Locally Stationary Processes with Stable and Tempered Stable Innovations

Shu Wei Chou Chen

Thesis presented
to
Institute of Mathematics and Statistics
of
University of São Paulo
to
Obtain the title
of
Doctorate in Science

Program: Statistics

Advisor: Prof. Dr. Pedro Alberto Morettin

During the development of this work the author received financial support from CAPES/CNPq

São Paulo, February of 2020

Processos Localmente Estacionários com Inovações Estáveis e Estáveis Temperadas

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 20/02/2020. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Pedro Alberto Morettin (orientador) - IME-USP
- Profa. Dra. Airlane Pereira Alencar - IME-USP
- Prof. Dr. Reinaldo Castro de Souza - PUC-RJ
- Profa. Dra. Thelma Sáfadi - UFLA
- Prof. Dr. João Ricardo Sato - UFACB

Acknowledgements

I wish to express my deepest gratitude to my advisor, Prof. Pedro Morettin, for the guidance and motivation in my thesis, and also, for his patience and support during my stay in Brazil. Besides my advisor, I would like to thank the rest of my thesis committee, for their comments and suggestions from different perspectives.

I am very thankful to the professors and staff of the Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, for their professional examples, guidance and support during my PhD. Also, I would like to acknowledge the University of Costa Rica, for supporting my academic training in Statistics from distance.

I am also very grateful to my parents and my brother for supporting me spiritually throughout this period of writing my thesis and my life in general. Finally, I would like to thank my friends and everyone that I encountered during my stay in Brazil who helped me academically and emotionally in adapting to this beautiful country, for their support, friendship, and for their contributions in different theoretical and simulating discussions.

Abstract

Chou-Chen, S.W. **Locally Stationary Processes with Stable and Tempered Stable Innovations.** 2020. Thesis (PhD) - Instituto de Matemática e Estatística, Universidade de São Paulo, 2020.

In the literature, the class of locally stationary processes assumes that there is a time-varying spectral representation, i.e. the existence of finite second moment. In this work, we first propose the α-stable locally stationary process by modifying the innovations into stable distributions, which has heavy tail, and the indirect inference to estimate this type of model. Due to the infinite variance, some of interesting properties such as time-varying autocorrelation cannot be defined. However, since the α-stable family of distributions, as a generalization of the Gaussian distribution, is closed under linear combination, which includes the possibility of handling asymmetry and thicker tails, the proposed model presents the same tail behavior throughout the time. We carry out simulations to study the performance of the indirect inference and compare it to the existing methodology, blocked Whittle estimation. When the process has stable innovations, the indirect inference presents more promising results than the existing methodology because of infinite variance. Next, we consider the locally stationary process with tempered stable innovations, whose center is similar to that of a stable distribution, but its tails are lighter (semi-heavy tail) and all moments are finite. We present some theoretical results of this model and propose a two-step estimation to estimate the parametric form of the model. Simulations suggest that the time-varying structure can be estimated well, but the parameters related to the innovation are biased for small time series length. However, the bias disappears when time series length increases. Finally, an empirical application is illustrated.

Keywords: Locally stationary process, stable distribution, tempered stable distribution, indirect inference, two-step estimation.

Resumo

Chou-Chen, S.W. **Processos Localmente Estacionários com Inovações Estáveis e Estáveis Temperadas.** 2020. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, 2020.
Na literatura, a classe dos processos localmente estacionários supõe que existe a representação espectral variando no tempo, i.e. a existência do segundo momento finito. Neste trabalho, propomos primeiro o processo localmente estacionário $\alpha-$estável modificando as inovações em distribuição estável, a qual tem cauda pesada, e a inferência indireta para estimar este tipo de modelo. Devido à variância infinita, algumas propriedades interessantes como as autocorrelações variando no tempo não podem ser definidas. Contudo, como a família das distribuições $\alpha-$estáveis, como uma generalização da distribuição Gaussiana, é fechada sob combinações lineares, na qual inclui a possibilidade de manipular assimetria e cauda mais pesada, o modelo proposto apresenta o mesmo comportamento de cauda ao longo do tempo. Simulações são feitas para estudar o desempenho da inferência indireta e para compará-lo com uma metodologia existente, estimação de Whittle em blocos. Quando o processo tem inovações estáveis, a inferência indireta apresenta resultados promissores que os métodos existentes porque o modelo tem variância infinita. Em seguida, consideramos o processo localmente estacionário com inovações estáveis temperadas, do qual o centro é similar ao caso estável, mas suas caudas são mais leves (cauda semi-pesada) e todos os seus momentos são finitos. Apresentamos alguns resultados teóricos desse modelo e propomos a estimação em dois passos para estimar a forma paramétrica do modelo. Simulações sugerem que a estrutura variando no tempo pode ser estimada satisfatoriamente, mas os parâmetros relacionados às inovações são viesados para séries temporais curtas. Porém, o viés desaparece quando o comprimento da série aumenta. Finalmente, uma aplicação empírica é ilustrada.

Palavras-chave: Processo localmente estacionário, distribuição estável, distribuição estável temperada, inferência indireta, estimação em dois passos.

Contents

Abbreviation

tvAR	time-varying Autoregressive model
tvMA	time-varying Moving average model
tvARMA	time-varying Autoregressive moving average model
AM	Auxiliary Model
IM	Model of Interest
kur	Kurtosis
skw	Skewness

List of Symbols

$\mathbb{N}$	The set of natural number
$\mathbb{R}$	The set of real number
$X_{t,T}$	A real valued time series with discrete time $t \in \{1, ..., T\}$
$V(g)$	Total variation of a function g
$\mathcal{L}$	Likelihood function
$\mathcal{N}$	Normal distribution
$S_\alpha(\sigma, \beta, \mu)$	Stable distribution with parameters α, σ, β and μ
$S\alpha S$	Symmetric α Stable distribution
$CTS(\alpha, \lambda_+, \lambda_-, C_+, C_-, \mu)$	Classical tempered stable distribution with parameters $\alpha, \lambda_+, \lambda_-, , C_+, C_-$ and μ
$stdCTS(\alpha, \lambda_+, \lambda_-)$	Standardized classical tempered stable distribution with parameters α, λ_+ and λ_-
$\Phi_{t,T}(B)$	Autoregressive operator
$\Theta_{t,T}(B)$	Moving average operator
$G(t, s)$	one-sided Green's matrix
$g(t, s)$	one-sided Green's function

List of Figures

List of Tables

Chapter 1

Introduction

Dahlhaus (1996a, 1997) introduced the class of locally stationary processes that describes the types of processes that are approximately stationary in a neighborhood of each time point but its structure, such as covariances and parameters, gradually changes throughout the time period. This type of processes has been proved to achieve meaningful asymptotic theory by applying infill asymptotics. The idea of this approach is that the time varying parameters are rescaled to the unit interval, and thus, more available observations imply obtaining more contribution for each local structure. Consequently, statistical asymptotic results such as consistency, asymptotic normality, efficiency, locally asymptotically normal expansions, etc. are obtained. There is an extensive literature about estimation and hypothesis testing methods, e.g. Chan and Palma (2019) cites recent advanced papers on this topic.

Most results of locally stationary processes assume innovations with finite second moment. However, different areas, such as actuarial science, biostatistics, computer science, finance and physics, have been observed phenomena with heavy tail distributions and/or infinite variance (Grabchak, 2016a). In this thesis, we consider that the locally stationary process assume heavy-tailed innovations. Specifically, two classes of distributions are considered: $\alpha-$stable distribution and the tempered stable distribution.

The advantage of assuming $\alpha-$stable distributions is its flexibility for asymmetry and thick tails. Additionally, it is closed under linear combinations and includes the Gaussian distribution as a special case. In this case, the process is $\alpha-$stable and we call this kind of process $\alpha-$stable locally stationary process (strictly sense because of the infinite variance). However, its estimation is difficult since the density function does not have a closed-form and its moments of order greater than two do not exist. Therefore, the usual estimation methods such as maximum likelihood and method of moments do not work.

On the other hand, the tempered stable distribution is obtained by changing the tail behavior of a stable distribution. Its center is similar to stable distribution, but its tails are lighter, which is called semi-heavy tails. This distribution keeps most of the attractive properties and still has all finite moments. However, it is closed under linear combinations only under some restrictive conditions.

In the stable innovation case, alternative estimation approaches such as methods based on quantiles (McCulloch, 1986) or on the empirical characteristic function (Koutrouvelis, 1981) are proposed. However, those methods are only useful for the estimation of the $\alpha-$stable distribution parameters and, therefore, they are difficult to apply for more complex models.

The strategy to estimate this kind of process is the indirect inference proposed by Gourieroux *et al.* (1993) and Gallant and Tauchen (1996). Since $\alpha-$stable distributions can be easily simulated, the indirect approach, which is an intensive computationally simulation based method, can be a solution to overcome the estimation problem. Models involving stable distribution were successfully implemented in indirect inference for independent samples from the α-stable distributions and $\alpha-$stable ARMA processes (Lombardi and Calzolari (2008)). Moreover, some time series models involving stable distributions are also successfully implemented using indirect inference (Calzolari and Halbleib, 2018; Calzolari *et al.* , 2014; Sampaio and Morettin, 2015, 2018).

In the tempered stable innovation case, the model presents less attractive properties, such as closeness under linear combinations. Nevertheless, since its moments of all orders are finite, time series models involving tempered stable innovations can be estimated using traditional methods with weakly stationary assumption, e.g. Feng and Shi (2017) investigated Fractional integrated GARCH model with tempered stable distribution. Important properties of tempered stable distributions and their associated processes are covered in Grabchak (2016a); Küchler and Tappe (2013).

There are different extensions or subclasses of tempered stable distribution, e.g. mixed tempered stable distribution (Hitaj *et al.*, 2018; Rroji and Mercuri, 2015), modified tempered stable distribution (Kim *et al.*, 2006) and KR distribution (Kim *et al.*, 2008). Despite of these variations, we will focus on the standardized classical tempered stable distribution, which is implemented in GARCH models in Kim *et al.* (2008). The attractive feature of this distribution is that it has zero mean and unit variance, and the parameters can be estimated using two-step estimation. In our case, the parametric time varying structure can be estimated by using blocked Whittle likelihood, proposed by Dahlhaus (1997). Next, by supposing independent standardized classical tempered stable innovations, recovering from the residuals of the model, consistent estimation related to the tempered stable distribution can be obtained by maximum likelihood estimation.

Our contribution in the $\alpha-$stable innovation case is twofolds. Firstly, we present properties of the $\alpha-$stable locally stationary processes. Second, we carried out simulations by using indirect inference for this type of models with linear coefficients throughout the time. When a time series has infinite variance, simulations suggests that indirect inference performs better than the blocked Whittle estimation in term of bias and standard error. However, estimation is more time consuming. Furthermore, in the tempered stable innovation case, we performed simulation studies in order to evaluate the consistency of the estimation. Biased estimators in the second step are detected for relatively small time series and the bias is reduced when the time series length increases. Finally, we illustrate an application with models assuming

$\alpha-$stable and tempered stable innovations. The program and routines were performed in R version 3.5.3 R Core Team (2019).

1.1 Organization of the thesis

We organize this thesis as follows. The first part consists of a brief review of important concepts in this work: locally stationary process (Chapter 2), $\alpha-$stable and tempered stable distribution (Chapter 3), and the indirect inference (Chapter 4). Then, we present the properties of the $\alpha-$stable locally stationary processes and some examples in Chapter 5. In Chapter 6, we perform several simulations to study the indirect inference and illustrate an application by assuming known index of stability α. In the same way, we carried out simulations for the unknown index of stability α and present an application in Chapter 7. Next, the locally stationary processes with tempred stable innovations is covered in Chapter 8. Finally, conclusions and future works are presented in Chapter 9.

Chapter 2

Locally Stationary Processes

In time series, stationary processes have been well studied due to important properties such as invariant mean and variance and covariance structure depending on difference of times. However, most real world time series data are not stationary and fitting a stationary process to a nonstationary time series could be inappropriate and it will usually lead to wrong conclusions.

To deal with nonstationarity, there are some well-known techniques that convert nonstationary time series into stationary ones, such as differentiation, trend removal and regression analysis based on other input variables.

Dahlhaus (2012) pointed out some difficulties on developing nonstationary processes:

1. There is no natural generalization from stationary to nonstationary time series, except those nonstationary models which are generated by a time invariant generation mechanism such as integrated or cointegrated models.

2. It is not clear how to set down a meaningful asymptotic theory for nonstationary processes.

One way to generalize the stationary process is the idea of locally stationarity. This is the case where a stochastic process X_t might be stationary over small periods of time, but this property of stationarity changes slowly over the longer period of time. Priestley (1965) introduced the processes with a time-varying spectral representation

$$X_t = \int_{-\pi}^{\pi} e^{i\lambda t} A_t(\lambda) d\xi(\lambda), \quad t \in \mathbb{Z},$$

where $\xi(\lambda)$ is an orthogonal increment process and $A_t(\lambda)$ is a time-varying transfer function. Note that when $A_t(\lambda)$ is constant with respect to t, then we obtain the special case when X_t is globally stationary.

Since future observations of a nonstationary process may not contain any information at the probabilistic structure of the process at present, the theory of locally stationary processes is based on infill asymptotic approach. As in nonparametric statistics, the idea of infill asymptotic is that the functions over time are rescaled to the unit interval in order to achieve some meaningful asymptotic theory.

We will introduce the theory of locally stationary process [1] by giving a simple example of time varying AR(1) process,

$$X_t + \alpha_t X_{t-1} = \sigma_t \varepsilon_t, \quad \text{with } \epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1). \tag{2.1}$$

Applying infill asymptotics means that α_t and σ_t are replaced by $\alpha(\frac{t}{T})$ and $\sigma(\frac{t}{T})$ with curves $\alpha(\cdot) : [0,1] \to (-1,1)$ and $\sigma(\cdot) : [0,1] \to (0,\infty)$. Note that if we fit the parametric model $\alpha_{\theta,t} = b+ct+dt^2$ to the non-rescaled model (2.1), it is easy to construct estimators such as maximum likelihood estimator; however, it is nearly impossible to derive the finite sample properties of these estimators. Moreover, classical non-rescaled asymptotic considerations have no sense since $\alpha_{\theta,t} \to \infty$, if $t \to \infty$ while $|\alpha_t| < 1$ in the observed segment.

On the other hand, using infill asymptotic, as $T \to \infty$, we obtain more and more available observations for contributing each local structure, and there are statistical asymptotic results such as consistency, asymptotic normality, efficiency, locally asymptotically normal (LAN) expansions, etc. for non-stationary processes. Moreover, note that classical asymptotics for stationary processes arise as a special case of this infill asymptotics in case where all parameter curves are constant.

2.1 Linear locally stationary processes

The formal definition of a linear locally stationary processes is as follows.

Definition 2.1. *(Linear locally stationary processes) The sequence of stochastic processes $X_{t,T}$ $(t = 1, ..., T)$ is a linear locally stationary processes if $X_{t,T}$ has a representation*

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \sum_{j=-\infty}^{\infty} a_{t,T}(j)\varepsilon_{t-j}, \tag{2.2}$$

where some regularity conditions are satisfied for μ, $\alpha_{t,T}$ and ε_t.

Note that if the ε_t are stationary, it can be represented as

$$\varepsilon_t = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{i\lambda t} d\xi(\lambda),$$

where $\xi(\lambda)$ is a process with zero mean and orthonormal increments. Then, the representation (2.2) is basically equivalent to

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} e^{i\lambda t} A_{t,T}(\lambda) d\xi(\lambda), \tag{2.3}$$

[1] this chapter is based on Dahlhaus (2012), which contains an overview of the theory of locally stationary processes.

with the transfer function $A_{t,T}(\lambda) := \sum_{j=-\infty}^{\infty} a_{t,T}(j)e^{-i\lambda j}$.

To continue with the regularity conditions of the linear locally stationary process defined in (2.2), let $V(g)$ be the total variation of a function g on $[0,1]$, that is

$$V(g) = \sup\left\{\sum_{k=1}^{m} |g(x_k) - g(x_{k-1})| : 0 \leq x_0 < ... < x_m \leq 1, m \in \mathbb{N}\right\},$$

and for some $\kappa > 0$ let

$$\ell(j) := \begin{cases} 1, & |j| \leq 1 \\ |j| \log^{1+\kappa} |j| & |j| > 1. \end{cases} \tag{2.4}$$

Assumption 2.1. *Suppose that the sequence of stochastic process $X_{t,T}$ has a representation as in (2.2) and satisfies the following conditions:*

(i)

$$\sup_t |a_{t,T}(j)| \leq \frac{K}{\ell(j)}, \quad \text{with } K \text{ independent of } T; \tag{2.5}$$

(ii) there exist funcions $a(\cdot, j) : (0,1] \to \mathbb{R}$ with

$$\sup_u |a(u,j)| \leq \frac{K}{\ell(j)}, \tag{2.6}$$

$$\sup_j \sum_{t=1}^{T} \left| a_{t,T}(j) - a\left(\frac{t}{T}, j\right)\right| \leq K, \tag{2.7}$$

$$V(a(\cdot, j)) \leq \frac{K}{\ell(j)}, \tag{2.8}$$

(iii) μ has finite total variation, and the ε_t are i.i.d. with $E[\varepsilon_t] = 0$, $E[\varepsilon_s, \varepsilon_t] = 0$ for $s \neq t$ and $E[\varepsilon_t^2] = 1$.

For some local results, stronger smoothness assumptions have to be imposed. For example, for some i,

$$\sup_u \left|\frac{\partial^i \mu(u)}{\partial u^i}\right| \leq K, \tag{2.9}$$

$$\sup_u \left|\frac{\partial^i a(u,j)}{\partial u^i}\right| \leq \frac{K}{\ell(j)}, \quad \text{for } j = 0, 1, ... \tag{2.10}$$

and a stronger assumption than (2.7) is

$$\sup_{t,T} \left| a_{t,T}(j) - a\left(\frac{t}{T}, j\right)\right| \leq \frac{K}{T\ell(j)}. \tag{2.11}$$

Consequently, the stationary approximation of (2.2) can be constructed

$$\tilde{X}_t(u) = \mu(u) + \sum_{j=-\infty}^{\infty} a(u,j)\,\varepsilon_{t-j}.$$

Definition 2.2. *(Time-varying spectral density and covariance) Let $X_{t,T}$ be a stochastic process with representation as in (2.2). The function*

$$f(u,\lambda) := \frac{1}{2\pi} |A(u,\lambda)|^2 \tag{2.12}$$

is the time-varying spectral density of $X_{t,T}$, where $A(u,\lambda) := \sum_{j=-\infty}^{\infty} a(u,j)e^{-i\lambda j}$ is the time-varying transfer function, and

$$c(u,k) := \int_{-\pi}^{\pi} f(u,\lambda)e^{i\lambda k}d\lambda = \sum_{j=-\infty}^{\infty} a(u,k+j)a(u,j) \tag{2.13}$$

is the time-varying covariance of lag k at rescaled time u.

Under Assumption 2.1 with the condition (2.11), it can be shown that

$$cov\left(X_{[uT],T}, X_{[uT]+k,T}\right) = c(u,k) + O(T^{-1}), \tag{2.14}$$

uniformly in u and k. Moreover, the condition (2.11) implies that

$$\sup_{t,\lambda} \left| A_{t,T}(\lambda) - A\left(\frac{t}{T},\lambda\right)\right| \leq KT^{-1}. \tag{2.15}$$

Consider the Wigner-Ville spectrum for fixed T (Martin and Flandrin, 1985)

$$f_T(u,\lambda) := \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} cov\left(X_{[uT-s/2],T}, X_{[uT+s/2],T}\right) e^{-i\lambda s}. \tag{2.16}$$

Dahlhaus (1996a) showed that under Assumption 2.1 and the condition (2.10) for all j, that $f_T(u,\lambda)$ tends in squared mean to $f(u,\lambda)$ as defined in (2.12), for all $u \in (0,1)$, that is

$$\int_{-\pi}^{\pi} |f_T(u,\lambda) - f(u,\lambda)|^2 d\lambda = o(1), \quad \text{for all } u \in (0,1). \tag{2.17}$$

This result justifies the so called instantaneous spectrum. As the time-varying spectral density $f(u,\lambda)$ is uniquely defined, $A(u,\lambda)$, $a(u,j)$, and $\mu(t/T)$ are also uniquely determined (Dahlhaus, 2012).

