estimators of $\alpha$.

**Table 3.7** Bias of the estimators of $\gamma$ for each of the estimation methods with 1000 samples of size $N$ drawn from a $\mathscr{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ as in (2.7).(MLE, Jeffreys with Expected Information, Firth with Expected Information, Jeffreys with Observed Information, Bootstrap with $10*N$ replications).

| Id | $n$ | $\alpha$ | $N$ | $\gamma$ | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | 18.15 | 2.16e+06 | −1.56e+01 | 8.87e−01 | −9.77e+00 | −9.08e+00 |
| 2 | 1 | −15 | 121 | 18.15 | 1.10e+06 | −1.35e+01 | 1.41e+01 | −4.96e+00 | −4.13e+00 |
| 3 | 1 | −5 | 49 | 5.42 | 1.53e+06 | −3.44e+00 | 3.68e+00 | 1.66e+00 | 5.95e−01 |
| 4 | 1 | −5 | 121 | 5.42 | 1.47e+05 | −2.22e+00 | 3.27e+00 | 2.56e+00 | 1.61e+00 |
| 5 | 1 | −3 | 49 | 2.88 | 3.00e+05 | −1.35e+00 | 1.82e+00 | 2.61e+00 | 1.18e+00 |
| 6 | 1 | −3 | 121 | 2.88 | 2.03e+04 | −6.96e−01 | 6.85e−01 | 1.30e+00 | 9.84e−01 |
| 7 | 1 | −1 | 49 | 0.41 | 6.04e+02 | −3.59e−02 | 4.26e+00 | 2.74e−01 | 1.71e−01 |
| 8 | 1 | −1 | 121 | 0.41 | 7.68e−02 | −2.22e−05 | 4.75e+00 | 1.40e−02 | 6.91e−02 |
| 9 | 2 | −15 | 49 | 16.13 | 1.73e+06 | −1.19e+01 | 1.29e+01 | −3.69e+00 | −3.01e+00 |
| 10 | 2 | −15 | 121 | 16.13 | 5.31e+05 | −9.15e+00 | 1.42e+01 | 2.17e+00 | 5.89e−01 |
| 11 | 2 | −5 | 49 | 4.82 | 2.41e+05 | −2.04e+00 | 3.05e+00 | 2.57e+00 | 1.67e+00 |
| 12 | 2 | −5 | 121 | 4.82 | 8.06e+03 | −1.01e+00 | 9.15e−01 | 2.63e+00 | 1.20e+00 |
| 13 | 2 | −3 | 49 | 2.56 | 3.47e+04 | −6.59e−01 | 8.84e−01 | 1.46e+00 | 8.82e−01 |
| 14 | 2 | −3 | 121 | 2.56 | 3.99e+02 | −2.75e−01 | 1.55e−01 | 4.60e−01 | 3.80e−01 |
| 15 | 2 | −1 | 49 | 0.36 | 1.02e−01 | 8.76e−03 | 7.27e+00 | 1.99e−02 | 1.04e−01 |
| 16 | 2 | −1 | 121 | 0.36 | 4.36e−02 | 1.00e−02 | 5.09e+00 | 1.79e−02 | 4.54e−02 |
| 17 | 3 | −15 | 49 | 15.48 | 1.27e+06 | −9.89e+00 | 1.68e+01 | 3.39e−01 | −1.10e−01 |
| 18 | 3 | −15 | 121 | 15.48 | 2.57e+05 | −6.84e+00 | 1.18e+01 | 2.09e+00 | 3.06e+00 |
| 19 | 3 | −5 | 49 | 4.63 | 7.92e+04 | −1.34e+00 | 2.24e+00 | 2.97e+00 | 1.64e+00 |
| 20 | 3 | −5 | 121 | 4.63 | 1.14e+00 | −5.83e−01 | 3.60e−01 | 2.42e+00 | 8.83e−01 |
| 21 | 3 | −3 | 49 | 2.46 | 3.59e+03 | −4.02e−01 | 3.16e−01 | 9.59e−01 | 7.20e−01 |
| 22 | 3 | −3 | 121 | 2.46 | 2.40e−01 | −1.52e−01 | 8.43e−02 | 3.20e−01 | 2.54e−01 |
| 23 | 3 | −1 | 49 | 0.35 | 6.68e−02 | −7.38e−04 | 1.79e+00 | 1.31e−02 | 7.14e−02 |
| 24 | 3 | −1 | 121 | 0.35 | 2.57e−02 | −9.86e−04 | 1.61e+00 | 7.86e−03 | 1.91e−02 |
| 25 | 8 | −15 | 49 | 14.70 | 2.28e+05 | −4.92e+00 | 7.92e+00 | 4.08e+00 | 4.48e+00 |
| 26 | 8 | −15 | 121 | 14.70 | 3.01e+03 | −2.36e+00 | 1.93e+00 | 3.11e+00 | 3.18e+00 |
| 27 | 8 | −5 | 49 | 4.39 | 3.18e+01 | −4.18e−01 | 2.19e−01 | 3.10e+00 | 7.95e−01 |
| 28 | 8 | −5 | 121 | 4.39 | 3.00e−01 | −1.34e−01 | 1.00e−01 | 1.04e+00 | 2.38e−01 |
| 29 | 8 | −3 | 49 | 2.34 | 3.08e−01 | −1.09e−01 | 7.52e−02 | 2.69e−01 | 2.75e−01 |
| 30 | 8 | −3 | 121 | 2.34 | 1.03e−01 | −4.81e−02 | 1.96e−02 | 4.66e−02 | 1.24e−01 |
| 31 | 8 | −1 | 49 | 0.33 | 3.08e−02 | −5.59e−03 | 7.41e−02 | 3.37e−03 | 2.76e−02 |
| 32 | 8 | −1 | 121 | 0.33 | 1.20e−02 | −2.18e−03 | 2.09e−03 | −1.55e−03 | 1.48e−02 |

**Table 3.8** Mean squared error (MSE) of the estimators of $\gamma$ for each of the estimation methods for 1000 samples of size $N$ drawn from a $\mathscr{G}_A^0(\alpha,\gamma,n)$ distribution with $\gamma$ as in (2.7).(MLE, Jeffreys with Expected Information, Firth with Expected Information, Jeffreys with Observed Information, Bootstrap with $10*N$ replications).

| Id | $n$ | $\alpha$ | $N$ | $\gamma$ | MLE | Jeff. Exp. | Firth Exp. | Jeff. Obs. | Boot |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −15 | 49 | 18.146 | 1.482e+14 | 245.603 | 1647.249 | 118.579 | 113.521 |
| 2 | 1 | −15 | 121 | 18.146 | 7.240e+13 | 183.686 | 28520.138 | 55.220 | 74.049 |
| 3 | 1 | −5 | 49 | 5.421 | 4.300e+15 | 12.452 | 1074.292 | 34.928 | 16.714 |
| 4 | 1 | −5 | 121 | 5.421 | 9.174e+11 | 6.235 | 245.823 | 38.147 | 22.926 |
| 5 | 1 | −3 | 49 | 2.882 | 5.171e+12 | 2.291 | 155.740 | 44.498 | 11.038 |
| 6 | 1 | −3 | 121 | 2.882 | 9.836e+11 | 1.143 | 38.130 | 19.035 | 7.813 |
| 7 | 1 | −1 | 49 | 0.405 | 8.159e+08 | 0.118 | 10983.381 | 5.248 | 0.336 |
| 8 | 1 | −1 | 121 | 0.405 | 1.474e−01 | 0.097 | 4216.242 | 0.123 | 0.119 |
| 9 | 2 | −15 | 49 | 16.130 | 1.012e+14 | 144.792 | 7032.374 | 42.434 | 77.452 |
| 10 | 2 | −15 | 121 | 16.130 | 5.906e+12 | 88.657 | 7629.381 | 49.661 | 92.028 |
| 11 | 2 | −5 | 49 | 4.818 | 1.834e+12 | 5.305 | 504.460 | 36.393 | 23.580 |
| 12 | 2 | −5 | 121 | 4.818 | 2.413e+10 | 2.624 | 36.467 | 35.038 | 13.552 |
| 13 | 2 | −3 | 49 | 2.562 | 2.343e+11 | 1.046 | 320.862 | 20.570 | 6.713 |
| 14 | 2 | −3 | 121 | 2.562 | 7.261e+08 | 0.557 | 2.423 | 4.853 | 1.592 |
| 15 | 2 | −1 | 49 | 0.360 | 1.763e−01 | 0.122 | 72888.920 | 0.123 | 0.179 |
| 16 | 2 | −1 | 121 | 0.360 | 6.568e−02 | 0.058 | 25271.553 | 0.067 | 0.069 |
| 17 | 3 | −15 | 49 | 15.485 | 1.605e+14 | 101.428 | 9729.160 | 38.406 | 93.492 |
| 18 | 3 | −15 | 121 | 15.485 | 1.209e+12 | 54.422 | 3426.698 | 66.793 | 133.474 |
| 19 | 3 | −5 | 49 | 4.626 | 1.223e+12 | 3.352 | 1870.670 | 37.952 | 21.399 |
| 20 | 3 | −5 | 121 | 4.626 | 5.370e+02 | 1.791 | 5.762 | 27.689 | 7.262 |
| 21 | 3 | −3 | 49 | 2.459 | 1.425e+10 | 0.787 | 8.927 | 10.921 | 4.358 |
| 22 | 3 | −3 | 121 | 2.459 | 7.837e−01 | 0.396 | 0.595 | 2.836 | 0.856 |
| 23 | 3 | −1 | 49 | 0.346 | 7.984e−02 | 0.055 | 559.968 | 0.072 | 0.093 |
| 24 | 3 | −1 | 121 | 0.346 | 2.573e−02 | 0.017 | 1371.225 | 0.036 | 0.015 |
| 25 | 8 | −15 | 49 | 14.704 | 1.828e+13 | 36.071 | 3037.727 | 96.794 | 178.854 |
| 26 | 8 | −15 | 121 | 14.704 | 2.974e+10 | 18.661 | 145.902 | 64.796 | 94.470 |
| 27 | 8 | −5 | 49 | 4.392 | 9.643e+06 | 1.867 | 4.832 | 38.297 | 6.266 |
| 28 | 8 | −5 | 121 | 4.392 | 1.313e+00 | 0.860 | 1.085 | 12.242 | 1.152 |
| 29 | 8 | −3 | 49 | 2.335 | 9.962e−01 | 0.478 | 0.683 | 3.342 | 0.852 |
| 30 | 8 | −3 | 121 | 2.335 | 2.401e−01 | 0.190 | 0.210 | 0.554 | 0.258 |
| 31 | 8 | −1 | 49 | 0.328 | 1.306e−02 | 0.012 | 28.160 | 0.020 | 0.011 |
| 32 | 8 | −1 | 121 | 0.328 | 3.811e−03 | 0.003 | 0.003 | 0.005 | 0.004 |

CHAPTER 4

# Application of the MLE, Firth, and Bootstrap Estimators to Real Data

## 4.1 Estimation of the equivalent number of looks

In the Monte Carlo simulations of Chapter 3 we took the number of looks, $n$, as a known parameter. When dealing with real data this parameter occasionally needs to be estimated. This may seem strange since one supposedly knows how many looks were used to form an image. However, what one knows is the nominal number of looks. If all looks were independent, the nominal number of looks would be equal to the parameter $n$ in the $\mathcal{G}_A^0$ distribution, which we call the equivalent number of looks.

When only one look is used the nominal number of looks and equivalent number of looks are both equal to one. When the nominal number of looks is greater than one, estimation of the equivalent number of looks is necessary. The amplitude return signal is $Z_A = Y_A \cdot X_A$, where $Y_A$ represents the speckle noise and $X_A$ represents the terrain backscatter. We follow Frery *et al.* (1997) and assume that $Y_A \sim \Gamma^{1/2}(n,n)$. In order to estimate $n$ from our observations of $Z_A$ we need more information about $X_A$.

We followed the procedure in Yanasse *et al.* (1994), which was to identify homogeneous regions of the image to be studied and assume that $X_A = c$ with $c$ an unknown constant. Under these assumptions the density function of $Y_A$ is given by

$$f_{Y_A}(y) = \frac{2n^n}{\Gamma(n)} y^{2n-1} \exp(-ny^2) \quad y\ ,\ n > 0.$$

We can derive the density function for $Z_A$ by noting that if $Y_A \sim f(y)$ then $Z_A = cY_A \sim \frac{1}{c} f(y/c)$

and therefore

$$f_{Z_A}(z) = \frac{2n^n}{\Gamma(n)}\frac{1}{c^{2n}}z^{2n-1}\exp(-nz^2/c^2) \quad z\ ,\ n\ ,\ c > 0.$$

With the density of $Z_A$ in hand we can derive the reduced log-likelihood for $n$ and $c$:

$$\ell(n,c) = \sum_{i=1}^{N} n\ln(n) - \ln(\Gamma(n)) - 2n\ln(c) + (2n-1)\ln(z_i) - \frac{nz_i^2}{c^2}.$$

The first order condition obtained from equating $\partial\ell/\partial c = 0$ yields $c^2 = \overline{z^2}$. If we transform the data to $z^* = z/(\overline{z^2})^{1/2}$ the solution to $\partial\ell/\partial n = 0$ is

$$\ln(n) - \psi(n) = -2\overline{\ln(z^*)}. \tag{4.1}$$

The solution of this equation for data from homogeneous regions yields the equivalent number of looks for the entire image.

## 4.2 A simple single look image

First, some notes on the figures in this chapter. Version 2.6.1 of the R computing environment (R Development Core Team, 2007) was used to perform all the analyses of the number of looks and to generate all the figures in this chapter. The amplitude data has a very large range. If one tries to plot all values using a linear scale from black to white the images appear as almost entirely black with a few small white regions. Therefore we find it necessary to clip the images. The clipping criteria used for the plots was the $75^{th}$ percentile of the amplitude. Hence all amplitude values greater than the third quartile received the value of the third quartile. Note that this was only for plotting purposes, all estimations used the unclipped data. Similarly, the estimated values for the roughness parameter, $\alpha$, range from approximately $-0.5$ to $-10^6$. A plot using a linear scale would not capture the variations between $-0.5$ to $-15$, which most interests us. Hence, all estimated values of $\alpha$ less than $-15$ are replaced by $-15$ in plots. In a final note, results from using the Jeffreys prior with observed information are not reported. This

is because the observed information matrix was not invertible for many cases and the BFGS algorithm either never started or soon failed to converge.

Figure 4.1 shows the first image analyzed. This image corresponds to the L-Band, was formed using only one look, and was encoded in byte format. Since it is in byte format, the full range of amplitudes runs from 0 to 255, hence we should expect limited amplitude variation. The image was divided into 2116 blocks of 11 pixels by 11 pixels. The 121 observations in each block were used to simultaneously determine the estimated values of $\alpha$ and $\gamma$ in that block.

The divergence criteria was satisfied for 47.5% of the blocks. As can be seen in Table 3.2, in the Monte Carlo experiment 32.7% of samples satisfied the divergence criteria for $(n,\alpha,N) = (1,-15,121)$, which implies that our image is quite homogeneous and has many blocks with $\alpha < -15$. The MLE failed to converge in more than 51% of the blocks. Even the bootstrap estimator, which converged in nearly all the Monte Carlo simulations, failed to converge in 7% of the blocks. The estimator based on the Jeffreys prior with expected information converged for all of the blocks, however, its average estimate of $\alpha$ was $-4.7$ with the lowest estimated value being $-13.19$, implying intermediate homogeneity, whereas the image is quite homogeneous as evidenced by the large percentage of blocks satisfying the divergence criteria. Hence, the Jeffreys prior with expected information estimator may be poorly suited for identifying homogeneous regions.

Figure 4.2 overlays the MLE estimates of $\alpha$ over the original image from Figure 4.1. Since our motivation was whether or not a pilot in an unknown area can safely land his plane, heterogeneous regions have been identified by red colors, while homogeneous regions are identified by white and intermediate regions by yellow. The borders between regions are mixtures of different $\mathscr{G}_A^0$ distributions and are identified as heterogeneous regions. When MLE does not converge the estimates for $\alpha$ are usually quite negative, implying homogeneity, hence the existence of many white blocks (MLE failed to converge in 51% of the blocks).

A similar figure for the bootstrap estimator can be found in Figure 4.3. Note that whereas Figure 4.2 contained many contiguous white squares (signifying homogeneous areas) many of these white squares are yellow in Figure 4.3 indicating intermediate roughness. Since we know

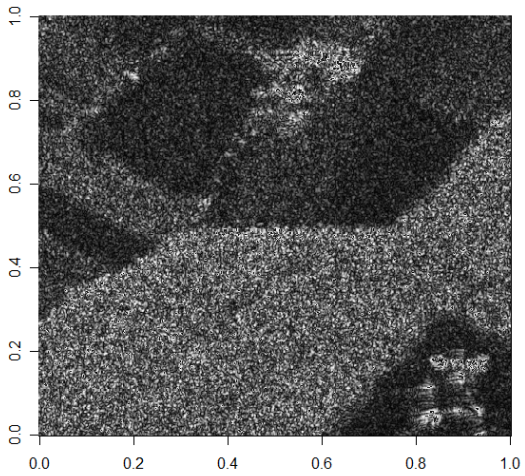


**Figure 4.1** Simple L-Band SAR image used to analyze real data. The amplitude data are in byte format and go from black (0 amplitude) to white (amplitude of 255).

the MLE estimates are biased toward homogeneity, a pilot would be wise to avoid areas with many yellow squares in this figure. Of course we can only confirm the validity of our results by comparing the true ground conditions with our estimates. Unfortunately, we do not have this information.

Finally, we include Figure 4.4 which provides an overlay of the difference between the MLE and bootstrap estimates of the roughness parameter. In all cases where there was a significant difference between the estimators (when the blocks are white), the bootstrap estimator indicated a more heterogeneous block.

## 4.3 A multi-look image with heterogeneous targets

Since the image under analysis in this section was obtained from multiple looks, we must estimate the equivalent number of looks before proceeding to the roughness estimation.

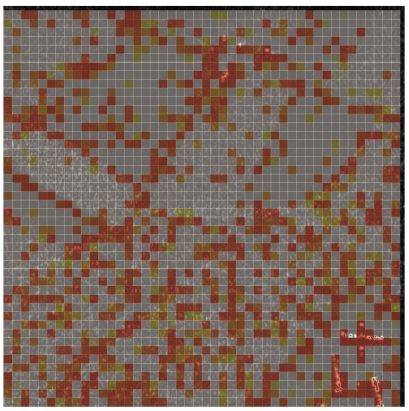

**Figure 4.2** MLE estimated $\alpha$ values for Figure 4.1. Heterogeneous areas are colored red while more homogeneous areas are white.

### 4.3.1 Estimation of the equivalent number of looks

Here, we follow the procedure described in Section 4.1. Figure 4.5 shows the image to be studied and the regions used to identify the equivalent number of looks. The image was taken over a suburb of Munich and was formed with four nominal looks. We consider the data from channel one of four. The amplitude data was stored in single precision float format and can range from zero to numbers much larger than any observed intensity, allowing much greater variation. One can immediately see the much greater complexity of this image.

The equivalent number of looks was estimated for each of the 38 regions in the figure by solving Equation (4.1). A histogram of the amplitudes in the region was plotted on the same graph as the density implied by the estimate of $n$ to ensure that the model adequately described the data. These results can be found in Figures 4.6–4.10. The last figure also contains a histogram of the estimated value of $n$. The data are very well described by the $\Gamma^{1/2}$ model. The mean value of the estimates of $n$ is 3.26 and the weighted mean using the number of observations in each region as a weight is 3.23. We chose to use $n = 3.2$ to perform our estimates of the roughness parameter.

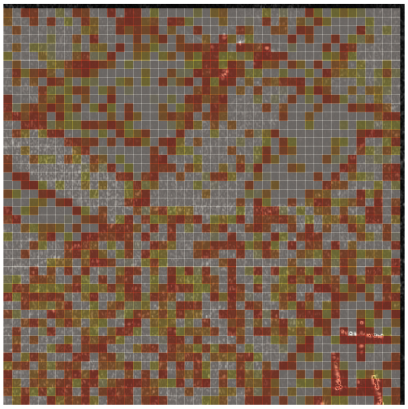

**Figure 4.3** Bootstrap estimator estimated $\alpha$ values for Figure 4.1. Heterogeneous areas are colored red while more homogeneous areas are white.

