"Science Stays True Here"
Journal of Mathematics and Statistical Science, 295-318 | Science Signpost Publishing

A Bayesian approach to Generalized Signed-Rank Estimation for Nonlinear Models with Multidimensional Indices

Eddy Kwessi [*], **Brice M. Nguelifack** , **Guy-vanie Miakonkana**

Trinity University, 1 Trinity Place, San Antonio, TX 78258, USA

United States Naval Academy, Annapolis, MD 21402 , USA

African School of Economics, Princeton University, Princeton, NJ 08544, USA

Abstract

In this paper, we propose a Bayesian estimation method for generalized signed-rank estimates in nonlinear models with multidimensional indices. Simulations of the Bayesian posterior parameters using Markov Chain Monte Carlo approach are given.

Keywords: Bayesian; Generalized signed-rank; Nonlinear; Multidimensional; Estimation.

AMS Subject Classification: 62G30; 62G35; 94A12; 62J02.

[*]Corresponding author

1. Introduction

There has been an increased focus on spatio-temporal models with multidimensional indices over the past decades. Applications of these models span many areas of scientific research: In signal processing and spatio temporal modelling: Rao et al. (1994); McClellan (1982), in texture modeling: Francos et al. (1993), Yuan and Rao (1992), Zhang and Mandrekar (2001), in acoustics: Irizary (2011), in electronics: Sánchez et al. (2011); Zavala and Messina (2014), in bioinformatics: Yang and Su (2010), in statistics: Kundu (1993), Dunson and Taylor (2005) (variable selections), Brown (1990) (asymptotic distributions of parameters).

Most of these models are single-index and nonlinear, and an extensive literature on these type of models can be found for example in Wu (1981); Jennrich (1969), Gallant (1987). In this paper, we focus on multidimensional nonlinear models given classically by the equation

$$y_{\mathbf{t}} = f(x_{\mathbf{t}}, \boldsymbol{\theta}) + \epsilon_{\mathbf{t}}, \quad \mathbf{t} \leq \mathbf{n}, \tag{1.1}$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_K)$ and $\mathbf{n} = (n_1, n_2, \ldots, n_K) \in \mathbb{N}^K$ which is the space of K dimensional non-negative integer values; $\leq$ denotes the partial ordering on $\mathbb{N}^K$, that is, for $\mathbf{m} = (m_1, m_2, \ldots, m_K) \in \mathbb{N}^K$ and $\mathbf{n} = (n_1, n_2, \ldots, n_K) \in \mathbb{N}^k$, $\mathbf{m} \leq \mathbf{n}$ if $m_k \leq n_k$ for $k = 1, 2, \ldots, K$. The set $\{\epsilon_{\mathbf{t}}, \mathbf{t} \in \mathbb{N}^K\}$ is a φ-mixing (a weakly dependent) field of random variables with mean 0, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ is a parameter vector, $\{x_{\mathbf{t}}, \mathbf{t} \in \mathbb{N}^K\}$ a set of known field of constant vectors, and f is a known nonlinear link function. We will assume in this paper that the link function $f(\cdot, \cdot)$ is real-valued. We note however that there are areas such signal processing (see Rao et al. (1994)) where $f(\cdot, \cdot)$ is considered complex-valued. Our goal in this paper is to find Bayesian estimators for $\boldsymbol{\theta}$. Indeed, strongly consistent and asymptotically normal (as $|\mathbf{n}| = \Pi_{i=1}^k n_i \to \infty$) signed-rank estimators for $\boldsymbol{\theta}$ were proposed by Nguelifack et al. (2015) with sufficient conditions on the nonlinear link function given. The present paper

is a continuation of that endeavor but into a Bayesian setting. It is already well-established in the literature that nonparametric methods, in some situations, can be more efficient than traditional parametric methods in particular when the true distribution of the data is heavy-tailed or skewed (see Hettmansperger and Mckean (2011)). With the advent of computers powerful enough to handle long and complicated programs the Bayesian approach has been intensively used in theoretical as well as practical applications. This approach is built on the assumption that there is a state $\boldsymbol{\theta}$, observed data $x_{\mathbf{t}}$, a prior function $p(\boldsymbol{\theta})$, and the likelihood function $p(x_{\mathbf{t}}|\boldsymbol{\theta})$ over possible observations. One of the goals is to find the posterior function $p(\boldsymbol{\theta}|x_{\mathbf{t}})$ as a function proportional to $p(x_{\mathbf{t}}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ using Bayes' theorem. It is often the case that $p(\boldsymbol{\theta}|x_{\mathbf{t}})$ can be difficult to compute in a closed form, so we will rely, as it is often the case in practice, on approximations and simulations.

The Bayesian nonparametric inference based on ranks has been studied by Zhan and Hettmansperger in Zhan and Hettmansperger (2005), Zhan and Hettmansperger (2007), Zhan and Hettmansperger (2009). The framework they proposed is these series of papers works well for linear link functions and single-index parameter spaces. In this paper, we propose a Bayesian estimation method for nonlinear link functions $f(x_{\mathbf{t}}, \boldsymbol{\theta})$ and for multi-indices spaces. These two distinctions allow for more flexibility and a broader ranges of possible applications. The remainder of the paper is organized as follows: in section 2.1 and section 2.2, we review Generalized Signed-Rank estimators (GSR) and their asymptotic normality, laying work the ground for a frequentist approach. In section 3, we derive an asymptotic posterior distribution for the unknown parameter $\boldsymbol{\theta}$. In section 3.3, we propose simulations results using the asymptotic distribution derived earlier as well as a Markov Chain Monte Carlo method. In Section 4, we provide a discussion about the relevance of this line of research.

2. Frequentist approach

In the following two subsections, we discussed the relevant details about a frequentist approach to the generalized signed-rank estimation for models with multidimensional indices.

2.1 Generalized Signed-Rank Estimation

Consider model (1.1). We shall assume that $\boldsymbol{\theta}$ is in the parameter space Θ, $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ which is an interior point of Θ, and $x \in \mathbb{R}^p$. We define the GSR estimator of $\boldsymbol{\theta}_0$ to be any vector $\widehat{\boldsymbol{\theta}}_S$ minimizing

$$D_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \leq \mathbf{n}} a_{\mathbf{n}}(\mathbf{t}) \rho(|z(\boldsymbol{\theta})|_{(|\mathbf{t}|)}) \tag{2.1}$$

where $z_{\mathbf{t}}(\boldsymbol{\theta}) = y_{\mathbf{t}} - f(x_{\mathbf{t}}, \boldsymbol{\theta})$ and $|z(\boldsymbol{\theta})|_{(|\mathbf{t}|)}$ is the $\mathbf{t}^{th}$ ordered value among $|z_{\mathbf{1}}(\boldsymbol{\theta})|, \ldots, |z_{\mathbf{n}}(\boldsymbol{\theta})|$. The function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, convex and strictly increasing. The numbers $a_{\mathbf{n}}(\mathbf{t})$ are scores generated as $a_{\mathbf{n}}(\mathbf{t}) = \varphi(|\mathbf{t}|/(|\mathbf{n}|+1))$, for some bounded nondecreasing score function $\varphi : (0,1) \to \mathbb{R}^+$ that has at most a finite number of discontinuities. Since $D_{\mathbf{n}}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, Lemma 2 in Jennrich (1969) implies the existence of a minimizer of $D_{\mathbf{n}}(\boldsymbol{\theta})$.