As stated in Dahlhaus (2012), all theorems that will be presented here use "under suitable regularity conditions" and they have slightly different conditions, but all results are conjecturable to be prove under Assumption 2.1.

At the end of this section, we present the following proposition which presents the time-varying ARMA as an example of locally stationary process (see Dahlhaus and Polonik, 2009, Proposition 2.4).

Proposition 2.1. *(tvARMA). Consider the system of difference equations*

$$\sum_{j=0}^{p} \alpha_j \left(\frac{t}{T}\right) X_{t-j,T} = \sum_{k=0}^{q} \beta_k \left(\frac{t}{T}\right) \sigma \left(\frac{t-k}{T}\right) \varepsilon_{t-k} \tag{2.18}$$

where ε_t are i.i.d. with $E\varepsilon_t = 0$, $E|\varepsilon_t| < \infty$, $\alpha_0(u) \equiv \beta_0(u) \equiv 1$ and $\alpha_j(u) = \alpha_j(0)$, $\beta_k(u) = \beta_k(0)$ for $u < 0$. If all $\alpha_j(\cdot)$ and $\beta_k(\cdot)$, as well as $\sigma^2(\cdot)$, are of bounded variation and $\sum_{j=0}^{p} \alpha_j(u) z^j \neq 0$ for all u and all $0 < |z| \leq 1 + \delta$ for some $\delta > 0$, then there exists a solution of the form

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j) \varepsilon_{t-j}$$

which fullfills (2.6), (2.7) and (2.8) of Assumption 2.1. If the parameters are differentiable with bounded derivatives, then also (2.9), (2.10) and (2.11) are fulfilled (for $i = 1$). Moreover, the time-varying spectral density is given by

$$f(u, \lambda) = \frac{\sigma^2(u)}{2\pi} \frac{\left| \sum_{k=0}^{q} \beta_k(u) e^{i\lambda k} \right|^2}{\left| \sum_{j=0}^{p} \alpha_j(u) e^{i\lambda j} \right|^2}. \tag{2.19}$$

2.2 Time-varying autoregressive processes

In this section, we will present the time-varying autoregressive process as a special case of linear locally stationary process. The time-varying autoregressive process (tvAR(p)) has the following representation

$$X_{t,T} + \sum_{j=1}^{p} \alpha_j \left(\frac{t}{T}\right) X_{t-j,T} = \sigma \left(\frac{t}{T}\right) \varepsilon_t, \quad t \in \mathbb{Z}, \tag{2.20}$$

where the ε_t are independent random variables with mean zero and variance 1. We assume $\sigma(u) = \sigma(0)$, $\alpha_j(u) = \alpha_j(0)$ for $u < 0, j = 1, ..., p$, and $\sigma(u) = \sigma(1)$, $\alpha_j(u) = \alpha_j(1)$ for $u > 1, j = 1, ..., p$. Moreover, we also assume some smoothness conditions on $\sigma(\cdot)$ and $\alpha_j(\cdot), j = 1, ..., p$.

The idea of *locally stationary process* is that given a fixed time point $u_0 = \frac{t_0}{T}$, the process $X_{t,T}$ can be approximated by a stationary process $\tilde{X}_t(u_0)$ defined by

$$\tilde{X}_t(u_0) + \sum_{j=1}^{p} \alpha_j(u_0) \tilde{X}_{t-j}(u_0) = \sigma(u_0) \varepsilon_t, \quad t \in \mathbb{Z}. \tag{2.21}$$

Dahlhaus (2012) states that under some suitable regularity conditions, it can be shown that

$$\left| X_{t,T} - \tilde{X}_t(u_0) \right| = O_P \left(\left| \frac{t}{T} - u_0 \right| + \frac{1}{T} \right),$$

which justifies the notation "locally stationary process". Moreover, $X_{t,T}$ has unique time-varying spectral density which is locally the same as the spectral density of $\tilde{X}_t(u)$,

$$f(u,\lambda) = \frac{\sigma^2(u)}{2\pi}\left|1+\sum_{j=1}^{p}\alpha_j(u)e^{-ij\lambda}\right|^{-2}. \tag{2.22}$$

Furthermore, $X_{t,T}$ has locally the same autocovariance than $\tilde{X}_t(u)$ which is

$$c(u,j) := \int_{-\pi}^{\pi} e^{ij\lambda} f(u,\lambda)d\lambda, \quad j\in\mathbb{Z}.$$

2.2.1 Parametric Whittle-type estimates

A well-known estimation approach is assuming that the $(p+1)$-dimensional parameter curve can be parameterized by a finite-dimensional parameter, that is

$$\boldsymbol{\theta}(\cdot) = \big(\alpha_1(\cdot),...,\alpha_p(\cdot),\sigma^2(\cdot)\big) = \boldsymbol{\theta}_\eta(\cdot), \quad \eta\in\mathbb{R}^q.$$

Then, the stationary Whittle estimate, introduced by Whittle (1953), can be applied for the parameter curve $\boldsymbol{\theta}(\cdot)$ on a segment about u_0

$$\hat{\boldsymbol{\theta}}_T^W(u_0) := \operatorname*{argmin}_{\theta\in\Theta} \mathcal{L}_T^W(u_0,\boldsymbol{\theta}) \tag{2.23}$$

where $\mathcal{L}_T^W(u_0,\boldsymbol{\theta})$ is the Whittle likelihood

$$\mathcal{L}_T^W(u_0,\boldsymbol{\theta}) := \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\theta(\lambda) + \frac{I_T(u_0,\lambda)}{f_\theta(\lambda)}\right\}d\lambda, \tag{2.24}$$

with the tapered periodogram on a segment about u_0 given by

$$I_T(u_0,\lambda) := \frac{1}{2\pi H_N}\left|\sum_{s=1}^{N} h\left(\frac{s}{N}\right)X_{[u_0T]-N/2+s,T}e^{-i\lambda s}\right|^2, \tag{2.25}$$

where $h:[0,1]\to\mathbb{R}$ is a data taper with $h(x)=h(1-x)$, and $H_N := \sum_{j=0}^{N-1} h^2\left(\frac{j}{N}\right) \sim N\int_0^1 h^2(x)dx$ is the normalizing factor.

For fitting globally the parametric model $\boldsymbol{\theta}_\eta(\cdot)$ with time-varying spectrum $f_\eta(u,\lambda) := f_{\boldsymbol{\theta}_\eta(u)}(\lambda)$, Dahlhaus (1997) considered the block Whittle estimates

$$\hat{\boldsymbol{\eta}}_T^{BW}(u_0) := \operatorname*{argmin}_{\eta\in\Theta_\eta} \mathcal{L}_T^{BW}(\eta) \tag{2.26}$$

where

$$\mathcal{L}_T^{BW}(\boldsymbol{\eta}) := \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\eta(u_j,\lambda) + \frac{I_T(u_j,\lambda)}{f_\eta(u_j,\lambda)}\right\}d\lambda \tag{2.27}$$

is the block Whittle likelihood with $u_j := t_j/T$, $t_j := S(j-1) + N/2$, $j = 1, ..., M$ and $T = S(M-1) + N$. More details of this approach is presented in Section 2.6.

2.2.2 Inference for nonparametric tvAR models

If the time series is short or there is a specific parametric model, parametric estimates for tvAR(p) that were presented before are a good option. However, nonparametric models are preferred due to its flexibility.

Consider the (negative) conditional log-likelihood at time t of a tvAR(p)

$$\ell_{t,T}(\boldsymbol{\theta}) := -\log f_\theta(X_{t,T}|X_{t-1,T}, ..., X_{1,T}) \tag{2.28a}$$

$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\left(X_{t,T} + \sum_{j=1}^{p} \alpha_j X_{t-j,T}\right)^2, \tag{2.28b}$$

where $\boldsymbol{\theta} = (\alpha_1, ..., \alpha_p, \sigma^2)'$. Based on this conditional likelihood, several estimates can be constructed:

1. A Kernel estimate defined by

$$\hat{\boldsymbol{\theta}}(u_0) = \underset{\theta}{\operatorname{argmin}} \frac{1}{bT}\sum_{t=1}^{T} K\left(\frac{u_0 - t/T}{b}\right)\ell_{t,T}(\boldsymbol{\theta}), \tag{2.29}$$

   where $K : \mathbb{R} \to [0, \infty)$ is a kernel with $K(x) = K(-x)$, $\int K(x)dx = 1, K(x) = 0$ for $x \notin [-\frac{1}{2}, \frac{1}{2}]$ and b is the bandwidth.

2. A local polynomial fit defined by $\hat{\boldsymbol{\theta}}(u_0) = \hat{\boldsymbol{c}}$ with

$$(\hat{\boldsymbol{c}}_0, ..., \hat{\boldsymbol{c}}_d)' = \underset{\boldsymbol{c}_0,...,\boldsymbol{c}_d}{\operatorname{argmin}} \frac{1}{bT}\sum_{t=1}^{T} K\left(\frac{u_0 - t/T}{b}\right)\ell_{t,T}\left(\sum_{j=0}^{d} \boldsymbol{c_j}\left(\frac{t}{T} - u_0\right)^j\right), \tag{2.30}$$

   which is investigated by Kim (2001).

3. An orthogonal series estimate defined by

$$\bar{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} \ell_{t,T}\left(\sum_{j=0}^{J(T)} \boldsymbol{\beta}_j \psi_j\left(\frac{t}{T}\right)\right), \tag{2.31}$$

   where $\{\psi_j(\cdot)\}$ forms an orthonormal basis for functions in some function space.

4. A nonparametric maximum likelihood estimate defined by

$$\hat{\boldsymbol{\theta}}(\cdot) = \underset{\boldsymbol{\theta}(\cdot)\in\Theta}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} \ell_{t,T}\left(\theta\left(\frac{t}{T}\right)\right), \tag{2.32}$$

where Θ is an adequate function space.

5. Finally, a parametric fit for the curves $\theta(\cdot) = \theta_\eta(\cdot)$ with $\eta \in \mathcal{R}^q$ defined by

$$\hat{\boldsymbol{\eta}} = \operatorname*{argmin}_{\boldsymbol{\eta}} \frac{1}{T} \sum_{t=1}^{T} \ell_{t,T} \left(\theta_\eta \left(\frac{t}{T} \right) \right). \tag{2.33}$$

2.3 Local likelihood

In this section, a more general theoretical framework of nonparametric inference for time series with time-varying finite-dimensional parameters $\boldsymbol{\theta}(\cdot)$ is presented. The idea is that at each time point $u_0 \in (0,1)$, we approximate a stationary process $\tilde{X}_t(u_0)$ to the original process $X_{t,T}$.

Suppose that we estimate the multivariate parameter curve $\boldsymbol{\theta}(\cdot)$ by minimizing the local (negative) conditional log-likelihood

$$\hat{\boldsymbol{\theta}}_T^C(u_0) := \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} \mathcal{L}_T^C(u_0, \boldsymbol{\theta})$$

where

$$\mathcal{L}_T^C(u_0, \boldsymbol{\theta}) := \frac{1}{T} \sum_{t=1}^{T} \frac{1}{b} K \left(\frac{u_0 - t/T}{b} \right) \ell_{t,T}(\boldsymbol{\theta}). \tag{2.34}$$

and

$$\ell_{t,T}(\boldsymbol{\theta}) := -\log f_\theta \left(X_{t,T} | X_{t-1,T}, ..., X_{1,T} \right)$$

where K is a kernel defined as in (2.29). We assume that $b = b_T \to 0$ and $bT \to \infty$ as $T \to \infty$.

We approximate $\mathcal{L}_T^C(u_0, \boldsymbol{\theta})$ with $\tilde{\mathcal{L}}_T^C(u_0, \boldsymbol{\theta})$ which is defined by

$$\tilde{\mathcal{L}}_T^C(u_0, \boldsymbol{\theta}) := \frac{1}{T} \sum_{t=1}^{T} \frac{1}{b} K \left(\frac{u_0 - t/T}{b} \right) \tilde{\ell}_{t,T}(\boldsymbol{\theta}) \tag{2.35}$$

with the local (negative) conditional log-likelihood for the process $\tilde{X}_t(u_0)$

$$\tilde{\ell}_t(u_0, \boldsymbol{\theta}) := -\log f_\theta(\tilde{X}_t(u_0) | \tilde{X}_{t-1}(u_0), ..., \tilde{X}_1(u_0)).$$

To continue, we present the derivation of the asymptotic bias, mean-squared error, consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_T(u_0)$ for local minimum-distance function $\mathcal{L}_T(u_0, \boldsymbol{\theta})$ such as (2.34). Typically, both $\mathcal{L}_T(u_0, \boldsymbol{\theta})$ and $\tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta})$ will converge to the same limit-function $\mathcal{L}(u_0, \boldsymbol{\theta})$. Let

$$\boldsymbol{\theta_0}(u_0) := \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(u_0, \boldsymbol{\theta}).$$

and let

$$\mathcal{B}_T(u_0, \boldsymbol{\theta}) := \mathcal{L}_T(u_0, \boldsymbol{\theta}) - \tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta})$$

Theorem 2.1.

(i) Suppose that Θ is compact with $\boldsymbol{\theta}_0(u_0) \in Int(\Theta)$, the function $\mathcal{L}(u_0, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ and the minimum $\boldsymbol{\theta}_0(u_0)$ is unique. If

$$\sup_{\boldsymbol{\theta}\in\Theta} \left| \tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta}) - \mathcal{L}(u_0, \boldsymbol{\theta}) \right| \xrightarrow{P} 0, \quad \text{and} \tag{2.36}$$

$$\sup_{\boldsymbol{\theta}\in\Theta} |\mathcal{B}_T(u_0, \boldsymbol{\theta})| \xrightarrow{P} 0 \tag{2.37}$$

then

$$\hat{\boldsymbol{\theta}}_T(u_0) \xrightarrow{P} \boldsymbol{\theta}_0(u_0). \tag{2.38}$$

(ii) Suppose in addition that $\mathcal{L}(u, \theta)$ and $\boldsymbol{\theta}_0(u)$ are uniformly continuous in u and $\boldsymbol{\theta}$, and the convergence in (2.36) and (2.37) is uniformly in $u_0 \in [0, 1]$. Then

$$\sup_{u_0\in[0,1]} \left| \hat{\boldsymbol{\theta}}_T(u_0) - \boldsymbol{\theta}_0(u_0) \right| \xrightarrow{P} 0. \tag{2.39}$$

To continue, the results on asymptotic normality is stated. Denote ∇ the derivatives with respect to the θ_i, i.e., $\nabla := \left(\frac{\partial}{\partial \theta_i}\right)_{i=1,\ldots,d}$.

Theorem 2.2. *Let $\boldsymbol{\theta}_0 := \boldsymbol{\theta}_0(u_0)$. Suppose that $\mathcal{L}_T(u_0, \boldsymbol{\theta})$, $\tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta})$, and $\mathcal{L}(u_0, \boldsymbol{\theta})$ are twice continuously differentiable in $\boldsymbol{\theta}$ with nonsingular matrix $\Gamma(u_0) := \nabla^2 \mathcal{L}(u_0, \boldsymbol{\theta}_0)$. Let further*

$$\sqrt{bT} \nabla \tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V(u_0))$$

with some sequence $b = b_T$, where $b \to 0$ and $bT \to \infty$ and

$$\sup_{\boldsymbol{\theta}\in\Theta} \left| \nabla^2 \tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta}) - \nabla^2 \mathcal{L}(u_0, \boldsymbol{\theta}) \right| \xrightarrow{P} 0.$$

If in addition

$$\sqrt{bT} \left(\Gamma(u_0)^{-1} \nabla \mathcal{B}_T(u_0, \boldsymbol{\theta}_0) - \frac{b^2}{2} \boldsymbol{\mu}^0(u_0) \right) = o_P(1)$$

with some $\boldsymbol{\mu}^0(\cdot)$ and

$$\sup_{\boldsymbol{\theta}\in\Theta} \left| \nabla^2 \mathcal{B}_T(u_0, \boldsymbol{\theta}) \right| \xrightarrow{P} 0,$$

then

$$\sqrt{bT} \left(\hat{\boldsymbol{\theta}}_T(u_0) - \boldsymbol{\theta}(u_0) + \frac{b^2}{2} \boldsymbol{\mu}^0(u_0) \right) \xrightarrow{D} \mathcal{N} \left(0, \Gamma(u_0)^{-1} V(u_0) \Gamma(u_0)^{-1}\right). \tag{2.40}$$

2.4 Kullback-Leibler information for Gaussian processes

In this section we present some results from the linear local stationary processes (2.2) in the Section 2.1 with Gaussian innovations.

Consider the exact Gaussian maximum likelihood estimate

$$\hat{\eta}_T^{ML} := \underset{\eta \in \Theta_\eta}{\operatorname{argmin}} \, \mathcal{L}_T^E(\eta) \tag{2.41}$$

where η is a finite-dimensional parameter such as in (2.26), and

$$\mathcal{L}_T^E(\eta) = \frac{1}{2}\log(2\pi) + \frac{1}{2T}\log\det\Sigma_\eta + \frac{1}{2T}\left(\boldsymbol{X} - \mu_\eta\right)' \Sigma_\eta^{-1}\left(\boldsymbol{X} - \mu_\eta\right), \tag{2.42}$$

with $\boldsymbol{X} = (X_{1,T}, ..., X_{T,T})'$, $\mu_\eta = (\mu_\eta(1/T), ..., \mu_\eta(T/T))'$, and Σ_η the covariance matrix of the model. Under certain regularity conditions $\hat{\eta}_T^{ML}$ will converge to

$$\eta_0 := \underset{\eta \in \Theta_\eta}{\operatorname{argmin}} \, \mathcal{L}(\eta) \tag{2.43}$$

where

$$\mathcal{L}(\eta) := \lim_{T\to\infty} E\mathcal{L}_T^E(\eta). \tag{2.44}$$

The following theorem states that $\mathcal{L}(\eta)$ is equivalent to the calculation of the Kullback-Leibler information divergence.

Theorem 2.3. *Let $X_{t,T}$ be a locally stationary process with true mean and spectral density curves $\mu(\cdot)$, $f(u,\lambda)$ and model curves $\mu_\eta(\cdot)$, $f_\eta(u,\lambda)$, respectively. Under suitable regularity conditions, we have*

$$\begin{aligned}\mathcal{L}(\eta) &= \lim_{T\to\infty} \mathcal{L}_T^E(\eta) \\ &= \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\eta(u,\lambda) + \frac{f(u,\lambda)}{f_\eta(u,\lambda)}\right\}d\lambda du + \frac{1}{4\pi}\int_0^1 \frac{(\mu_\eta(u)-\mu(u))^2}{f_\eta(u,0)}du.\end{aligned} \tag{2.45}$$

Note that if we suppose a stationary model, i.e., $f_\eta(\lambda) := f_\eta(u,\lambda)$ and $m := \mu_\eta(u)$ does not depend on u, the Kullback-Leibler information divergence for stationary processes is obtained from the above theorem:

$$\mathcal{L}(\eta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\eta(\lambda) + \int_0^1 \frac{f(u,\lambda)du}{f_\eta(\lambda)}\right\}d\lambda + \frac{1}{4\pi}f_\eta(0)^{-1}\int_0^1 (m-\mu(u))^2 du. \tag{2.46}$$

Based on (2.45), consider a quasi-likelihood criterion

$$\mathcal{L}_T^{QL}(\eta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\eta(u,\lambda) + \frac{\hat{f}(u,\lambda)}{f_\eta(u,\lambda)}\right\}d\lambda du + \frac{1}{4\pi}\int_0^1 \frac{(\mu_\eta(u)-\hat{\mu(u)})^2}{f_\eta(u,0)}du. \tag{2.47}$$

where $\hat{f}(u,\lambda)$ and $\hat{\mu(u)}$ are suitable nonparametric estimates of $f(u,\lambda)$ and $\mu(u)$, respectively.