### 4.3.2 Second stage of multi-look estimation

We estimated the roughness parameter, $\alpha$, for each of 95256 blocks of 11 by 11 pixels. The results from the bootstrap estimator can be found in Figure 4.11. For this image, the divergence criteria was satisfied for approximately 9% of the blocks analyzed. In Table 3.2, 8.6% of the samples satisfy the divergence criteria for $(n, \alpha, N) = (3, -15, 121)$. Given that this image has many very heterogeneous areas, the implication is that the homogeneous areas have values of $\alpha$ much less than $-15$ and that most blocks in these areas satisfy the divergence criteria. The MLE failed to converge in more than 11% of the samples. The bootstrap estimator failed to converge in less than 4% of the blocks while the Jeffreys and Firth estimators with expected information didn't converge for 1.5% of the samples. Examination of Figure 4.11 reveals that a high intensity homogeneous area was successfully identified (near column 1000, row 100). Many of the otherwise homogeneous regions have heterogeneous paths crossing them.

Figures 4.12–4.14 compare the results from the bootstrap estimator with the MLE, Jeffreys with expected information, and Firth with expected information estimators, respectively. Once again, the bootstrap estimator tends to identify more heterogeneity than the MLE and less than

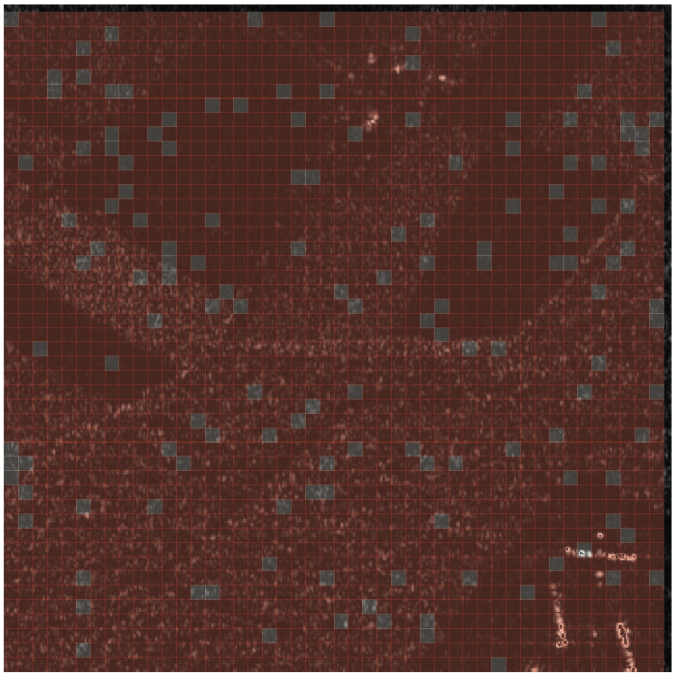

**Figure 4.4** Difference between the estimated $\alpha$ values for the bootstrap estimator and MLE for Figure 4.1. White areas represent blocks with significant differences. In all these cases the bootstrap estimator gives smaller estimates of $\alpha$ suggesting more heterogeneous blocks.

both Jeffreys and Firth with expected information. This becomes even more clear when one examines the difference between the clipped estimated values of $\alpha$ in Figures 4.15–4.17.

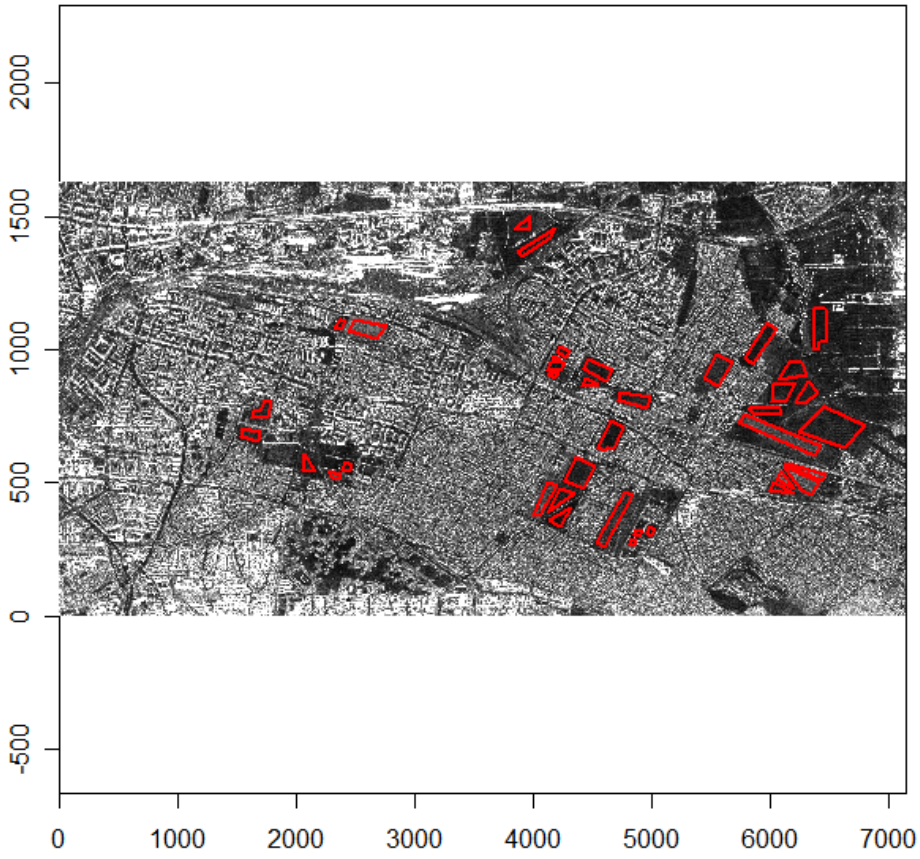


**Figure 4.5** Heterogeneous multi-look image with the regions used to estimate the number of looks marked by red polygons. White is high amplitude and black is low amplitude (DLR polarimetric).