It is clear that the the least squares (LS) and the least absolute deviation (LAD) estimators are particular cases of GSR estimators. In fact the LS is obtained by taking $\varphi^+ \equiv 1$ and $\rho(t) = t^2, t \geq 0$ while the LAD is obtained by taking $\varphi^+ \equiv 1$ and $\rho(t) = t$. The case of the LS has been discussed by N. Bansal et al. Bansal et al. (1999) and S. Nandi Nandi (2012) among others.

We will however use an equivalent form

$$D_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \leq \mathbf{n}} a_{\mathbf{n}}(\mathbf{t}) (\rho \circ \tilde{G}_{\boldsymbol{\theta}}^{-1})(\xi_{(|\mathbf{t}|)}), \tag{2.2}$$

where $\xi_{(|\mathbf{t}|)}$ are order statistics from the uniform $U(0,1)$ distribution.

Put $\Gamma_{\boldsymbol{\theta}}(s) = \rho[\tilde{G}_{\boldsymbol{\theta}}^{-1}(s)]$ for $s \in [0,1]$ and $\lambda_{\mathbf{t}} = a_{\mathbf{n}}(R(\xi_{|\mathbf{t}|}))$, where $R(\xi_{|\mathbf{t}|})$ is the rank of $\xi_{|\mathbf{t}|}$ among $\xi_{\mathbf{1}}, \ldots, \xi_{\mathbf{n}}$. Then (2.1) can be rewritten as

$$D_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \leq \mathbf{n}} a_{\mathbf{n}}(\mathbf{t})(\rho \circ \tilde{G}_{\boldsymbol{\theta}}^{-1})(\xi_{|\mathbf{t}|}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \leq \mathbf{n}} \lambda_{\mathbf{t}} \Gamma_{\boldsymbol{\theta}}(\xi_{|\mathbf{t}|}).$$

Since φ^+ is bounded, we have $\|\lambda_{\mathbf{t}}\| < \infty$. Set $S_{\mathbf{n}}(\boldsymbol{\theta}) = \mathcal{D}_{\boldsymbol{\theta}}^{\beta} D_{\mathbf{n}}(\boldsymbol{\theta})$ and $\Phi_{\boldsymbol{\theta}}(s) = \mathcal{D}_{\boldsymbol{\theta}}^{\beta} \Gamma_{\boldsymbol{\theta}}(s)$ for $|\beta| = 1$, where $\mathcal{D}_{\boldsymbol{\theta}}^{\beta}$ is the differential operator defined by $\mathcal{D}_{\boldsymbol{\theta}}^{\beta} = \frac{\partial^{|\beta|}}{\partial \theta_1^{\beta_1} \ldots \partial \theta_p^{\beta_p}}$ with $\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{N}_0^n, \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ the multi-index and $|\beta| = \sum_{i=1}^{p} \beta_i$. Let the $|\mathbf{n}| \times p$ matrix $\boldsymbol{X}^*$ be the matrix of $\Phi_{\boldsymbol{\theta}}$ evaluated at all $|\mathbf{n}|$ residuals $z(\boldsymbol{\theta})$ and $h_{\mathbf{tt}}^{\mathbf{n}}$ be the $|\mathbf{t}|th$ diagonal element of the hat-matrix $\boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}$. It follows that $\widehat{\boldsymbol{\theta}}_S$ is a zero of

$$S_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \leq \mathbf{n}} \lambda_{\mathbf{t}} \Phi_{\theta}(\xi_{|\mathbf{t}|}). \tag{2.3}$$

We will let $W^{m,p}(\mathrm{U})$ be the usual Sobolev space on an open neighborhood U of $\mathbb{R}^{|\mathbf{n}|}$ defined as

$$W^{m,p}(\mathrm{U}) = \left\{ \Gamma \in L^p(\mathrm{U}) : \ \mathcal{D}_{\boldsymbol{\theta}}^{\beta} \Gamma \in L^p(\mathrm{U}) \text{ with } |\beta| \leq m \right\},$$

where $L^p(\mathrm{U})$ is the spaces of functions g such that g^p is Lebesgue integrable on U.

2.2 Consistency and Asymptotic normality

In this section, we recall without proof two important results from Nguelifack et al. (2015) on the consistency and asymptotic normality of the GSR estimate $\widehat{\boldsymbol{\theta}}_S$. This two results are established under the following assumptions:

A1 : The parameter space Θ is compact and the function $f(\cdot, \cdot)$ is continuous with continuous derivatives.

A2 : G has Lebesgue density g that is symmetric about 0 and strictly decreasing on $\mathbb{R}^+$.

A3: $P(f(\mathbf{x},\boldsymbol{\theta}) = f(\mathbf{x},\boldsymbol{\theta}_0)) < 1$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

A4: For $1 < q < \infty$, there is an integrable function h not depending on $\boldsymbol{\theta}$ such that $|\rho(\tilde{G}_{\boldsymbol{\theta}}^{-1}(v))| \leq h(v)$, for all $\boldsymbol{\theta} \in \Theta$ with $E[h^q(Y)] < \infty$.

A5: Let $\{M_{\mathbf{n}}, \mathbf{n} \in \mathbb{N}^k\}$ be a field of $k \times k$ non-singular matrices such that

$$\frac{1}{|\mathbf{n}|} M_{\mathbf{n}} \sum_{\mathbf{t} \leq \mathbf{n}} \{\nabla_{\boldsymbol{\theta}} f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta})\}^T \{\nabla_{\boldsymbol{\theta}} f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta})\} M_{\mathbf{n}}^T$$

converges to a positive definite matrix $\Sigma_{\boldsymbol{\theta}_0}$ uniformly as $|\mathbf{n}| \to \infty$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \to 0$.

A6: $\lim_{|\mathbf{n}| \to \infty} \max_{\mathbf{1} \leq \mathbf{t} \leq \mathbf{n}} h_{\mathbf{tt}}^{\mathbf{n}} = 0$.

A7: $\boldsymbol{\theta} \to \Gamma_{\boldsymbol{\theta}}(t)$ is a map in $W^{3,p}(B)$, where B is a neighborhood of $\boldsymbol{\theta}_0$ for every fixed t.

A8: $A_{\boldsymbol{\theta}_0} = E[\varphi(\xi)[\mathcal{D}_{\boldsymbol{\theta}}^{\beta} \Phi_{\boldsymbol{\theta}}(\xi)]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}]$, where $\xi \sim U(0,1)$, is a symmetric positive definite matrix for $|\beta| = 1$.

A9: There exist functions $\psi_\beta \in W^{2,p}(\mathbb{R})$ independent of θ such that $|\mathcal{D}_{\boldsymbol{\theta}}^{\beta} \Phi_{\boldsymbol{\theta}}(s)| \leq \psi_\beta(s)$ for every $\boldsymbol{\theta} \in B$ and $|\beta| \leq 2$.

Remark 1. The above assumptions are the same as in Nguelifack et al. (2015). Assumption **A3** is needed for $\boldsymbol{\theta}_0$ to be identified. In our proof all we need is that the space defined by

$$\Omega_{\boldsymbol{\theta},\boldsymbol{\theta}_0}^{\varphi^+} = \{s \in (0,1) : \tilde{G}_{\boldsymbol{\theta}}(s) \neq \tilde{G}_{\boldsymbol{\theta}_0}(s) \text{ and } \varphi^+(s) > 0\},$$

has positive measure for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Assumption **A2** admits a wide variety of error distribution examples of which are the normal, the logistic, and the Cauchy distributions with location parameter equal to 0. Under assumptions **A1**–**A3**, we can use a similar strategy as in Hossjer (1994) to show that for any $s > 0$, for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and for all $\mathbf{t} \leq \mathbf{n}$,

$$\tilde{G}_{\theta}(s) = P(|\epsilon_{\mathbf{t}} - \{f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta}) - f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta}_0)\}| \leq s) < E_X\{P_{\epsilon_{\mathbf{t}}}(|\epsilon_{\mathbf{t}}| \leq s)\} = \tilde{G}_{\boldsymbol{\theta}_0}(s).$$

The following theorems will be essential for the Bayesian approach and for sake of brevity, will be given without proof. The proof can be found in Nguelifack et al. (2015).