In order to study efficiency of parameter estimates, we define the Fisher information matrix as

$$\Gamma := \lim_{T\to\infty} T\ E_{\eta_0}\left\{(\nabla\mathcal{L}_T^E(\eta_0))(\nabla\mathcal{L}_T^E(\eta_0))'\right\},$$

and its following representation (for more details see Dahlhaus, 1996b).

Theorem 2.4. *Let $X_{t,T}$ be a locally stationary process with correctly specified mean curve $\mu_\eta(u)$ and time-varying spectral density $f_\eta(u,\lambda)$. Under suitable regularity conditions, we have*

$$\begin{aligned}\Gamma =&\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}(\nabla\log f_{\eta_0})(\nabla\log f_{\eta_0})'d\lambda du\\ &+\frac{1}{2\pi}\int_0^1(\nabla\mu_{\eta_0}(u))(\nabla\mu_{\eta_0}(u))'f_{\eta_0}^{-1}(u,0)du.\end{aligned} \tag{2.48}$$

Finally, we will present how the time-varying spectral density can be estimated. Let

$$\hat{f}_T(u,\lambda) := \frac{1}{b_f}\int K_f\left(\frac{\lambda-\mu}{b_f}\right)I_T(u,\mu)d\mu \tag{2.49}$$

where K_f is a symmetric kernel with $\int K_f(x)dx = 1$, b_f is the bandwidth in frequency direction, and $I_T(u,\lambda)$ is the tapered periodogram on a segment of length N about u as defined in (2.25).

Theorem 2.5. *Let $X_{t,T}$ be a locally stationary process with $\mu(\cdot)\equiv 0$. Under suitable regularity conditions, we have*

(i) $E\left\{I_T(u,\lambda)\right\} = f(u,\lambda)+\frac{1}{2}b_t^2\int_{-1/2}^{1/2}x^2K_t(x)dx\frac{\partial^2}{\partial u^2}f(u,\lambda)+o(b_t^2)+O\left(\frac{\log(b_tT)}{b_tT}\right)$;

(ii) $E\left\{\hat{f}_T(u,\lambda)\right\} = f(u,\lambda)+\frac{1}{2}b_t^2\int_{-1/2}^{1/2}x^2K_t(x)dx\frac{\partial^2}{\partial u^2}f(u,\lambda)+\frac{1}{2}b_f^2\left(\int_{-1/2}^{1/2}x^2K_f(x)dx\right)\frac{\partial^2}{\partial\lambda^2}f(u,\lambda)+$ $o(b_t^2+b_f^2+\frac{\log(b_tT)}{b_tT})$;

(iii) $Var\left\{\hat{f}_T(u,\lambda)\right\} = (b_tb_fT)^{-1}2\pi f(u,\lambda)^2\int_{-1/2}^{1/2}K_t(x)^2dx\int_{-1/2}^{1/2}K_f(x)^2dx(1+\delta_{\lambda 0})$.

Finally, minimizing the relative mean-squared error

$$\text{RMSE}(\hat{f}) := E(\hat{f}(u,\lambda)/f(u,\lambda-1)^2 \tag{2.50}$$

with respect to b_t, b_f, K_f and K_t, Dahlhaus (1996c) proved that with

$$\Delta_u := \frac{\frac{\partial^2}{\partial u^2}f(u,\lambda)}{f(u,\lambda)} \quad\text{and}\quad \Delta_\lambda := \frac{\frac{\partial^2}{\partial\lambda^2}f(u,\lambda)}{f(u,\lambda)},$$

the optimal RMSE is obtained with

$$b_t^{opt} = T^{-1/6}(576\pi)^{1/6}\frac{\Delta_\lambda}{\Delta_u^5}^{1/12} \quad\text{and}\quad b_f^{opt} = T^{-1/6}(576\pi)^{1/6}\frac{\Delta_u}{\Delta_\lambda^5}^{1/12},$$

and optimal kernels $K_t^{opt}(x) = K_f^{opt}(x) = 6(1/4 - x^2)$ with optimal rate $T^{-2/3}$.

2.5 Gaussian likelihood theory for locally stationary processes

In case of stationary processes, the Whittle likelihood is an approximation of the negative log Gaussian likelihood (2.42) (cf. Dzhaparidze and Kotz, 2012), that is,

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\eta(\lambda) + \frac{I_T(\lambda)}{f_\eta(\lambda)}\right\} d\lambda, \tag{2.51}$$

where $I_T(\lambda)$ is the periodogram. In order to introduce the Generalized Whittle Likelihood, the periodogram can be decomposed as follows

$$I_T(\lambda) = \frac{1}{2\pi T}\left|\sum_{r=1}^{T} X_r e^{-i\lambda r}\right|^2 = \frac{1}{T}\sum_{t=1}^{T} J_T\left(\frac{t}{T}, \lambda\right), \tag{2.52}$$

with the preperiodogram

$$J_T(u,\lambda) := \frac{1}{2\pi}\sum_{1\le[uT+0.5+k/2],[uT+0.5-k/2]\le T} X_{[uT+0.5+k/2],T}X_{[uT+0.5-k/2],T}e^{-i\lambda k}. \tag{2.53}$$

Observe that the preperiodogram can be interpreted as a local version of the periodogram at time t (for more details see Dahlhaus, 2012).

Then, we can define the generalized Whittle likelihood by replacing $I_T(\lambda)$ in (2.51) by the average of the preperiodogram and the model spectral density $f_\eta(\lambda)$ by the time-varying spectral density $f_\eta(u,\lambda)$ of a nonstationary model, that is

$$\mathcal{L}_T^{GW}(\eta) := \frac{1}{T}\sum_{j=1}^{T}\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\eta(\frac{t}{T},\lambda) + \frac{J_T(\frac{t}{T},\lambda)}{f_\eta(\frac{t}{T},\lambda)}\right\} d\lambda. \tag{2.54}$$

Observe that if the fitted model is stationary, then it is identical to the Whittle likelihood, thus the classical Whittle estimator is obtained. Let

$$\hat{\eta}_T^{GW} := \operatorname*{argmin}_{\eta\in\Theta_\eta} \mathcal{L}_T^{GW}(\eta) \tag{2.55}$$

be the quasi-likelihood estimate, and $\hat{\eta}_T^{ML}$ be the Gaussian maximum likelihood estimator (MLE) defined in (2.41). In the following theorem, the asymptotic normality result in the parametric case is presented.

Theorem 2.6. *Let $X_{t,T}$ be a locally stationary process. Under suitable regularity conditions* *we have in the case* $\mu(\cdot) = \mu_\eta(\cdot) = 0$

$$\sqrt{T}(\hat{\eta}_T^{GW} - \eta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma^{-1}V\Gamma^{-1})$$

and

$$\sqrt{T}(\hat{\eta}_T^{ML} - \eta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma^{-1}V\Gamma^{-1})$$

where

$$\Gamma_{ij} = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}(f - f_{\eta_0})\nabla_{ij}f_{\eta_0}^{-1}d\lambda du + \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}(\nabla_i \log f_{\eta_0})(\nabla_j \log f_{\eta_0})d\lambda du$$

and

$$V_{ij} = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}(\nabla_i f_\eta^{-1})(\nabla_j f_\eta^{-1})d\lambda du.$$

If the model is correctly specified, then $V = \Gamma$ *which is defined as in* (2.48)*. This means that both estimates are asymptotically Fisher efficient. Even more the sequence of experiments is locally asymptotically normal (LAN) and both estimates are locally asymptotically minimax.*

Finally, the following theorem states the properties of the different likelihoods discussed above.

Theorem 2.7. *Under suitable regularity conditions, we have for* $k = 0, 1, 2$*,*
(i) $\sup_{\eta\in\Omega_\eta}\left|\nabla^k\left\{\mathcal{L}_T^{GW}(\eta) - \mathcal{L}_T^{E}(\eta)\right\}\right| \xrightarrow{P} 0,$
(ii) $\sup_{\eta\in\Omega_\eta}\left|\nabla^k\left\{\mathcal{L}_T^{GW}(\eta) - \mathcal{L}(\eta)\right\}\right| \xrightarrow{P} 0,$
(iii) $\sup_{\eta\in\Omega_\eta}\left|\nabla^k\left\{\mathcal{L}_T^{E}(\eta) - \mathcal{L}(\eta)\right\}\right| \xrightarrow{P} 0.$

Under stronger assumptions, we can obtain $\hat{\eta}_T^{GW} - \hat{\eta}_T^{ML} = O_P(T^{-1+\varepsilon})$ (for more details see Dahlhaus, 2000).

2.6 Blocked Whittle estimation

In this section, we present a general estimation approach proposed by Dahlhaus (1997). Suppose that we are interested in fitting a locally stationary model with time-varying spectral density $f_\theta(u, \lambda), \theta \in \Theta \subset \mathbb{R}^P$ to observations $X_{1,T}, ..., X_{T,T}$. Let $h : \mathbb{R} \to \mathbb{R}$ be a data taper with $h(x) = 0$ for $x \notin [0, 1)$ and (for N even),

$$d_N(u, \lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right) X_{[uT]-N/2+s+1,T}\exp(-i\lambda s), \tag{2.56}$$

$$H_{k,N}(\lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right)^k \exp(-i\lambda s), \tag{2.57}$$

$$I_N(u, \lambda) := \frac{1}{2\pi H_{2,N}(0)}\left|d_N(u, \lambda)\right|^2. \tag{2.58}$$

$I_N(u,\lambda)$ is the periodogram over a segment of length N with midpoint $[uT]$. We shift by Q from segment to segment and calculate I_N over segments with midpoints $t_j := Q(j-1)+N/2,\ j=1,...,M$ where $T=Q(M-1)+N$, and $u_j := t_j/T$ is the rescaled midpoint. Next, define the blocked Whittle estimator

$$\hat{\boldsymbol{\theta}}_T^{BW} = \operatorname*{argmin}_{\theta\in\Theta} \mathcal{L}_T^{BW}(\theta) \tag{2.59}$$

where

$$\mathcal{L}_T^{BW}(\boldsymbol{\theta}) := \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left\{\log f_\theta(u_j,\lambda)+\frac{I_T(u_j,\lambda)}{f_\theta(u_j,\lambda)}\right\}d\lambda \tag{2.60}$$

is the blocked Whittle likelihood.

The justification of the blocked Whittle likelihood comes as follows. Let $\bar{f}$ and f be the probability density of the observations and the true spectral density, respectively, and $\bar{f}_\theta$ and f_θ be the corresponding probability density and the density of the proposed model, respectively. In Gaussian case, Dahlhaus (1996b) proved that the asymptotic Kullback-Leibler information divergence is

$$\lim_{T\to\infty}\frac{1}{T}E_{\bar{f}}\log(\bar{f}/\bar{f}_\theta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log f_\theta(u,\lambda)+\frac{f(u,\lambda)}{f_\theta(u,\lambda)}\right\}d\lambda du + constant. \tag{2.61}$$

Therefore,

$$\mathcal{L}(\theta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log f_\theta(u,\lambda)+\frac{f(u,\lambda)}{f_\theta(u,\lambda)}\right\}d\lambda du \tag{2.62}$$

may be considered as a distance between the true process with spectral density $f(u,\lambda)$ and the proposed model with spectral density $f_\theta(u,\lambda)$. If the model is correct,

$$\theta_0 = \operatorname*{argmin}_{\theta\in\Theta}\mathcal{L}(\theta) \tag{2.63}$$

is the true parameter of the model.

Theorem 2.8. *Suppose that we observe the realization $X_{1,T},...,X_{T,T}$ from a locally stationary process of Definition 2.1 with $\mu(u)=0$ satisfying the Assumption 2.1 and the equation (2.11). Moreover, suppose that the Assumption 3.1 from Dahlhaus (1997) holds. Then*

$$\hat{\theta}_T^{BW} \to \theta_0$$

in probability. Moreover,

$$\sqrt{T}\left(\hat{\theta}_T^{BW}-\theta_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, c_h\Gamma^{-1}(V+W)\Gamma^{-1}) \tag{2.64}$$

with

$$\begin{aligned}\Gamma =&\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left(f(u,\lambda)-f_{\theta_0}(u,\lambda)\right)\nabla^2 f_{\theta_0}(u,\lambda)^{-1}d\lambda du\\&+\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left(\nabla\log f_{\theta_0}(u,\lambda)\right)\left(\nabla\log f_{\theta_0}(u,\lambda)\right)' d\lambda du\\V =&\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} f(u,\lambda)^2\nabla f_{\theta_0}(u,\lambda)^{-1}\nabla f_{\theta_0}(u,\lambda)^{-1}d\lambda du\\W =&\frac{1}{8\pi}\int_0^1\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(u,\lambda)f(u,\mu)\nabla f_{\theta_0}^{-1}(u,\lambda)\nabla f_{\theta_0}^{-1}(u,\mu)' g_4(\lambda,-\lambda,\mu)d\lambda d\mu du,\end{aligned}$$

and $c_h = H_4/H_2^2$ *if* $S = N$, $c_h = 1$ *if* $S/N \to 0$, *and* g_4 *is defined in the Definition 2.1 in Dahlhaus (1997).*

2.7 Prediction

There are few works that address the problem for predicting and forecasting locally stationary processes, since the infill asymptotics is employed and the interest of researchers is generally focused on the behavior of the observed time period.

Van Bellegem and von Sachs (2004) apply locally stationary processes to predict economic data by considering the observed values $X_{0,T},\cdots,X_{T-h-1,T}$ and rescaling the time interval to $\left[0,1-\frac{h+1}{T}\right]$, where h is the forecasting horizon and the ratio h/T tends to zero as T tends to infinity. On the other hand, Palma *et al.* (2013) proposed a state space framework for estimating, prediction and making statistical inferences for Gaussian locally stationary processes with short and long memory, and the possibility of handling missing values. The Kalman filter proposed makes the possibility of obtaining the exact and approximate maximum likelihood estimates, one-step and multi-step predictors along with their error bands.

On the other hand, Bardet and Doukhan (2017) introduced a new class of time-varying AR(1) which is locally stationary but with periodic parameter functions. We will explore the possibility of predicting this process because of the periodic feature of the present time period could be extended to the future time period.

Chapter 3

Stable and tempered stable distributions

Most statistical models assume Gaussian error distribution due to the fact that it is the domain of attraction from all distributions with the finite variance, and thus, its theoretical results are easier to handle. However, different areas, such as actuarial science, biostatistics, computer science, finance and physics, have been observed phenomena with heavy tail distributions and/or infinite variance (Grabchak, 2016a). Hence, considering heavy tail distributions could be an alternative to model this kind of phenomena.

Stable distribution presents attractive theoretical properties, such as the extremely heavy tails and stability under linear combinations, but the fact that moments of order greater than two do not exist is a restrictive assumption in real-world applications. However, it is closed under linear combination which includes the possibility of handling asymmetry and thicker tails. On the other hand, a tempered stable distribution is obtained by changing the tail behavior of a stable distribution. As a result, its center is similar to that of the stable distribution, but its tails are lighter, which is called semi-heavy tails. In contrast to the stable distributions, tempered stable distribution keeps most of the attractive properties and still has all finite moments.

In this chapter, we will present the relevant results on the theory of stable distributions and tempered stable distributions.

3.1 Stable distribution

3.1.1 Stable distribution definition and its characteristic function

In this section, following the approach of Samorodnitsky and Taqqu (1994), four equivalent definitions of a stable distribution are presented, in which the last one is based on its characteristic function.

Definition 3.1. *A random variable X is said to have a stable distribution if for any positive numbers A and B, there is a positive number C and a real number D such that*

$$AX_1 + BX_2 \stackrel{d}{=} CX + D, \tag{3.1}$$

where X_1 and X_2 are independent copies of X, and where $\stackrel{d}{=}$ denotes equality in distribution.

A random variable X is called strictly stable if in (3.1) holds with $D = 0$, and a stable random variable X is called symmetric stable if its distribution is symmetric, i.e. if X and $-X$ have the same distribution. As a result, a symmetric stable random variable is strictly stable, but strictly stable random variable is not necessarily symmetric.

Definition 3.2. *A random variable X is said to have a stable distribution if for any $n \geq 2$, there is a positive number C_n and a real number D_n such that*

$$X_1 + X_2 + ... + X_n \stackrel{d}{=} C_n X + D_n, \tag{3.2}$$

where $X_1, X_2, ..., X_n$ are independent copies of X.

Definition 3.3. *A random variable X is said to have a stable distribution if it has a domain of attraction, i.e., if there is a sequence of i.i.d. random variables $Y_1, Y_2, ...$ and a sequence of positive numbers $\{d_n\}$ and real numbers $\{a_n\}$, such that*

$$\frac{Y_1 + Y_2 + ... + Y_n}{d_n} + a_n \stackrel{d}{\Rightarrow} X. \tag{3.3}$$

The notation $\stackrel{d}{\Rightarrow}$ denotes convergence in distribution.

Definition 3.4. *A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$ and μ real such that its characteristic function has the following form:*

$$\phi_X(u) := E\left(e^{i\theta X}\right) = \begin{cases} \exp\left\{-\sigma^\alpha|\theta|^\alpha(1 - i\beta(\operatorname{sign}\theta)\tan\frac{\pi\alpha}{2}) + i\mu\theta\right\} & \text{if } \alpha \neq 1, \\ \exp\left\{-\sigma|\theta|(1 + i\beta\frac{2}{\pi}(\operatorname{sign}\theta)\log|\theta|) + i\mu\theta\right\} & \text{if } \alpha = 1. \end{cases} \tag{3.4}$$

The parameter α is the index of stability and

$$\operatorname{sign}\theta = \begin{cases} 1 & \text{if } \theta > 0, \\ 0 & \text{if } \theta = 0, \\ -1 & \text{if } \theta < 0. \end{cases} \tag{3.5}$$

The definitions 3.1, 3.2, 3.3, 3.4 are equivalent. Especially, for the Definition 3.1, it can be showed that $C^\alpha = A^\alpha + B^\alpha$. In addition, for the Definition 3.2, $C_n = n^{1/\alpha}$ (for more details see Samorodnitsky and Taqqu, 1994).

Note that the characteristic function of a stable function in (3.4) is characterized by four parameters where $\alpha \in (0, 2]$ is the index of stability. Then, we will refer a random variable with stable distribution as α-stable random variable, and we will denote it by $S_\alpha(\sigma, \beta, \mu)$.

The probability densities of α-stable random variable are continuous, but they do not have closed form with three exceptions:

(i) (Gaussian distribution) When $\alpha = 2$ and $\beta = 0$, $X \sim S_2(\sigma, 0, \mu) = \mathcal{N}(\mu, 2\sigma^2)$ has density

$$f(x) = \frac{1}{2\sigma\sqrt{\pi}} e^{\frac{-(x-\mu)^2}{4\sigma^2}}.$$

(ii) (Cauchy distribution) When $\alpha = 1$ and $\beta = 0$, $X \sim S_1(\sigma, 0, \mu)$ has density

$$f(x) = \frac{\sigma}{\pi((x-\pi)^2 + \sigma^2)}.$$

(iii) (Lévy distribution) When $\alpha = 1/2$ and $\beta = 1$, $X \sim S_{1/2}(\sigma, 1, \mu)$ has density

$$f(x) = \left(\frac{\sigma}{2\pi}\right)^{1/2} e^{-\frac{\sigma}{2(x-\mu)}} I_{(x>\mu)}.$$

(iv) $X \sim S_\alpha(0, 0, \mu)$ has the degenerate distribution for any $0 < \alpha \leq 2$, which is not the interest of this research.

Figure 3.1 compares the standard Gaussian distribution, standard Cauchy distribution, standard Lévy distribution and t-distribution ($\nu = 3$). Note that the tail behavior of the Cauchy distribution is much heavier than the t-distribution and normal distribution. On the other hand, the Lévy distribution is asymmetric, while others are symmetric.