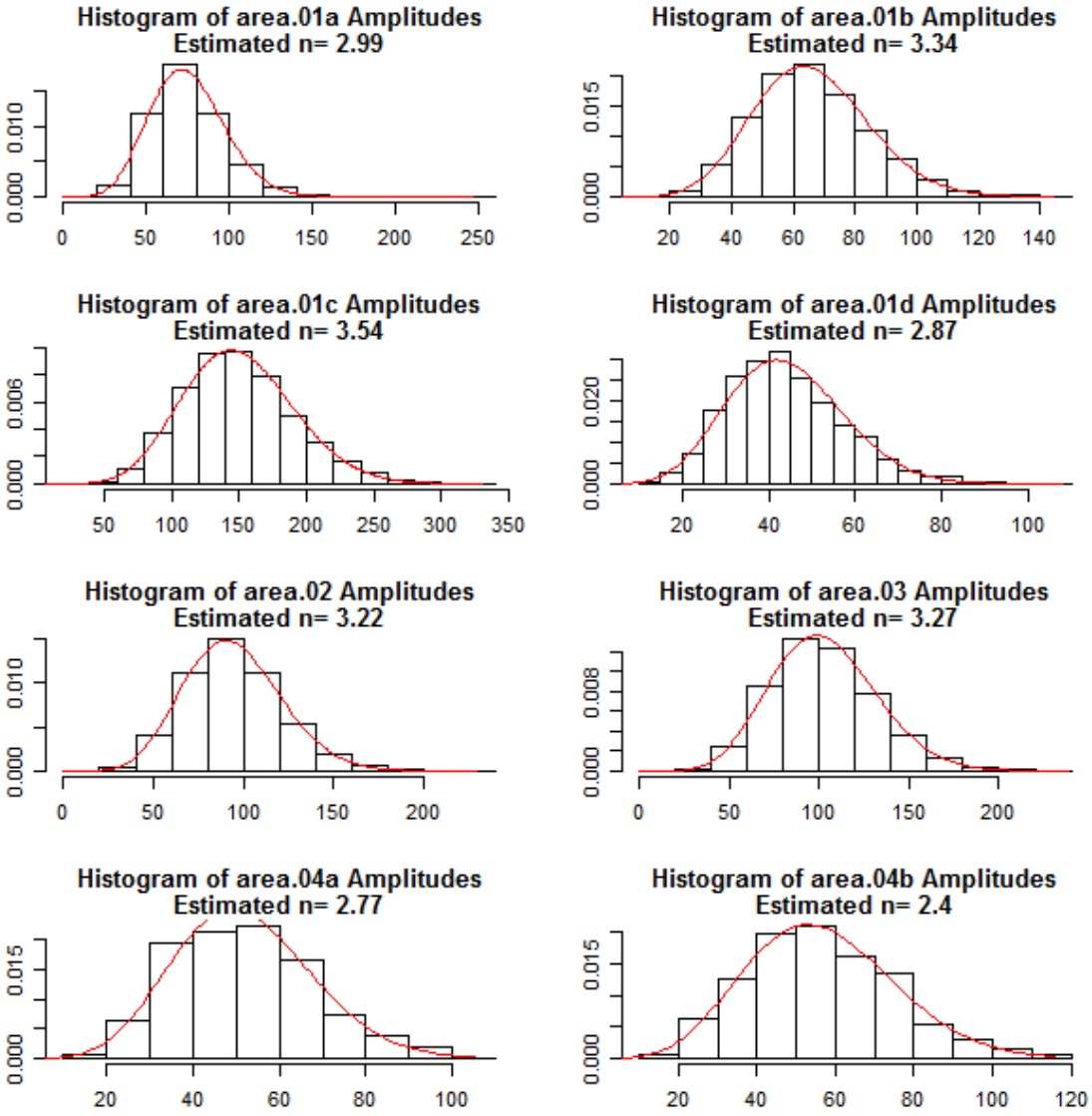


**Figure 4.6** Estimates of the equivalent number of looks, $n$, for the DLR image, 1.

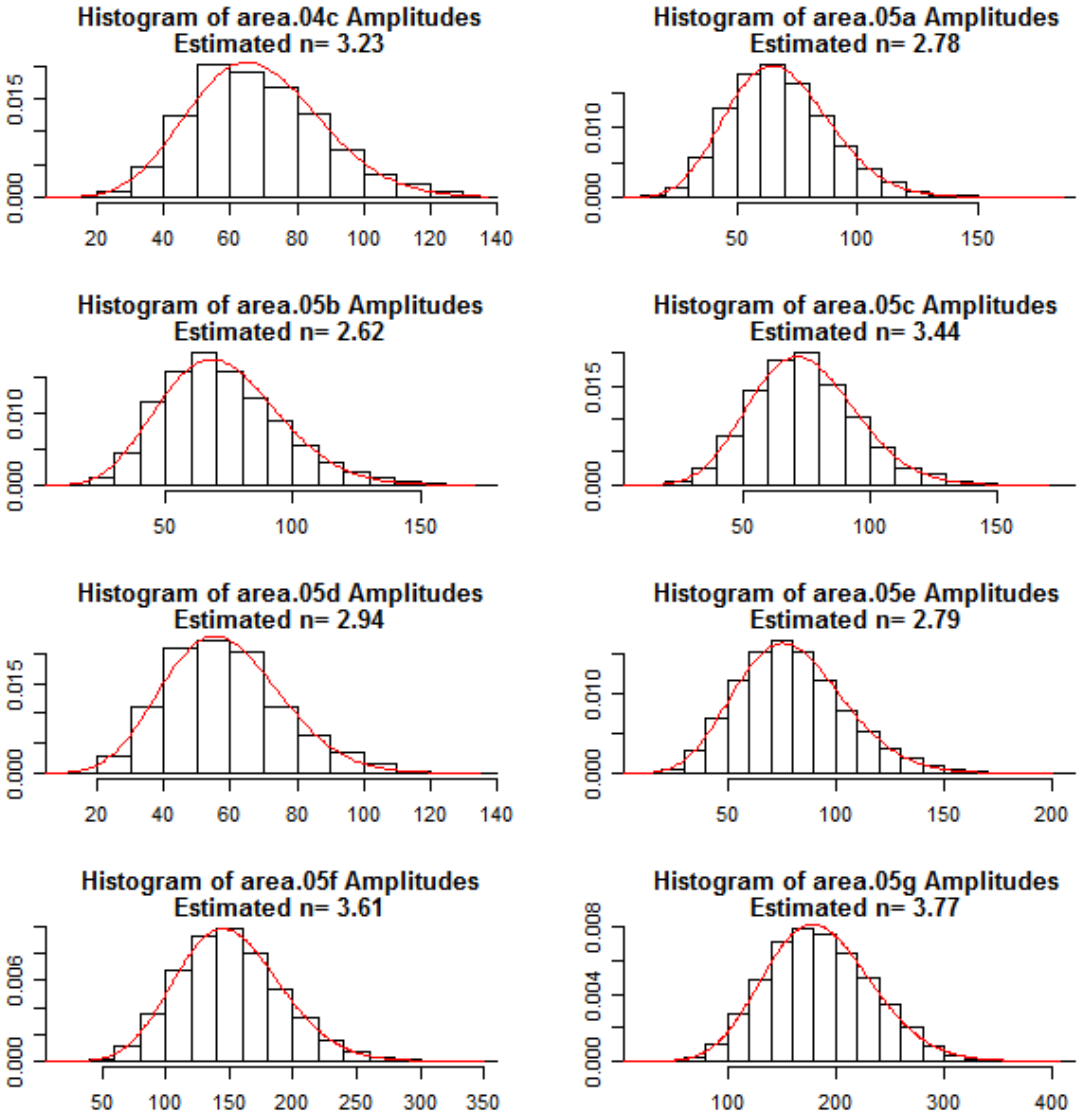


**Figure 4.7** Estimates of the equivalent number of looks, $n$, for the DLR image, 2.

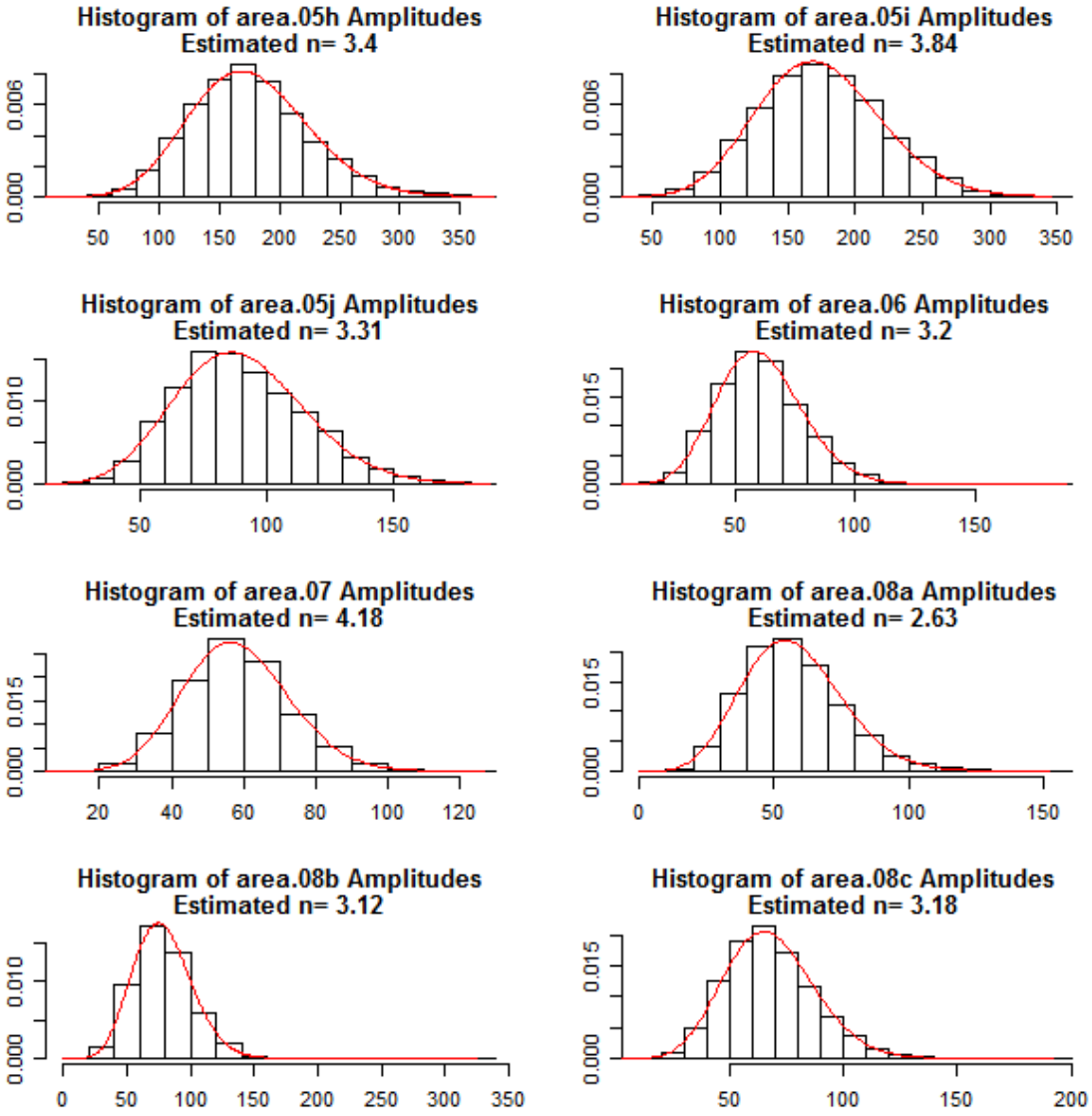


**Figure 4.8** Estimates of the equivalent number of looks, $n$, for the DLR image, 3.

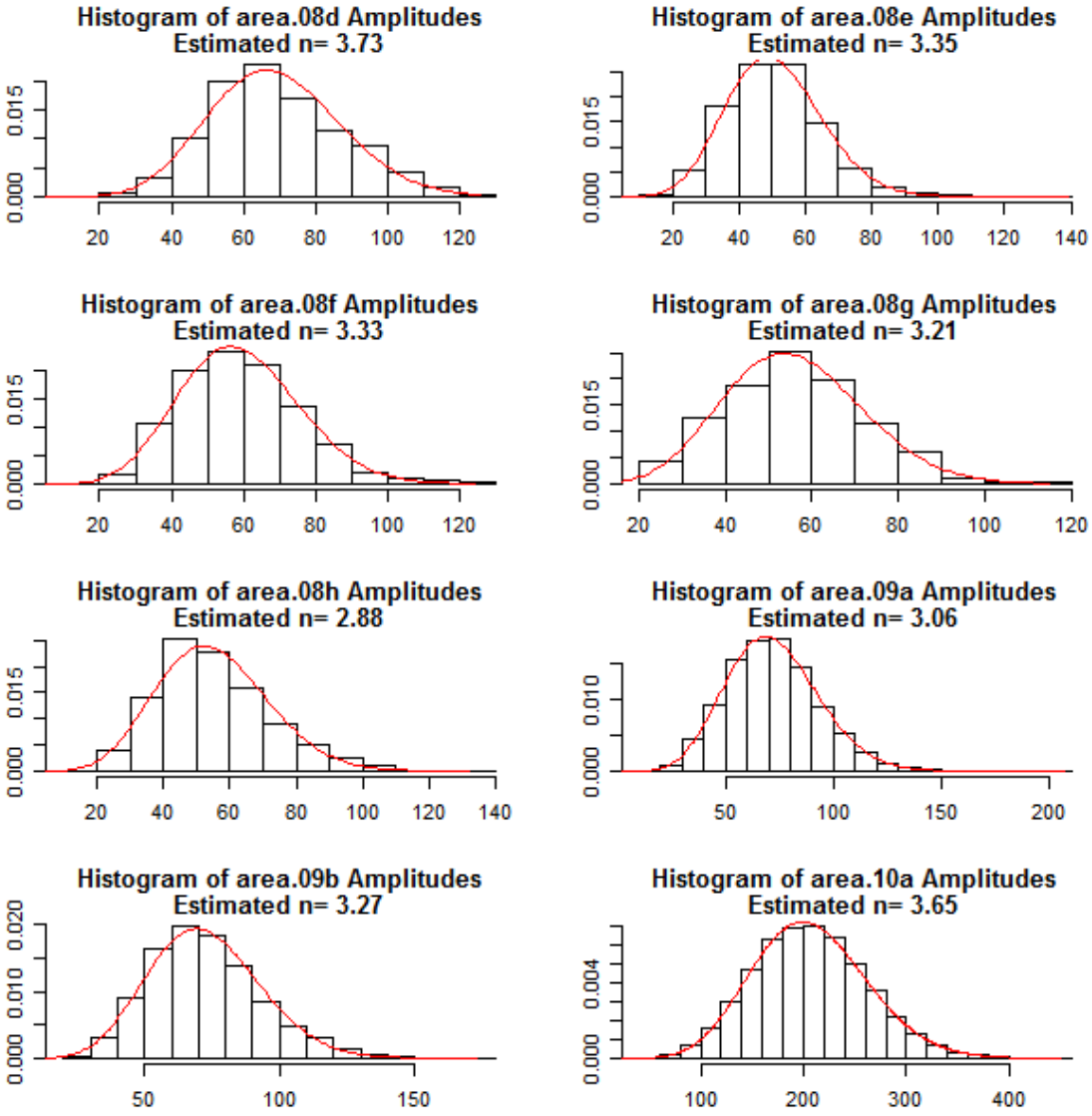


**Figure 4.9** Estimates of the equivalent number of looks, $n$, for the DLR image, 4.

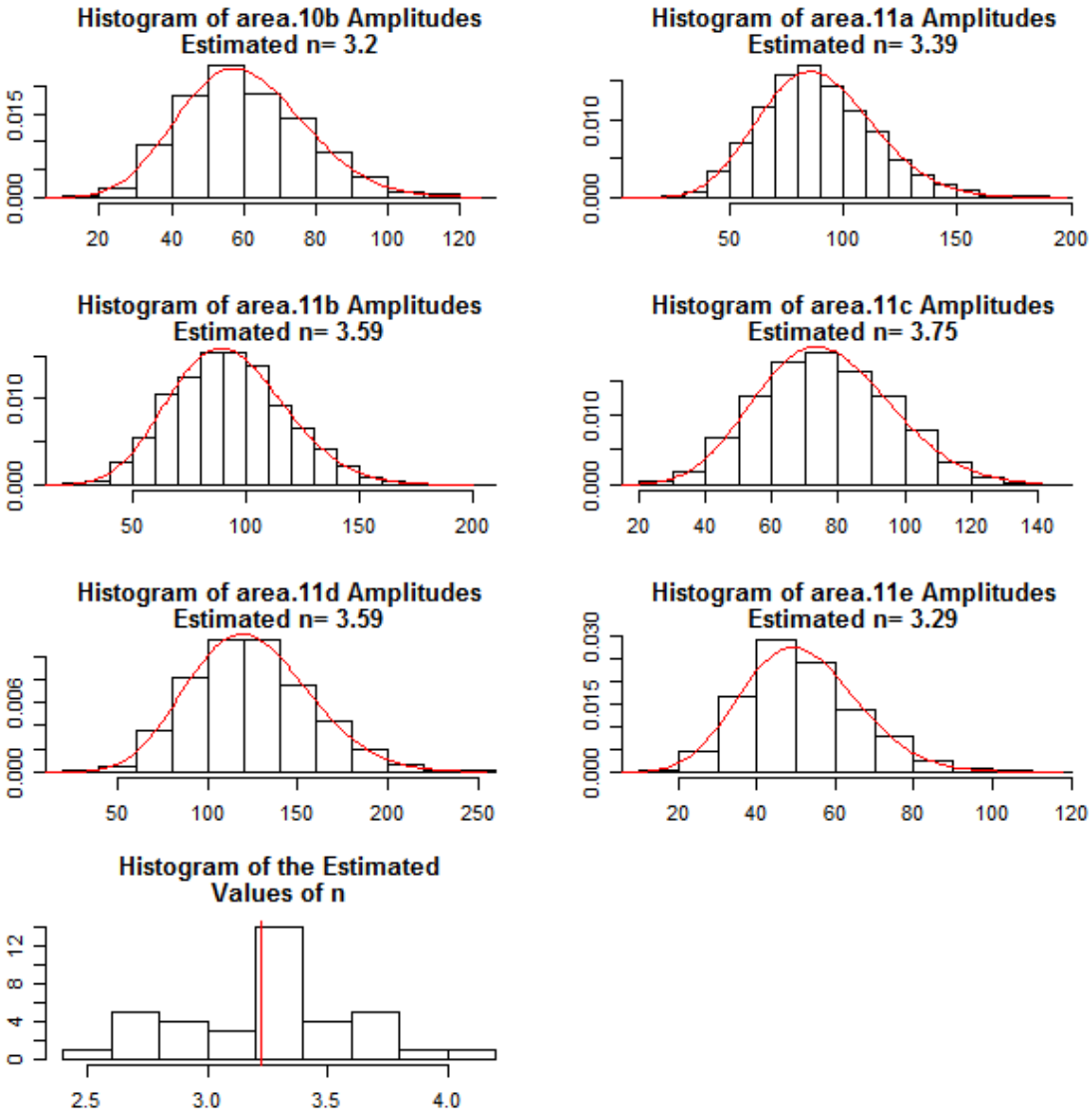


**Figure 4.10** Estimates of the equivalent number of looks, $n$, for the DLR image, 5. The last graph is a histogram of the estimated values for all 38 regions.

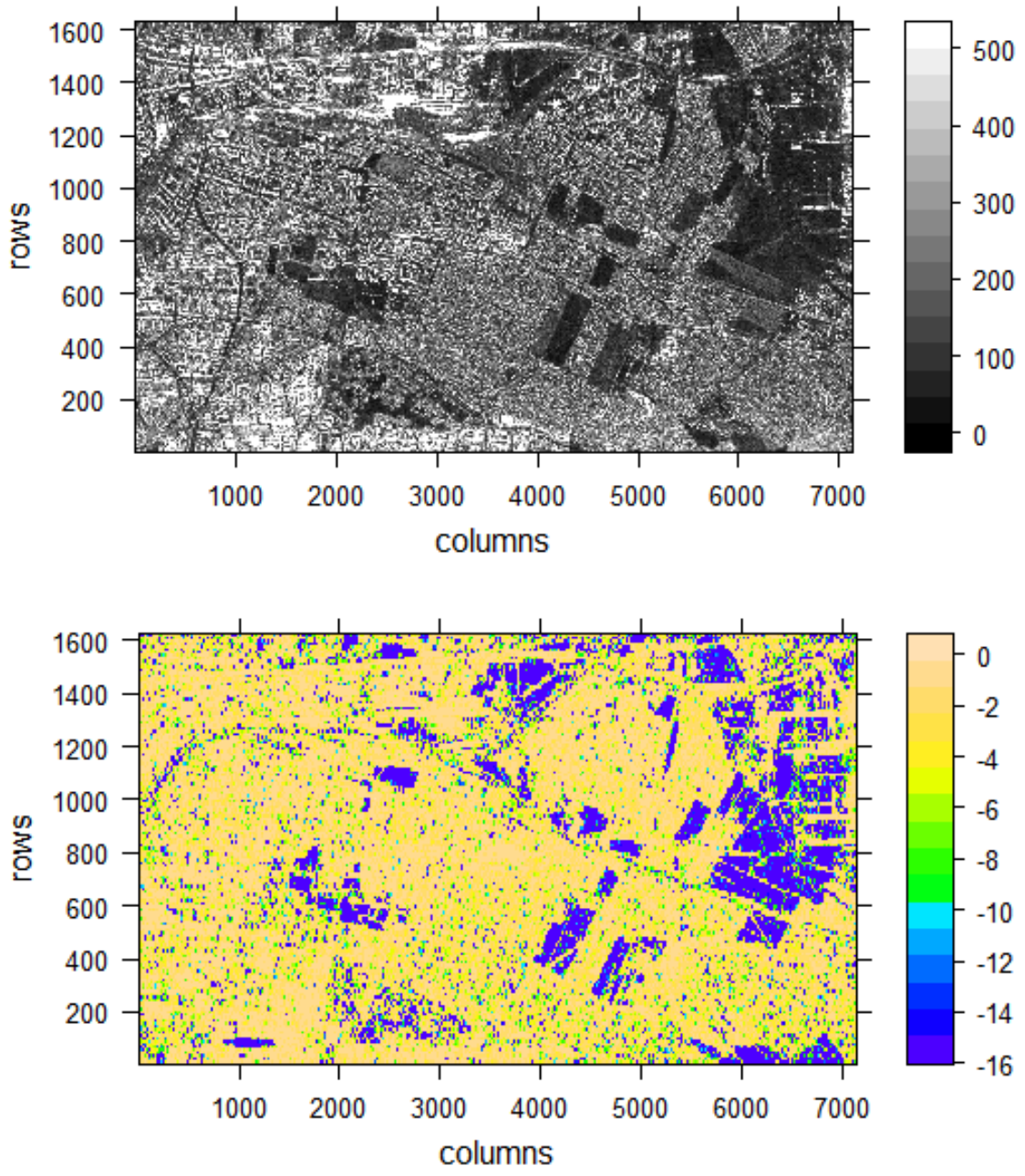


**Figure 4.11** Panel of the DLR image and the bootstrap estimates of the roughness parameter, $\alpha$.

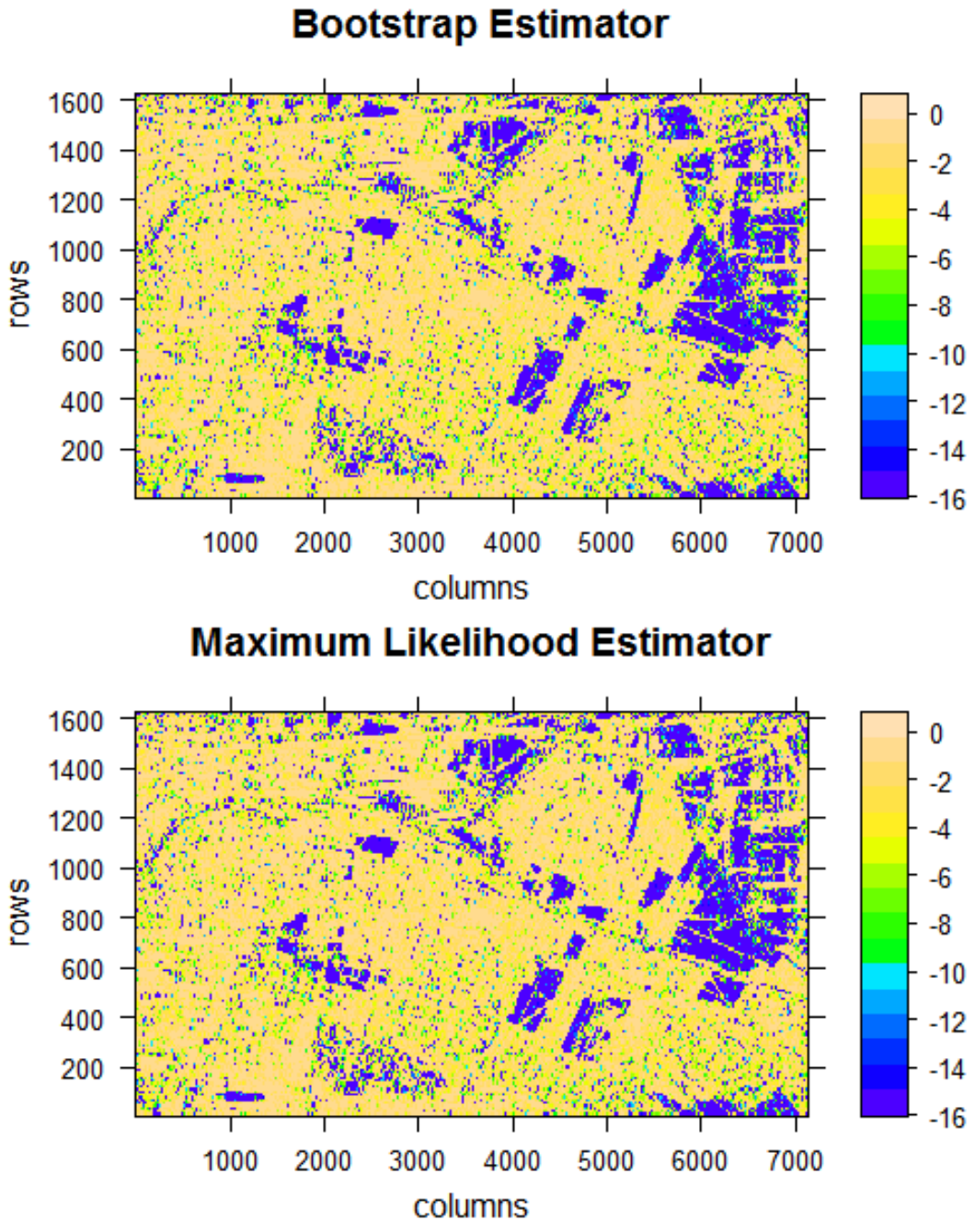


**Figure 4.12** Panel of the bootstrap and MLE estimates of the roughness parameter, $\alpha$, for the DLR image.

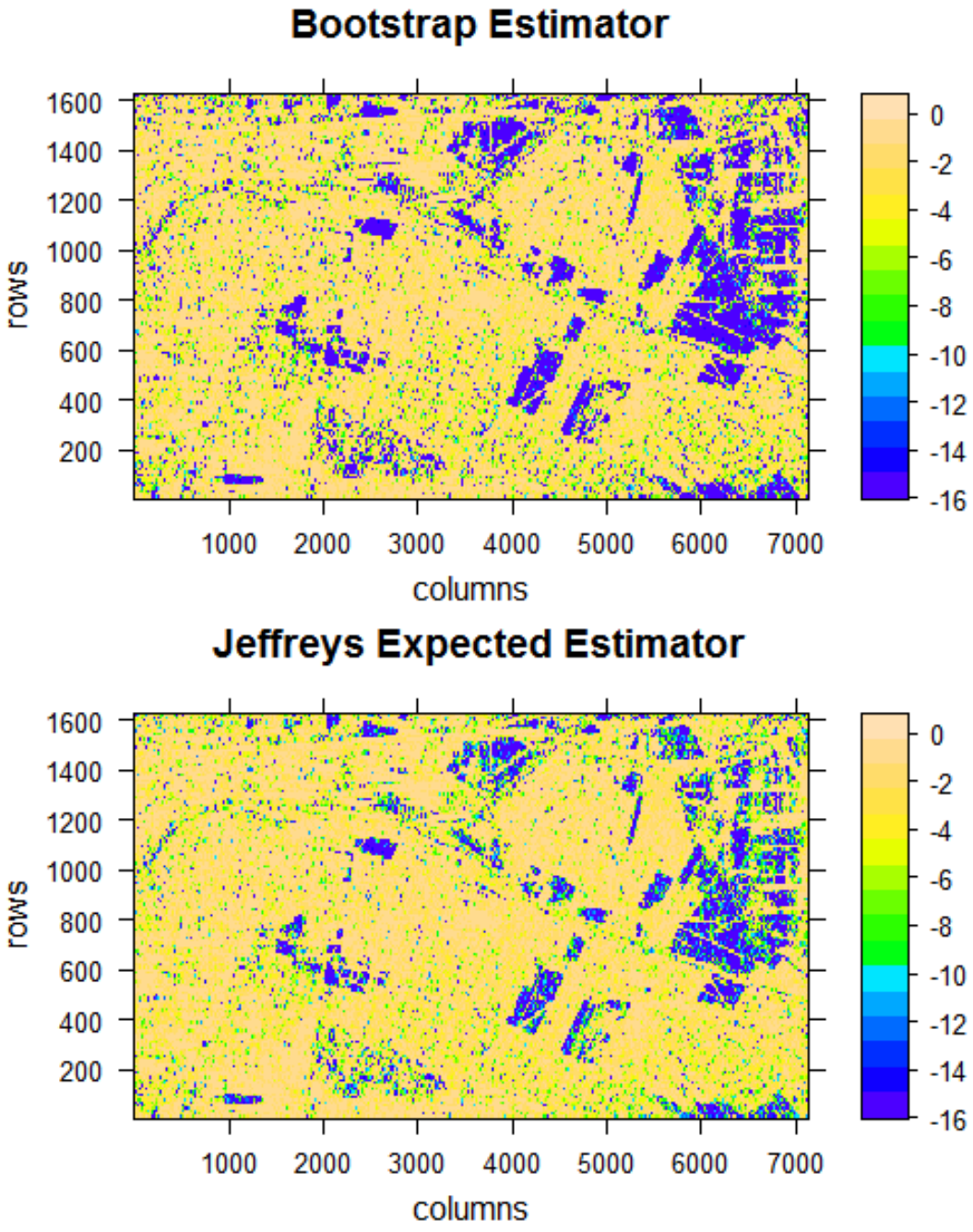


**Figure 4.13** Panel of the bootstrap and Jeffreys expected estimates of the roughness parameter, $\alpha$, for the DLR image.

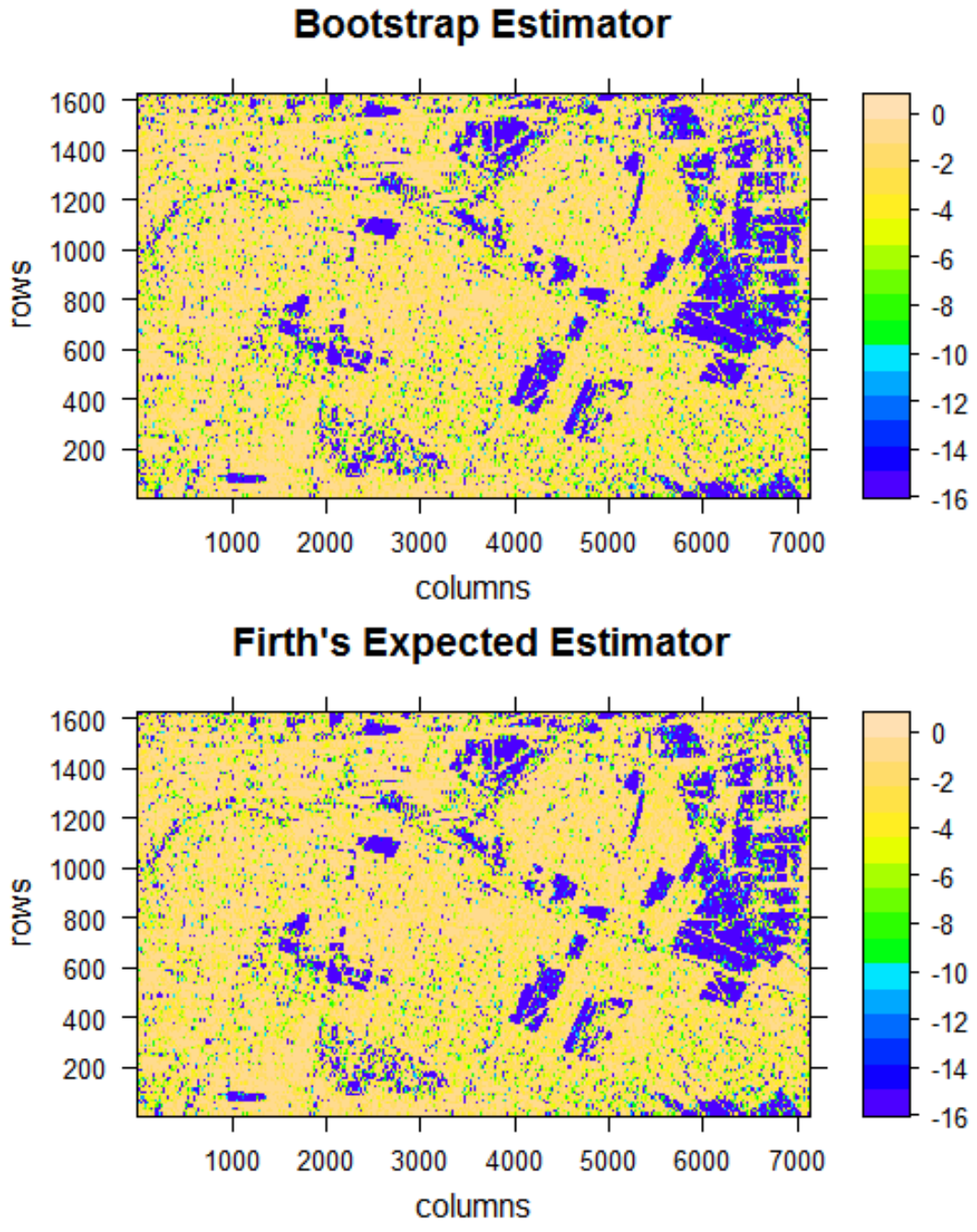


**Figure 4.14** Panel of the bootstrap and Firth expected estimates of the roughness parameter, $\alpha$, for the DLR image.