Theorem 2.1. Under **A1**–**A4**, we have

$$\widehat{\boldsymbol{\theta}}_S \to \boldsymbol{\theta}_0 \quad a.s. \quad \text{as} \quad |\mathbf{n}| \to \infty.$$

Theorem 2.2. Under assumptions **A1**–**A9**, we have the following

$$\begin{aligned} \sqrt{|\mathbf{n}|} M_{\mathbf{n}}^{-1} S_{\mathbf{n}}(\boldsymbol{\theta}_0) &\underset{\mathcal{D}}{\to} N_p\left(0, \Sigma_{\boldsymbol{\theta}_0}\right) \\ \sqrt{|\mathbf{n}|} M_{\mathbf{n}}^{-1} A_{\boldsymbol{\theta}_0}^{-1} S_{\mathbf{n}}(\boldsymbol{\theta}_0) &= \sqrt{|\mathbf{n}|} M_{\mathbf{n}}^{-1} (\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0) + o_p(1) \end{aligned}$$

where $\Sigma_{\boldsymbol{\theta}_0} = E[\varphi^+(\xi)\Phi_{\boldsymbol{\theta}_0}(\xi)(\Phi_{\boldsymbol{\theta}_0}(\xi))^T]$.

3. Bayesian approach

3.1 Derivation of the asymptotic posterior distribution

We will proceed as and Zhan and Hettmansperger (2009). The distribution of $S_{\mathbf{n}}(\boldsymbol{\theta})$ will be considered as a pseudo-likelihood. Therefore, choosing $\boldsymbol{\theta}$ that maximizes $p(\boldsymbol{\theta}|S_{\mathbf{n}}(\boldsymbol{\theta}))$ is the same as solving $S_{\mathbf{n}}(\boldsymbol{\theta}) = 0$ for $\boldsymbol{\theta}$. We will summarize the information about the data in the asymptotic distribution of $S_{\mathbf{n}}(\boldsymbol{\theta})$, given by Theorem 2.2. The prior of $\boldsymbol{\theta}$ for a given $\boldsymbol{\theta}_0$, is then

$$\pi(\boldsymbol{\theta}) = (2\pi)^{-p/2} |\Sigma_{\boldsymbol{\theta}_0}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \Sigma_{\boldsymbol{\theta}_0}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right].$$

We let $\Omega_{\boldsymbol{\theta}} = \frac{1}{|\mathbf{n}|} M_{\mathbf{n}} A_{\boldsymbol{\theta}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} A_{\boldsymbol{\theta}}^{-1} M_{\mathbf{n}}^T$. We have the following theorem:

Theorem 3.1. *Let $\boldsymbol{\theta}_0 \in \Theta$ be given. Under assumptions A1-A14, the asymptotic posterior*

distribution $p(\boldsymbol{\theta}|S_{\boldsymbol{n}}(\boldsymbol{\theta}))$ *of* $\boldsymbol{\theta}$ *(with prior* $\pi(\boldsymbol{\theta})$*) given* $S_{\boldsymbol{n}}(\boldsymbol{\theta})$ *is proportional to the quantity*

$$\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)^{T}\widehat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)\right],$$

where

$$\widehat{\boldsymbol{\Sigma}}=(\Omega_{\boldsymbol{\theta}}^{-1}+\Sigma_{\boldsymbol{\theta}_0}^{-1})^{-1}\quad \textit{and}\quad \widehat{\boldsymbol{\theta}}=\widehat{\boldsymbol{\Sigma}}(\Omega_{\boldsymbol{\theta}}^{-1}\widehat{\theta}_S+\Sigma_{\boldsymbol{\theta}_0}^{-1}\boldsymbol{\theta}_0). \tag{3.1}$$

Proof. It follows from Theorem 2.2 that, given θ_0, we have $(\widehat{\boldsymbol{\theta}}_S-\boldsymbol{\theta})\sim N_k(0,\Omega_{\boldsymbol{\theta}})$. Let the pseudo-likelihood function be defined as

$$p(S_{\mathbf{n}}(\boldsymbol{\theta})|\boldsymbol{\theta})=\exp\left[-\frac{1}{2}\left(\left[\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta})\right]^{T}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\left[\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta})\right]\right)\right].$$

Then the asymptotic posterior of $\boldsymbol{\theta}$ given $S_{\mathbf{n}}(\boldsymbol{\theta})$ is

$$\begin{aligned}p(\boldsymbol{\theta}|S_{\mathbf{n}}(\boldsymbol{\theta}))\propto\exp\left[-\frac{1}{2}\left(\left[\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta})\right]^{T}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\left[\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta})\right]\right)\right]\\ \cdot\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}-\theta_0\right)^{T}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)\right].\end{aligned} \tag{3.2}$$

where the constant of proportionality is given by $C=\displaystyle\int_{\Omega}p(S_{\mathbf{n}}(\boldsymbol{\theta})|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$. We note from Theorem 2.2 that, $\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}(\widehat{\boldsymbol{\theta}}_S-\boldsymbol{\theta})=\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}A_{\boldsymbol{\theta}}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta})+o_p(1)$. Since $\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta})\sim N_p(0,\boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, then $(\widehat{\boldsymbol{\theta}}_S-\boldsymbol{\theta})\sim N_k(0,\Omega_{\boldsymbol{\theta}})$. It follows that the first expression in the right-hand side of equation (3.2) yields

$$\exp\left[-\frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_S-\boldsymbol{\theta})^{T}\Omega_{\boldsymbol{\theta}}^{-1}\left[(\widehat{\boldsymbol{\theta}}_S-\boldsymbol{\theta})\right]\right)\right].$$

The posterior asymptotic distribution of $\boldsymbol{\theta}$ given $S_{\mathbf{n}}(\boldsymbol{\theta}))$ is therefore proportional to

$$\exp\left[-\frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_S-\boldsymbol{\theta})^{T}\Omega_{\boldsymbol{\theta}}^{-1}(\widehat{\boldsymbol{\theta}}_S-\boldsymbol{\theta})+\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^{T}\Sigma_{\boldsymbol{\theta}_0}^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)\right)\right]. \tag{3.3}$$

Equation (3.3) is equivalent to

$$\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)^{T}\widehat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)\right],$$

after completion of the quadratic form with $\widehat{\boldsymbol{\Sigma}} = (\Omega_{\boldsymbol{\theta}}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^{-1})^{-1}$, and $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\Sigma}}(\Omega_{\boldsymbol{\theta}}^{-1}\widehat{\theta}_S + \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^{-1}\boldsymbol{\theta}_0)$. □

Remark 2. There is a similarity between the expression asymptotic prior of the GSR estimate in the linear case and its counterpart in the nonlinear case, in the sense that in both cases, the posterior mean $\widehat{\boldsymbol{\theta}}$ is a weighted average of the GSR estimate $\widehat{\boldsymbol{\theta}}_S$ and prior mean $\boldsymbol{\theta}_0$, with weights concentrated in data matrix $\Omega_{\boldsymbol{\theta}}$ and the prior precision matrix $\Sigma_{\boldsymbol{\theta}_0}$. This is possible because the nonlinear expression of the dispersion function $D_{\mathbf{n}}$ given by either equation (2.1) or (2.2) is similar to that of the linear dispersion function obtained when $f(\cdot,\cdot)$ is linear.