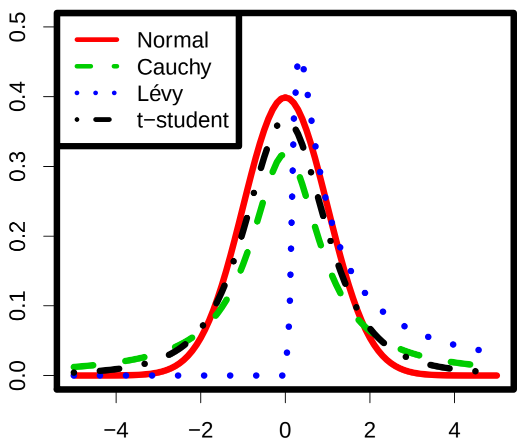


Figure 3.1: *Density function of standard Normal* $S_2(1/\sqrt{2}, 0, 0)$*, standard Cauchy* $S_1(1, 0, 0)$*, standard Lévy* $S_{1/2}(1, 1, 0)$ *and t-distribution* ($\nu = 3$)*.*

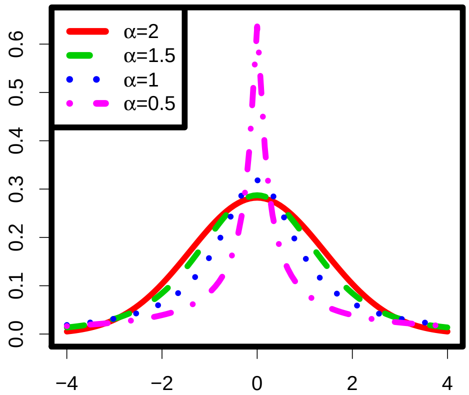
Figure 3.2 presents how varying each parameter can change the density function behavior: α indicates the tail heaviness, β controls the skewness, σ is the scale parameter, and μ is the shift parameter. In the next section, we will present properties of α-stable distribution to justify the theoretical meaning of these parameters.

3.1.2 Properties

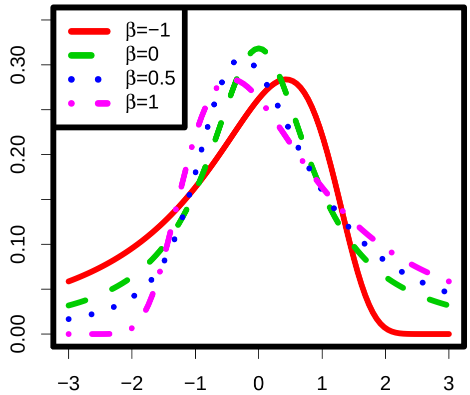
Some important properties of stable random variables are presented.

Proposition 3.1. *Let* X_1 *and* X_2 *be independent random variables with* $X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i), i = 1, 2$*. Then* $X_1 + X_2 \sim S_\alpha(\sigma, \beta, \mu)$*, with*

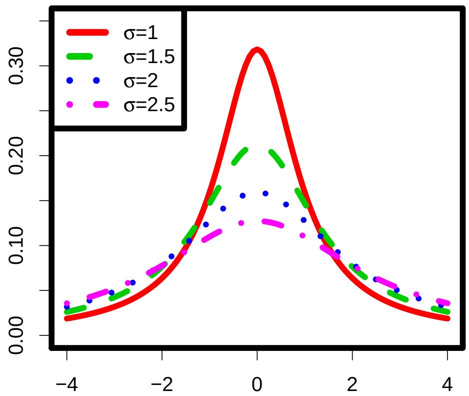
$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \quad \beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2.$$



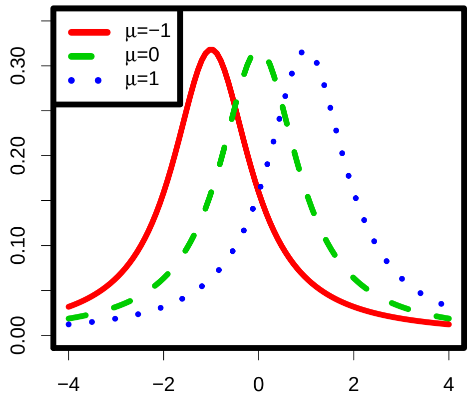
(a) $S_\alpha(1, 0, 0)$ varying α.



(b) $S_1(1, \beta, 0)$ varying β.



(c) $S_1(\sigma, 0, 0)$ varying σ.



(d) $S_1(1, 0, \mu)$ varying μ.

Figure 3.2: *Density functions of α-stable distributions varying their parameters.*

Proposition 3.2. *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ and let a be a real constant. Then $X+a \sim S_\alpha(\sigma, \beta, \mu+a)$.*

Proposition 3.3. *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ and let a be a non-zero real constant. Then*

$$\begin{aligned} aX &\sim S_\alpha(|a|\sigma, \text{sign}(a)\beta, a\mu) & \text{if } \alpha \neq 1 \\ aX &\sim S_1\left(|a|\sigma, \text{sign}(a)\beta, a\mu - \tfrac{2}{\pi}a(\ln|a|)\sigma\beta\right) & \text{if } \alpha = 1 \end{aligned} \tag{3.6}$$

Thus, the parameter σ is called the scale parameter. However, observe that when $\alpha = 1$ and $\beta \neq 0$, this name does not make sense since the multiplication by a constant affects the shift parameter in a non-linear way. When $\mu = 0$, we have the following proposition.

Proposition 3.4. *For any $0 < \alpha < 2$. Then*

$$X \sim S_\alpha(\sigma, \beta, 0) \Leftrightarrow -X \sim S_\alpha(\sigma, -\beta, 0). \tag{3.7}$$

Proposition 3.5. *Let $X \sim S_\alpha(\sigma, \beta, \mu)$.*

(i) If $\alpha \neq 1$, then X is strictly stable if and only if $\mu = 0$.

(ii) X with $\alpha = 1$ is strictly stable if and only if $\beta = 0$.

The next proposition describes α as a tail heaviness parameter. For α close to 2 indicates that the tail is thinner (closer to the normal distribution), while decreasing α means that the tail will get heavier.

Proposition 3.6. *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $\alpha < 2$. Then*

$$\begin{cases} \lim\limits_{\lambda\to\infty} \lambda^\alpha P\{X > \lambda\} = C_\alpha \left(\frac{1+\beta}{2}\right)\sigma^\alpha, \\ \lim\limits_{\lambda\to\infty} \lambda^\alpha P\{X > -\lambda\} = C_\alpha \left(\frac{1-\beta}{2}\right)\sigma^\alpha, \end{cases} \tag{3.8}$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha}\sin x dx\right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)cos(\pi\alpha/2)} & \text{if } \alpha \neq 1, \\ 2/\pi & \text{if } \alpha = 1. \end{cases}$$

Proposition 3.6 describes the tail behavior and it leads to the following property related to its moments.

Proposition 3.7. *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$. Then*

$$\begin{aligned} E|X|^p < \infty \quad &\textit{for any } 0 < p < \alpha, \\ E|X|^p = \infty \quad &\textit{for any } p \geq \alpha \end{aligned} \tag{3.9}$$

Proposition 3.8. *Let $X \sim S_\alpha(\sigma, \beta, 0)$ with $0 < \alpha < 2$ and $\beta = 0$ in the case $\alpha = 1$. Then, for every $0 < p < \alpha$, there is a constant $c_{\alpha,\beta}(p)$ such that*

$$(E|X|^p)^{1/p} = c_{\alpha,\beta}(p)\sigma. \tag{3.10}$$

The constant $c_{\alpha,\beta}(p) = (E|X_0|^p)^{1/p}$ where $X_0 \sim S_\alpha(1, \beta, 0)$ and

$$(c_{\alpha,\beta}(p))^p = \frac{2^{p-1}\Gamma\left(1 - \frac{p}{\alpha}\right)}{p\int_0^\infty u^{-p-1}\sin^2 u du}\left(1 + \beta^2 \tan^2 \frac{\alpha\pi}{2}\right)^{p/2\alpha} \cos\left(\frac{p}{\alpha}\arctan\left(\beta\tan\frac{\alpha\pi}{2}\right)\right). \tag{3.11}$$

Proposition 3.9. *When $1 < \alpha \leq 2$, the shift parameter μ equals the mean.*

Proposition 3.10. *(Laplace transform) The Laplace transform $Ee^{-\gamma X}$, $\gamma \geq 0$, of a random variable $X \sim S_\alpha(\sigma, 1, 0)$, $0 < \alpha \leq 2$, $\sigma > 0$, is*

$$Ee^{-\gamma X} = \begin{cases} \exp\left\{-\frac{\sigma^\alpha}{\cos\frac{\pi\alpha}{2}}\gamma^\alpha\right\} & \textit{if } \alpha \neq 1, \\ \exp\left\{\sigma\frac{2}{\pi}\gamma\ln\gamma\right\} & \textit{if } \alpha = 1. \end{cases} \tag{3.12}$$

3.1.3 Symmetric α-stable random variables

In this section, we describe some important results of a symmetric random variable and then some main properties of a symmetric α-stable random variable are stated. A random variable X is called symmetric if $X \stackrel{d}{=} -X$.

Proposition 3.11. *Let $X \stackrel{d}{=} F$. Then, F is symmetric if, and only if, its characteristic function $\Phi_X(t)$ is real.*

Proposition 3.12. *If U and V are independent and symmetric, then $U - V$ and $U + V$ are symmetric.*

Proposition 3.13. *Let $X_1, \cdots, X_n$ be independent copies of X. If X is symmetric, then $S_n = X_1 + \cdots + X_n$ is symmetric.*

Proposition 3.14. *If X is symmetric and $E(X)$ exists, then $E(X) = 0$.*

In case of a stable random variable, we have:

Proposition 3.15. *If X is stable and independent of Y with $Y \stackrel{d}{=} X$, then $X - Y$ is stable and symmetric.*

Proposition 3.16. *$X \sim S_\alpha(\sigma, \beta, \mu)$ is symmetric if and only if $\beta = 0$ and $\mu = 0$. It is symmetric about μ if and only if $\beta = 0$.*

Since X is $S\alpha S$ (symmetric α-stable) if and only if $X \sim S_\alpha(\sigma, 0, 0)$, then its characteristic function is given by

$$E\left(e^{i\theta X}\right) = e^{-\sigma^\alpha |\theta|^\alpha}. \tag{3.13}$$

The distribution $S \sim S_\alpha(\sigma, \beta, 0)$ is said to be skewed to the right if $\beta > 0$ and skewed to the left if $\beta < 0$. Moreover, it is said to be totally skewed to the right if $\beta = 1$ and totally skewed to the left if $\beta = -1$. Therefore, the parameter β is identified as a skewness parameter.

3.1.4 Dependence structure

In case of Gaussian random variables, the existence of the second moment allows the definition of the covariance function and, hence, the study of dependence structure. In the case of stable random variable, however, the second moment does not exist (Proposition 3.7). The concept of covariation and codifference are basically measures of dependence for infinite variance random variables in order to replace the covariance (Kokoszka and Taqqu, 1994, 1995). The first one is defined only for $1 < \alpha < 2$ and it is not a useful tool compared to the codifference, which is defined for all $0 < \alpha \leq 2$. The advantage of this measure is that it does not require conditions on moments of the random variables and is defined in term of the characteristic functions.

First consider two random variables X and Y, and define a function $I_{X,Y}(\xi_1, \xi_2)$

$$I_{X,Y}(\xi_1, \xi_2) = -\ln E\left[e^{i(\xi_1 X + \xi_2 Y)}\right] + \ln E\left[e^{i\xi_1 X}\right] + \ln E\left[e^{i\xi_2 Y}\right]. \tag{3.14}$$

If both random variables are Gaussian ($\alpha = 2$), that is, $X \sim \mathcal{N}(\mu_X, \sigma_X)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$ and $Cov(X,Y) = \sigma_{X,Y}$. Then,

$$\begin{aligned} E\left[e^{i\xi_1 X}\right] &= e^{i\xi_1\mu_X} e^{-\frac{1}{2}\sigma_X\xi_1^2} \\ E\left[e^{i\xi_2 Y}\right] &= e^{i\xi_1\mu_Y} e^{-\frac{1}{2}\sigma_Y\xi_2^2} \\ E\left[e^{i(\xi_1 X+\xi_2 Y)}\right] &= e^{i(\xi_1\mu_X+\xi_2\mu_Y)} e^{-\frac{1}{2}\left(\xi_1^2\sigma_X+\xi_2^2\sigma_Y+2\xi_1\xi_2\sigma_{X,Y}\right)} \end{aligned}$$

Substituting in the equation (3.14), it is straightforward to obtain

$$I_{X,Y}(\xi_1,\xi_2) = \xi_1\xi_2\sigma_{X,Y}.$$

This means that $I_{X,Y}$ is proportional to the covariance. Kokoszka and Taqqu (1995) suggest to define the codifference by taking $\xi_1 = 1$ and $\xi_2 = -1$ and then,

$$Cov(X,Y) = \sigma_{X,Y} = -I_{X,Y}(1,-1).$$

Definition 3.5. *The codifference between two random variables X and Y is defined as*

$$\tau(X,Y) = -I_{X,Y}(1,-1) = \ln E\left[e^{i(X-Y)}\right] - \ln E\left[e^{iX}\right] - \ln E\left[e^{-iY}\right]. \tag{3.15}$$

3.2 Tempered stable distributions

In this section, we present the tempered stable distribution and some of their relevant properties. A random variable X follows a tempered stable distribution, first introduced by Koponen (1995), if its Lèvy measure[1] is given by:

$$M(dx) = \left(\frac{C_+e^{-\lambda_+ x}}{x^{1+\alpha}}\mathbb{1}_{(x>0)} + \frac{C_-e^{-\lambda_-|x|}}{|x|^{1+\alpha}}\mathbb{1}_{(x<0)}\right)dx, \tag{3.16}$$

where $\alpha < 2$, and $C_+, C_-, \lambda_+, \lambda_- \in (0,+\infty)$. The reason that it is called tempered stable distribution is because its Lèvy measure can be expressed as

[1] For more details see Appendix A.

$$M(dx) = q(x)M_{stable}(dx),$$

where

$$M_{stable}(dx) = \left(\frac{C_+}{x^{1+\alpha}} \mathbb{1}_{(x>0)} + \frac{C_-}{|x|^{1+\alpha}} \mathbb{1}_{(x<0)} \right) dx$$

is the Lévy measure of an $\alpha-$stable distribution and $q : \mathbb{R} \to \mathbb{R}_+$ is a tempering function

$$q(x) = e^{-\lambda_+ x} \mathbb{1}_{(x>0)} + e^{-\lambda_- |x|} \mathbb{1}_{(x<0)}.$$

Remark 3.1. *Cont and Tankov (2015) define the generalized tempered stable distribution with six parameters $\alpha_+, \alpha_- < 2$, and $C_+, C_-, \lambda_+, \lambda_- \in (0, +\infty)$ with Lévy measure given by:*

$$M(dx) = \left(\frac{C_+ e^{-\lambda_+ x}}{x^{1+\alpha_+}} \mathbb{1}_{(x>0)} + \frac{C_- e^{-\lambda_- |x|}}{|x|^{1+\alpha_-}} \mathbb{1}_{(x<0)} \right) dx, \tag{3.17}$$

where $\alpha_+, \alpha_- < 2$, and $C_+, C_-, \lambda_+, \lambda_- \in (0, +\infty)$. Moreover, the characteristic function is obtained by solving the integral:

$$\phi_X(u) := E\left(e^{iuX}\right) = \exp\left[iu\gamma_0 + \int_{\mathbb{R}} \left(e^{iux} - 1 - iux\right) \nu(dx) \right]. \tag{3.18}$$

Moreover, by setting $\gamma_0 = \mu - \Gamma(1-\alpha)\left(C_+ \lambda_+^{\alpha-1} - C_- \lambda_-^{\alpha-1}\right)$, the tempered stable distribution has mean μ.

Proposition 3.17. *Let X be a generalized tempered stable random variable. If $\alpha_\pm \neq 1$ and $\alpha_\pm \neq 0$, its characteristic function is*

$$\begin{aligned} \phi_X(u) = \exp\Bigg\{ iu\mu + \Gamma(-\alpha_+)\lambda_+^{\alpha_+} C_+ \left[\left(1 - \frac{iu}{\lambda_+}\right)^{\alpha_+} - 1 + \frac{iu\alpha_+}{\lambda_+} \right] \\ + \Gamma(-\alpha_-)\lambda_-^{\alpha_-} C_- \left[\left(1 + \frac{iu}{\lambda_-}\right)^{\alpha_-} - 1 - \frac{iu\alpha_-}{\lambda_-} \right] \Bigg\}. \end{aligned} \tag{3.19}$$

If $\alpha_+ = \alpha_- = 1$,

$$\phi_X(u) = \exp\left\{iu(\mu + C_+ - C_-) + C_+(\lambda_+ - iu)\log\left(1 - \frac{iu}{\lambda_+}\right)\right. \\ \left. + C_-(\lambda_- + iu)\log\left(1 + \frac{iu}{\lambda_-}\right)\right\}, \tag{3.20}$$

and if $\alpha_+ = \alpha_- = 0$*,*

$$\phi_X(u) = \exp\left\{iu\mu + C_+\left\{\frac{iu}{\lambda_+} + \log\left(1 - \frac{iu}{\lambda_+}\right)\right\} - C_-\left\{-\frac{iu}{\lambda_-} + \log\left(1 + \frac{iu}{\lambda_-}\right)\right\}\right\}. \tag{3.21}$$

Using the proposition 3.17, the cumulant of order n can be obtained by

$$c_n(X) = \frac{1}{i^n}\frac{\partial^n}{\partial u^n}\log\left(E\left[e^{iuX}\right].\right)\Big|_{u=0} \tag{3.22}$$

Thus, the first cumulants of the generalized tempered stable distributions are:

$$\begin{cases} c_1(X) = E(X) = \mu, \\ c_2(X) = Var(X) = \Gamma(2-\alpha_+)C_+\lambda_+^{\alpha_+-2} + \Gamma(2-\alpha_-)C_-\lambda_-^{\alpha_--2}, \\ c_3(X) = \Gamma(3-\alpha_+)C_+\lambda_+^{\alpha_+-3} - \Gamma(3-\alpha_-)C_-\lambda_-^{\alpha_--3}, \\ c_4(X) = \Gamma(4-\alpha_+)C_+\lambda_+^{\alpha_+-4} + \Gamma(4-\alpha_-)C_-\lambda_-^{\alpha_--4}. \end{cases} \tag{3.23}$$

The generalized tempered stable distribution includes several particular cases in the literature.

- For $\alpha_+ = \alpha_-$, the KoBol distribution is obtained (Boyarchenko and Levendorskiĭ , 2000).
- For $C_+ = C_- = C$ and $\alpha_+ = \alpha_- = \alpha$, the CGMY distribution is obtained (Carr and Geman , 2002).
- For $\alpha_+ = \alpha_-$ and $\lambda_+ = \lambda_-$, the truncated Lévy flight is obtained (Koponen, 1995).
- For $\alpha_+ = \alpha_- = 0$, we have the Bilateral gamma distribution (Küchler and Tappe , 2008).
- For $\alpha_+ = \alpha_- = 0$, $C_+ = C_- = C$ and $\lambda_+ = \lambda_-$, we have the variance gamma

distribution (Madan and Seneta, 1990).

Note that unlike the stable processes, the tempered stable distribution is well defined for $\alpha < 0$. In this case, the compound Poisson models are obtained. However, we only consider the case of $\alpha > 0$ so that the small jumps do have stable-like behavior[2].