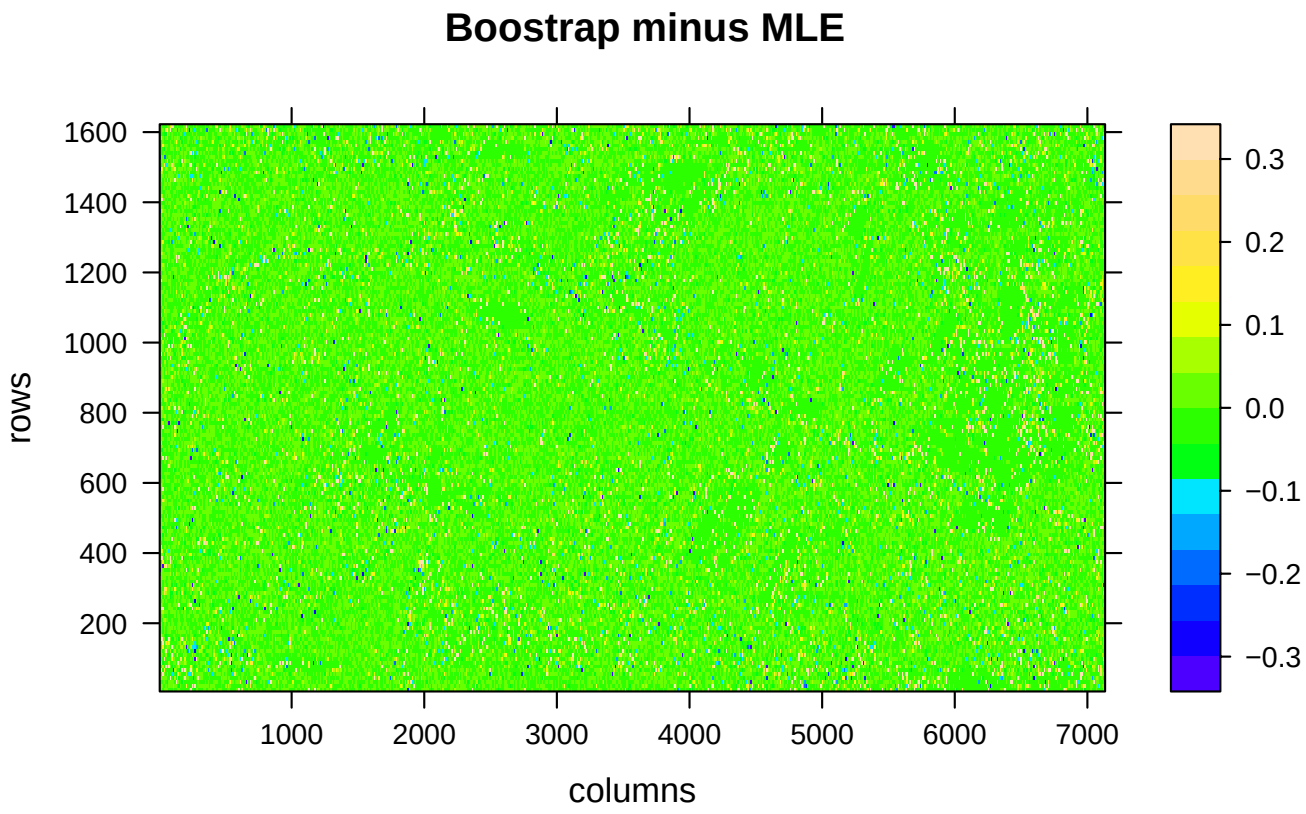


**Figure 4.15** Difference between the bootstrap and MLE estimates of the roughness parameter, $\alpha$, for the DLR image. The estimates were clipped before taking the difference and clipped once again afterwords.

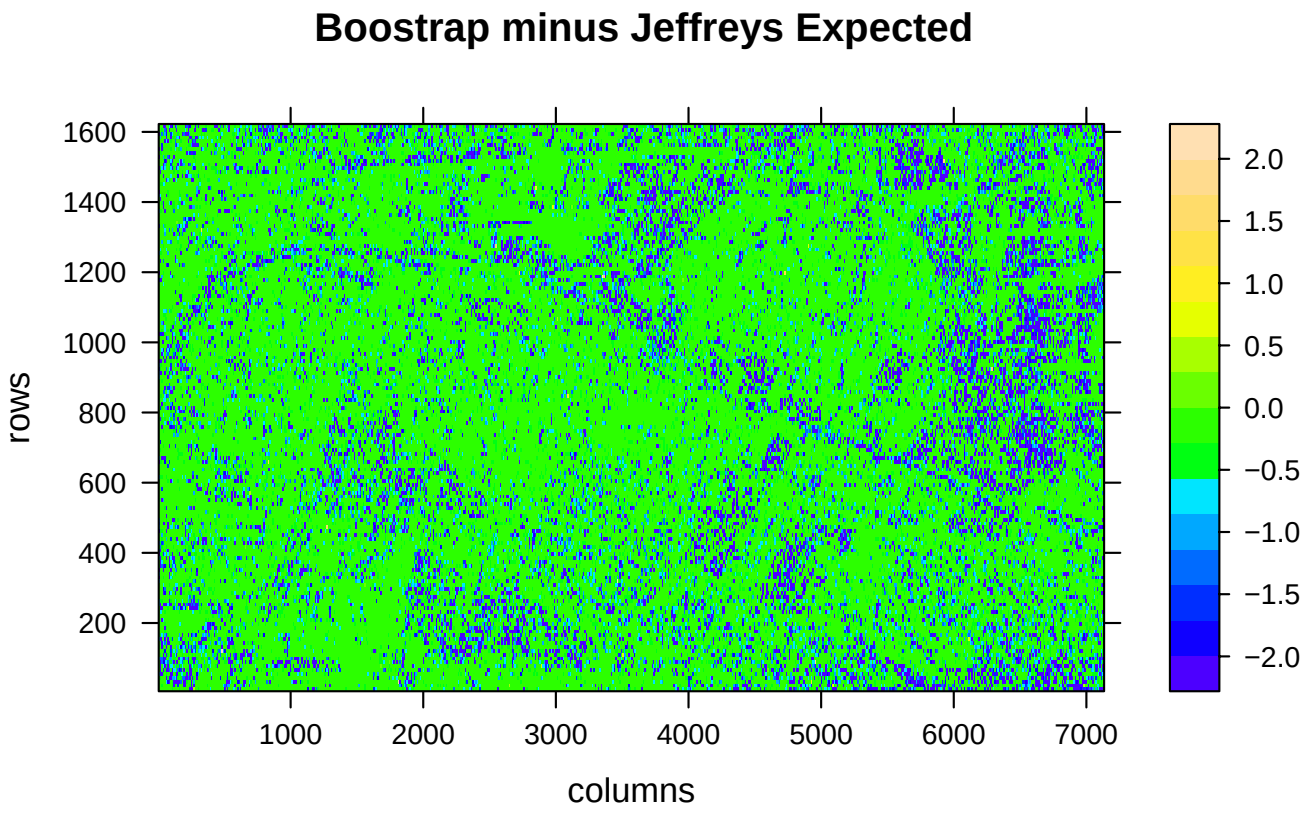


**Figure 4.16** Difference between the bootstrap and Jeffreys expected estimates of the roughness parameter, $\alpha$, for the DLR image. The estimates were clipped before taking the difference and clipped once again afterwords.

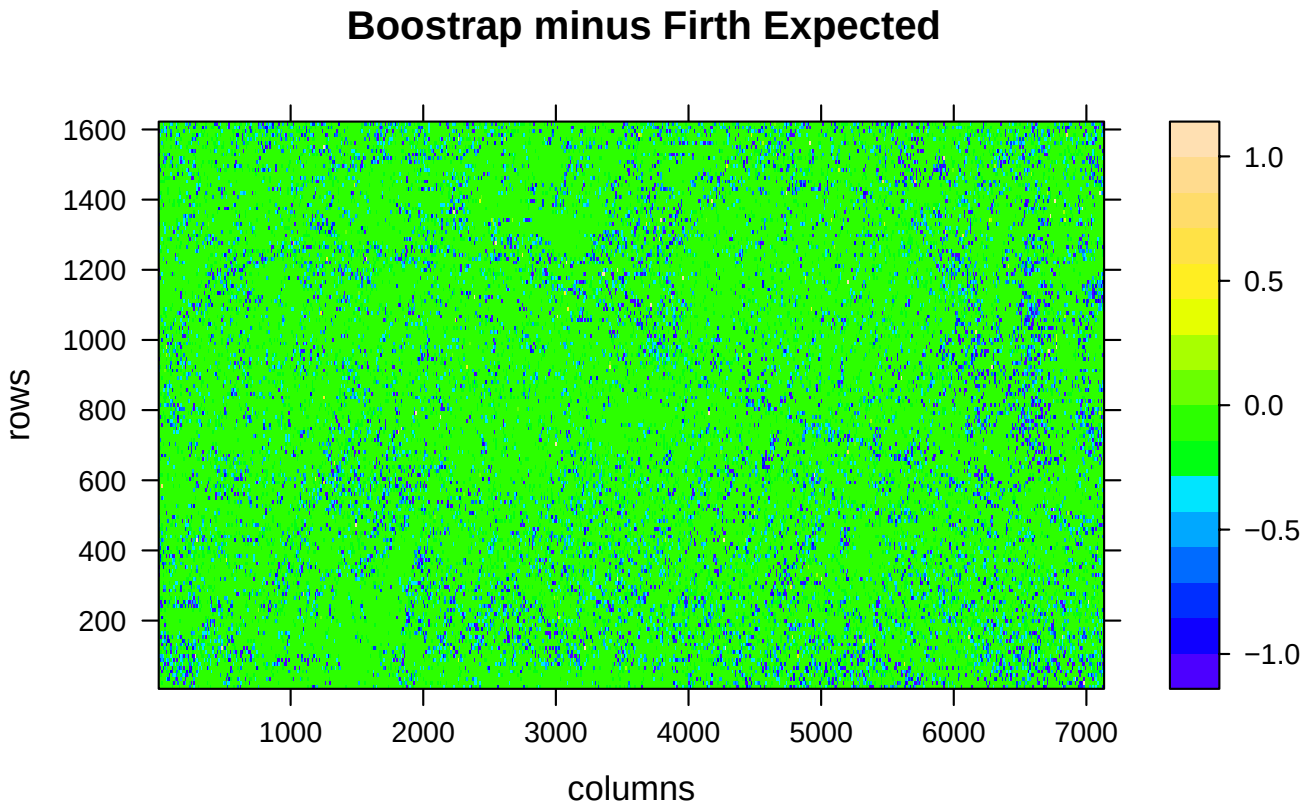


**Figure 4.17** Difference between the bootstrap and Firth expected estimates of the roughness parameter, $\alpha$, for the DLR image. The estimates were clipped before taking the difference and clipped once again afterwords.

## 4.4 A heterogeneous image with single and multiple looks

In this section we analyze data that was collected as single look complex data. Such data can be collapsed over rows or columns to form multi-look data. In this way we can study the performance of the bootstrap and other estimators on the same image with different numbers of equivalent looks. The real and imaginary components of the single look complex data are each stored as single precision float. Hence the maximum amplitude is not limited by storage considerations and one can expect large amplitude variations. We analyze the HV channel of this image.

### 4.4.1 Single look results

Figure 4.18 displays a panel of the amplitude of the original image and the bootstrap estimator's estimates of the roughness parameter $\alpha$. No large region is considered uniformly homogeneous and there are many regions with intermediate roughness. About 47% of the 37044 blocks of 11 by 11 pixels satisfy the divergence criteria. In Table 3.2 one can see that for $(n,\alpha,N) = (1,-15,121)$ about 33% of the samples diverged. Therefore homogeneous areas in the image are probably quite homogeneous and likely correspond to regions with $\alpha$ much less than $-15$. The bootstrap estimator failed to converge in only 7% of the blocks whereas the MLE failed to converge for 60% of the blocks. Both the Firth estimator with expected information and the Jeffreys prior with expected information converged for all samples.

Figures 4.19–4.21 display panels with the bootstrap estimates in the top panel and the MLE, Jeffreys, and Firth estimates in the lower panel. The MLE estimates are biased toward smoothness and the image shows this as there are many blue blocks (recall that estimates of $\alpha$ less than $-15$ are plotted as $-15$). The Jeffreys estimates are biased toward heterogeneity and are quite poor. The Firth estimates are similar to the bootstrap estimates although slightly more heterogeneous.

### 4.4.2 Multi-look results

Since we have single look complex data, we can generate multi-look data by collapsing rows or columns. Since we have many more columns than rows, we collapsed along the columns. We chose a nominal number of looks of eight. To collapse the data, assume that we have a matrix of amplitudes, $m_{i,j}$. To generate $n$-look data by collapsing over columns we generate the matrix $p_{i,k} = \sqrt{m^2_{i,(k-1)n+1} + \cdots + m^2_{i,kn}}$.

The image resulting from this matrix is shown in Figure 4.22 with the areas used to estimate the equivalent number of looks marked by red polygons. The resulting image is much more homogeneous, but still contains speckle noise. The high and low amplitude areas can be much more clearly identified.

Figures 4.23–4.25 contain histograms of the amplitudes for each of the polygons in Figure 4.22 and the density implied by the estimated values of the number of looks. Once again the data are well described by the $\Gamma^{1/2}$ model. The mean of the 17 estimates of the number of looks is 5.93 and the weighted mean with weight being the number of observations in each polygon is 5.87. We used 5.9 as our equivalent number of looks in the following analysis.

Figure 4.26 displays the multi-look image in the upper panel and the estimates of the roughness parameter using the bootstrap estimator in each of 11352 blocks of 7 by 7 pixels. There are now contiguous smooth regions and contiguous rough regions whereas with the single look data there was much more variation in the roughness estimates.

For this multi-look image, using 7 by 7 blocks, the divergence criteria was satisfied in 32% of the blocks. Table 3.2 shows that for $(n,\alpha,N) = (3,-15,49)$ and $(8,-15,49)$, 23.9% and 3.6% of the samples satisfied the divergence criteria. Given the equivalent number of looks equal to 5.9, we see that the real data contains some extremely homogeneous regions. The MLE failed to converge in 41% of the blocks, the bootstrap estimator in 13% (nearly twice the percentage as in the single look case), and the Jeffreys and Firth estimators in only one block.

Using 4620 blocks of 11 by 11 pixels, the divergence criteria was satisfied in 19% of the blocks. Again, Table 3.2 shows that for $(n,\alpha,N) = (3,-15,121)$ and $(8,-15,121)$, 8.6% and 0.1% of the samples satisfied the divergence criteria. For the equivalent number of looks equal

to 5.9, the extreme homogeneity of some regions is confirmed. The MLE failed to converge in 28% of the blocks, the bootstrap estimator in 12%, and the Jeffreys and Firth estimators in 2%.

Figures 4.27–4.29 compare the bootstrap estimator using 49 and 121 observations with the MLE, Jeffreys and Firth estimators, respectively. The MLE continues to be biased toward smoother estimates of $\alpha$, although it is quite similar to the bootstrap for $N = 121$. The bootstrap itself does not change much from $N = 49$ to $N = 121$, a desirable property if one assumes that more data gives more accurate results. The Jeffreys estimator remains strongly biased toward rougher estimates of $\alpha$ with $N = 49$, and less so for $N = 121$, with contiguous regions of smoothness and roughness forming. The bootstrap and Firth estimates are extremely similar with the Firth estimates slightly more toward roughness than the bootstrap estimates. Still the bootstrap estimates are more stable between $N = 49$ and $N = 121$.

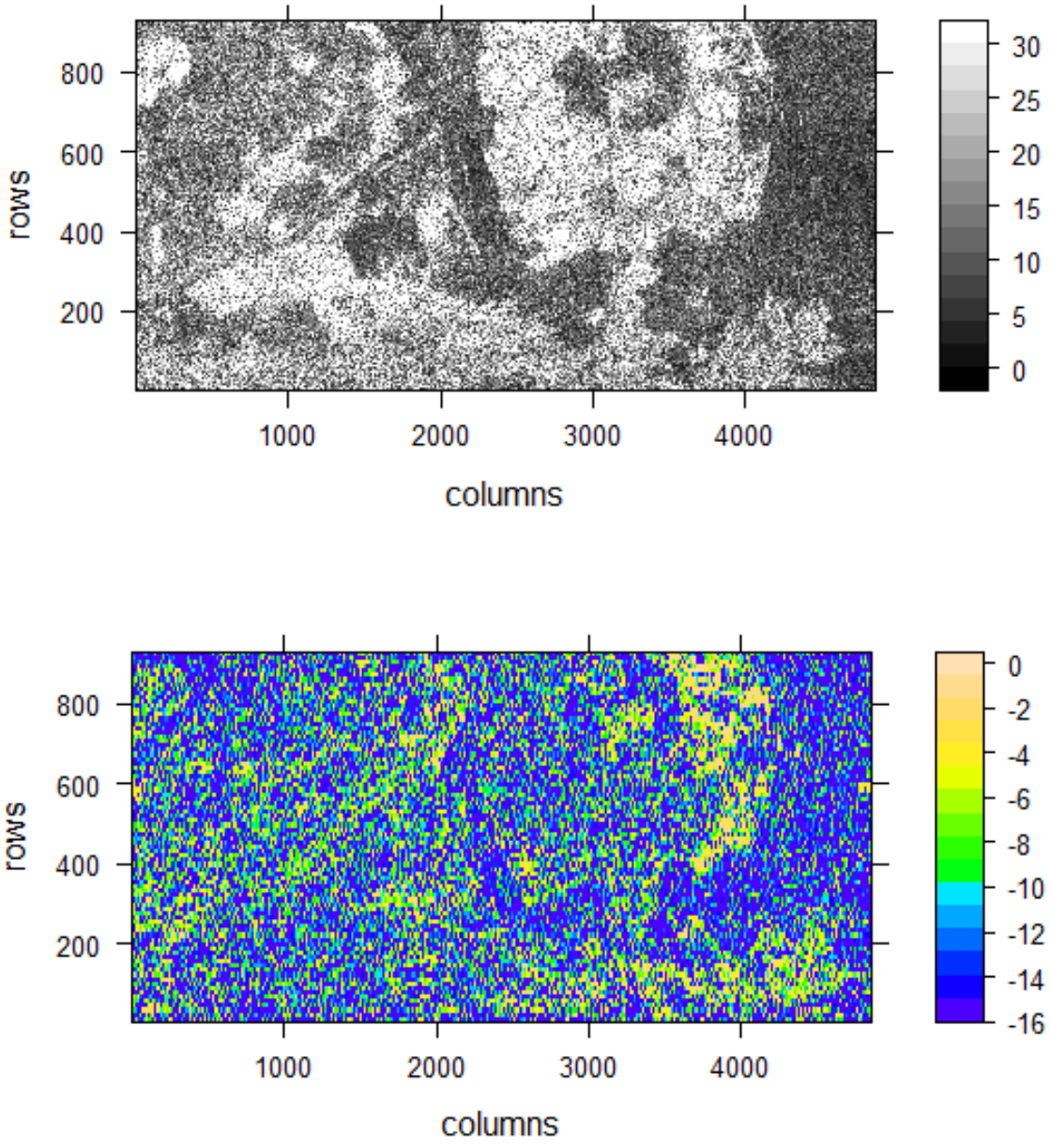


**Figure 4.18** Panel of a heterogeneous image in single look complex form (PA image) with one look and the bootstrap estimates of the roughness parameter, $\alpha$.