Remark 3. The results of Theorem 3.1 generalize the ones obtain by Zhan and Hettmansperger (2009). Indeed, it is not difficult to see that when $f(\mathbf{x},\cdot)$ is linear, and $k = 1$ (single-index case), then the matrix $\Omega_{\boldsymbol{\theta}}^{-1}$ is reduced to $n\tau_{\varphi}^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \lim_{n\to\infty} n^{-1}\boldsymbol{X}^T\boldsymbol{X}$. To see why, take $\varphi \equiv 1, \rho(s) = s$, and $\mathbf{n} = n$.

It was shown on page 25 in McKean and Hettemansperger (2011) that $A_{\boldsymbol{\theta}} = 2G'(\boldsymbol{\theta}) = \tau_{\varphi}$ where G' represents the derivative of the distribution function G of the errors satisfying assumption **A2**. Assumption **A5** becomes $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \lim_{n\to\infty} n^{-1}\boldsymbol{X}^T\boldsymbol{X}$. Therefore with the choice of $M_n = \tau_{\varphi}^{-2}\boldsymbol{\Sigma}_{\theta}$, we have $\Omega_{\boldsymbol{\theta}}^{-1} = nM_n^T A_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}A_{\boldsymbol{\theta}}M_n = n\tau_{\varphi}^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$

3.2 Examples and Simulations

Example 1: We consider the following model

$$y_{\mathbf{t}} = \alpha\cos(\lambda_1 x_{t_1} + \lambda_2 x_{t_2}) + \varepsilon_{\mathbf{t}}, \tag{3.4}$$

where α is a real unknown parameter and $\lambda_i, i = 1, 2$ are unknown parameters with $\mathbf{t} = (t_1, t_1)$. This model is a nonlinear sinusoidal used in textures modeling Rao et al. (1994), Zhang and Mandrekar (2001). Assumptions **A1**-**A6** were checked in Nguelifack et al. (2015). We start with a true value $\boldsymbol{\theta} = (\alpha, \lambda_1, \lambda_2) \equiv (4, 1.886, 1.1)$. For our simulations, we will choose $n_1 = n_2 = n = 40, 50, 80, 100$, and we assume $\varepsilon_{\mathbf{t}} \sim N(0, 1), \varepsilon_{\mathbf{t}} \sim t(3), \varepsilon_{\mathbf{t}} \sim$ Cauchy$(0, 1)$. A frequentist method for the generalized signed-rank estimation was proposed in Nguelifack et al. (2015) and it yields an average estimate of $\widehat{\boldsymbol{\theta}}_S = (\widehat{\alpha}_S, \widehat{\lambda}_{1,S}, \widehat{\lambda}_{2,S}) \equiv (3.902, 1.885, 1.113)$ after 1000 runs. If we choose a starting parameter $\boldsymbol{\theta}_0 = (0.05, 0.05, 0.01)$ and $M_n = \text{diag}\left(1, \frac{1}{n}, \frac{1}{n}\right)$, then from Nguelifack et al. (2015)

$$\Sigma_{\boldsymbol{\theta}_0} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{8} \\ 0 & \frac{1}{8} & \frac{1}{6} \end{pmatrix} \quad \text{and} \quad \Sigma_{\boldsymbol{\theta}} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{\alpha^2}{6} & \frac{\alpha^2}{8} \\ 0 & \frac{\alpha^2}{8} & \frac{\alpha^2}{6} \end{pmatrix}.$$

From the same reference, we have $A_{\boldsymbol{\theta}} = \text{diag}(a, b, c) \equiv \text{diag}(2g(\alpha), 2g(\lambda_1), 2g(\lambda_2))$, where $g = G'$ is the density function of respectively the standard normal distribution, the Student-t distribution with 3 degrees of freedom, and the the Cauchy distribution with parameters 0 and 1. We obtain after calculations the matrix

$$\Omega_{\boldsymbol{\theta}} = \begin{pmatrix} \frac{1}{2n^2a^2} & 0 & 0 \\ 0 & \frac{\alpha^2}{6n^4b^2} & \frac{\alpha^2}{8n^4bc} \\ 0 & \frac{\alpha^2}{8n^4bc} & \frac{\alpha^2}{6n^4c^2} \end{pmatrix} \Longrightarrow \Omega_{\boldsymbol{\theta}}^{-1} = \begin{pmatrix} 2n^2a^2 & 0 & 0 \\ 0 & \frac{97n^4b^2}{7\alpha^2} & -\frac{72n^4bc}{7\alpha^2} \\ 0 & -\frac{72n^4bc}{7\alpha^2} & \frac{96n^4c^2}{7\alpha^2} \end{pmatrix}.$$

Hence

$$\widehat{\Sigma}^{-1} = \Omega_{\boldsymbol{\theta}}^{-1} + \Sigma_{\boldsymbol{\theta}_0}^{-1} = \begin{pmatrix} 2(n^2a^2 + 1) & 0 & 0 \\ 0 & \frac{96}{7}\left(\frac{n^4b^2}{\alpha^2} + 1\right) & -\frac{72}{7}\left(\frac{n^4bc}{\alpha^2} + 1\right) \\ 0 & -\frac{72}{7}\left(\frac{n^4bc}{\alpha^2} + 1\right) & \frac{97}{7}\left(\frac{n^4c^2}{\alpha^2} + 1\right) \end{pmatrix}.$$

Finally, we obtain

$$\widehat{\Sigma} = \begin{pmatrix} \frac{1}{2(n^2a^2+1)} & 0 & 0 \\ 0 & \frac{7}{6}\frac{\alpha^2(n^4c^2+\alpha^2)}{D} & \frac{7}{8}\frac{\alpha^2(n^4bc+\alpha^2)}{D} \\ 0 & \frac{7}{8}\frac{\alpha^2(n^4bc+\alpha^2)}{D} & \frac{7}{6}\frac{\alpha^2(n^4b^2+\alpha^2)}{D} \end{pmatrix},$$

where $D = 7n^8b^2c^2 + 16n^4b^2\alpha^2 + 16\alpha^2n^4c^2 + 7\alpha^2 - 18n^4bc\alpha^2$.

Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\lambda}}_1, \widehat{\boldsymbol{\lambda}}_2)$ be the asymptotic posterior mean. It follows from equation (3.1) that

$$\begin{aligned} \widehat{\boldsymbol{\alpha}} &= \frac{n^2a^2\widehat{\alpha}_S + 1}{n^2a^2 + 1}, \\ \widehat{\boldsymbol{\lambda}}_1 &= \frac{n^4(n^4c^2 + \alpha^2)(96b^2\widehat{\lambda}_{1,S} - 72bc\widehat{\lambda}_{2,S})}{6D} + \frac{n^4(n^4bc + \alpha^2)(-72bc\widehat{\lambda}_{1,S} + 96c^2\widehat{\lambda}_{2,S})}{8D}, \\ \widehat{\boldsymbol{\lambda}}_2 &= \frac{n^4(n^4bc + \alpha^2)(96b^2\widehat{\lambda}_{1,S} - 72bc\widehat{\lambda}_{2,S})}{8D} + \frac{n^4(n^4b^2 + \alpha^2)(-72bc\widehat{\lambda}_{1,S} + 96c^2\widehat{\lambda}_{2,S})}{6D}. \end{aligned} \tag{3.5}$$

The table 1 below summarizes the estimates obtained for the bayesian posterior mean for different sample sizes n and different errors distributions. This table shows that as the sample sizes increases, then the bayesian estimates get closer to the GSR estimates, as discussed in remark 4.