As in Rroji and Mercuri (2015) and Kim *et al.* (2008), we consider the same restrictions: $\alpha_+ = \alpha_- = \alpha \in (0, 2)$ and $\gamma_0 = \mu - \Gamma(1-\alpha)\left(C_+\lambda_+^{\alpha-1} - C_-\lambda_-^{\alpha-1}\right)$. In this way, this random variable is called classical tempered stable and it is denoted by $X \sim CTS(\alpha, \lambda_+, \lambda_-, C_+, C_-, \mu)$. For this random variable, by applying Proposition 3.17 and setting $\alpha_+ = \alpha_-$, its characteristic function is as follows:

$$\phi_X(u) = \begin{cases} \exp\left\{iu\mu - iu\Gamma(1-\alpha)\left(C_+\lambda_+^{\alpha-1} - C_-\lambda_-^{\alpha-1}\right) + C_+\Gamma(-\alpha)\left[(\lambda_+ - iu)^\alpha - \lambda_+^\alpha\right] \right. \\ \qquad \left. + C_-\Gamma(-\alpha)\left[(\lambda_- + iu)^\alpha - \lambda_-^\alpha\right]\right\}, & \text{if } \alpha \neq 1, \\ \exp\left\{iu(\mu + C_+ - C-) + C_+(\lambda_+ - iu)\log\left(1 - \frac{iu}{\lambda_+}\right) \right. \\ \qquad \left. + C_-(\lambda_- + iu)\log\left(1 + \frac{iu}{\lambda_-}\right)\right\}, & \text{if } \alpha = 1. \end{cases} \tag{3.24}$$

Moreover, it is straightforward to obtain the cumulant of order n by derivating the characteristic exponent. We obtain $c_1(X) = \mu$, and for $n \geq 2$:

$$c_n(X) = \Gamma(n-\alpha)\left(C_+\lambda_+^{\alpha-n} + (-1)^n C_-\lambda_-^{\alpha-n}\right). \tag{3.25}$$

As consequences, the first four moments of the distribution are:

$$\begin{cases} E(X) = c_1(X) = \mu \\ Var(X) = c_2(X) = \Gamma(2-\alpha)\left[C_+\lambda_+^{\alpha-2} + C_-\lambda_-^{\alpha-2}\right] \\ \gamma_1 = \frac{c_3(X)}{c_2^{3/2}(X)} = \frac{\Gamma(3-\alpha)\left[C_+\lambda_+^{\alpha-3} - C_-\lambda_-^{\alpha-3}\right]}{c_2^{3/2}(X)} \\ \gamma_2 = 3 + \frac{c_4(X)}{c_2^2(X)} = 3 + \frac{\Gamma(4-\alpha)\left[C_+\lambda_+^{\alpha-4} + C_-\lambda_-^{\alpha-4}\right]}{c_2^2(X)} \end{cases} \tag{3.26}$$

Note that the sign of the skewness depends on the difference between $C_+\lambda_+^{\alpha-3}$ and

[2] Stable and tempered stable distributions are infinitely divisible distributions and they are closely related to Lévy processes. This class of stochastic process has interesting properties and is not covered in this thesis (For more details cf. Cont and Tankov, 2015).

$C_-\lambda_-^{\alpha-3}$. On the other hand, similar to the stable distributions, there are linear combinations of tempered stable distributions that are still tempered stable distributions under some parameter restrictions.

Proposition 3.18 (Lemma 4.1. from Küchler and Tappe, 2013)**.**

(1) Suppose that $X_i \sim CTS(\alpha, \lambda_+, \lambda_-, C_{+i}, C_{-i}, \mu_i)$, $i = 1, 2$ are independent. Then,

$$X_1 + X_2 \sim CTS(\alpha, \lambda_+, \lambda_-, C_{+1} + C_{+2}, C_{-1} + C_{-2}, \mu_1 + \mu_2) \tag{3.27}$$

(2) For $X \sim CTS(\alpha, \lambda_+, \lambda_-, C_+, C_-, \mu_i)$ and $\rho > 0$, we have

$$\rho X \sim CTS(\alpha, \lambda_+/\rho, \lambda_-/\rho, C_+\rho^\alpha, C_-\rho^\alpha, \mu) \tag{3.28}$$

Proof. By applying equation (3.24), it can be verified for (1) $\phi_{X_1+X_2}(u) = \phi_{X_1}(u)\phi_{X_2}(u)$ and for (2) $\phi_{\rho X}(u) = \phi_X(\rho u)$. □

Proposition 3.19. *For $\lambda_+ = \lambda_- = \lambda$, $\mu = 0$ and $C_+ = C_- = C$, the tempered stable distribution converges to the symmetric stable distribution when λ goes to zero.*

3.2.1 Standardized classical tempered stable distributions

In this section, we introduce the standardized classical tempered stable distribution and study some of important properties.

Definition 3.6. *Let $X \sim CTS(\alpha, \lambda_+, \lambda_-, C_+, C_-, \mu)$. By setting $\mu = 0$ and*

$$C = C_+ = C_- = \frac{1}{\Gamma(2-\alpha)\left(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}\right)}, \tag{3.29}$$

X has zero mean and unit variance. It is called standardized classical tempered stable distribution, which will be denoted by $X \sim stdCTS(\alpha, \lambda_+, \lambda_-)$.

By applying the proposition 3.17, its characteristic function has the following form:

$$E\left(e^{iuX}\right) = \begin{cases} \exp\left\{\frac{(\lambda_+ - iu)^\alpha - \lambda_+^\alpha + (\lambda_- + iu)^\alpha - \lambda_-^\alpha}{\alpha(\alpha-1)(\lambda_+^{\alpha-2}+\lambda_-^{\alpha-2})} + \frac{iu(\lambda_+^{\alpha-1}-\lambda_-^{\alpha-1})}{(\alpha-1)(\lambda_+^{\alpha-2}+\lambda_-^{\alpha-2})}\right\}, & \text{if } \alpha \neq 1, \\ \exp\left\{\frac{1}{\lambda_+^{-1}+\lambda_-^{-1}}\left[(\lambda_+ - iu)\log\left(1-\frac{iu}{\lambda_+}\right) + (\lambda_- + iu)\log\left(1+\frac{iu}{\lambda_-}\right)\right]\right\}, & \text{if } \alpha = 1. \end{cases} \tag{3.30}$$

Remark 3.2. *Let $X \sim stdCTS(\alpha, \lambda_+, \lambda_-)$. X converges to the standard normal distribution when $\alpha \to 2$.*

By substituting (3.29) in the equation (3.22), the cumulants of a stdCTS distribution are $c_1(X) = 0$ and for $n \geq 2$

$$c_n(X) = \frac{\Gamma(n-\alpha)\left[\lambda_+^{\alpha-n} + (-1)^n \lambda_-^{\alpha-n}\right]}{\Gamma(2-\alpha)\left(\lambda_+^{\alpha-2}+\lambda_-^{\alpha-2}\right)}. \tag{3.31}$$

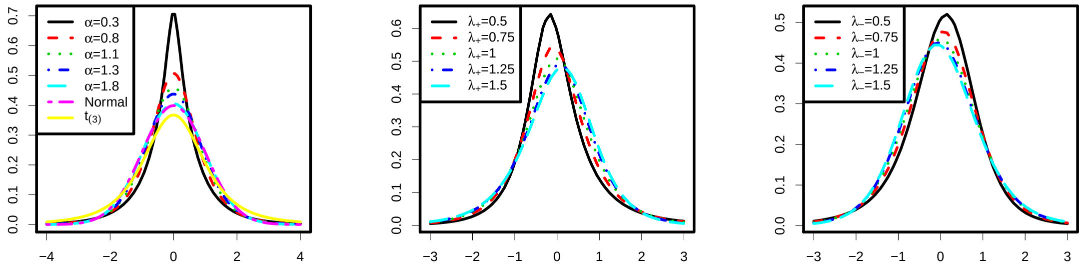
As consequences, the first four moments of the stdCTS distribution are:

$$\begin{cases} E(X) = c_1(X) = 0 \\ Var(X) = c_2(X) = 1 \\ \gamma_1 = c_3(X) = \frac{(2-\alpha)\left(\lambda_+^{\alpha-3}-\lambda_-^{\alpha-3}\right)}{\left(\lambda_+^{\alpha-2}+\lambda_-^{\alpha-2}\right)} \\ \gamma_2 = 3 + c_4(X) = 3 + \frac{(3-\alpha)(2-\alpha)\left(\lambda_+^{\alpha-4}+\lambda_-^{\alpha-4}\right)}{\left(\lambda_+^{\alpha-2}+\lambda_-^{\alpha-2}\right)} \end{cases} \tag{3.32}$$

In this case, the sign of the skewness depends on the difference between $\lambda_+^{\alpha-3}$ and $\lambda_-^{\alpha-3}$. For $\lambda_+ > \lambda_-$, it is positively skewed, while $\lambda_+ < \lambda_-$, it is negatively skewed. On the other hand, the stdCTS distribution always has kurtosis greater than 3.

Figure 3.3 presents how each parameter can change the density function behavior. By varying α, they are similar to stable distributions and still more leptokurtic than the standard Gaussian distribution and t-distribution ($\nu = 3$). By varying λ_+ and λ_-, it can be observed that the asymmetry changes.

In order to understand the skewness and the kurtosis of the stdCTS distribution, Figures 3.4 and 3.5 present how the skewness and kurtosis changes by varying λ_+ and λ_- for $\alpha = 0.5$ and 1.5. For lower α, the asymmetry and leptokurtisis away from the Gaussian case are more noticeable.

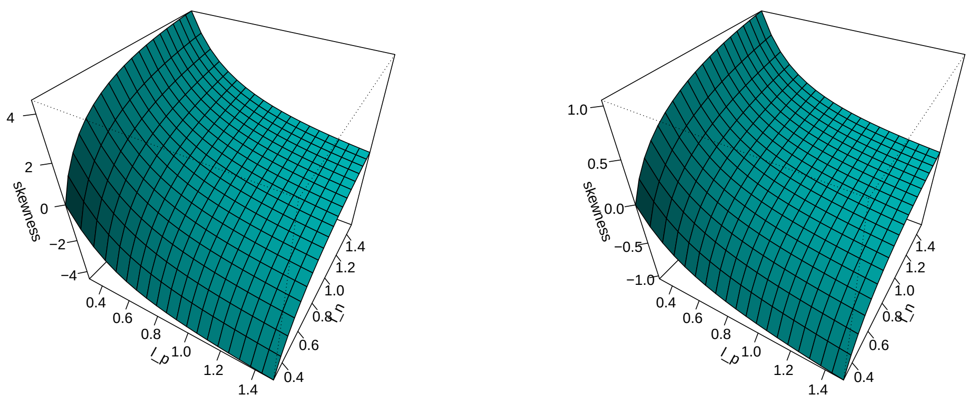


(a) $stdCTS(\alpha, 1, 1)$, standardized normal and t distribution.

(b) $stdCTS(0.8, \lambda_+, 1)$.

(c) $stdCTS(1.1, 1, \lambda_-)$.

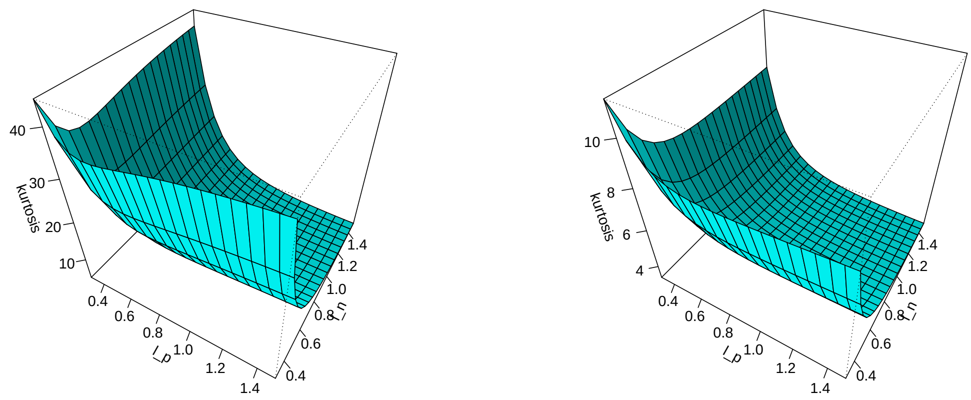
Figure 3.3: *Density functions of standardized tempered stable distributions varying each of their parameters with other parameters fixed.*



(a) $\alpha = 0.5$.

(b) $\alpha = 1.5$.

Figure 3.4: *Skewness of standardized tempered stable distributions with $\alpha = 0.5$ and 1.5 varying λ_+ and λ_-.*



(a) $\alpha = 0.5$.   (b) $\alpha = 1.5$.

Figure 3.5: *Kurtosis of standardized tempered stable distributions with* $\alpha = 0.5$ *and* 1.5 *varying* λ_+ *and* λ_-.

3.3 Simulation

3.3.1 Stable distribution

When a random variable has the density function and distribution function, its simulation is an easy task. However, the general stable random variables do not have closed form. Weron and Weron (1995) proposed an algorithm to generate α-stable distribution. In order to simulate a random variable $X \sim S_\alpha(1, \beta, 0)$,

1. generate a random variable U uniformly distributed on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, and an independent exponential random variable W with mean 1, then

2. let $B_{\alpha,\beta} = \frac{\arctan\left(\beta \tan \frac{\pi\alpha}{2}\right)}{\alpha}$ and $S_{\alpha,\beta} = \left[1 + \beta^2 \tan^2 \frac{\pi\alpha}{2}\right]^{1/(2\alpha)}$, and compute

$$X = \begin{cases} S_{\alpha,\beta} \frac{\sin\left(\alpha(U + B_{\alpha,\beta})\right)}{(\cos U)^{1/\alpha}} \left[\frac{\cos(U - \alpha(U + B_{\alpha,\beta}))}{W}\right]^{\frac{1-\alpha}{\alpha}} & \text{if } \alpha \neq 1, \\ \frac{2}{\pi}\left[\left(\frac{\pi}{2} + \beta U\right)\tan U - \beta \log\left(\frac{\frac{\pi}{2} W \cos U}{\frac{\pi}{2} + \beta U}\right)\right] & \text{if } \alpha = 1. \end{cases} \tag{3.33}$$

Next, we can obtain a random variable $Y \sim S_\alpha(\sigma, \beta, \mu)$ by means of the standardization formula:

$$Y = \begin{cases} \sigma X + \mu & \text{if } \alpha \neq 1, \\ \sigma X + \frac{2}{\pi}\beta\sigma \log \sigma + \mu & \text{if } \alpha = 1. \end{cases} \tag{3.34}$$

3.3.2 Tempered stable distribution

Kawai and Masuda (2011) studied different approaches which are based on acceptance-rejection sampling, Gaussian approximation of a small jump component, and infinite shot noise series representations, and they concluded that the acceptance-rejection sampling proposed by Baeumer and Meerschaert (2010) is both most efficient and handiest in terms of computational issues.

The acceptance-rejection sampling method is carried out by the following steps. For $\alpha \in (0, 1)$, the centered and totally positively skewed (one-sided) tempered stable distribution X with parameter vector (α, λ, C) can be simulated exactly as follows:

1. Generate a random variable U uniformly distributed on $(0, 1)$.
2. Let $\sigma = \left(-C\Gamma(-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)\right)^{1/\alpha}$ and simulate $V \sim S_\alpha(\sigma, 1, 0)$.
3. If $U \leq e^{-\lambda V}$, exit with $V - \Gamma(1-\alpha)C\lambda^{\alpha-1}$. Otherwise, return to Step 1.

For $\alpha \in (1, 2)$, the above exact acceptance-rejection method cannot be applied due to the fact that the support of the one-sided tempered stable distribution is whole $\mathbb{R}$ instead of $\mathbb{R}_+$. The approximative acceptance-rejection sampling, proposed by Baeumer and Meerschaert (2010), consists of the follow steps:

1. Fix $c > 0$.
2. Generate a random variable U uniformly distributed on $(0, 1)$.
3. Let $\sigma = \left(-C\Gamma(-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)\right)^{1/\alpha}$ and simulate $V \sim S_\alpha(\sigma, 1, 0)$.
4. If $U \leq e^{-\lambda(V+c)}$, exit with $V - \Gamma(1-\alpha)C\lambda^{\alpha-1}$. Otherwise, return to Step 1.

For both cases, the shift term $-\Gamma(1-\alpha)C\lambda^{\alpha-1}$ is included in order to center the distribution. Note that the simulation for $\alpha \in (1, 2)$ is done by choosing $c > 0$ which acts as a truncation due to the support of the whole real line $\mathbb{R}$. Baeumer and Meerschaert (2010) presents the basic properties of this algorithm. The simulated distribution converges in $L^1(\mathbb{R})$ to its target density as $c \to \infty$. However, it is not viable to obtain a smaller distribution error by taking a large c, since the algorithm becomes extremely inefficient because of the low acceptance rate.

Finally, the bilateral tempered stable distribution $X \sim CTS(\alpha, \lambda_+, \lambda_-, C_+, C_-, 0)$ is simulated by implementing the algorithm at least twice; once for the positive component and the other for the negative, i.e. $X = X_+ - X_-$ where X_+ is the one-sided tempered stable distribution with parameter vector (α, λ_+, C_+) and X_- is the one with (α, λ_-, C_-).

Chapter 4

Indirect Inference

The advantage of using the class of α-stable distributions is their flexibility for asymmetry and heavy tails, and moreover, they are closed under linear combinations, which includes the Gaussian distribution as a particular case. However, its estimation is difficult since its density function does not have a closed-form and the moments of order greater than two do not exist. Therefore, the usual estimation methods such as the maximum likelihood and method of moments do not work.

Alternative estimation approaches such as methods based on quantiles (McCulloch, 1986) or on the empirical characteristic function (Koutrouvelis, 1981) are proposed. However, they are only useful for the estimation of the stable distributions parameters and are difficult to apply for more complex models.

Since stable distributions can be easily simulated, the indirect approaches proposed by Gourieroux *et al.* (1993) and Gallant and Tauchen (1996) could be the solution to more complex models involving stable distributions. The indirect inference was proposed by Gourieroux *et al.* (1993) in the context of econometric models with latent variables, but it has been proved to be useful in situations where the direct maximization of the likelihood function is not available. For instance, Lombardi and Calzolari (2008) employed this approach to estimate α-stable parameters and ARMA models with α-stable distributions with a constrained approach; and Sampaio and Morettin (2015, 2018) also used this method to estimate the parameters of randomized generalized autoregressive conditional heteroskedastic models.

4.1 The principle of the approach

In this section, the principle of the indirect inference is presented based on the fourth chapter of Gouriéroux and Monfort (1997). Suppose we have a sample of T observations $\boldsymbol{y}$ and a *model of interest* (IM) whose likelihood function $\mathcal{L}_T^*(\boldsymbol{y};\theta)$ is difficult to handle and maximize (the model could depend on a matrix of explanatory variable $\boldsymbol{X}$). Consequently, the maximum likelihood of $\theta \in \Theta$, given by

$$\hat{\theta}_T = \underset{\theta\in\Theta}{\text{argmax}}\, \mathcal{L}_T^*(\theta; \boldsymbol{y}) = \underset{\theta\in\Theta}{\text{argmax}} \sum_{t=1}^{T} \ln \ell^*(\theta; y_t), \tag{4.1}$$

is unavailable. In addition, consider an alternative model, depending on a parameter vector $\lambda \in \Lambda$, called *auxiliary model* (AM). Suppose that the likelihood function $\mathcal{L}_T(\boldsymbol{y}; \lambda)$ of the AM is easier to handle. However, its estimator

$$\hat{\lambda}_T = \underset{\lambda\in\Lambda}{\text{argmax}}\, \mathcal{L}_T(\lambda; \boldsymbol{y}) = \underset{\lambda\in\Lambda}{\text{argmax}} \sum_{t=1}^{T} \ln \ell_T(\lambda; y_t), \tag{4.2}$$

is not necessarily consistent since the model is misspecified. The idea is to perform simulations under the IM to correct the bias of the estimator $\hat{\lambda}$.

To continue, we describe the general procedure of the indirect inference.

Step 1 : Compute the maximum likelihood estimate of λ based on T observations $\boldsymbol{y}$, which will be denoted as $\hat{\lambda}_T$.

Step 2 : Simulate a set of S vectors of size T from the IM on the basis of an arbitrary parameter vector $\hat{\theta}^{(0)}$. Let us denote each of those vectors as $\tilde{\boldsymbol{y}}^s(\hat{\theta}^{(0)}) = \left\{\tilde{y}_1^s(\hat{\theta}^{(0)}), \cdots, \tilde{y}_T^s(\hat{\theta}^{(0)})\right\}$ for $s = 1, \cdots, S$.