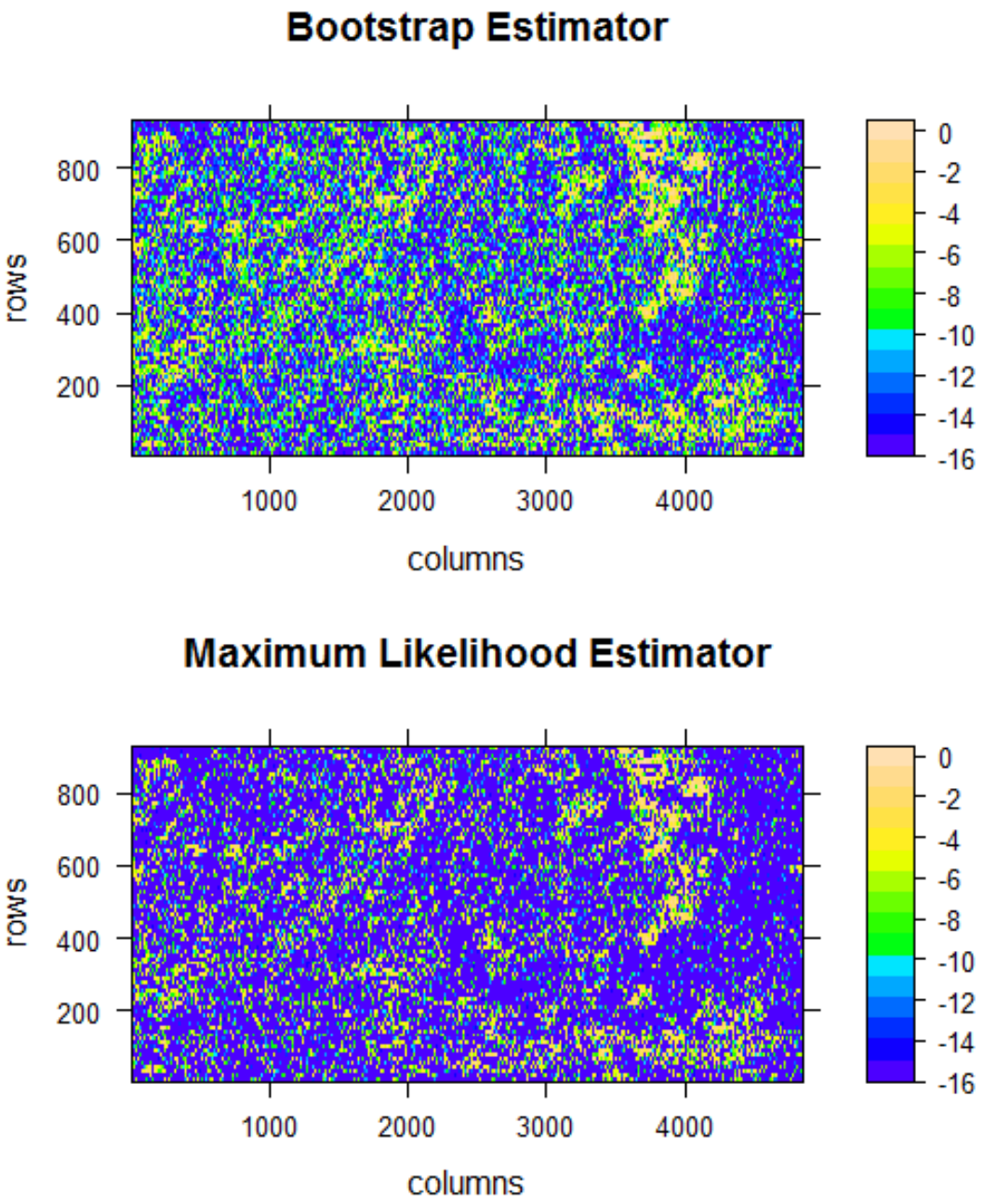


**Figure 4.19** Panel of the bootstrap and MLE estimates of the roughness parameter, $\alpha$, for the PA image with one look.

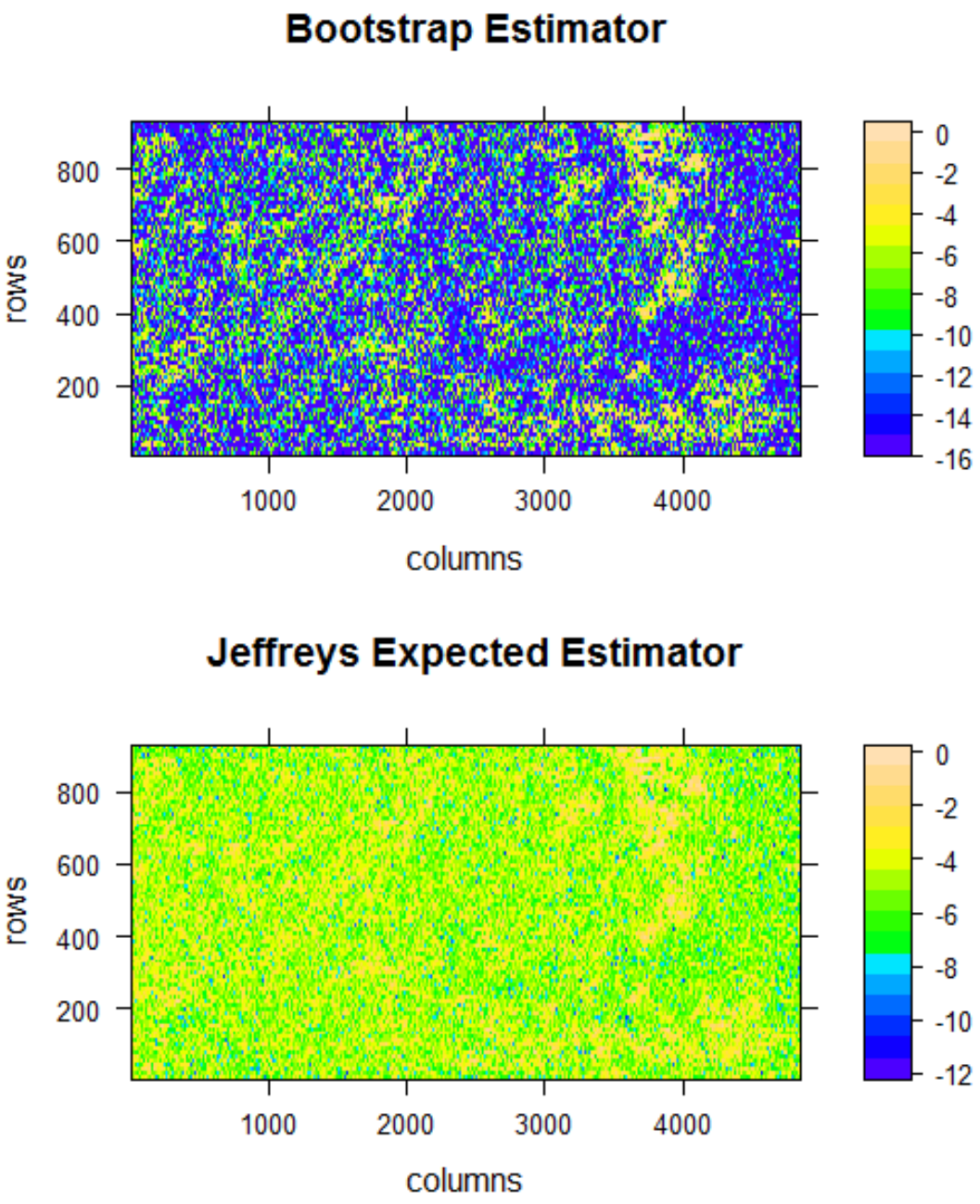


**Figure 4.20** Panel of the bootstrap and Jeffreys expected estimates of the roughness parameter, $\alpha$, for the PA image with one look.

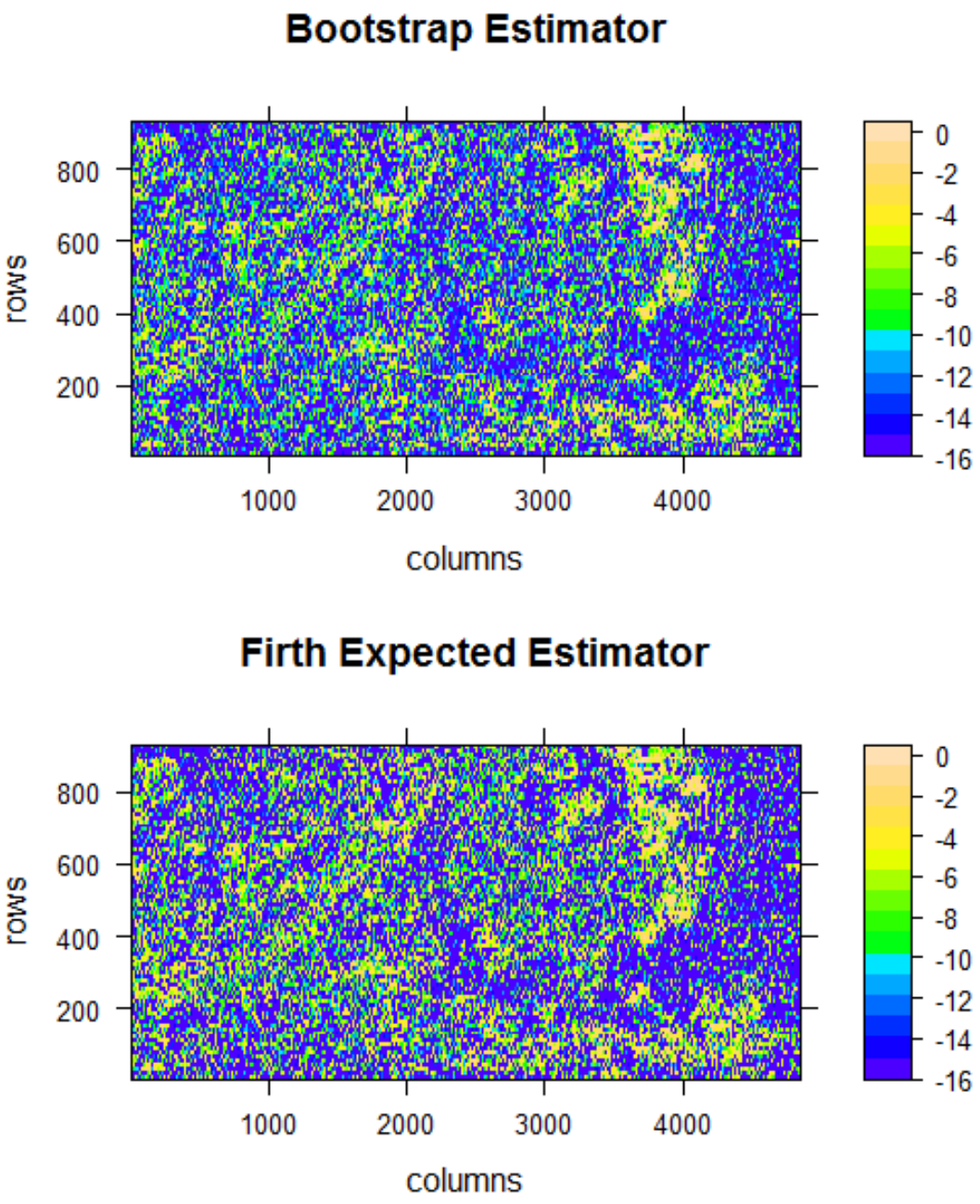


**Figure 4.21** Panel of the bootstrap and Firth expected estimates of the roughness parameter, $\alpha$, for the PA image with one look.

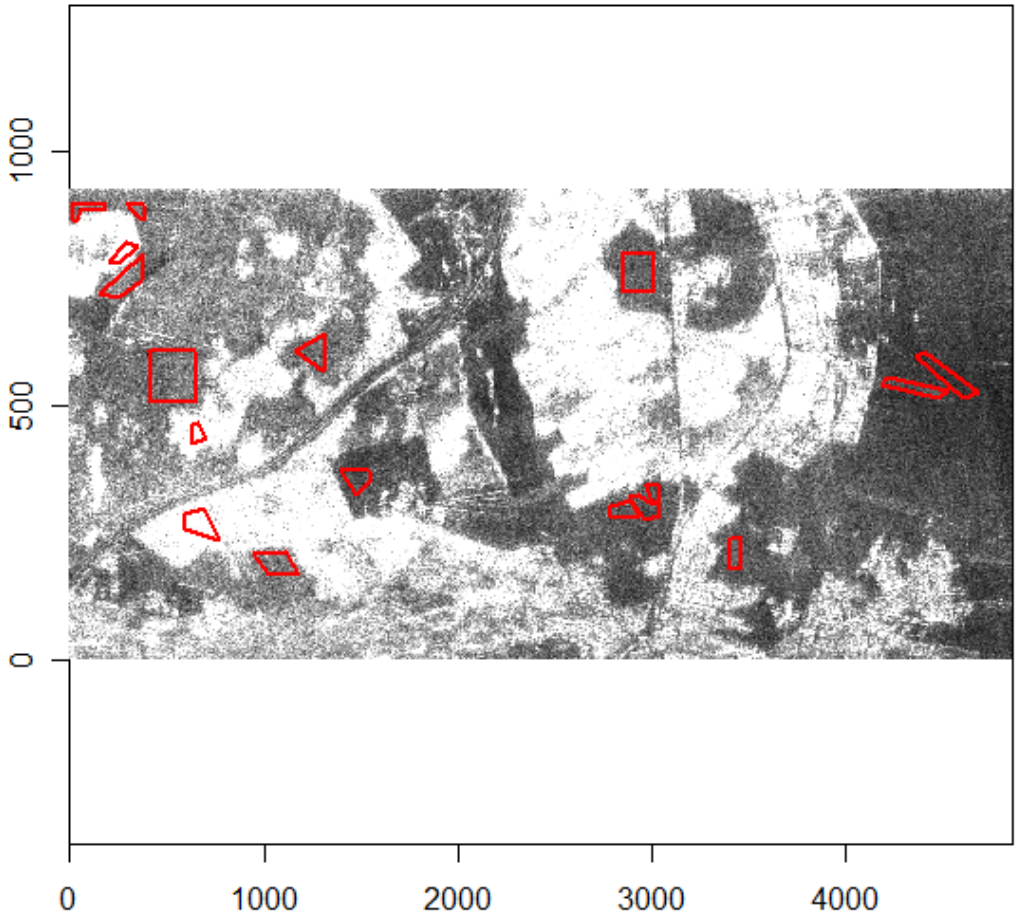


**Figure 4.22** Heterogeneous multi-look image with the regions used to estimate the number of looks marked by red polygons. White is high amplitude and black is low amplitude. (PA image)