Distribution of $\epsilon_{\mathbf{t}}$	Sample size	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\lambda}}_1$	$\widehat{\boldsymbol{\lambda}}_2$
$N(0,1)$	$n = 40$	3.8917	1.7300	1.1108
	$n = 60$	3.8974	1.8501	1.1125
	$n = 100$	3.9003	1.8803	1.1129
$t(3)$	$n = 40$	3.8879	1.8783	1.1124
	$n = 60$	3.8957	1.8836	1.1128
	$n = 100$	3.8997	1.8848	1.1129
Cauchy$(0,1)$	$n = 40$	3.8783	1.8816	1.1125
	$n = 60$	3.8914	1.8843	1.1129
	$n = 100$	3.8982	1.8849	1.1129

Table 1: Simulation Results

Remark 4. The example above highlights the fact that $\widehat{\boldsymbol{\theta}}$ is essentially a weighted average between the prior mean and the GSR estimate.

It also highlights a very important point about the bayesian mean: from Equation (3.5), it follows clearly that as $n \to \infty$,

1. $\widehat{\boldsymbol{\alpha}} \to \widehat{\alpha}_S$,
2. $\widehat{\boldsymbol{\lambda}}_1 \sim \frac{n^4(n^4c^2)(96b^2\widehat{\lambda}_1 - 72bc\widehat{\lambda}_2)}{42n^8b^2c^2} + \frac{n^4(n^4bc)(-72bc\widehat{\lambda}_1 + 96c^2\widehat{\lambda}_2)}{56n^8b^2c^2} \to \widehat{\lambda}_{1,S}$,
3. $\widehat{\boldsymbol{\lambda}}_2 \sim \frac{n^4(n^4bc)(96b^2\widehat{\lambda}_1 - 72bc\widehat{\lambda}_2)}{56n^8b^2c^2} + \frac{n^4(n^4b^2)(-72bc\widehat{\lambda}_1 + 96c^2\widehat{\lambda}_2)}{42n^8b^2c^2} \to \widehat{\lambda}_{2,S}$.

Therefore, for large values of n, the GSR estimate $\widehat{\boldsymbol{\theta}}_S$ is the dominant factor in the bayesian mean $\widehat{\boldsymbol{\theta}}$, and consequentially, the information matrix $\Omega_{\boldsymbol{\theta}}^{-1}$ quickly dominates the prior matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ in the expression of $\widehat{\boldsymbol{\Sigma}}$. This phenomenon is general and not limited to this example, and was also observed in the linear case by Zhan and Hettmansperger (2009). Therefore, in practice, it is often recommended to estimate the posterior mode $\widetilde{\boldsymbol{\theta}}$ instead of the posterior mean $\widehat{\boldsymbol{\theta}}$ if there is a reason to believe that the target distribution is skewed. We

note also that the constant of proportionality can be estimated using Laplace method as in Kass and Raftery (1995) and the references therein. There is also a new algorithm that seem to be efficient in calculating the constant of proportionality, see Lindgren and Rue (2015).

3.3 Markov Chain Monte Carlo Simulations

Due to the complex expression we obtained for the asymptotic posterior distribution of θ, numerical methods such as the Markov Chain Monte Carlo (MCMC) can be useful in general to obtain better approximations posterior parameters. We start with a brief overview of how MCMC works. The idea of MCMC was first introduced by Metropolis et al. (1953) as a method for the efficient simulation of the energy levels of atoms in a crystalline structure and was adapted in statistics by Hastings (1970). Say we need to know the important features of $\pi^*(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta$, the true distribution of $\boldsymbol{\theta}$. Since we know that π^* has a complex expression so that we cannot sample directly from it, MCMC is an indirect method that consists of constructing an aperiodic and irreducible Markov chain with state space Θ whose stationary (or invariant) distribution is π^*. Consequently, if the chain is run sufficiently long, then simulated values of the chain can be considered as a dependent sample from π^* and can be used to summarize key features about it. There is a rich literature on the subject of MCMC and its implementation (see for a head-start Metropolis et al. (1953); Hastings (1970); Geman and Geman (1984); Brooks (1988); Meyn and Tweedie (1993); Gelman et al. (1995)). The main theorem underpinning the MCMC method is that any chain which is irreducible and aperiodic has a unique stationary distribution, and that the t-step transition kernel will "converge" to that stationary distribution as $t \to \infty$. (See Meyn and Tweedie (1993)). Thus, to generate a chain with stationary distribution π^*, we need only to find transition kernels T that satisfy these conditions and for which $\pi^* T = \pi^*$, that is, T is such that given an observation $\boldsymbol{\theta}_1 \sim \pi^*(\boldsymbol{\theta}_1), T(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, then $\boldsymbol{\theta}_2 \sim \pi^*(\boldsymbol{\theta}_2)$. There are many choices for the transition kernels in practice, but the most popular is the Metropolis-Hastings

updating schemes (Hastings (1970)) with many variants among which the Gibbs Sampler kernel (introduced by Geman and Geman (1984)). We will now propose an algorithm based on an adaptation of Metropolis-Hastings updating schemes.

1. Consider K indices. Fix $k \in \{1, 2, \cdots K\}$,

2. Take $J \geq 2$ samples $\{X_k^{1,j}, X_k^{2,j}, \cdots\}_{j=2}^{J}$ from a given distribution D (referred to in practice as the proposal distribution). We note that for the j-th sample, $X_k^{i,j}$ is the i-th p-dimensional vector from D, where $i = 1, 2, \cdots$.

3. For a fixed j, select a starting point $\boldsymbol{\theta}_k^{0,j}$ from $X_k^{1,j}, X_k^{2,j}, \cdots$.

4. Let $\boldsymbol{\theta}_k^{*,j} \sim D$. For $i = 1, 2, \cdots$,

   (i) Calculate $r_k = \frac{p(\boldsymbol{\theta}_k^{*,j} | S_{n_k}(\boldsymbol{\theta}_k^{*,j}))}{p(\boldsymbol{\theta}_k^{i-1,j} | S_{n_k}(\boldsymbol{\theta}_k^{i-1,j}))}$.

   (ii) Let

$$\boldsymbol{\theta}_k^{i,j} = \begin{cases} \boldsymbol{\theta}_k^{*,j} & \text{with probability } \min(r_k, 1) \\ \boldsymbol{\theta}_k^{i-1,j} & \text{with probability } 1 - \min(r_k, 1) \end{cases} .$$

5. Let L_k be the length of each sequence after discarding a certain percentage, referred to in the literature as Burn-In percentage (Geman and Geman (1984)). We note that $\boldsymbol{\theta}_k^{i,j} = (\boldsymbol{\theta}_{k,l}^{i,j})_{l=1,2,\cdots,p}$ is a p-dimensional vector. We define the between and within sequence averages in index k as the quantities

$$\overline{\boldsymbol{\theta}_{k,l}^{\cdot,j}} = \frac{1}{L_k} \sum_{i=1}^{L_k} \boldsymbol{\theta}_{k,l}^{i,j} \quad \text{and} \quad \overline{\boldsymbol{\theta}_{k,l}^{\cdot,\cdot}} = \frac{1}{J} \sum_{j=1}^{J} \overline{\boldsymbol{\theta}_{k,l}^{\cdot,j}}.$$