Step 3 : Then, estimate parameters of the AM using simulated values from the IM

$$\tilde{\lambda}_T^s(\hat{\theta}^{(0)}) = \underset{\lambda\in\Lambda}{\text{argmax}}\, \mathcal{L}_T\left(\lambda; \tilde{\boldsymbol{y}}^s(\hat{\theta}^{(0)})\right) \tag{4.3}$$

for each $s = 1, \cdots, S$ and compute the mean

$$\tilde{\lambda}_{TS}(\hat{\theta}^{(0)}) = \frac{1}{S}\sum_{s=1}^{S} \tilde{\lambda}_T^s(\hat{\theta}^{(0)}). \tag{4.4}$$

Step 4 : Numerically update the initial guess $\hat{\theta}^{(0)}$ in order to minimize the distance

$$\left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right]' \Omega \left[\hat{\lambda}_T - \tilde{\lambda}_{TS}, (\theta)\right] \tag{4.5}$$

where Ω is a symmetric nonnegative matrix defining the metric.

As a result, the indirect inference estimator is defined by

$$\hat{\theta}_{TS} = \hat{\theta}_{TS}(\Omega) = \underset{\theta\in\Theta}{\text{argmin}} \left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right]' \Omega \left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right]. \tag{4.6}$$

Generally, the estimation step is performed with a numerical algorithm, such as Newton-Raphson. Then, for a given estimate $\hat{\theta}^{(p)}$, the procedure yields $\hat{\theta}^{(p+1)}$ and the process will be repeated until the series of $\hat{\theta}^{(p)}$ converges. The estimator is then given by

$$\hat{\theta} = \lim_{p\to\infty} \hat{\theta}^{(p)}. \tag{4.7}$$

Similarly, we can consider another indirect inference estimator. In Step 2, 3 and 4 above, we can change the procedures by computing

$$\check{\theta}_{TS} = \check{\theta}_{TS}(\Omega) = \underset{\theta\in\Theta}{\operatorname{argmin}} \left[\hat{\lambda}_T - \tilde{\tilde{\lambda}}_{TS}(\theta)\right]' \Omega \left[\hat{\lambda}_T - \tilde{\tilde{\lambda}}_{TS}(\theta)\right], \tag{4.8}$$

where

$$\tilde{\tilde{\lambda}}_{TS}(\theta) = \underset{\lambda\in\Lambda}{\operatorname{argmin}} \mathcal{L}_{TS}\left(\lambda, \tilde{\boldsymbol{y}}^{TS}(\theta)\right) \tag{4.9}$$

and $\tilde{\boldsymbol{y}}^{TS}(\theta)$ is the simulated path of length TS given θ. These two indirect inference estimators have the same asymptotic properties.

Finally, an alternative but similar method proposed by Gallant and Tauchen (1996) considers the score function of the AM:

$$\sum_{t=1}^{T} \frac{\partial \ell(\lambda; y_t)}{\partial \lambda}, \tag{4.10}$$

which is zero for the quasi-maximum likelihood estimator of λ. The idea is similar: we minimize the distance of score computed on the simulated observations

$$\min_{\theta} \left\{ \sum_{s=1}^{S} \sum_{t=1}^{T} \frac{\partial \ln \ell(\hat{\lambda}; y_t^s(\hat{\theta}))}{\partial \lambda} \right\}' \Sigma \left\{ \sum_{s=1}^{S} \sum_{t=1}^{T} \frac{\partial \ln \ell(\hat{\lambda}; y_t^s(\hat{\theta})}{\partial \lambda} \right\}. \tag{4.11}$$

where Σ is a symmetric nonnegative definite matrix. The same numerical algorithm can be applied to obtain the estimate, and it is given by

$$\check{\theta}_* = \lim_{p\to\infty} \check{\theta}_*^{(p)}. \tag{4.12}$$

This approach is useful when an analytic expression for the gradient of the AM is available, since in this case, the numerical optimization routine for the $\hat{\lambda}_S$ can be avoided.

Finally, it is important that the dimension of the AM parameter λ must be larger or equal to the dimension of θ in order to guarantee the uniqueness of the solution. When the dimension of the parameter vectors agrees, i.e. $\dim\lambda = \dim\theta$, these three types of indirect inference are consistent for any Ω and Σ.

Proposition 4.1. *If* $\dim\lambda = \dim\theta$ *and for* T *sufficiently large:*

(i) $\hat{\theta}(\Omega) = \hat{\theta}$ *and* $\check{\theta}(\Omega) = \check{\theta}$ *are independent of* Ω*;*

(ii) $\check{\theta}_*(\Sigma) = \check{\theta}_*$ *is independent of* Σ*;*

(iii) $\hat{\theta} = \check{\theta}_*$.

In this way, these three approaches are equivalent, and choosing which method to use will depend on the practical problem to be analyzed.

4.2 Asymptotic properties

In this section, we present the asymptotic properties of indirect inference estimators. This section is similar and it is based on Gourieroux *et al.* (1993) but we modify some conditions that exclude covariables X and we state the conditions which include the possibility of infill asymptotics.

Suppose that we have a sample of T observations $\boldsymbol{y}$ and a model of interest (IM) whose (negative) likelihood function[1] $\mathcal{L}_T^*(\boldsymbol{y};\theta)$ with $\theta \in \Theta$ is difficult to handle. Consider an auxiliary model (AM) with (negative) likelihood function $\mathcal{L}_T(\lambda;\boldsymbol{y})$ with $\lambda \in \Lambda$ and the maximum likelihood estimator is given by

$$\hat{\lambda}_T = \underset{\lambda\in\Lambda}{\operatorname{argmin}}\, \mathcal{L}_T(\lambda;\boldsymbol{y}). \tag{4.13}$$

Suppose that we have the following conditions:

(C1) The likelihood function $\mathcal{L}_T$ of the AM tends almost surely, as $T \to \infty$, to a non-stochastic limit, which depends on the unknown auxiliary parameter λ and the true parameter θ_0, that is

$$\lim_{T\to\infty} \mathcal{L}_T(\lambda;\boldsymbol{y}) = \mathcal{L}(\theta_0,\lambda). \tag{4.14}$$

(C2) The limit of the likelihood function is continuous with respect to λ and has an unique minimum

$$\lambda_0 = \underset{\lambda\in\Lambda}{\operatorname{argmin}}\, \mathcal{L}(\theta_0;\lambda). \tag{4.15}$$

(C3) Define the binding function

$$b(\theta) = \underset{\lambda\in\Lambda}{\operatorname{argmin}}\, \mathcal{L}(\theta;\lambda), \tag{4.16}$$

which is a one-to-one mapping from Θ onto Λ and its first derivative with respect to θ is of full column rank.

(C4) The negative of the Hessian matrix of the likelihood function of the AM converges to a non-stochastic limit J_0, that is

$$J_0 = \lim_{T\to\infty} -\frac{\partial^2}{\partial\lambda\partial\lambda^T}\mathcal{L}_T(\lambda_0;\boldsymbol{y}), \tag{4.17}$$

[1] The negative likelihood function is considered here because of the Definition in Section 2.2.2, but the optimization is the same since maximization of the likelihood function is equivalent to minimize the negative likelihood function.

where λ_0 is defined in (4.15).

(C5) The gradient of the likelihood function of the AM converges in distribution to a Gaussian law. Let I_0 be the asymptotic variance-covariance matrix

$$I_0 = \lim_{T\to\infty} Var\left[\sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\boldsymbol{y})\right].$$

(C6) The asymptotic covariance between the gradients of the likelihood function of the AM at two units s_1 and s_2 from the simulated sample is constant, that is, for $s_1 \neq s_2$

$$\lim_{T\to\infty} Cov\left\{\sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\tilde{\boldsymbol{y}}^{s_1}(\theta_0)), \sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\tilde{\boldsymbol{y}}^{s_2}(\theta_0))\right\} = K. \tag{4.18}$$

In our case, we have $K = 0$ since we are not dealing with exogenous variable.

Given θ, simulate a set of S vectors of size T which will be denoted by $\tilde{\boldsymbol{y}}^s(\theta) = \{\tilde{y}_1^s(\theta), \cdots, \tilde{y}_T^s(\theta)\}$ for $s = 1, \cdots, S$. Next, we can compute

$$\tilde{\lambda}_T^s(\theta) = \underset{\lambda\in\Lambda}{\operatorname{argmin}}\, \mathcal{L}_T\left(\lambda;\tilde{y}^s(\theta)\right), \tag{4.19}$$

for each $s = 1, \cdots, S$ and

$$\tilde{\lambda}_{TS}(\theta) = \frac{1}{S}\sum_{s=1}^{S}\tilde{\lambda}_T^s(\theta). \tag{4.20}$$

Under usual regularity conditions, we have

$$\lim_{T\to\infty}\tilde{\lambda}_{TS}(\theta) = b(\theta), \text{ and } \quad \lim_{T\to\infty}\tilde{\lambda}_T = b(\theta_0) = \lambda_0,$$

where $\tilde{\lambda}_T$ is defined in (4.13).

Then, the indirect inference estimator is defined by

$$\hat{\theta}_{TS} = \hat{\theta}_{TS}(\Omega) = \underset{\theta\in\Theta}{\operatorname{argmin}}\left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right]'\Omega\left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right], \tag{4.21}$$

where Ω is a symmetric nonnegative matrix defining the metric.

Proposition 4.2. *Under the conditions (C1), (C2) and (C3), the indirect inference estimator $\hat{\theta}_{TS}$ is consistent for fixed S and $T \to \infty$.*

Proposition 4.3. *Under the conditions (C1)-(C6), the indirect inference estimator $\hat{\theta}_{TS}$ is asymptotically normal for fixed S and $T \to \infty$, that is*

$$\sqrt{T}(\hat{\theta}_{TS} - \theta) \xrightarrow{d} \mathcal{N}\left[0, W(S,\Omega)\right], \tag{4.22}$$

where the asymptotic variance-covariance matrix is

$$
\begin{aligned}
W(S,\Omega) = \left(1+\frac{1}{S}\right)\left\{\frac{\partial b'(\theta_0)}{\partial \theta}\Omega\frac{\partial b(\theta_0)}{\partial \theta'}\right\}^{-1}\left[\frac{\partial b'(\theta_0)}{\partial \theta}\right]\Omega \\
\times J_0^{-1}\left[I_0 - K\right]J_0^{-1}\Omega\left[\frac{\partial b(\theta_0)}{\partial \theta'}\right]\left\{\frac{\partial b'(\theta_0)}{\partial \theta}\Omega\frac{\partial b(\theta_0)}{\partial \theta'}\right\}^{-1}.
\end{aligned}
\tag{4.23}
$$

Proof. The consistency of the estimator follows by the limit of the optimization problem:

$$
\min_{\theta\in\Theta}\left[b(\theta_0)-b(\theta)\right]'\Omega\left[b(\theta_0)-b(\theta)\right]. \tag{4.24}
$$

To continue with the distribution of $\hat{\theta}_{TS}$, we consider the first derivative of $D_{TS}(\theta) = \left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right]'\Omega\left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right]$, which is

$$
\frac{\partial D_{TS}(\theta)}{\partial \theta} = 2\left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta)\right]'\Omega\left[-\frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^S(\theta)}{\partial \theta}\right]. \tag{4.25}
$$

Then, the first-order condition is

$$
\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^S(\hat{\theta}_{TS})'}{\partial \theta}\right]\Omega\left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\hat{\theta}_{TS})\right] = 0. \tag{4.26}
$$

Note that the first-order expansion of $\tilde{\lambda}_{TS}(\tilde{\theta})$ around θ_0 is

$$
\tilde{\lambda}_{TS}(\tilde{\theta}) = \frac{1}{S}\sum_{s=1}^{S}\tilde{\lambda}_T^s(\tilde{\theta}) \approx \frac{1}{S}\sum_{s=1}^{S}\tilde{\lambda}_T^s(\theta_0) + \frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^s(\theta_0)}{\partial \theta'}\left(\tilde{\theta}-\theta_0\right).
$$

Plugging it into (4.26), we obtain

$$
\begin{aligned}
&\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^S(\hat{\theta}_{TS})'}{\partial \theta}\right]\Omega\left\{\hat{\lambda}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\lambda}_T^s(\theta_0) - \frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^s(\theta_0)}{\partial \theta'}\left(\hat{\theta}_{TS}-\theta_0\right)\right\} \approx 0 \\
&\Rightarrow \left\{\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^S(\hat{\theta}_{TS})'}{\partial \theta}\right]\Omega\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^s(\theta_0)}{\partial \theta'}\right]\right\}\left(\hat{\theta}_{TS}-\theta_0\right) \approx \\
&\qquad \left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial \tilde{\lambda}_T^s(\hat{\theta}_{TS})}{\partial \theta}\right]\Omega\left[\hat{\lambda}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\lambda}_T^s(\theta_0)\right].
\end{aligned}
$$

For sufficiently large T,

$$
\sqrt{T}\left(\hat{\theta}_{TS}-\theta_0\right) \approx \left\{\left[\frac{\partial b'(\theta_0)}{\partial \theta}\right]\Omega\left[\frac{\partial b(\theta_0)}{\partial \theta'}\right]\right\}^{-1}\left[\frac{\partial b'(\theta_0))}{\partial \theta}\right]\Omega\sqrt{T}\left(\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta_0)\right). \tag{4.27}
$$

The asymptotic distribution of $\sqrt{T}\left(\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta_0)\right)$ is given as follows.

Since $\frac{\partial}{\partial\lambda}\mathcal{L}_T(\hat{\lambda}_T;\boldsymbol{y}) = 0$, expanding it around λ_0 we have

$$\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\boldsymbol{y}) + \frac{\partial^2}{\partial\lambda\partial\lambda^T}\mathcal{L}_T(\lambda_0;\boldsymbol{y})\left(\hat{\lambda}_T-\lambda_0\right) \approx 0$$

$$\Rightarrow \sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\boldsymbol{y}) + \frac{\partial^2}{\partial\lambda\partial\lambda^T}\mathcal{L}_T(\lambda_0;\boldsymbol{y})\sqrt{T}\left(\hat{\lambda}_T-\lambda_0\right) \approx 0$$

$$\begin{aligned}\Rightarrow \sqrt{T}\left(\hat{\lambda}_T-\lambda_0\right) &\approx \left[-\frac{\partial^2}{\partial\lambda\partial\lambda^T}\mathcal{L}_T(\lambda_0;\boldsymbol{y})\right]^{-1}\left[\sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\boldsymbol{y})\right]\\ &\approx J_0^{-1}\left[\sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\boldsymbol{y})\right].\end{aligned}$$

Similarly, for $\tilde{\lambda}_T^s(\theta_0)$ we have

$$\sqrt{T}\left(\hat{\lambda}_T^s(\theta_0)-\lambda_0\right) \approx J_0^{-1}\left[\sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\tilde{\boldsymbol{y}}^s(\theta_0))\right].$$

Then, we have

$$\begin{aligned}\sqrt{T}\left[\hat{\lambda}_T-\tilde{\lambda}_{TS}(\theta_0)\right] &= \sqrt{T}\left[\left(\hat{\lambda}_T-\lambda_0\right)-\left(\tilde{\lambda}_{TS}(\theta_0)-\lambda_0\right)\right]\\ &= \sqrt{T}\left(\hat{\lambda}_T-\lambda_0\right)-\sqrt{T}\left[\frac{1}{S}\sum_{s=1}^{S}\left(\tilde{\lambda}_T^s(\theta_0)-\lambda_0\right)\right]\\ &\approx J_0^{-1}\left[\sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\boldsymbol{y})-\sqrt{T}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial}{\partial\lambda}\mathcal{L}_T\left(\lambda_0;\tilde{\boldsymbol{y}}^s(\theta_0)\right)\right].\end{aligned}$$

Define $\Delta_T := \sqrt{T}\frac{\partial}{\partial\lambda}\mathcal{L}_T(\lambda_0;\boldsymbol{y})-\sqrt{T}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial}{\partial\lambda}\mathcal{L}_T\left(\lambda_0;\tilde{\boldsymbol{y}}^s(\theta_0)\right)$.

Under the conditions (C1)-(C6), Δ_T is asymptotically normal with zero mean and the asymptotic variance-covariance matrix given by

$$\begin{aligned}W &= \lim_{T\to\infty} Var(\Delta_T)\\ &= I_0+\frac{1}{S}I_0-2K+2\frac{S(S-1)}{2S^2}K\\ &= \left(1+\frac{1}{S}\right)I_0-2K+\frac{(S-1)}{S}K\\ &= \left(1+\frac{1}{S}\right)I_0-K\left(1+\frac{1}{S}\right)\\ &= \left(1+\frac{1}{S}\right)[I_0-K].\end{aligned}$$

Next, the asymptotic variance-covariance matrix of $\sqrt{T}\left[\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta_0)\right]$ is given by

$$\begin{aligned} Var\left[\sqrt{T}\left(\hat{\lambda}_T - \tilde{\lambda}_{TS}(\theta_0)\right)\right] &= J_0^{-1}\, Var\left[\Delta_T\right] J_0^{-1} \\ &= \left(1+\frac{1}{S}\right) J_0^{-1}\;\left[I_0 - K\right] J_0^{-1}. \end{aligned}$$

Finally, we can conclude from (4.27) that the asymptotic variance-covariance matrix of $\sqrt{T}\left(\hat{\theta}_{TS} - \theta_0\right)$ is

$$\begin{aligned} Var\left[\sqrt{T}\left(\hat{\theta}_{TS} - \theta_0\right)\right] = &\left(1+\frac{1}{S}\right)\left\{\frac{\partial b'(\theta_0)}{\partial\theta}\Omega\frac{\partial b(\theta_0)}{\partial\theta'}\right\}^{-1}\left[\frac{\partial b'(\theta_0)}{\partial\theta}\right]\Omega \\ &\times J_0^{-1}\left[I_0 - K\right] J_0^{-1}\Omega\left[\frac{\partial b(\theta_0)}{\partial\theta'}\right]\left\{\frac{\partial b'(\theta_0)}{\partial\theta}\Omega\frac{\partial b(\theta_0)}{\partial\theta'}\right\}^{-1}. \end{aligned}$$

□

Similarly, the alternative indirect inference estimator $\check{\theta}_{TS}$ has the same asymptotic properties as in Proposition 4.3.

Proof. Similarly to the previous proof, we have

$$\sqrt{T}\left(\check{\theta}_{TS} - \theta_0\right) \approx \left\{\left[\frac{\partial\tilde{\tilde{\lambda}}_{TS}(\check{\theta}_{TS})}{\partial\theta}\right]\Omega\left[\frac{\partial\tilde{\tilde{\lambda}}_{TS}(\theta_0)}{\partial\theta'}\right]\right\}^{-1}\left[\frac{\partial\tilde{\tilde{\lambda}}_{TS}(\check{\theta}_{TS}))}{\partial\theta}\right]\Omega\sqrt{T}\left(\hat{\lambda}_T - \tilde{\tilde{\lambda}}_{TS}(\theta_0)\right). \tag{4.28}$$

Following the same direction, the same asymptotic properties are obtained since

$$\sqrt{T}\left(\tilde{\tilde{\lambda}}_{TS}\left(\theta_0\right) - \lambda_0\right) \approx J_0^{-1}\sum_{s=1}^{S}\frac{\sqrt{T}}{S}\frac{\partial}{\partial\lambda}\mathcal{L}_T\left(\lambda_0; \tilde{\boldsymbol{y}}^{TS}\left(\theta_0\right)\right).$$

□

4.3 Constrained indirect estimation

One condition in the indirect estimation proposed by Gourieroux *et al.* (1993) is that the parameters of the auxiliary model are unrestricted. Therefore, the asymptotic distribution of the pseudo-maximum likelihood estimator is normal with a full rank covariance matrix under standard regularity conditions. This assumption is not realistic, and in most situations, it is necessary to add some restrictions to the parameters. Calzolari *et al.* (2004) generalized the indirect inference to include the possibility of handling equality or inequality restriction on the parameter $\lambda \in \Lambda$ of the auxiliary model. In this case, the maximum likelihood estimator of the auxiliary model under the constraints is optimizing the Lagrange function

$$\mathcal{Q}(\lambda) = \tilde{\mathcal{L}}(\lambda, \boldsymbol{y}) + \lambda h'(\lambda), \tag{4.29}$$

where $h'(\lambda)$ is a vector of functions summarizing the constraints and λ is a vector of Lagrange multipliers. Then, the binding function can be obtained by a constrained maximization of the likelihood function of the auxiliary model. Moreover, Calzolari *et al.* (2004) showed that the asymptotic normal distribution can be obtained by appropriate changes in three conditions on the unconstrained case.