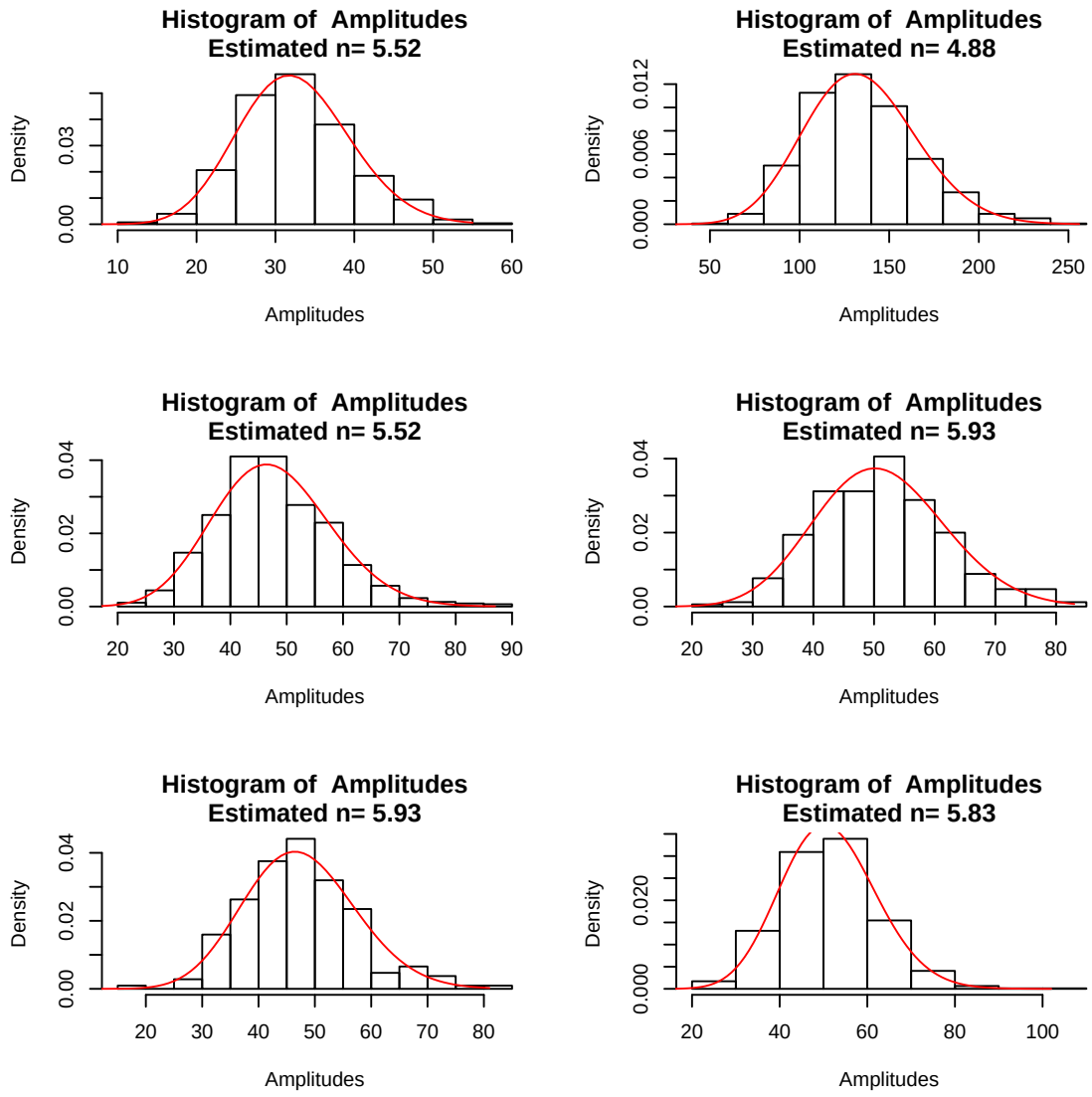


**Figure 4.23** Estimates of the equivalent number of looks, $n$, for the PA image with 8 nominal looks, 1.

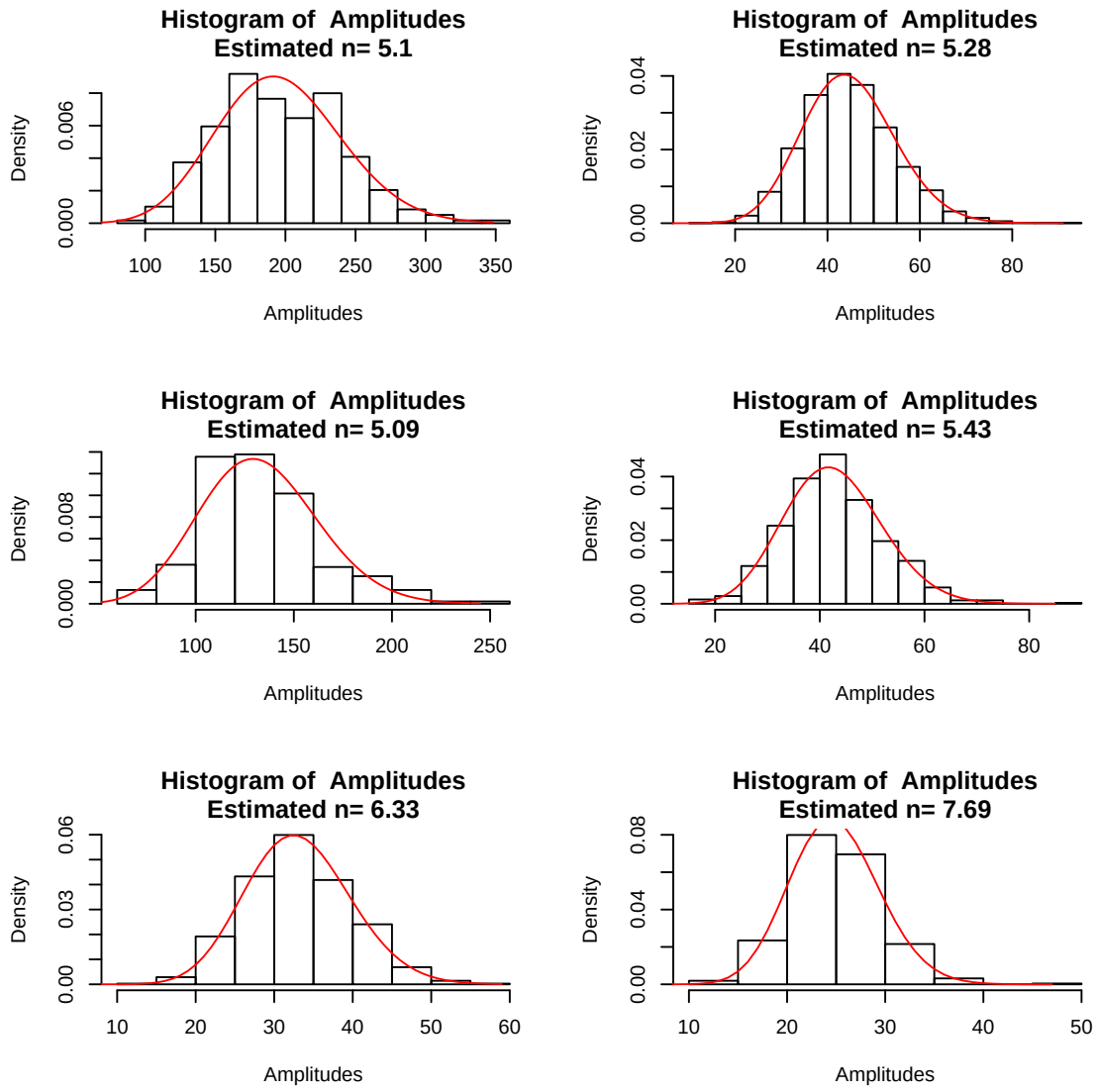


**Figure 4.24** Estimates of the equivalent number of looks, $n$, for the PA image with 8 nominal looks, 2.

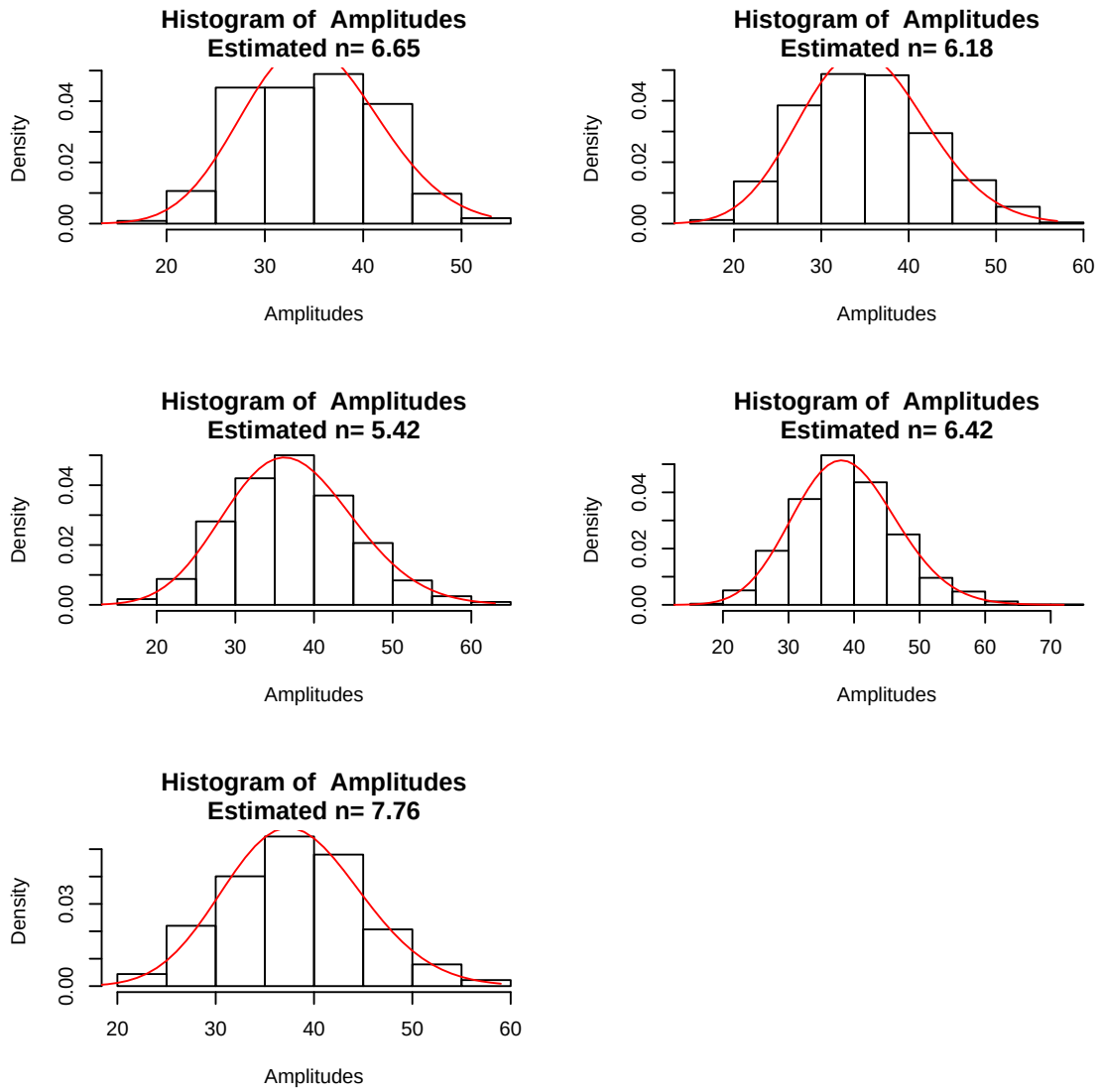


**Figure 4.25** Estimates of the equivalent number of looks, $n$, for the PA image with 8 nominal looks, 3.

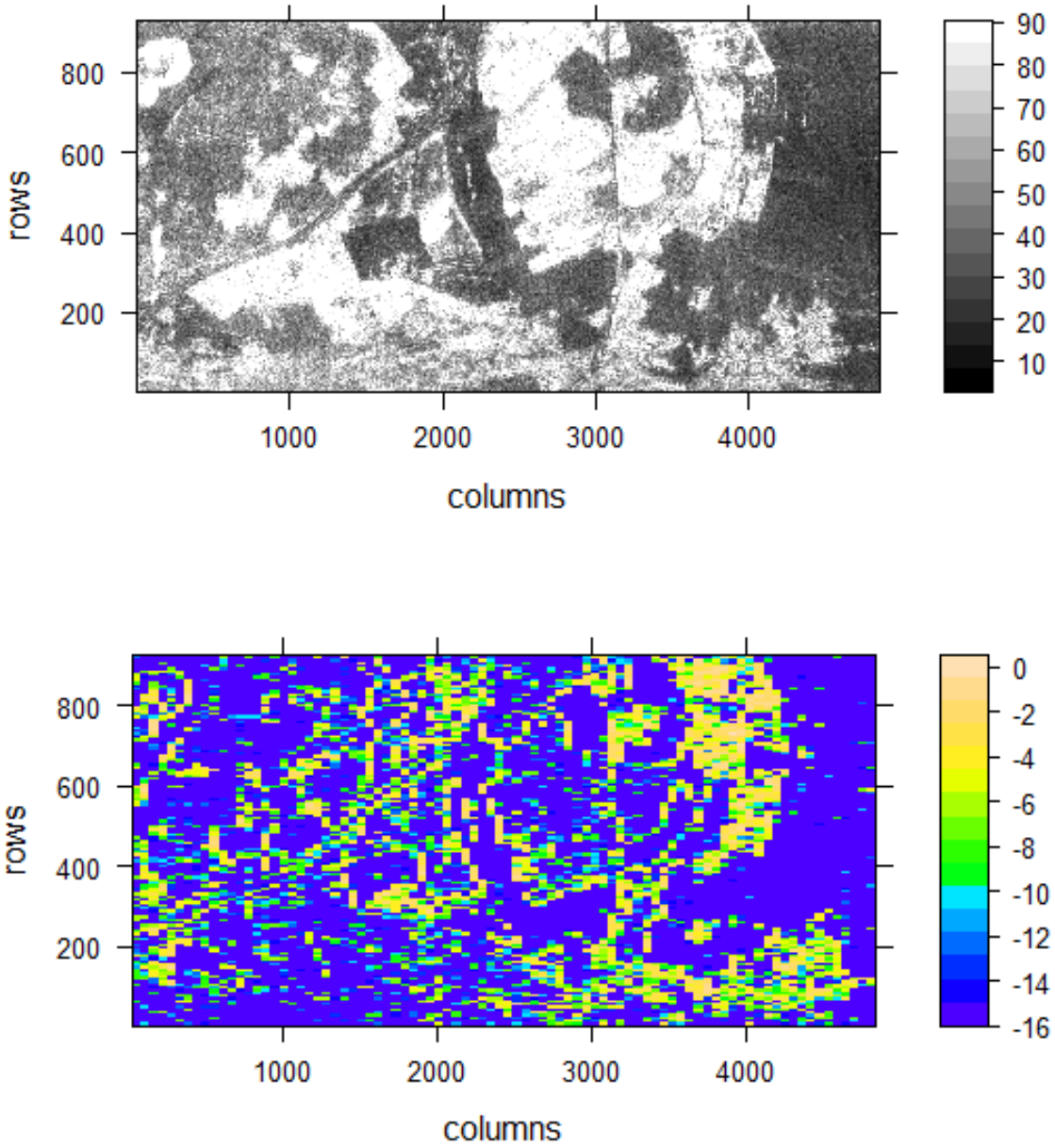


**Figure 4.26** Panel of the PA image with eight nominal looks and the bootstrap estimates of the roughness parameter, $\alpha$, using 7 by 7 windows.

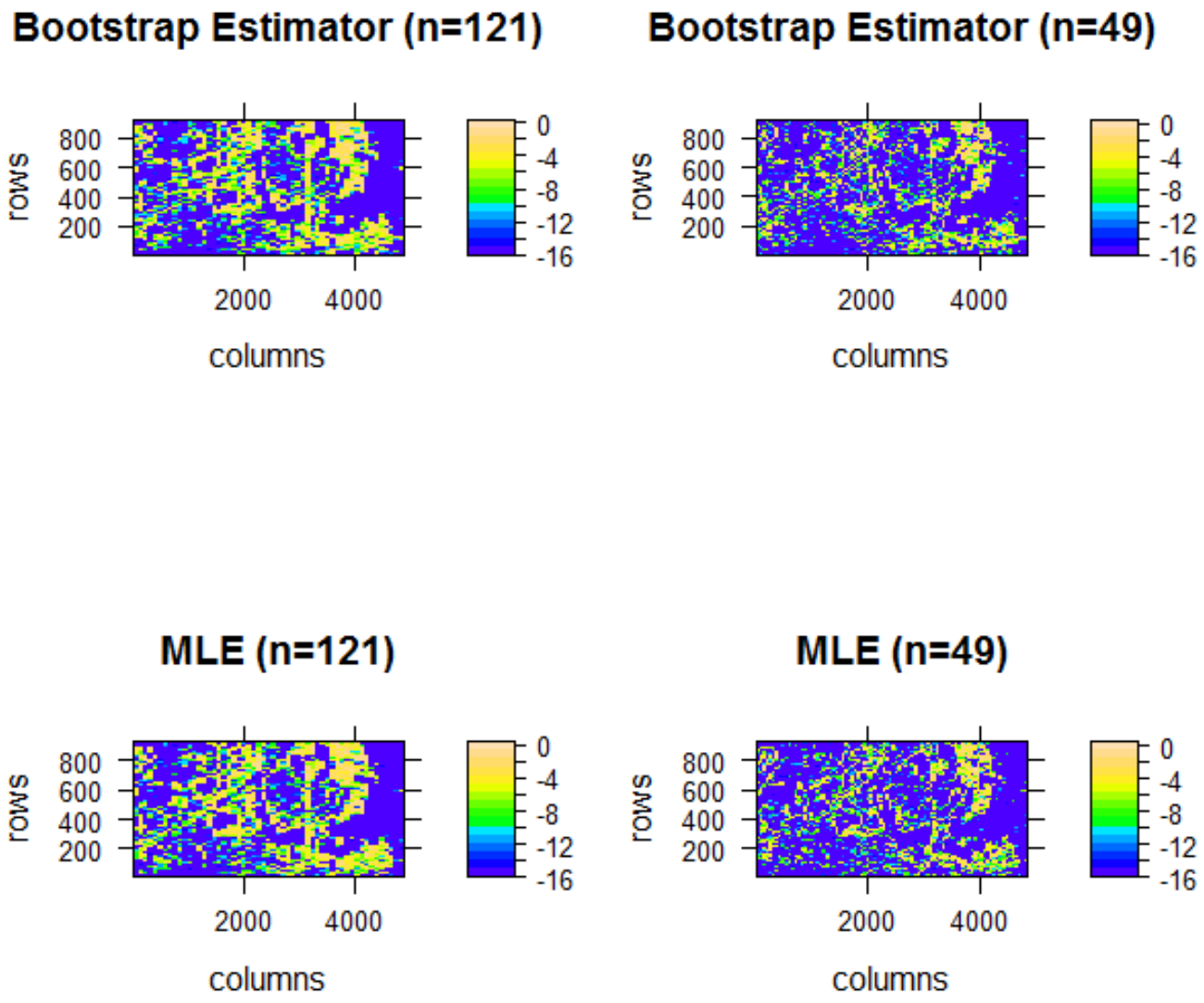


**Figure 4.27** Panel of the bootstrap and MLE estimates of the roughness parameter, $\alpha$, for the PA image with eight nominal looks.

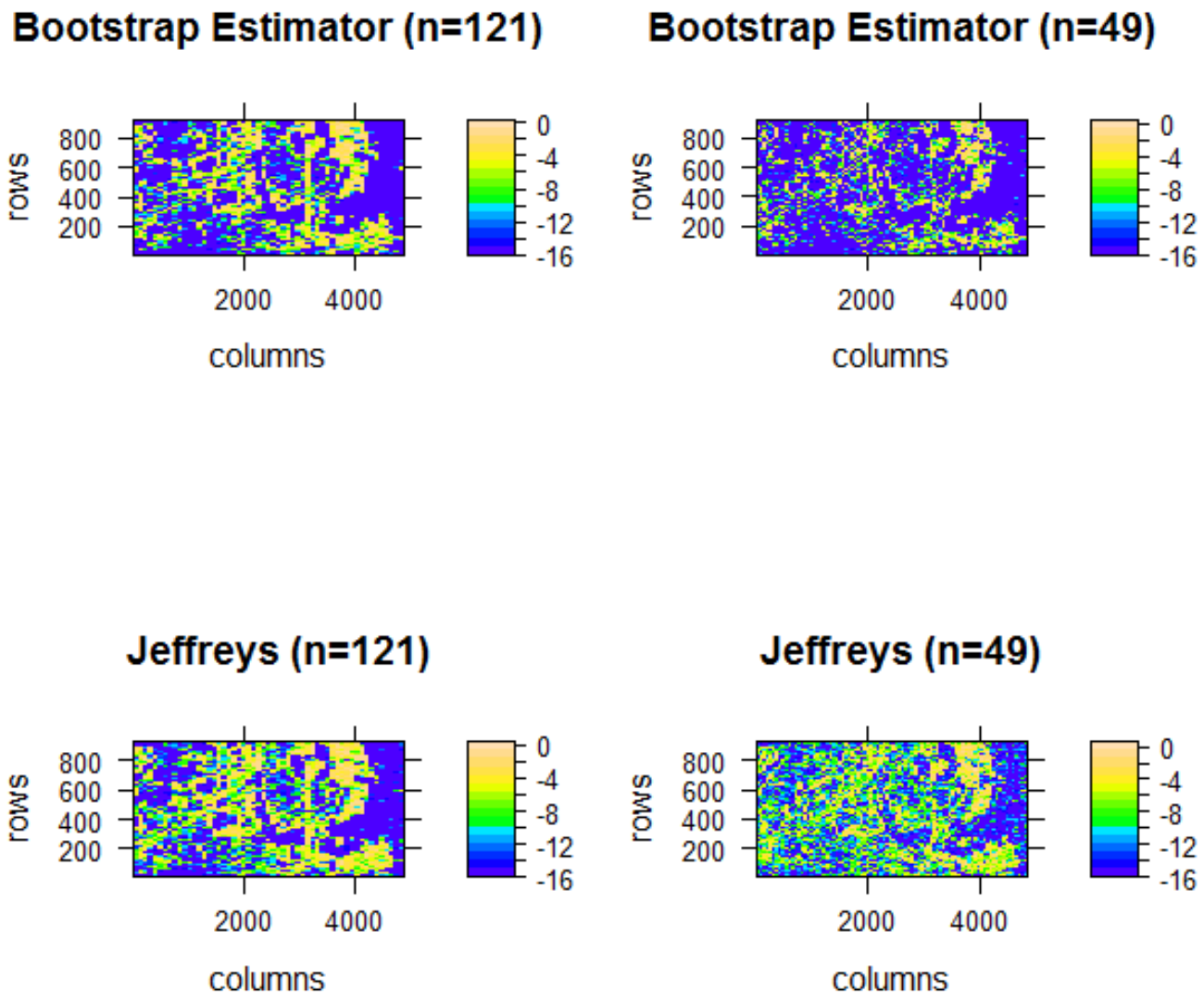


**Figure 4.28** Panel of the bootstrap and Jeffreys expected estimates of the roughness parameter, $\alpha$, for the PA image with eight nominal looks.

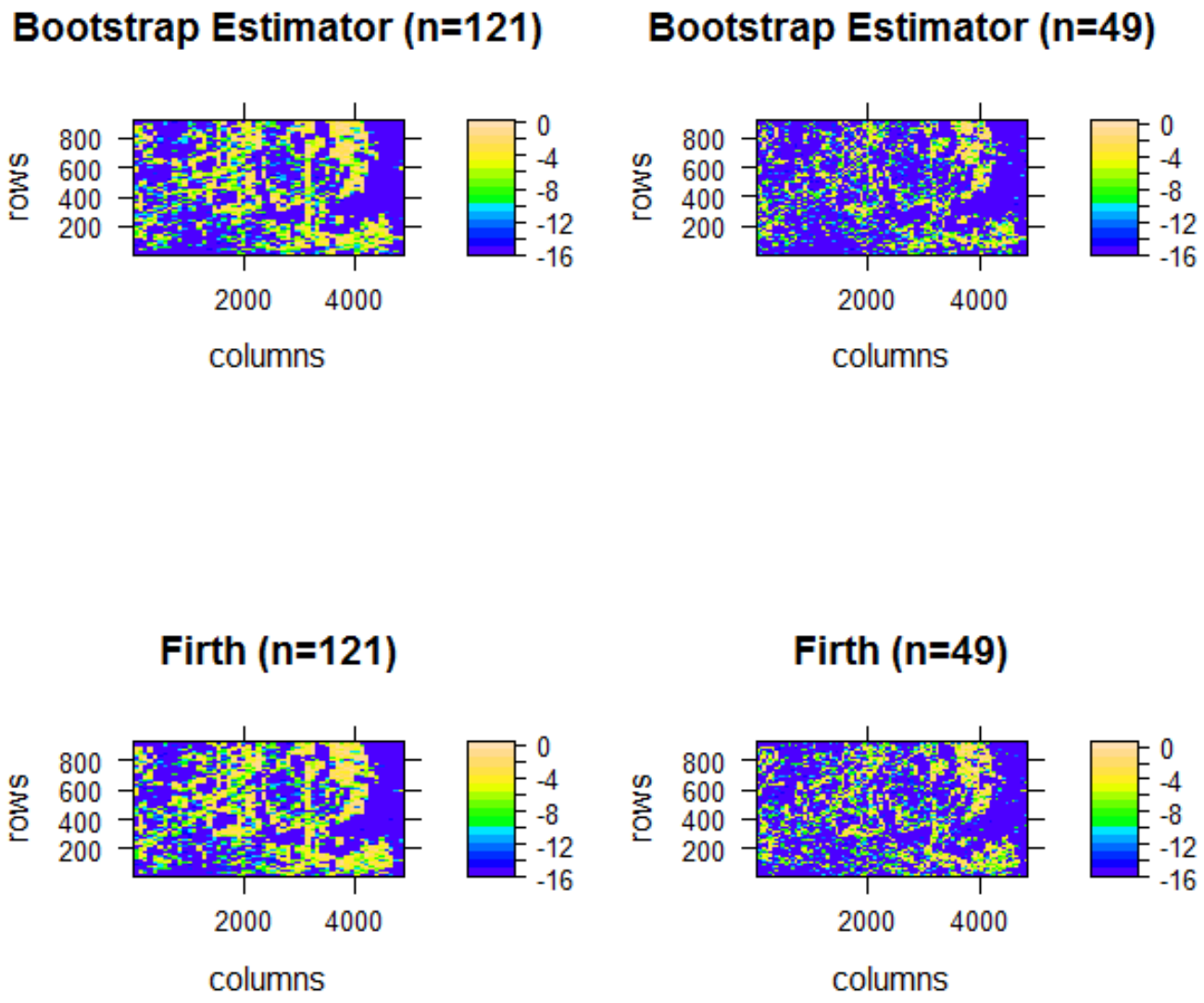


**Figure 4.29** Panel of the bootstrap and Firth expected estimates of the roughness parameter, $\alpha$, for the PA image with eight nominal looks.