   We also define the between and within sequence variances in index k as the quantities

$$B_k = \frac{L_k}{J-1} \sum_{j=1}^{J} \left(\overline{\boldsymbol{\theta}_{k,l}^{\cdot,j}} - \overline{\boldsymbol{\theta}_{k,l}^{\cdot,\cdot}} \right)^2 \quad \text{and} \quad W_k = \frac{1}{J} \frac{1}{L_k - 1} \sum_{j=1}^{J} \sum_{i=1}^{L_k} \left(\boldsymbol{\theta}_{k,l}^{i,j} - \overline{\boldsymbol{\theta}_{k,l}^{\cdot,j}} \right)^2.$$

Finally, letting $\boldsymbol{\theta}_{k,l}$ represent the l-th component of $\boldsymbol{\theta}$ in the k-th index, $Var(\boldsymbol{\theta}_{k,l}|S_{n_k}(\boldsymbol{\theta}))$ can be approximated as weighted average between B_k and W_k as

$$\widehat{Var}(\boldsymbol{\theta}_{k,l}|S_{n_k}(\boldsymbol{\theta})) = \frac{1}{J}B_k + \frac{L_k - 1}{L_k}W_k.$$

We then define the scale reduction coefficient in the k-th index as

$$R_k^2 := \frac{\widehat{Var}(\boldsymbol{\theta}_k^l|S_{n_k}(\boldsymbol{\theta}))}{W_k} = \left(\frac{1}{J}\frac{B_k}{W_k} + \frac{L_k - 1}{L_k}\right). \tag{3.6}$$

6. We repeat the steps above until the quantity R_k defined in equation (3.6) approaches 1 for each parameter and for each index k.

Remark 5. We observe that the proposal distribution D can be chosen to have any form and the Markov chain will converge eventually to the stationary distribution, see Gilk et al. (1997), page 7. This choice is crucial when one is concerned with mixing, that is, how quickly the chain moves around the support of the stationary distribution, and the convergence rate. For our applications, we will be concerned only with obtaining reliable estimates.

Example 2: We consider again the model 3.4 in Example 1. In this case, $K = 1, p = 3$. Now, we assume that $\boldsymbol{\theta} = (\alpha, \lambda_1, \lambda_2) \sim N_3((4, 1.886, 1.1), \Sigma)$, where $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.1 & 0.01 \\ 0 & 0.01 & 0.1 \end{pmatrix}$. Let $\boldsymbol{\theta}^+ = (1, 3, 1)$ and $\Sigma^+ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. We take $J = 100$ samples of length $L = 10000$ from proposal distributions $D = N_3(\boldsymbol{\theta}^+, \Sigma^+)$ or $D = Unif(A)$, where A is the cube $[-4, 4]^3$. We apply steps (1)-(6) above until $R_k \approx 1$ for each component of the estimated parameter vector. The table 2 below represents a comparison between asymptotic posterior mean calculated

for $n = 100$ and $\varepsilon_{\mathbf{t}} \sim N(0,1)$ and MCMC estimate of the mean. The values are averaged over the J samples. The Bias of the Bayesian estimator of a parameter is calculated as the estimator minus the true value, whereas the bias and MSE of the MCMC estimator is calculated from the equations below

$$\text{Bias}(\widehat{\boldsymbol{\theta}}) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}\right), \quad \text{MSE}(\widehat{\boldsymbol{\theta}}) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}\right)^2.$$

Method	Estimate, Bias, MSE	α	λ_1	λ_2
Asymptotic posterior	Estimate	3.9003	1.8803	1.1108
	Bias	0.1083	0.0057	-0.0129
	MSE	0.0117	3.25×10^{-6}	1.66×10^{-4}
MCMC	Estimate	4.0012	1.8839	1.1034
	Bias	0.0012	-0.0020	-0.0065
	MSE	0.0321	0.0020	0.0030

Table 2: Comparison between asymptotic posterior and MCMC estimates

We note that the Bias of the asymptotic posterior mean estimate is quite large compared to that of MCMC estimate. We know by remark 4 that for large values of n, the bayesian mean approaches the rank estimate which also approaches the true value, and therefore reducing the bias. In Figure 1 below, we show the effect of the proposal distributions on the paths and speed of convergence.

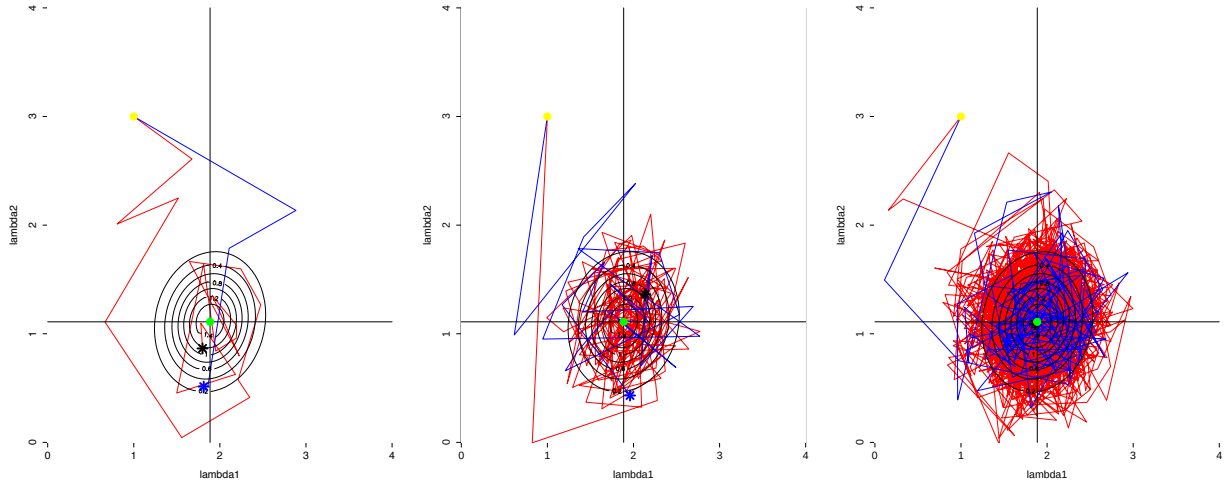

Figure 1: MCMC paths for the estimation of λ_1 and λ_2, for respective sequence lengths $L = 100, 1000, 10000$. The blue lines represent the paths with proposal distributions $D = Unif([-4,4]^2)$ and the red lines represent the paths with proposal distributions $D = N_2((1.886, 1.11), \Sigma^+|_{(\lambda_1,\lambda_2)})$. The yellow dot is the starting point $(1, 3)$ and the green dot is the true value (1.886, 1.11) of (λ_1, λ_2). The black and blue stars represents the end points of each path.