Chapter 5

tvARMA process with $\alpha-$stable innovations

In this chapter, we present theoretical results of a tvARMA process with α-stable innovations. Recalling the tvARMA process from (2.18) and assuming that the innovations are α-stable, the system of difference equations is defined by

$$\sum_{j=0}^{p} \alpha_j \left(\frac{t}{T}\right) X_{t-j,T} = \sum_{k=0}^{q} \beta_k \left(\frac{t}{T}\right) \gamma \left(\frac{t-k}{T}\right) \varepsilon_{t-k}, \tag{5.1}$$

where ε_t are i.i.d. and $\varepsilon_t \sim S_\alpha(1/\sqrt{2}, \beta, 0)$ with $\alpha \in (0, 2)$. Assume $\alpha_0(u) \equiv \beta_0(u) \equiv 1$ and $\alpha_j(u) = \alpha_j(0)$, $\beta_k(u) = \beta_k(0)$ for $u < 0$. Suppose also that all $\alpha_j(\cdot)$ and $\beta_k(\cdot)$, as well as $\gamma^2(\cdot)$[1], are of bounded variation. The reason that the scale parameter of the innovations is set to be $\sigma = 1/\sqrt{2}$ is when $\alpha = 2$, the standardized Gaussian innovation is obtained.

It is possible to define the equation (5.1) as:

$$\Phi_{t,T}(B) X_{t,T} = \Theta_{t,T}(B) z_{t,T}, \tag{5.2}$$

where $z_{t,T} = \gamma(\frac{t}{T})\varepsilon_t$; $\Phi_{t,T}(B) = 1 + \alpha_1(\frac{t}{T})B + \cdots + \alpha_p(\frac{t}{T})B^p$ and $\Theta_{t,T}(B) = 1 + \beta_1(\frac{t}{T})B + \cdots + \beta_q(\frac{t}{T})B^q$ are the autoregressive (AR) and moving average (MA) operators, respectively.

There are several works related to stable linear processes. For instance, chapter 7 of Embrechts *et al.* (1997) and Chapter 13 of Brockwell and Davis (1991) give a general review of stable linear processes. Kokoszka and Taqqu (1994) study the infinite variance stable ARMA processes and Kokoszka and Taqqu (1995) study fractional ARIMA with stable innovations. Mikosch *et al.* (1995) proposed a Whittle-type estimator to estimate the coefficients of the ARMA model. In the stable innovation and time-varying coefficient context, Shelton Peiris and Thavaneswaran (2001a,b) considered the univariate and multivariate case of the system (5.2) with symmetric stable innovations and assume $\gamma(\cdot) = 1$. However, they considered time-dependent coefficient without the local stationarity condition.

[1] We use $\gamma(\cdot)$ instead of $\sigma(\cdot)$ to avoid confusion with σ, the scale parameter of α-stable innovations.

5.1 Existence and Uniqueness of a Solution

Before we study the local stationarity conditions on the time-varying coefficients, we present a set of regularity conditions of existence and uniqueness of solution of the system based on the concepts defined by Shelton Peiris and Thavaneswaran (2001a,b).

Definition 5.1.

- *The process* (5.2) *is AR regular (or causal) if there exist* $a_{t,T}(j)$ *such that*

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j)\varepsilon_{t-j}. \tag{5.3}$$

satisfying $\sum_{j=0}^{\infty} |a_{t,T}(j)|^{\delta} < \infty$ *for all* t *and* $\delta = min\{1, \alpha\}$.

- *The process* (5.2) *is MA regular (or invertible) if there exist* $b_{t,T}(j)$ *such that*

$$\varepsilon_t = \sum_{j=0}^{\infty} b_{t,T}(j)X_{t-j,T}. \tag{5.4}$$

satisfying $\sum_{j=0}^{\infty} |b_{t,T}(j)|^{\delta} < \infty$ *for all* t *and* $\delta = min\{1, \alpha\}$.

We will show that the random series in (5.3) converges a.s. if and only if $\sum_{j=0}^{\infty} |a_{t,T}(j)|^{\alpha} < \infty$, and by applying the Proposition 13.3.1 in Brockwell and Davis (1991), it converges absolutely if and only if $\sum_{j=0}^{\infty} |a_{t,T}(j)|^{\delta} < \infty$ with $\delta = \min\{1, \alpha\}$. Similar arguments are applied to (5.4).

Proposition 5.1. *The random series in* (5.3) *converges a.s. if and only if*

$$\sum_{j=0}^{\infty} |a_{t,T}(j)|^{\alpha} < \infty.$$

Proof. Suppose that $\sum_{j=0}^{\infty} |a_{t,T}(j)|^{\alpha} < \infty$. Using the Propositions 3.3 and 3.8, we have

$$a_{t,T}(j)\varepsilon_{t-j} \sim S_{\alpha}\left(\frac{1}{\sqrt{2}}|a_{t,T}(j)|, \text{sign}\left[a_{t,T}(j)\right])\beta, 0\right), \tag{5.5}$$

and for $0 < p < \alpha$,

$$E|a_{t,T}(j)\varepsilon_{t-j}|^p = c_{\alpha,\beta}(p)^p \left(\frac{1}{\sqrt{2}}\right)^p |a_{t,T}(j)|^p,$$

where $c_{\alpha,\beta}(p)$ is a constant. Next,

$$\sum_{j=0}^{\infty} E|a_{t,T}(j)\varepsilon_{t-j}|^p = c_{\alpha,\beta}(p)^p \left(\frac{1}{\sqrt{2}}\right)^p \sum_{j=0}^{\infty} |a_{t,T}(j)|^p < \infty.$$

Since $\sum_{j=0}^{\infty} E|a_{t,T}(j)\varepsilon_{t-j}|^p < \infty$, it follows that (5.3) converges a.s. (see Chow and Teicher , 2003, Corollary 3 on p. 117).

Conversely, the random series in (5.3) converges a.s. implies that $a_{t,T}(j)\varepsilon_{t-j} \xrightarrow{a.s.} 0$. Then, since $\{a_{t,T}(j)\varepsilon_{t-j}\}$ is an independent sequence, it is straightforward that from the Borel-Cantelli lemma there exists $K_1 > 0$ such that

$$\sum_{j=0}^{\infty} P\left(|a_{t,T}(j)\varepsilon_{t-j}| > K_1\right) < \infty.$$

From the Proposition 3.6, we obtain

$$\lim_{\lambda\to\infty} \lambda^\alpha \sum_{j=1}^{\infty} P\left(|a_{t,T}(j)\varepsilon_{t-j}| > \lambda\right) = C_\alpha \left(\frac{1}{\sqrt{2}}\right) \sum_{j=1}^{\infty} |a_{t,T}(j)|^\alpha.$$

In other words, for all $\varepsilon > 0$ there exists K_2 such that for all $\lambda > K_2$,

$$\frac{C_\alpha\left(\frac{1}{\sqrt{2}}\right)\sum_{j=1}^{\infty}|a_{t,T}(j)|^\alpha - \varepsilon}{\lambda^\alpha} < \sum_{j=1}^{\infty} P\left(|a_{t,T}(j)\varepsilon_{t-j}| > \lambda\right) < \frac{C_\alpha\left(\frac{1}{\sqrt{2}}\right)\sum_{j=1}^{\infty}|a_{t,T}(j)|^\alpha + \varepsilon}{\lambda^\alpha}.$$

Let $K = \max(K_1, K_2)$ and we have

$$\frac{C_\alpha\left(\frac{1}{\sqrt{2}}\right)\sum_{j=1}^{\infty}|a_{t,T}(j)|^\alpha - \varepsilon}{K^\alpha} < \sum_{j=1}^{\infty} P\left(|a_{t,T}(j)\varepsilon_{t-j}| > K\right) < \infty.$$

Then, we conclude that $\sum_{j=1}^{\infty} |a_{t,T}(j)|^\alpha < \infty$. □

To continue, we omit the subscript T from above notation. Consider the homogeneous difference equation

$$\Phi_t(B)u_t = 0. \tag{5.6}$$

If $\alpha_p(\frac{t}{T}) \neq 0$ for any t, there exist p linearly independent solution $\psi_{1,t}, \psi_{2,t}, ..., \psi_{p,t}$ such that

$$\Psi(t) = \begin{bmatrix} \psi_{1,t} & \cdots & \cdots & \psi_{p,t} \\ \psi_{1,t-1} & \ddots & & \psi_{p,t-1} \\ \vdots & & \ddots & \vdots \\ \psi_{1,t-p+1} & \cdots & \cdots & \psi_{p,t-p+1} \end{bmatrix} \tag{5.7}$$

is invertible for any t (see Miller, 1968). Therefore, we can define

$$G(t,s) = \Psi(t)\,[\Psi(s)]^{-1}\,, \tag{5.8}$$

the one-sided Green's function matrix associated with the AR operator $\Phi_t(B)$. It can be showed that $G(t,s)$ is unique and invariant under different solutions $\Psi(t)$ obtained from the homogeneous difference equation (5.6). Furthermore, the one-sided Green's function associated with the AR operator $\Phi_t(B)$ is defined as the upper left-hand element in the matrix (5.8),

$$g(t,s) = [G(t,s)]_{11}\,, \tag{5.9}$$

which is also unique and invariant. Now, we are ready to establish the conditions for AR regularity and MA regularity.

Theorem 5.1. *Let $\{X_{t,T}\}$ be a sequence of stochastic process that satisfies (5.2). Suppose that $\alpha_p(\frac{t}{T}) \neq 0$ for all t, and $g(t,s)$, the one-sided Green's functions associated with $\Phi_t(B)$, is such that $\sum_{s=-\infty}^{t} |g(t,s)|^{\delta} < \infty$, for all t. Assume also that $\sum_{s=-0}^{q} |\beta_j(\cdot)|^2 < \infty$ for all t, and $\Phi_t(z)$ ($\Phi_t(z) \neq 0$ for $|z| \leq 1$) and $\Theta_t(z)$ have no common roots. Then, there is a valid solution, given by*

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j)\varepsilon_{t-j}, \tag{5.10}$$

to (5.2) with coefficients uniquely determined by

$$a_{t,T}(j) = \begin{cases} 0, & j<0,\\ \gamma(\frac{t-j}{T}), & j=0,\\ \gamma(\frac{t-j}{T}) \sum_{j=0}^{k} \beta_k(\frac{t-j+k}{T}) g(t,t-j+k), & 0 \leq j \leq q,\\ \gamma(\frac{t-j}{T}) \sum_{j=0}^{q} \beta_k(\frac{t-j+k}{T}) g(t,t-j+k), & j>q. \end{cases} \tag{5.11}$$

Proof. By setting $z_{t,T} = \gamma(\frac{t}{T})\varepsilon_t$, along with the absolute convergence conditions above, the proof is similar to Shelton Peiris and Thavaneswaran (2001b). □

Theorem 5.2. *Let $\{X_{t,T}\}$ be a sequence of stochastic process that satisfies (5.2). Suppose that $\beta_q(\frac{t}{T}) \neq 0$ for all t, and $h(t,s)$, the one-sided Green's function associated with $\Theta_t(B)$, is such that $\sum_{s=-\infty}^{t} |h(t,s)|^{\delta} < \infty$, for all t. Assume also that $\sum_{s=-0}^{p} |\alpha_j(\cdot)|^2 < \infty$ for all t, and $\Phi_t(z)$ and $\Theta_t(z)$ ($\Theta_t(z) \neq 0$ for $|z| \leq 1$) have no common roots. Then, the process (5.2) is invertible and its explicit inversion is given by*

$$\varepsilon_t = \sum_{j=0}^{\infty} b_{t,T}(j) X_{t-j,T}. \tag{5.12}$$

where $X_{t,T}$ denotes an arbitrary solution and the coefficients are uniquely determined by

$$b_{t,T}(j) = \begin{cases} 0, & j < 0, \\ \frac{1}{\gamma(\frac{t}{T})}, & j = 0, \\ \frac{1}{\gamma(\frac{t}{T})} \sum_{l=0}^{k} \alpha_k(\frac{t-j+k}{T}) h(t, t-j+k), & 0 \leq j \leq p, \\ \frac{1}{\gamma(\frac{t}{T})} \sum_{l=0}^{q} \alpha_k(\frac{t-j+k}{T}) h(t, t-j+k), & j > p. \end{cases} \tag{5.13}$$

Theorem 5.3. *Let $\{X_{t,T}\}$ be a sequence of stochastic process that satisfies (5.1) that is AR regular. The solution $X_{t,T}$ of the form (5.3) is strictly stable and $X_{t,T} \sim S_\alpha(\sigma^*, \beta^*, 0)$, with*

$$\sigma^* = \left(\frac{1}{\sqrt{2}}\right) \left\{ \sum_{j=0}^{\infty} |a_{t,T}(j)|^\alpha \right\}^{1/\alpha}, \text{ and } \beta^* = \beta \left\{ \frac{\sum_{j=0}^{\infty} \text{sign}\,[a_{t,T}(j)]\, |a_{t,T}(j)|^\alpha}{\sum_{j=0}^{\infty} |a_{t,T}(j)|^\alpha} \right\}.$$

Proof. The explicit form of the solution is straightforward using Propositions 3.1 and 3.3. Moreover, the Proposition 3.5 implies that for each t, the solution $X_{t,T}$ is strictly stable since each of them has location parameter equals to 0. □

5.2 Local Stationarity

Similar to the Proposition 2.1, we can present the corresponding version for stable innovations. Since it is not a second-order process, the time-varying spectral density does not exist.

Theorem 5.4. *Consider the system of difference equations in (5.1) satisfying the AR regular conditions stated above. Suppose that all $\alpha_j(\cdot)$ and $\beta_k(\cdot)$, as well as $\gamma^2(\cdot)$ are of bounded variation. Then, there exists a solution of the form*

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j) \varepsilon_{t-j},$$

which fullfills (2.6), (2.7) and (2.8) of Assumption 2.1.

Proof. We give the proof for tvAR(p) process (i.e. $q = 0$) and then the extension to tvARMA(p,q) is straightforward. Since the process (5.1) is AR regular, there exists a solution of the form

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j) \varepsilon_{t-j},$$

that is well defined and the coefficients are given by

$$a_{t,T}(j) = \left[\prod_{\ell=0}^{j-1} \boldsymbol{\alpha} \left(\frac{t-\ell}{T} \right) \right]_{11} \gamma \left(\frac{t-j}{T} \right)$$

with

$$\boldsymbol{\alpha}(u) = \begin{pmatrix} -\alpha_1(u) & -\alpha_2(u) & \cdots & \cdots & -\alpha_p(u) \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

(for more detail see Miller, 1968). Then, to prove the existence of the functions $\alpha(\cdot, j)$ satisfying (2.6), (2.7) and (2.8) of Assumption 2.1 follows the same proof to that of finite innovation case (see Appendix in Dahlhaus and Polonik, 2009). □

Remark 5.1.

(i) Note that since $a_{t,T}(j) \approx a\left(\frac{t}{T}, j\right)$, $X_{t,T}$ can be approximated by

$$\tilde{X}_{t,T} = \sum_{j=0}^{\infty} a\left(\frac{t}{T}, j\right) \varepsilon_{t-j}, \tag{5.14}$$

which converges a.s. if and only if $\sum_{j=0}^{\infty} \left| a\left(\frac{t}{T}, j\right)\right|^{\alpha} < \infty$.

Moreover, $\tilde{X}_{t,T} \sim S_\alpha(\sigma^+, \beta^+, 0)$, with

$$\sigma^+ = \frac{1}{\sqrt{2}} \left\{ \sum_{j=0}^{\infty} \left| a\left(\frac{t}{T}, j\right)\right|^{\alpha} \right\}^{1/\alpha}, \text{ and } \ \beta^+ = \beta \left\{ \frac{\sum_{j=0}^{\infty} \operatorname{sign}\left[a\left(\frac{t}{T}, j\right)\right] \left|a\left(\frac{t}{T}, j\right)\right|^{\alpha}}{\sum_{j=0}^{\infty} \left|a\left(\frac{t}{T}, j\right)\right|^{\alpha}} \right\}.$$

(ii) Since $X_{t,T}$ in (5.3) can be expressed as a linear combination of α-stable random variables, $X_{t,T}$ is strictly stable with the same index of stability α.

(iii) Observe that $X_{t,T}$ is not strictly stationary, but it can be approximated by $\tilde{X}_{t,T}$ which is locally (strictly) stationary and strictly stable with the same index of stability.

(iv) Weak stationarity does not make sense since the second moment does not exist. Consequently, (2.19) does not exist.

(v) Let $X_{1,T}, ..., X_{T,T}$ be the sequence of solutions defined in (5.3) and $\tilde{X}_{1,T}, ..., \tilde{X}_{T,T}$ be the sequence of the stochastic process defined in (5.14). Both processes are strictly $\alpha-$stable, since all linear combinations are strictly stable with the same index of stability. This means that the weak stationarity is lost but it is substituted by the same tail

behavior throughout the time. This is the reason we call this process $\alpha-$stable locally (strictly) stationary process.

5.3 tvARMA with symmetric $\alpha-$stable innovations

In this section, we will consider the special case with symmetric $\alpha-$stable ($S\alpha S$) innovations.

Corollary 5.1. *Consider the system of difference equations in (5.1) with i.i.d. $S\alpha S$ innovations, that is, $\varepsilon_t \sim S_\alpha(1/\sqrt{2}, 0, 0)$ with $\alpha \in (0,2)$, satisfying AR regular conditions stated above. Suppose that all $\alpha_j(\cdot)$ and $\beta_k(\cdot)$, as well as $\gamma^2(\cdot)$ are of bounded variation. Then, there exists a solution of the form*

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j)\varepsilon_{t-j},$$

which fullfills (2.6), (2.7) and (2.8) of Assumption 2.1. Moreover, this solution $X_{t,T}$ is $S\alpha S$ and $X_{t,T} \sim S_\alpha(\sigma^*, 0, 0)$, with

$$\sigma^* = \frac{1}{\sqrt{2}} \left\{ \sum_{j=0}^{\infty} |a_{t,T}(j)|^\alpha \right\}^{1/\alpha}.$$

Similarly to the general case, $X_{t,T}$ can be approximated by $\tilde{X}_{t,T} \sim S_\alpha(\sigma^+, 0, 0)$ as in (5.14), with

$$\sigma^+ = \left\{ \sum_{j=0}^{\infty} \left| a\left(\frac{t}{T}, j\right) \right|^\alpha \right\}^{1/\alpha}.$$

Remark 5.2.

(i) Note that in the case of $S\alpha S$ innovations, since $X_{t,T}$ in (5.3) can be expressed as a linear combination of $S\alpha S$ random variables, $X_{t,T}$ is $S\alpha S$ with the same index of stability α.

(ii) At the same time, $X_{t,T}$ is not strictly stationary, but it can be approximated by $\tilde{X}_{t,T}$ which is locally (strictly) stationary and $S\alpha S$ with the same index of stability.

(iii) Weak stationarity does not make sense since the second moment does not exist. Consequently, (2.19) does not exist.

(iv) In the same way, consider $X_{1,T}, ..., X_{T,T}$ be the sequence of solutions defined in (5.3) and $\tilde{X}_{1,T}, ..., \tilde{X}_{T,T}$ be the sequence of the stochastic process defined in (5.14). Both processes are symmetric $\alpha-$stable, since all linear combinations are symmetric stable with the same index of stability. This means that the weak stationarity is lost but it is substituted by the same tail behavior throughout the time.