CHAPTER 5

# Conclusions and Future Directions

We identified the existence of monotone likelihood in $\mathscr{G}_A^0$ samples which satisfy the divergence criteria, Equation (2.13). We have implemented three bias correcting estimators suggested by Firth (1993) and Jeffreys (1946) to resolve this problem as well as an estimator, based on resampling, suggested by the author. In a Monte Carlo simulation we determined that the estimators which are most robust to non-convergence because of monotone likelihood are the Jeffreys invariant prior based on the expected information, the bootstrap estimator of the author, and Firth's estimator using the expected information (only if we consider weak convergence).

The Jeffreys prior recalls the man who has lost his keys on the sidewalk, but is searching for them in the street because there is more light there. By penalizing by the information, the estimated results from the Jeffreys prior are much smaller in magnitude than all other estimators, since there is more information for values of the parameters closer to zero. As more information accumulates with larger sample sizes, the estimates become more diverse.

Both the Firth and bootstrap estimators suffer from large mean squared errors, however most of this imprecision occurs because of large negative estimates of $\alpha$. If one accepts that for $\alpha = -5$ an estimate of $-15$ is poor, then no more useful practical information is gained about the estimator by penalizing it more for an estimate of $-200$ or even less. Hence, an evaluation of the estimators based on their ability to classify regions correctly may be more informative than classical statistical measures. Perhaps something along the lines of Mejail *et al.* (2003) can be explored. Also, since one can determine *a priori* if the MLE estimate of $\alpha$ diverges by examining the sample moments and one also knows that divergence implies homogeneity of the target region (Frery *et al.*, 1997, equation (7)) and not a lack of information about the data, it may be that failure to converge or large negative estimates of $\alpha$ do not have to be seen as weaknesses of likelihood based estimators.

This interpretation is supported by the results of the analysis of real data in the previous chapter. The Jeffreys estimator with expected information in Figure 4.20 on page 78 demonstrates this. It may converge more and have a smaller MSE for most cases, but it is uninformative in this real world situation whereas the bootstrap estimator seems to provide more information. The final conclusion about which estimators are better should be drawn based on future studies which test the abilities of the estimators to classify data into useful categories, rather than precisely estimate the roughness parameter.

CAPÍTULO A

# Resumo dos Capítulos em Português

## A.1 Capítulo 1: Introdução

Imagens formadas com radiação coerente oferecem alta resolução pelo uso de informação de fase. Entretanto, a coerência também contamina a imagem com ruídos de *speckle*. Em geral, a intensidade do sinal refletido depende de dois componentes: a reflexão da superfície e o ruído *speckle*. No modelo multiplicativo, esses componentes são multiplicados e considerados independentes (Frery *et al.*, 1997). A intensidade da reflexão da superfície depende da rugosidade do terreno alvo, entre outros fatores. Como a Figura 1.1 (European Space Agency, 2008) sugere, superfícies lisas que não são perpendiculares ao sinal do radar refletem pouca energia ao sensor e aparecem escuras, enquanto se elas forem perpendiculares (como no exemplo da montanha) elas aparecem claras na imagem. Florestas refletem quantidades intermediárias de radiação e plantações pequenas quantidades. Mesmo que uma área apareça clara em uma imagem com base nas condições do terreno, o *speckle* pode gerar uma mancha escura, já que a radiação refletida de outros alvos pode interferir e cancelar o sinal. Similarmente, manchas claras podem ocorrer em áreas que deveriam ser escuras devido ao *speckle*.

Nesse trabalho, estudamos as propriedades de um modelo estatístico de dados com *speckle* (o modelo $\mathcal{G}_A^0$), o qual tem sido usado para classificar alvos como heterogêneos, homogêneos, ou intermediários. As imagens que estudamos provêm de radares de abertura sintética (SAR). A abertura é sintética porque a antena se move acima do alvo e o observa de vários pontos distintos, simulando uma antena maior (ver Figura 1.3).

Imagine um piloto com SAR em seu avião que necessita pousar em uma área desconhecida coberta por nuvens ou à noite. É de extrema importância que ele pouse em uma área homogênea. Os modelos e estimadores estudados nessa tese podem auxiliar esse piloto aflito.

Nessa tese encontramos verossimilhança monótona para o estimador MV dos parâmetros da distribuição $\mathcal{G}_A^0$. Também desenvolvemos um teste simples para determinar se uma dada amostra sofrerá de verossimilhança monótona. Implementamos três estimadores disponíveis na literatura (Firth, 1993) para corrigir o viés causado pela verossimilhança monótona e também sugerimos e implementamos um quarto estimador de nossa autoria com base em reamostragem bootstrap. Executamos experimentos de Monte Carlo para estudar todos os estimadores e também aplicamos todos os estimadores a dados reais.

## A.2 Capítulo 2: A distribuição $\mathcal{G}_A^0$ e inferência por MV

O modelo que usamos para estudar imagens com *speckle* é o modelo $\mathcal{G}_A^0$ introduzido por Frery *et al.* (1997). Tal modelo descreve a amplitude da radiação refletida, para imagens heterogêneas. A distribuição $\mathcal{G}_A^0$ possui três parâmetros $(\alpha, \gamma, n)$ e sua função de densidade é dada na Equação (2.1). O espaço paramétrico é dado por $\alpha < 0$, $\gamma > 0$ e $n \geq 1$. Quando o primeiro parâmetro, $\alpha$ (o parâmetro de rugosidade), é próximo de zero ($-1.1$, por exemplo), os dados correspondem a áreas altamente heterogêneas (como áreas urbanas). Nesse caso, a distribuição tem caudas grossas como pode ser visto na Figura 2.1. Valores de $\alpha$ próximos de $-5$ correspondem a áreas intermediárias, como florestas, e as caudas são mais finas para essa distribuição. Valores de $\alpha$ menores do que aproximadamente $-15$ correspondem a áreas extremamente homogêneas, como campos com plantações. A densidade nesse caso é bastante concentrada, com pequeno peso nas caudas.

O parâmetro $\gamma$ é um fator de escala, o qual usaremos para normalização. Se $Z' \sim \mathcal{G}_A^0(\alpha, 1, n)$, então $\sqrt{\gamma}Z' \sim \mathcal{G}_A^0(\alpha, \gamma, n)$. O parâmetro $n$ corresponde ao número de "*looks*" (visadas) que formaram a imagem. Os momentos de $Z \sim \mathcal{G}_A^0(\alpha, \gamma, n)$ são dados pela Equação (2.2). A função de distribuição cumulativa de $\mathcal{G}_A^0$ é relacionada à distribuição $\mathcal{F}$ de Snedecor que é usada para gerar nossas variáveis aleatórias nas simulações de Monte Carlo. Com base na Equação (2.1) pode-se derivar a Equação (2.3) para a função de log-verossimilhança reduzida de uma amostra independente e identicamente distribuída (iid) de tamanho $N$. Note que nessa equação tomamos

o parâmetro $n$ como dado.

Estimativas de máxima verossimilhança (MV) dos parâmetros são obtidas pela maximização da Equação (2.3). Isso pode ser feito igualando-se a primeira derivada com respeito a cada parâmetro a zero, resultando na Equação (2.4). Como a Equação (2.4) não tem solução fechada, faz-se necessário utilizar rotinas de maximização não-linear para obter as estimativas de nossos parâmetros.

O algoritmo iterativo mais usado para maximização de funções não-lineares é o BFGS (Press *et al.*, 1992). Em Frery *et al.* (2004) os autores reportam que o algoritmo BFGS freqüentemente não converge quando usado para estimar os parâmetros ($\alpha$ and $\gamma$) da Equação (2.3). Nós confirmamos esse resultado. A má performance do BFGS é atribuída à natureza extremamente plana da função de verossimilhança, que faz com que a matriz hessiana seja próxima de singular e, conseqüentemente, não contenha informação útil.

Visando resolver esse problema Frery *et al.* (2004) sugerem um algoritmo iterativo que não depende de nenhuma informação sobre a matriz hessiana, além de não utilizar a informação do gradiente para determinar a próxima direção na qual alterar os parâmetros. Esse algoritmo (FCS) se alterna na maximização da função de verossimilhança com respeito a cada um dos parâmetros. Os autores reportam que esse algoritmo *sempre* converge, mesmo nas situações em que o BFGS deixa de convergir. Esse resultado também é confirmado nesse trabalho.

O último algoritmo que consideramos é o *steepest descent* (SD). Ao escolher a próxima direção na qual maximizar a função, SD se move na direção do gradiente.

Um aspecto crítico de qualquer processo iterativo é a determinação da convergência ou não do algoritmo. Seja $\delta(t) = p(t) - p(t-1)$; seja $\varepsilon_1$ e $\varepsilon_2$ constantes (*e.g.* $\varepsilon_1 = 10^{-4}$ e $\varepsilon_2 = 5 \times 10^{-3}$). Com base nessas definições, a linguagem de programação Ox (Doornik, 2001) define o critério de convergência na Equação (2.6), onde está claro que para $\varepsilon_1 \leq \varepsilon_2$ convergência forte implica convergência fraca. Nós usamos esse critério para determinar convergência em nossos experimentos.

Todos os algoritmos foram implementados na linguagem de programação Ox versão 3.1. Todos os algoritmos alcançaram o máximo de 10000 iterações e testes para convergência forte

foram feitos após cada iteração. Se após 10000 iterações convergência forte não tivesse sido alcançada, testamos para convergência fraca.

Dividimos nossa análise Monte Carlo em 80 experimentos de Monte Carlo menores, cada um deles indexado pela tripla $(n,\alpha,N)$, onde $n \in \{1,2,3,8\}$, $\alpha \in \{-15,-5,-3,-1\}$ e $N \in \{9,25,49,81,121\}$. Um subconjunto dos resultados de nosso experimento pode ser encontrado nas Tabelas 2.1 e 2.2. Em ambas as tabelas a tripla $(n,\alpha,N)$ indexa o experimento e $\gamma^*$ é calculado com base nesta tripla como na Equação (2.7). O experimento revela que apesar de FCS convergir em todas as situações, ele não fornece o maior valor da função de log-verossimilhança. Os algoritmos que fornecem os maiores valores da função de log-verossimilhança também retornam valores grandes dos parâmetros.

Por isso, expandimos a função log-verossimilhança, Equação (2.3), para valores grandes dos parâmetros $-\alpha$ e $\gamma$, com $n$ e $N$ fixos na Equação (2.8). Após alguns cálculos encontramos que (para $\mathcal{O}(\min(-\alpha,\gamma)^{-2})$) a função log-verossimilhança reduzida é dada pela Equação (2.11). Os dois primeiros termos são maximizados para $\gamma/(-\alpha) = \overline{z^2}$. Se os dois últimos termos forem negativos, então as estimativas MV são $(\hat{\alpha},\hat{\gamma})_{\mathrm{MLE}} = (-\infty,\infty)$ com $\hat{\gamma}/(-\hat{\alpha}) = \overline{z^2}$. Quando isso ocorre tem-se uma "verossimilhança monótona" (Bryson & Johnson, 1981). Isso ocorre quando a Equação (2.13) é satisfeita. Isso está em consonância com nossos resultados do Monte Carlo. Quando o algoritmo BFGS diverge, a Equação (2.13) é satisfeita e $\hat{\gamma}/(-\hat{\alpha}) = \overline{z^2}$. Ademais Frery *et al.* (1997) mostram que quando $-\alpha,\gamma \to \infty$ com $-\alpha/\gamma = \beta_2$ a distribuição $\mathcal{G}_A^0$ converge em distribuição para $\Gamma^{1/2}(n,n\beta_2)$, a qual é um modelo para dados extrememente homogêneos. Assim sendo, a divergência do parâmetro estimado implica em um alvo de rugosidade desprezivel.

Firth (1993) propõe uma correção de viés que pode ser usada com verossimilhança monótona. A idéia é deslocar a equação escore para que a nova solução seja finita e não-viesada (ver Figura 2.2). Duas alternativas para o deslocamento da escore são $A^{(E)} = -i(\theta)b_1(\theta)/N$, a qual usa $i(\theta) = \kappa_{r,s}$ (a informação esperada), e $A^{(O)} = -I(\theta)b_1(\theta)/N$, que usa $I(\theta) = -U_{rs}$ (a informação observada).

Firth observa que para um modelo da família exponencial na forma canônica sua cor-

reção é dada pela Equação (2.22) e a correção de viés corresponde a encontrar a moda da distribuição posterior após utilizar a priori invariante de Jeffreys (1946) (isto é, maximizando-se $L^* = L|i(\theta)|^{1/2}$).

O modelo $\mathscr{G}_A^0$ não é um modelo da família exponencial (menos ainda na forma canônica). Mesmo assim, implementamos a priori invariante de Jeffreys usando ambas as matrizes de informação observada e esperada, já que essa priori penaliza as regiões paramétricas onde a informação é pequena (ver Equações (2.27), (2.28), e (2.29)). Também implementamos a correção de Firth com base na informação esperada, dada na Equação (2.30).

Discutimos brevemente estimativas simples de viés usando bootstrap não-paramétrico, uma "melhor" estimativa de víes de Efron (1990), e uma adaptação da estimativa de Efron por Cribari-Neto *et al.* (2002).

Seja $X \overset{\text{iid}}{\sim} f_\theta(x)$ uma amostra aleatória de tamanho $N$ de uma distribuição $F_\theta$. O bootstrap não-paramétrico aproxima $F$ por $\hat{F}$, a função de distribuição empírica baseada nos dados. Gera-se amostras de $\hat{F}$ por amostragem com reposição a partir dos dados. Uma estimativa de viés por bootstrap não-paramétrico é formada tirando-se a média das estimativas *bootstap* do parâmetro, $\hat{\theta}_i$, sobre muitas amostras de $\hat{F}$ e subtraindo-se a estimativa original. Uma estimativa bootstrap corrigida para viés (BBC) é então formada pela subtração desse viés da estimativa original como na Equação (2.35).

A sugestão de Efron (1990) requer um pouco de notação. Em cada uma das $B$ reamostragens de bootstrap acima, a amostra pode ser descrita como o peso que cada uma das observações recebem na nova função de distribuição empírica. Na amostra original, cada observação recebe peso $1/N$. Isso pode ser sucintamente guardado em um vetor $P^0 = 1/N(1,\ldots,1)$. Para a $i$-ésima amostra bootstrap esse vetor pode ser representado por

$$P^i = \frac{1}{N}\left(\#\{X_1\}_i, \ldots, \#\{X_N\}_i\right),$$

onde $\#\{X_j\}_i$ representa o número de vezes que $X_j$ aparece na $i$-ésima amostra bootstrap. O vetor é chamado de vetor de reamostragem. A idéia de Efron é baseada na possibilidade de escrever o parâmetro estimado como uma função fechada dos dados usando o vetor $P^0$. Por

exemplo, a estimativa da média pode ser escrita como $\bar{X} = T(P^0) = P^0 \cdot x$. Uma estimativa do segundo momento poderia ser escrita como $\bar{X^2} = T(P^0) = P^0 \cdot (x^2)$.

Visando acelerar a convergência da estimativa do viés de forma que menos repetições de bootstrap sejam necessárias, Efron sugere que ao se calcular a estimativa de viés subtraia-se a estimativa do parâmetro resultante do uso de

$$P^* = \frac{1}{B}\sum_{i=1}^{B} P_i,$$

$T(P^*)$. A "melhor" estimativa da correção de viés bootstrap (BBBC) é dada pela Equação (2.36).

Como Cribari-Neto *et al.* (2002) observam, os estimadores MV para o modelo $\mathscr{G}_A^0$ não possuem forma fechada. Contudo, para implementar o BBBC eles usam uma abordagem similar à de Firth acima. Eles escrevem as equações a serem estimadas como uma função de $P^0$ e então usam as estimativas obtidas substituindo $P^0$ por $P^*$ nas equações estimadas para corrigir para viés. Reescrevem a função de log-verossimilhança na Equação (2.3) como Equação (2.37). Podem então calcular a $T(P^*)$ obtida através da substituição de $P^0$ por $P^*$ na Equação (2.37) e a inserção desta na Equação (2.36) para gerar sua estimativa BBBC.

Aqui, sugerimos um estimador baseado em reamostragem bootstrap que funciona mesmo na presença de verossimilhança monótona. Primeiro definimos o estimador e então discutimos suas propriedades. Na Equação (2.13), na página 21, definimos os critérios que caso satisfeitos levam a estimativas infinitas do parâmetro de MV. Nossa sugestão é efetuar um bootstrap não-paramétrico dos dados, mantendo apenas as amostras bootstrap para as quais o critério de divergência não é satisfeito até que um certo número pré-determinado de amostras bootstrap não-divergentes seja obtido. Então, nossa estimativa é dada por $T(P^*)$, como calculada na Equação (2.37). Apesar desse estimador requerer um considerável número de reamostragens e verificações do critério de divergência, ela requer apenas uma maximização não-linear.

Para que o nosso estimador seja útil, ele deve ser consistente e deve existir e ser finito para a grande maioria das amostras. A seguir estão os possíveis problemas que podem ocorrer com o estimador: (i) nenhuma amostra bootstrap existe para a qual o critério de divergência não seja satisfeito, Equação (2.13); (ii) mesmo se tivermos $B$ amostras bootstrap válidas, a pseudo-

amostra correspondente a $P^*$ pode satisfazer Equação (2.13); (iii) o estimador pode não ser consistente (convergir em probabilidade para o verdadeiro valor do parâmetro). Nos parágrafos seguintes tratamos cada um dos problemas listados acima.

A solução do problema (i) é resumida na Tabela 2.3. Ali nós tabulamos a probabilidade de gerar uma amostra para a qual todas as amostras bootstrap tenham verossimilhança monótona. A tabela mostra que o problema: piora para grandes valores de $-\alpha$; varia não-monotonicamente com $n$; e melhora com um maior número de observações, $N$. Em nossos experimentos Monte Carlo usando o estimador bootstrap, usamos o menor tamanho amostral de $N = 49$. Nesse caso, a maior probabilidade de que nosso estimador não existirá é aproximadamente $1 \times 10^{-8}\%$.

O problema (ii) é a possibilidade de que a pseudo-amostra correspondente a $P^*$ satisfará o critério de divergência, Equação (2.13), mesmo que cada amostra bootstrap usada para construi-la não satisfaça. Na tese, demonstramos algebricamente que isso não é possível.

O problema (iii) (que o estimador pode não ser consistente) é solucionado pela demonstração de que nosso estimador converge para o estimador MV, o qual é consistente e assim sendo nosso estimador também é consistente.

## A.3 Capítulo 3: Um estudo numérico dos estimadores MV, Firth e bootstrap

Para avaliar as correções descritas nas Seções 2.4.3 e 2.4.4, implementamos um experimento Monte Carlo com 32 sub-experimentos onde, para cada tripla $(n, \alpha, N)$ com $n \in \{1,2,3,8\}$, $\alpha \in \{-15,-5,-3,-1\}$, e $N \in \{49,121\}$, geramos 10000 amostras com $N$ observações de uma distribuição $\mathscr{G}_A^0(\alpha,\gamma,n)$ com $\gamma$ satisfazendo Equação (2.7). Para cada amostra, estimativas de $\alpha$ e $\gamma$ foram obtidas usando MV, MV com a priori invariante de Jeffreys calculado usando a matriz de informação esperada e observada, e o estimador de Firth (quatro estimadores). As primeiras três estimações são feitas usando o algoritmo BFGS (Press *et al.*, 1992). A última é

feita usando-se um solucionador de equações não-lineares.

Os resultados da Tabela 3.1 mostram que a porcentagem de amostras que satisfazem o critério de divergência é um pouco maior que a porcentagem de vezes que o algoritmo BFGS não convergiu fortemente quando maximizando a log-verossimilhança não ajustada.