4. Discussion

In this paper, we have derived a Bayesian method to the generalized signed-rank estimates for nonlinear models with multidimensional indices. The method relies upon obtaining good estimates from the frequentist GSR method and it uses asymptotic normality results to converge to the true value of the unknown the parameter. The most difficult part in applications is to check assumptions, particularly assumption **A2**. But when this task is accomplished, implementation is not very difficult. In fact, in this paper, we show that asymptotic posterior means are as accurate as MCMC estimates. The main drawback of the MCMC method is that it is not computationally efficient. It is our belief that for real life applications, both estimation approaches can be used, that is, the MCMC can be used to verify the accuracy of the asymptotic posterior estimate. The overall goal of proposing a Bayesian extension to nonlinear models with multidimensional indices is achieved, even if implementation can be

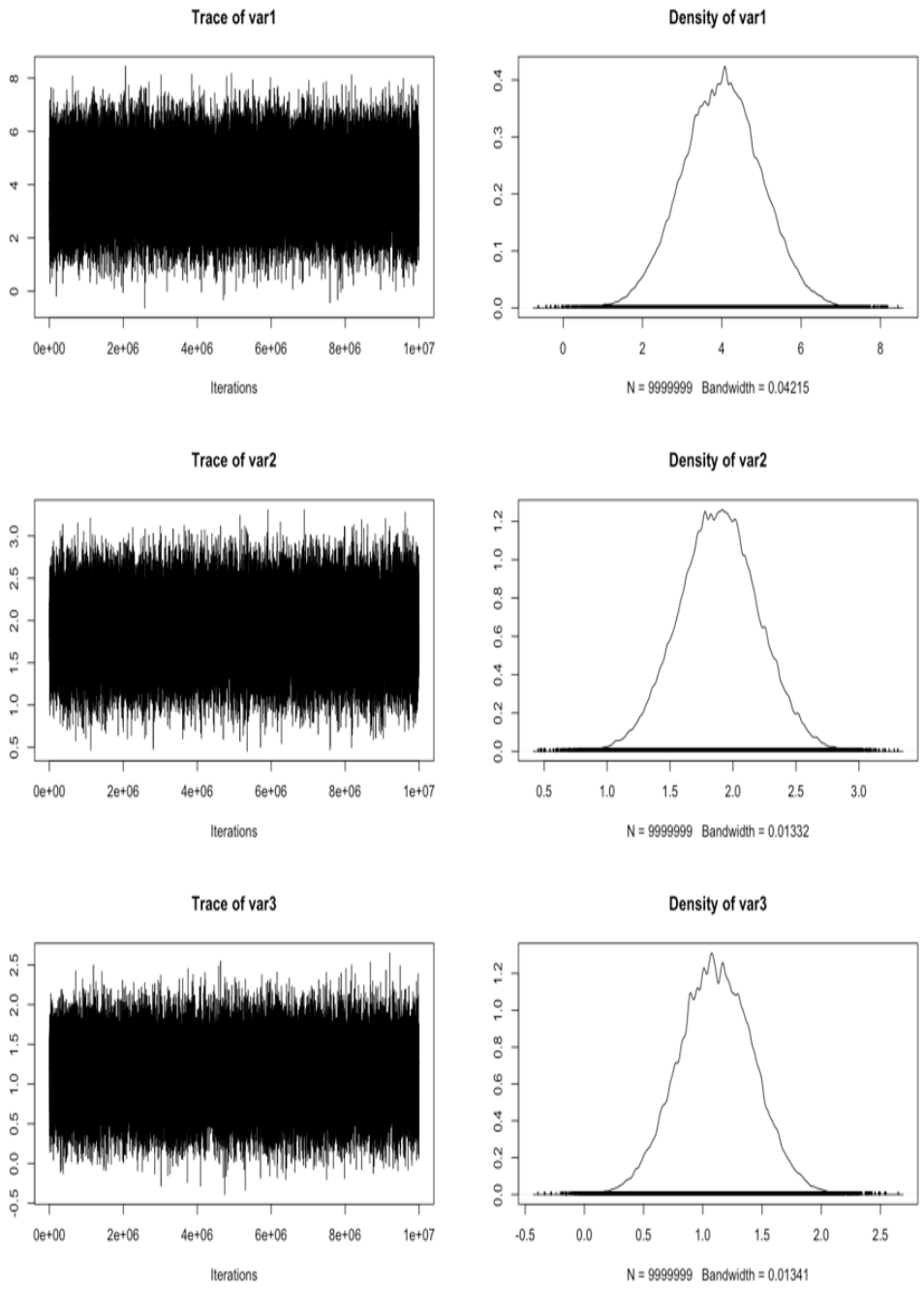

Figure 2: The left panels show trace plots of each variables on Example 2. It is clear based on the random pattern that there is good mixing in the chains. The right panels show an estimate density for the distribution of each variable. A bell-shaped pattern shows that the estimates are consistent with the hypotheses

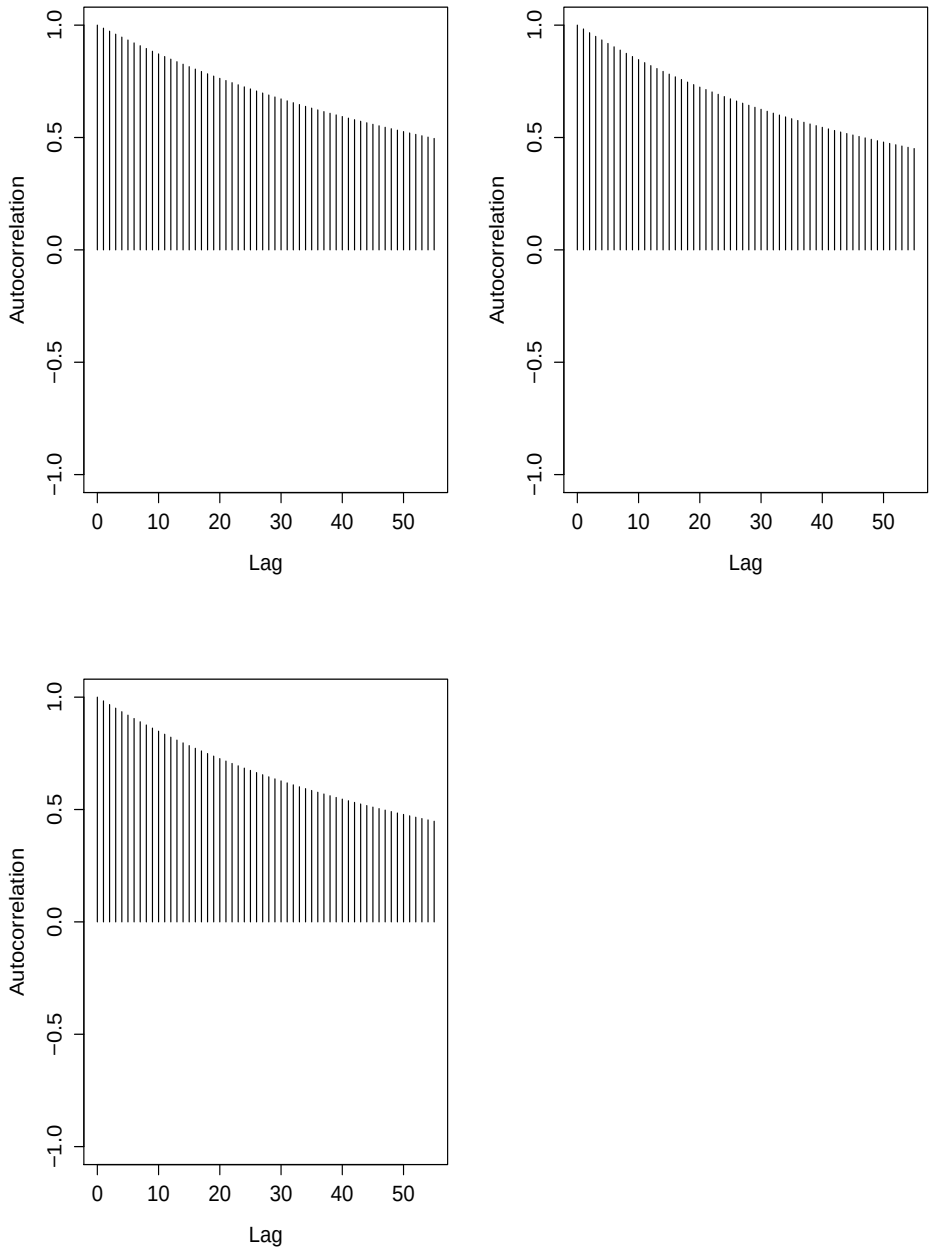

Figure 3: This plot was obtained for proposal distribution $D \sim Unif([-4,4]^3)$. It shows that the correlation among the draws decreases as the number of draws increases which implies a good mixing among the chains. If chains are run sufficiently long, the correlation will decrease even more

difficult. This model could be used in the context of generalized linear models, but it will be restricted to symmetric errors, which can be a limiting factors in some applications. Possible extensions of this method in actuarial science are feasible where severity and frequency of claims are often estimated using generalized linear models. Since the link between claim severity and policy characteristic is by nature nonlinear and depending on whether there was an increase or decrease in claims, it our belief that by modeling the frequency with a distribution of choice and the severity through segmentation of policies by drivers, age, gender, credit scores etc, there is a possibility of using this model in that context by considering severity categories as different dimensions. Such an endeavor is still under investigation

References

E. A. Baklanov. *The Strong Law Of Large Numbers For L-statistics With Dependent Data*, Siberian Mathematical Journal, Vol. 47, No. 6, pp. 975–979, 2006.

N. K. Bansal, G. G. Hamedani and H. Zhang. *Non-linear regression with multidimensional indices*, Statistics and Probability Letters, Vol. 45, pp. 175–186, 1999.

D. R. Brillinger. Fitting Cosines: Some Procedures and Some Physical Examples in I. B. MacNeill, G. J. Umphrey, D. R. Bellhouse, R. J. Kulperger and D. R. Brillinger (Eds) *Advances in the Statistical Sciences: Applied Probability, Stochastic Processes, and Sampling Theory*. Springer Netherlands, 1986.

P. Brooks. *Markov Chain Monte Carlo Method and Its Application*, Journal of the Royal Statistical Society. Series D (The Statistician), Vol. 47, No. 1, pp. 69-100, 1998.

E. N. Brown. *A note on the asymptotic distribution of parameter estimates for the harmonic regression model*, Biometrika, Vol 77, 3, 653-656, 1990.

Dunson, D. B. and Taylor, J. A. *Approximate Bayesian inference for quantiles*, Journal of Nonparametric Statistics, Vol. 17, pp. 385–400, 2005.

J. M. Francos, A. Z. Meiri and B. Porat. *A united texture model based on a 2-D Wold like decomposition*, IEEE Transactions on Signal Processing, Vol. 41, pp. 2665–2678, 1993.

A. R. Gallant. *Asymptotic Theory of Nonlinear least Square Estimator*, Annals of Statistics, vol. 9, No. 3, pp. 501-513, 1981.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, D. B. Bayesian data analysis. Chapman & Hall/CRC, London, 1995. MR1385925

S. Geman and D. Geman. *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Trans. Pattn Anal. Mach. Intell., 6, 721-741, 1984.

E. J. Hannan. *The estimation of frequency*, Journal of Applied Probability, Vol. 10, pp. 510–519, 1973.

W. R. Gilks, S. Richacrdson, D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1997.

W. K. Hastings. *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, 57, 97- 109, 1970.

T. P. Hettmansperger and J. W. McKean. *Robust Nonparametric Statistical Methods*, Second edition, CRC Press, Boca Raton, 2011.

O. Hossjer. *Rank-Based Estimates in the Linear Model With High Breakdown Point*, Journal of American Statistical Association. March, Vol. 89, No. 425, 1994.

R. Irizarry. *Local Harmonic estimation in Musical Sound Signals*, J. Amer. Stat. Assoc., 96:454, 357-367, 2011.

R. I. Jennrich. *Asymptotic Theory of Nonlinear least Square Estimator*, Annals of Statistics, Vol. 9, No. 3, pp. 501-513, 1969.

R. E. Kass and A. E. Raftery. *Bayes factors*, J. Amer. Stat. Assoc., Vol 90, No. 430, 773-795, 1995.

S. Kay. *Modern Spectral Estimation: Theory and Application*, Prentice Hall, NJ, 1999.

D. Kundu. *Asymptotic theory of least squares estimator of a particular nonlinear regression model*, Statistics and Probability Letters, Vol. 18, pp. 13–17, 1993.

D. Kundu and S. Nandi. *Determination of Discrete Spectrum in a Random Field*, Statistica Neerlandica, Vol. 57, No. 2, pp. 258–283, 2003.

F, Lindgren and H. Rue, *Bayesian Spatial Modelling with R-INLA*, J. Stat. Software, Vol 63, Issue 19. 2015.

J. H. McClellan. *Multidimensional spectral estimation*, Proc. IEEE, Vol. 70, pp. 1029–1039, 1982.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller *Equations of state calculations by fast computing machines*. J Chem. Phys., 21, pp. 1087–1091,1953.

S. P. Meyn and R. L.Tweedie. *Markov Chains and Stochastic Stability*. New York: Springer, 1993.

S. Nandi. *Estimation of parameters of two-dimensional sinusoidal signal in heavy-tailed errors*, Journal of Statistical Planning and Inference, Vol. 142, pp. 2799–2808, 2012.

B. M. Nguelifack, E. A. Kwessi, A. Abebe. *Generalized Signed-Rank Estimator for Nonlinear Models with Multidimensional Indices*, Journal of Nonparametric Statistics, Vol. 40, pp. 5012–5034, 2015.

C. R. Rao and L. C. Zhao, *Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals*, IEEE Transactions on Signal Processing, Vol. 41, pp. 1461–1464, 1993.

C. R. Rao, L. C. Zhao and B. Zhou. *Maximum likelihood estimation of 2-D superimposed exponential signals*, IEEE Transactions on Signal Processing, Vol. 42, pp. 1795–1802, 1994.

J. A. Rice and M. Rosenblatt. *On frequency estimation*, Biometrika, Vol. 75, pp. 477–484, 1988.

J. I. Sánchez, E. Barocio, P. Zuniga, R. Betancourt. *Instantaneous Modal Estimation of Sub-synchronous Oscillations Using Simplified Recursive Newton Type Algorithm*, Power and Energy Society General Meeting, IEEE, pp.1–8, 2011.

C.-F. Wu. *Asymptotic Theory of Nonlinear least Square Estimator*, Annals of Statistics, Vol. 9, No. 3, pp. 501–513, 1981.

R. Yang and Z. Su. *Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation*, Bioinformatics, Vol 26, pp.168–174, 2010.

T. Yuan and T. S. Rao. *Spectrum estimation for random fields with applications to Markov modeling and texture classification*, in Markov Random Fields: Theory and Applications, R. Chellappa and A.K. Jain, eds., American Mathematical Society, 1992.

A. J. Zavala and A. R. Messina. *A Dynamic Regression Approach to Power System Modal identification and Prediction*, Electric Power Comp. Syst. 42:13, pp. 1474–1483, 2014.

H. Zhang and V. Mandrekar. *Estimation of hidden frequencies for 2D stationary processes*, Journal of Time Series Analysis, Vol. 22, pp. 613–629, 2001.

Xiaojiang Zhan, Thomas P. Hettmansperger. *Bayesian R-estimates in linear models*, Journal of Statistical Research, Vol. 39, pp. 55–69, 2005.

Xiaojiang Zhan, Thomas P. Hettmansperger. *Bayesian R-estimates in linear models*, Journal of Computational Statistics and Data Analysis, Vol. 51, pp. 5077–5089, 2007.

Xiaojiang Zhan, Thomas P. Hettmansperger. *Bayesian R-estimates in linear models*, Journal of Statistics and Its Interface, Vol. 2, pp. 247–254, 2009.

Published: Volume 2018, Issue 8 / August 25, 2018