O uso da priori invariante de Jeffreys com informação esperada tem o maior impacto na falta de convergência forte. Em quase todos os casos BFGS convergiu fortemente com esta verossimilhança. A correção de Firth com informação esperada apresentou as piores propriedades de convergência forte após MV. Isso pode ser atribuído ao uso do solucionador não-linear. As taxas de convergência forte para o estimador usando a priori invariante de Jeffreys com informação observada foram baixas, especialmente para $n$ pequeno.

Os resultados para convergência fraca estão na Tabela 3.2. Uma comparação com a tabela anterior revela que os estimadores que usam as prioris invariantes de Jeffreys ou convergem fortemente ou não convergem. Os estimadores MV e Firth convergem fracamente freqüentemente, mas não fortemente. O estimador MV e o que usa a priori invariante de Jeffreys com informação observada têm taxas altas de não convergência.

Nas Tabelas 3.3 e 3.4 apresentamos o viés estimado para cada estimador de $\alpha$ calculado usando todas as amostras e somente as amostras para as quais o critério de divergência não é violado, respectivamente. MV melhora com a exclusão das amostras divergentes. O estimador que usa a priori de Jeffreys com informação esperada quase não é afetado quando amostras divergentes são excluídas. O estimador de Firth melhora seu viés quando as amostras divergentes são excluídas. O estimador que usa a priori de Jeffreys com informação observada é pouco afetado pela exclusão de amostras divergentes. A Tabela 3.3 e a Figura 3.1 não sugerem que uma das três modificações seja consistentemente menos viesada.

Consideramos o erro quadrático médio (EQM) dos estimadores de $\alpha$ nas Tabelas 3.5 e 3.6, novamente com e sem as amostras divergentes, respectivamente. O estimador MV e a correção de Firth sofrem aumentos nos EQMs quando amostras divergentes são incluídas. O EQM do estimador obtido numericamente através da priori esperada de Jeffreys quase não é afetado enquanto o EQM daquela com informação observada é levemente afetado. Examinando a

Figura 3.3 podemos ver que não existe uma escolha clara entre esses dois estimadores porque não sabemos de antemão qual valor do parâmetro está sendo estimado. Contudo, para $n = 8$ (que é informação conhecida) a priori de Jeffreys com informação esperada é claramente melhor que todas as outras alternativas. Em quase todos os casos, quando o tamanho da amostra aumenta, os estimadores melhoram seus viés e EQM.

Reportamos os resultados do mesmo experimento Monte Carlo descrito acima para o estimador bootstrap com $10 \times N$ amostras bootstrap. Na Tabela 3.1 verificamos que a verossimilhança monótona quase não afeta a convergência do estimador bootstrap. Uma comparação com a Tabela 3.2 mostra que o estimador bootstrap ou converge fortemente ou não converge.

A comparação das últimas colunas das Tabelas 3.3 e 3.4, mostra que o viés do estimador bootstrap é pouco afetado pela inclusão das amostras divergentes. Considerando somente a Tabela 3.3 e a Figura 3.1 podemos ver que o viés do estimador bootstrap é consistentemente entre os menores vieses para todos os valores de $\alpha$ (com exceção de quando $n = 8$). A Figura 3.2 mostra uma comparação direta entre o estimador que usa a priori invariante de Jeffreys com informação esperada e o estimador bootstrap. Eles são comparáveis, mas o estimador que usa a priori de Jeffreys continua dominante quando $n = 8$, enquanto o estimador bootstrap domina quando $\alpha = -15$.

Comparando as últimas colunas das Tabelas 3.5 e 3.6, observamos que o EQM do estimador bootstrap é inflado pela inclusão de amostras divergentes. Esta inflação é pior quando $\alpha = -15$, fazendo o estimador ficar impreciso. Mesmo assim, podemos observar na Figura 3.3, que para $n = 1$ ele se compara favoravelmente com o estimador que usa a priori invariante de Jeffreys com informação observada e não é nem pior nem melhor que o que usa a informação esperada. Para $n = 2$ o EQM geralmente fica entre as prioris. Para valores de $n$ maiores, o EQM grande quando $\alpha = -15$ torna o uso do estimador pouco recomendável.

## A.4 Capítulo 4: Aplicação a dados reais dos estimadores MV, Firth e bootstrap

Nos experimentos Monte Carlo do Capítulo 3 tratamos o número de *looks*, $n$, como conhecido. Quando lidamos com dados reais, este parâmetro precisa ser estimado quando $n > 1$ e neste caso é chamado do número equivalente de *looks*. Para estimar este parâmetro usamos o fato que a amplitude do sinal é $Z_A = Y_A \cdot X_A$ onde $Y_A$ representa o ruído *speckle* e $X_A$ representa o sinal da superfície. Seguimos Frery *et al.* (1997) e assumimos que $Y_A \sim \Gamma^{1/2}(n,n)$. Para estimar $n$ de nossas observações de $Z_A$ seguimos o procedimento de Yanasse *et al.* (1994). O procedimento consta de identificar regiões homogêneas da imagem a ser analisada e assumir que $X_A = c$ com $c$ uma constante desconhecida. Com essa informação derivamos a Equação (4.1). A solução desta equação em áreas homogêneas resulta na estimativa do número equivalente de *looks* para a imagem inteira.

Para as imagens neste capítulo todos os valores de amplitude maiores que o terceiro quartil receberam o valor do terceiro quartil. Note que isto foi feito somente para os gráficos e não para as estimações. Similarmente, os valores estimados do parâmetro de rugosidade, $\alpha$, que são menores que $-15$ receberam o valor $-15$ nos gráficos. Além disso, resultados usando a priori invariante de Jeffreys com informação observada não são reportados. Isto porque a matriz de informação observada foi singular para muitos casos e o algoritmo BFGS ou não se iniciou ou logo não convergiu.

A Figura 4.1 mostra uma imagem de um único *look* com pouca variação. A imagem foi dividida em 2116 blocos de 11 pixels por 11 pixels. As 121 observações em cada bloco foram usadas para estimar $\alpha$ e $\gamma$ naquele bloco.

O critério de divergência foi satisfeito por 47.5% dos blocos e o estimador MV não convergiu em mais de 51% dos blocos. O estimador bootstrap não convergiu em 7% dos blocos. O estimador baseado na priori invariante de Jeffreys com informação esperada convergiu para todos os blocos, mas gerou estimativas pequenas (em magnitude) de $\alpha$. Dado que a figura é homogênea, esperamos maiores (em magnitude) valores de $\alpha$. Assim sendo, a priori invariante

de Jeffreys com informação esperada parece fraca para a identificação de regiões homogêneas.

A Figura 4.2 sobrepõe as estimativas de MV para $\alpha$ sobre a imagem original da Figura 4.1. Como nossa motivação foi a de um piloto conseguir pousar numa área desconhecida, áreas heterogêneas são identificadas com cores vermelhas enquanto regiões homogêneas são identificadas por branco e regiões intermediárias por amarelo. As fronteiras entre regiões são misturas de distribuições $\mathscr{G}_A^0$ diferentes e são identificadas como heterogêneas. Quando o estimador MV não converge as estimativas de $\alpha$ usualmente são bem negativas, implicando homogeneidade. Por isso existem muitos blocos brancos, pois o estimador MV não convergiu em 51% dos blocos.

Uma representação semelhante para o estimador bootstrap se encontra na Figura 4.3. Muitos dos blocos brancos da figura anterior agora são amarelos, indicando rugosidade intermediária. Como sabemos que as estimativas por MV são viesadas para homogeneidade, um piloto deveria evitar as áreas com muitos blocos amarelos nesta figura. Claro, não podemos confirmar a validade de nossos resultados sem comparar com as condições na superfície. Infelizmente, não temos esta informação.

Agora analisamos uma imagem com *looks* múltiplos. Precisamos estimar o número equivalente de *looks*, como descrito acima, antes de proceder à estimação de rugosidade. A Figura 4.5 mostra a imagem e as regiões usadas para identificar o número equivalente de *looks*. Podemos ver a grande complexidade desta imagem.

O número equivalente de *looks* foi estimado para cada uma das 38 regiões na figura solucionando-se a Equação (4.1). Resultados comparando histogramas com as densidades estimadas estão nas Figuras 4.6–4.10. Os dados são bem descritos pelo modelo $\Gamma^{1/2}$. Usamos $n = 3.2$ para estimar $\alpha$.

Estimamos $\alpha$ para cada uma dos 95256 blocos de 11 por 11 pixels. Os resultados do estimador bootstrap estão na Figura 4.11. Para esta imagem, o critério de divergência foi satisfeito em 9% dos blocos e o estimador MV não convergiu em 11% das amostras. O estimador bootstrap não convergiu em menos que 4% dos blocos enquanto os estimadores Jeffreys e Firth com informação esperada não convergiram para 1.5% das amostras. Uma área homogênea de alta

intensidade foi identificada perto da coluna 1000 e linha 100.

As Figuras 4.12–4.14 comparam os resultados do estimador bootstrap com os estimadores MV, Jeffreys com informação esperada e Firth com informação esperada, respectivamente. O estimador bootstrap identifica mais heterogeneidade que o MV e menos que o Jeffreys e o Firth.

Agora, analisamos dados no formato *look* único complexo. Estes dados podem ser compactados por coluna ou por linha para criar dados de *looks* múltiplos. Dessa maneira podemos estudar o desempenho do estimador bootstrap e os outros estimadores na mesma imagem com números diferentes de *looks*.

A Figura 4.18 mostra um painel da amplitude da imagem original e das estimativas de $\alpha$ do estimador bootstrap. Nenhuma região é considerada uniformemente homogênea e existem muitas regiões com rugosidade intermediária. Aproximadamente 47% dos 37044 blocos de 11 por 11 pixels respeitam o critério de divergência. O estimador bootstrap não convergiu em 7% dos blocos enquanto o estimador MV não convergiu em 60% dos blocos. Ambos os estimadores Firth e Jeffreys com informação esperada convergiram para todas as amostras.

As Figuras 4.19–4.21 contêm painéis com as estimativas bootstrap no painel de cima e as estimativas MV, Jeffreys e Firth no painel de baixo. O estimador MV é viesado para homogeneidade e a imagem mostra isso porque existem muitos blocos azuis. As estimativas de Jeffreys são viesadas para heterogeneidade e são ruins. As estimativas Firth são semelhantes às estimativas bootstrap, mas um pouco mais heterogêneas.

Combinamos nossos dados em grupos de oito colunas para gerar dados de *looks* múltiplos. O resultado se encontra na Figura 4.22. As áreas usadas para estimar o número equivalente de *looks* são demarcadas por curvas vermelhas. A imagem resultante é muito mais homogênea, mas ainda contém ruído *speckle*. As áreas de alta e baixa amplitude podem ser identificadas mais claramente.

As Figuras 4.23–4.25 contêm histogramas das amplitudes para cada uma das áreas vermelhas na Figura 4.22 e a densidade estimada. Novamente, os dados são bem descritos pelo modelo $\Gamma^{1/2}$. Usamos 5.9 como o número equivalente de *looks* na análise que segue.

A Figura 4.26 mostra a imagem de *looks* múltiplos no painel superior e as estimativas boot-

strap de $\alpha$ em cada um dos 11352 blocos de 7 por 7 pixels no painel inferior. Existem regiões contíguas planas e regiões contíguas ásperas enquanto na imagem com dados de somente um *look* existia muito mais variação nas estimativas de rugosidade.

Para esta imagem de *looks* múltiplos, usando blocos de 7 por 7, o critério de divergência foi satisfeito em 32% dos blocos. O estimador MV não convergiu em 41% dos blocos, o estimador bootstrap em 13% (quase duas vezes a porcentagem do caso de um *look*), e os estimadores Jeffreys e Firth em somente um bloco.

Usando 4620 blocos de 11 por 11 pixels, o critério de divergência foi satisfeito em 19% dos blocos. O estimador MV não convergiu em 28% dos blocos, o estimador bootstrap em 12% e os estimadores Jeffreys e Firth em 2%.

As Figuras 4.27–4.29 comparam o estimador bootstrap usando 49 e 121 observações com os estimadores MV, Jeffreys e Firth, respectivamente. O estimador MV continua sendo viesado para homogeneidade, mas é bem semelhante ao bootstrap para $N = 121$. As estimativas do estimador bootstrap não mudam muito de $N = 49$ para $N = 121$, uma propriedade desejável se mais dados fornecem resultados melhores. O estimador Jeffreys continua fortemente viesado para rugosidade quando $N = 49$ e nem tanto para $N = 121$ com a formação de regiões contíguas de homogeneidade. As estimativas bootstrap e Firth são muito semelhantes, com as estimativas Firth um pouco mais para rugosidade que as estimativas bootstrap.

## A.5 Capítulo 5: Conclusões e direções futuras

Identificamos a existência de verossimilhança monótona em amsotras $\mathscr{G}_A^0$ que satisfazem o critério de divergência, Equação (2.13). Implementamos três estimadores para correção de viés propostos por Firth (1993) e Jeffreys (1946) para resolver esse problema, bem como um quarto estimador, com base em reamostragem, proposto pelo autor. Em uma simulação de Monte Carlo, determinamos que os estimadores que são mais robustos a não-convergência devido à verossimilhança monótona são a priori invariante de Jeffreys (com base na informação esperada), o estimador bootstrap do autor e o estimdor de Firth usando a informação esperada

(apenas quando se considera convergência fraca).

Ambos os estimadores de Firth e bootstrap apresentam erros quadráticos médios elevados, contudo a maior parcela dessa imprecisão ocorre devido a grandes estimativas negativas de $\alpha$. Se concordarmos que para $\alpha = -5$ uma estimativa de $-15$ é pobre, então não se ganha nenhuma informação prática sobre o estimador através de uma penalização ainda maior do mesmo para uma estimativa de $-200$ ou até menos. Assim, uma avaliação dos estimadores que tenha como base sua habilidade para classificar regiões corretamente pode ser mais informativa do que medidas estatísticas clássicas. Talvez algo na linha de Mejail *et al.* (2003) possa ser explorado. Ademais, como é possível determinar *a priori* se uma estimativa MV de $\alpha$ diverge examinando-se os momentos amostrais e também é sabido que divergência implica homogeneidade da região alvo (Frery *et al.*, 1997, Equation (7)), e não uma falta de informação sobre os dados, pode ser que a incapacidade de convergir ou estimativas negativas grandes de $\alpha$ não precisem ser vistas como uma fraqueza dos estimadores com base na verossimilhança.

Essa interpretação é sustentada pelos resultados da análise de dados reais no capítulo precedente. O estimador de Jeffreys com informação esperada na Figura 4.20 da página 78 demonstra isso. Ele pode convergir mais e ter um erro quadrático médio menor para a maioria dos casos, mas não é informativo naquela situação de mundo real, enquanto o estimador bootstrap parece fornecer mais informação. A conclusão final sobre que estimadores são melhores pode ser tirada com base em estudos futuros que testem a capacidade dos estimadores de classificar os dados em categorias úteis, ao invés de estimar precisamente o parâmetro de rugosidade.

# Bibliography


Barndorff-Nielsen, O. E. 1983. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

Bryson, M. C., & Johnson, M. E. 1981. The Incidence of Monotone Likelihood in the Cox Model. *Technometrics*, **23**, 381–383.

Clarkson, D. B., & Jennrich, R. I. 1991. Computing Extended Maximum Likelihood Estimates for Linear Parameter Models. *Journal of the Royal Statistical Society B*, **53**, 417–426.

Cordeiro, G. M., & Cribari-Neto, F. 1998. On Bias Reduction in Exponential and Non-exponential Family Regression Models. *Communications in Statistics, Simulation and Computation*, **27**, 485–500.

Cordeiro, G. M., & McCullagh, P. 1991. Bias Correction in Generalized Linear Models. *Journal of the Royal Statistical Society B*, **53**, 629–643.

Cox, D. R., & Reid, N. 1987. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**, 1–39.

Cox, D. R., & Reid, N. 1989. On the stability of maximum-likelihood estimators of orthogonal parameters. *The Canadian Journal of Statistics*, **17**(2), 229–233.

Cribari-Neto, F., Frery, A. C., & Silva, M. F. 2002. Improved Estimation of Clutter Properties in Speckled Imagery. *Computational Statistics & Data Analysis*, **40**, 801–824.

Doornik, J. A. 2001. *Ox: An Object-Oriented Matrix Programming Language*. Fourth edn. London: Timberlake Consultants and Oxford. http://www.doornik.com.

Efron, B. 1990. More Efficient Bootstrap Computations. *Journal of the American Statistical Association*, **85**, 79–89.

Efron, B., & Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall/CRC.

European Space Agency. 2008. *Advanced Synthetic Aperture Radar User Guide*. European Space Agency. http://envisat.esa.int/handbooks/asar/.

Firth, D. 1993. Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, **80**, 27–38.

Fraser, D. A. S., & Reid, N. 1995. Ancillaries and third-order significance. *Utilitas Mathematica*, **47**, 33–53.

Fraser, D. A. S., Reid, N., & Wu, J. 1999. A simple formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, **86**, 655–661.

Frery, A. C., Müller, H.-J., Yanasse, C.C.F., & Sant'Anna, S. J. S. 1997. A Model for Extremely Heterogeneous Clutter. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 648–659.

Frery, A. C., Cribari-Neto, F., & Souza, M. O. 2004. Analysis of Minute Features in Speckled Imagery with Maximum Likelihood Estimation. *EURASIP Journal on Applied Signal Processing*, **16**, 2476–2491.

Fried, D. L. 1966. Optical Resolution Through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures. *Journal of the Optical Society of America*, **56**, 1372–1379.

Heinze, G., & Schemper, M. 2001. A Solution to the Problem of Monotone Likelihood in Cox Regression. *Biometrics*, **57**, 114–119.

Jeffreys, H. 1946. An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society A*, **186**, 453–461.

Kolassa, J. E. 1997. Infinite Parameter Estimates in Logistic Regression, with Application to Approximate Conditional Inference. *Scandanavian Journal of Statistics*, **24**, 523–530.

Labeyrie, A. 1970. Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images. *Astronomy and Astrophysics*, **6**, 85.

Leung, D. H.-Y., & Wang, Y.-G. 1998. Bias Reduction using Stochastic Approximation. *Australian and New Zealand Journal of Statistics*, **40**, 43–52.

Lins, L. D., & Pianto, D. M. 2004. *Maximization of $\mathscr{G}_A^0$ Likelihood Functions*. Mimeo.

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

McCullagh, P. 1987. *Tensor Methods in Statistics*. London: Chapman and Hall.

Mejail, M. E., Jacobo-Berlles, J., Frery, A. C., & Bustos, O. H. 2000. Parametric Roughness Estimation in Amplitude SAR Images Under the Multiplicative Model. *Revista de Teledetección*, **13**, 37–49.

Mejail, M. E., Frery, A. C., Jacobo-Berlles, J., & Bustos, O.H. 2001. Approximation of Distributions for SAR Images: Proposal, Evaluation and Practical Consequences. *Latin American Applied Research*, **31**, 83–92.

Mejail, M. E., Jocobo-Berlles, J., Frery, A. C., & Bustos, O. H. 2003. Classification of SAR images using a general and tractable multiplicative model. *International Journal of Remote Sensing*, **24**(18), 3565–3582.

Press, W. H., Teukolosky, S. A., Vetterling, W. T., & Flannery, B. P. 1992. *Numerical Recipes in C: the Art of Scientific Computing*. Second edn. Cambridge: Cambridge University Press.

Quenouille, M. H. 1957. Notes on Bias in Estimation. *Biometrika*, **43**, 353–360.

R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Sartori, N. 2006. Bias Prevention of Maximum Likelihood Estimates for Scalar Skew Normal and Skew *t* Distributions. *Journal of Statistical Planning and Inference*, **136**, 4259–4275.

Schaefer, R. L. 1983. Bias Correction in Maximum Likelihood Logistic Regression. *Statistics in Medicine*, **2**, 71–78.

Silva, M. F., Cribari-Neto, F., & Frery, A. C. 2007. Improved Likelihood Inference for the Roughness Parameter of the $\mathscr{G}_A^0$ Distribution. *Environmetrics*. Forthcoming.

Vasconcellos, K. L. P., Frery, A. C., & Silva, L. B. 2005. Improving Estimation in Speckled Imagery. *Computational Statistics*, **20**, 503–519.

Yanasse, C. C. F., Frery, A. C., Sant'Anna, S. J. S., Filho, P. H., & Dutra, L. V. 1994. Statistical analysis of SAREX data over Tapajos—Brazil. *Pages 25–40 of: Workshop Proceedings, SAREX-92 South American Radar Experiment*.