5.4 Some examples

Example 5.1. The tvMA(q) model with stable innovations:

$$X_{t,T} = \sum_{k=0}^{q} \beta_k \left(\frac{t}{T}\right) \gamma \left(\frac{t-k}{T}\right) \varepsilon_{t-k}. \tag{5.15}$$

Example 5.2. Consider the tvAR(p) model with stable innovations

$$\sum_{j=0}^{p} \alpha_j \left(\frac{t}{T}\right) X_{t-j,T} = \gamma \left(\frac{t}{T}\right) \varepsilon_t. \tag{5.16}$$

Under the regularity conditions, $X_{t,T}$ does not have a solution of the form

$$X_{t,T} = \sum_{k=0}^{\infty} a_k \left(\frac{t}{T}\right) \varepsilon_{t-k},$$

but only of the form (5.3) with

$$a_{t,T}(j) = \left[\prod_{\ell=0}^{j-1} \boldsymbol{\alpha} \left(\frac{t-\ell}{T}\right)\right]_{11} \gamma \left(\frac{t-j}{T}\right) \tag{5.17}$$

with

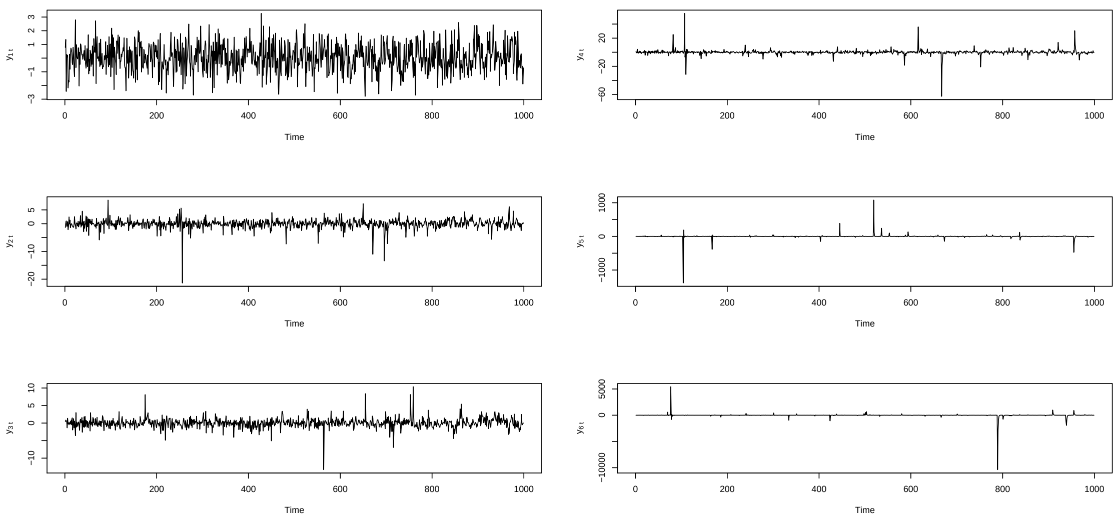
$$\boldsymbol{\alpha}(u) = \begin{pmatrix} -\alpha_1(u) & -\alpha_2(u) & \cdots & \cdots & -\alpha_p(u) \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

and $\boldsymbol{\alpha}(u) = \boldsymbol{\alpha}(0)$ for $u < 0$ (see Appendix in Dahlhaus and Polonik, 2009). Moreover, $a_{t,T}(j)$ can be approximated by $a(u, j) = (\boldsymbol{\alpha}(u)^j)_{11}\, \gamma(u)$ which satisfies Assumption 2.1, (ii).

Figure 5.1 presents simulated tvAR(1) process of $T = 1000$ observations with different innovation distribution (Gaussian, $t_{(3)}$ and symmetric stable innovations, $\alpha = 1.8, 1.6, 1.4$) and a linear coefficient $\alpha_1(u) = -0.2 + 0.6u$ and $\gamma(u) = 1$. We observe that for smaller α, the process seems to have more outliers.

5.5 Indirect inference for α-stable locally stationary processes

As we presented in Chapter 3, models involving α-stable random variable are difficult to estimate because its likelihood function is not available. However, the indirect inference can be employed due to the simulation easiness. As mentioned in Chapter 4, several works that



(a) Gaussian innvation (top), t innvation (middle) and $\alpha = 1.7$ (bottom).

(b) $\alpha = 1.4$ (top), $\alpha = 0.9$ (middle) and $\alpha = 0.6$ (bottom)

Figure 5.1: *Simulated tvAR(1) with Gaussian,* $t_{(3)}$ *and symmetric stable innovations* ($\alpha = 1.7, 1.4, 0.9, 0.6$)*, and time-varying coefficient* $\alpha_1(u) = -0.2 + 0.6u$ *and* $\gamma(u) = 1$.

include stationary processes with α-stable random variables have been proved useful, since their simulation is straightforward.

In the case of locally stationary processes, Chapter 2 presented the possibility of employing infill asymptotic approach for the auxiliary model to achieve meaningful asymptotic theory such as consistency and normality. In our case, the IM will be the α-stable locally stationary process and the AM will be the corresponding locally stationary process with skew-t innovations. We will study the case of a parametric model of α-stable locally stationary process with innovation $\varepsilon_t \sim S_\alpha(1/\sqrt{2}, \beta, 0)$ with $\alpha \in (0, 2)$.

5.6 Prediction

There are basically two problems in prediction of $\alpha-$stable locally stationary processes. First, as presented in the Section 2.7, there are few work related to prediction in locally stationary processes. Since the infill asymptotic is applied, the more observations are obtained, the more information are obtained in the time period $[\frac{1}{T}, ..., \frac{T}{T}]$. The interesting approach applied by Van Bellegem and von Sachs (2004) is by considering the observed values $X_{1,T}, \cdots, X_{T-h-1,T}$ and rescaling the time interval to $\left[0, 1 - \frac{h+1}{T}\right]$, where h is the forecasting horizon and the ratio h/T tends to zero as T tends to infinity. Second, the finite variance for traditional time series models allows the possibility of the best linear prediction based on the minimum mean square error.

Recalling that $\alpha-$stable tvARMA has infinite variance, prediction results based on stable ARMA processes with dependent coefficients are presented by Shelton Peiris and Thavaneswaran

(2001a,b). Then, it is possible to predict future values along with the approach applied by Van Bellegem and von Sachs (2004). Consider that the innovations are $S\alpha S$ random variables.

Suppose that we have the system of differential equation (5.1) that satisfies above regular conditions, and $X_{0,T}, \cdots, X_{T',T}$ with $T' = T - h - 1$ are observed. We are interested in prediction the h horizon, i.e. $X_{T-h,T}, ..., X_{T,T}$.

Since $X_{t,T}$ is AR regular, it can be expressed as

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j)\varepsilon_{t-j}. \tag{5.18}$$

Let $\hat{X}_{T'}(l)$ be the best linear predictor of $X_{T'+l,T}$ for $l = 1, ..., h$

$$\hat{X}_{T'}(l) = \sum_{j=0}^{\infty} A(T', T' - j)\varepsilon_{T'-j}, \tag{5.19}$$

where $A(T', T' - j)$ are some functions. Since the prediction error $e_{T'}(l) = \hat{X}_{T'+h,T} - X_{T'}(l)$ is also $S\alpha S$ random variable, it is possible to define its dispersion as $d = \sigma^{\alpha}$ with σ its scale parameter. The idea is to minimize the dispersion d. Note that

$$\begin{aligned} e_{T'}(l) &= X_{T'+l,T} - \hat{X}_{T'}(l) \\ &= \sum_{j=0}^{\infty} a_{T'+l,T}(j)\varepsilon_{T'+l-j} - \sum_{j=0}^{\infty} A(T', T' - j)\varepsilon_{T'-j} \\ &= \sum_{j=0}^{l-1} a_{T'+l,T}(j)\varepsilon_{T'+l-j} + \sum_{j=l}^{\infty} a_{T'+l,T}(j)\varepsilon_{T'+l-j} - \sum_{j=0}^{\infty} A(T', T' - j)\varepsilon_{T'-j} \\ &= \sum_{j=0}^{l-1} a_{T'+l,T}(j)\varepsilon_{T'+l-j} + \sum_{j=0}^{\infty} \left(a_{T'+l,T}(j + l) - A(T', T' - j)\right)\varepsilon_{T'-j}. \end{aligned} \tag{5.20}$$

Then, assuming $\varepsilon_t \sim S_{\alpha}(\frac{1}{\sqrt{2}}, 0, 0)$ and using properties of $S\alpha S$ random variables, its dispersion is

$$\operatorname{disp}[e_{T'}(l)] = \left(\frac{1}{\sqrt{2}}\right)^{\alpha} \sum_{j=0}^{l-1} |a_{T'+l,T}(j)|^{\alpha} + \left(\frac{1}{\sqrt{2}}\right)^{\alpha} \sum_{j=0}^{\infty} |\left(a_{T'+l,T}(j + l) - A(T', T' - j)\right)|^{\alpha}. \tag{5.21}$$

Minimizing the expression (5.21), we obtain the following theorem.

Theorem 5.5. *The minimum dispersion predictor is given by*

$$\hat{X}_{T'}(l) = \sum_{j=0}^{\infty} a_{T'+l,T}(j + l)\varepsilon_{T'-j}. \tag{5.22}$$

Proof. From (5.21), it is straightforward to obtain

$$\min \operatorname{disp}\left[e_{T'}(l)\right] = \left(\frac{1}{\sqrt{2}}\right)^{\alpha} \sum_{j=0}^{l-1} |a_{T'+l,T}(j)|^{\alpha},$$

with $a_{T'+l,T}(j+l) = A(T', T'-j)$ for $j = 0, 1, \ldots$. □

Chapter 6

Indirect inference for α-stable tvARMA process

In this chapter, we study the parameter estimation of a parametric tvARMA model with α-stable innovations. We consider the case with known parameters of the stable innovations.

6.1 tvARMA process with known parameters innovations

In this situation, the model of interest is the tvARMA with innovation $\varepsilon_t \sim S_\alpha(1/\sqrt{2}, \beta, 0)$ with known α and β. It means that the model of interest with the parameter curves can be parametrized by a finite-dimensional parameter θ. The estimation strategy is to consider an auxiliary model with the same parametric time-varying coefficient structure where the innovations ε_t are i.i.d. with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = 1$.

At first sight, it is convenient to consider the block Whittle likelihood discussed in Section 2.2.1 or the generalized Whittle likelihood in Section 2.5 due to the rigorous asymptotic properties such as consistency and normality. Moreover, the dimension of the parameter space of the model of interest and the auxiliary model is the same. However, we carried out several simulation experiments and they turned out to be inappropriate and the convergence of indirect estimation is either difficult or slow. The possible reason is that using these auxiliary estimators involves the time-varying spectral density and its estimates, the preperiodogram, but the α-stable tvARMA process do not have the second moment. Although the Whittle's likelihood has been widely used for non-Gaussian models because of its flexible estimation procedure, it has been proved to produce unreliable estimates in some non-Gaussian cases (Contreras-Cristán *et al.*, 2006).

Then, due to the fact that the skew-t distribution, introduced by Azzalini and Capitanio (2003), was successfully implemented in indirect inference for independent samples from the α-stable distributions and $\alpha-$stable ARMA processes from Lombardi and Calzolari (2008), we adapt the same methodology of assuming the skew-t distribution for the auxiliary model.

The probability density function of the skew-t distribution is defined by

$$\begin{aligned} f(x;\nu,\tilde{\beta},\sigma,\mu) &= \frac{2}{\sigma} f_t(z;\nu) F_t\left(\tilde{\beta} z \sqrt{\frac{\nu+1}{z^2+\nu}};\nu+1\right) \\ &= 2\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left[1+\frac{z^2}{\nu}\right]^{-\frac{\nu+1}{2}} F_t\left(\tilde{\beta} z \sqrt{\frac{\nu+1}{z^2+v}};\nu+1\right), \end{aligned} \tag{6.1}$$

where $z = \frac{x-\mu}{\sigma}$.

The advantage of using this distribution is that it has four parameters in which σ is scale parameter, μ is the location parameter, ν controls the heaviness of the tail and $\tilde{\beta}$ is in charge of the asymmetry of the distribution. Therefore, it is similar to the α-stable distribution and the likelihood function is available. Hence, we will use the standardized t-distribution with $\nu = 3$ degrees of freedom for the case of known parameters since its tail is heavier than the Gaussian one. We will discuss the case when α is unknown in Chapter 7.

6.2 Indirect inference for the α-stable tvAR(1)

Consider the case of tvAR(p) with $p = 1$ and $\gamma\left(\frac{t}{T}\right) = \gamma$

$$X_{t,T} + \alpha_1\left(\frac{t}{T}\right) X_{t-1,T} = \gamma\varepsilon_t, \tag{6.2}$$

where $\varepsilon_t \sim S_\alpha\left(1/\sqrt{2},\beta,0\right)$ with known α and β. Note that there is a solution of the form (5.3) with

$$a_{t,T}(j) = \gamma\prod_{\ell=0}^{j-1}\left[-\alpha_1\left(\frac{t-\ell}{T}\right)\right].$$

Consequently, there is a function $a(u,j) = \gamma\left(-\alpha_1(u)\right)^j$ which satisfies Assumption 2.1, (ii). In other words, $X_{t,T}$ can be locally approximated by a stationary process.

To continue, we illustrate how the indirect inference can be employed to the tvAR(1) in (6.2) with the linear parametric form of the time-varying coefficient $\alpha_1(u) = \theta_0 + \theta_1 u$, and we consider that $\varepsilon_t \sim S_\alpha(1/\sqrt{2},\beta,0)$ for known α and β. Therefore, the parameter vector of the model of interest is $\theta = (\theta_0,\theta_1,\gamma)$.

We use the same parametric form of the process with the scaled t-density with $\nu = 3$ as the auxiliary model with the likelihood function defined in (2.33).

The scaled t density function is defined by

$$f(x;\nu,\sigma) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left[1+\frac{\left(\frac{x}{\sigma}\right)^2}{\nu}\right]^{-\frac{\nu+1}{2}}. \tag{6.3}$$

Consequently, the vector of parameters of the auxiliary model is $\lambda = \left(\theta_0^{(A)},\theta_1^{(A)},\sigma^{(A)}\right)$.

6.2.1 Simulation results

The first simulation was performed by assuming two different scenarios. The first one assumes known parameters $\alpha = 1.3$ and $\beta = 0$ and unknown $(\theta_0, \theta_1, \gamma) = (0.35, 0.4, 1.2)$. In the second scenario, known parameters $\alpha = 1.9$ and $\beta = 0.9$ and unknown $(\theta_0, \theta_1, \gamma) = (-0.3, 0.8, 1)$ are assumed. We carried out simulations for $T = 500$, 1000 and 1500 observations based on $R = 1000$ independent replications each scenario. The indirect inference was carried out using $S = 100$. At the same time, we also performed the blocked Whittle estimation presented in the Section 2.2.1 considering the suggestion of block size $N = \lfloor T^{0.8} \rfloor$ and shifting each block by $Q = \lfloor 0.2N \rfloor$ time units from Dahlhaus and Giraitis (1998). It is important to report that in the first scenario ($\alpha = 1.3$) the results from the blocked Whittle estimates in the Table 6.1 are only based on $R = 945, 898$ and 891 replications included for $T = 500, 1000$, and 1500, respectively, since these excluded replications either presented convergence problem or fail to satisfy the locally stationary condition. At the same time, all replications for the blocked Whittle estimates for the second ($\alpha = 1.9$) converged. This outcome is expected since the innovation distributions approximate to the Gaussian distribution for α close to 2.

Table 6.1 reports the Monte Carlo mean and standard error of both estimation methods. We notice that the Monte Carlo mean from the indirect estimates seems to be consistent, that is, they approximate to the real parameters and present lower standard errors as T increases. On the other hand, the Monte Carlo mean of the blocked Whittle estimates are farther from the real parameters and they present higher standard errors compared to our estimation approach. Moreover, Table 6.2 presents the kurtosis and skewness of all estimates from both methods. In general, all indirect estimates present lower kurtosis and the skewness close to 0. It is important to notice that since the second moment of the process does not exist, the parameter γ estimates from the blocked Whittle likelihood present highly positive asymmetry and they subestimate the true parameter.

Scenario		Indirect estimates						Blocked Whittle estimates[1]		
		Model of Interest			Auxiliary model					
$(\alpha, \beta, \theta_0, \theta_1, \gamma)$	T	θ_0	θ_1	γ	$\theta_0^{(A)}$	$\theta_1^{(A)}$	$\gamma^{(A)}$	$\theta_0^{(W)}$	$\theta_1^{(W)}$	$\gamma^{(W)}$
$(1.3, 0, 0.35, 0.4, 1.2)$	500	0.3492 (0.0282)	0.4003 (0.0457)	1.1999 (0.0674)	0.3492 (0.0281)	0.4004 (0.0457)	1.4149 (0.1587)	0.3366 (0.1193)	0.3947 (0.2075)	7.2051 (7.0580)
	1000	0.3500 (0.0176)	0.3997 (0.0286)	1.1990 (0.0470)	0.3500 (0.0175)	0.3997 (0.0285)	1.4108 (0.1102)	0.3465 (0.0585)	0.3984 (0.1029)	8.5515 (10.5234)
	1500	0.3503 (0.0122)	0.3996 (0.0202)	1.1978 (0.0367)	0.3503 (0.0121)	0.3996 (0.0202)	1.4073 (0.0863)	0.3514 (0.0527)	0.3937 (0.0927)	9.0899 (9.0223)
$(1.9, 0.9, -0.3, 0.8, 1)$	500	-0.2952 (0.0881)	0.7897 (0.1523)	0.9966 (0.0366)	-0.2952 (0.0878)	0.7893 (0.1520)	0.6571 (0.0481)	-0.2880 (0.1172)	0.7825 (0.2216)	1.2086 (0.6352)
	1000	-0.2975 (0.0585)	0.7926 (0.1028)	0.9996 (0.0260)	-0.2972 (0.0584)	0.7923 (0.1023)	0.6603 (0.0343)	-0.2917 (0.0811)	0.7845 (0.1545)	1.2197 (0.4734)
	1500	-0.2974 (0.0494)	0.7958 (0.0793)	0.9997 (0.0209)	-0.2975 (0.0491)	0.7955 (0.0792)	0.6603 (0.0274)	-0.2940 (0.0639)	0.7926 (0.1162)	1.2709 (0.8738)

Table 6.1: *Monte Carlo mean and standard error (in parentheses) for different sample sizes ($T = 500, 1000, 1500$) using indirect estimators (both model of interest and auxiliary model) and blocked Whittle estimates assuming known α and β from α-stable tvAR(1) based on $R = 1000$ replications.*

[1] In tvAR(1) simulations, the blocked Whittle estimates did not converge in some cases. Therefore, ex-

Scenario			Indirect estimates			Blocked Whittle estimates[1]		
$(\alpha, \beta, \theta_0, \theta_1, \gamma)$	T		θ_0	θ_1	γ	$\theta_0^{(W)}$	$\theta_1^{(W)}$	$\gamma^{(W)}$
$(1.3, 0, 0.35, 0.4, 1.2)$	500	Kur	6.9415	4.9937	3.1057	9.7077	6.2082	38.8804
		Skw	-0.2709	-0.1038	0.0924	-1.6850	0.3139	4.8177
	1000	Kur	5.2264	5.1948	3.0838	7.4547	6.0051	79.2429
		Skw	-0.2613	-0.0110	0.1894	-0.7995	0.1690	7.2834
	1500	Kur	4.7557	4.1868	3.0163	25.1581	40.4760	76.0068
		Skw	-0.1881	-0.0876	0.1757	-0.7907	-2.0097	6.8018
$(1.9, 0.9, -0.3, 0.8, 1)$	500	Kur	3.0330	2.8565	3.1076	3.3783	3.0944	375.8563
		Skw	0.1388	-0.1354	0.1241	-0.0129	-0.0875	16.6778
	1000	Kur	3.1678	3.2835	2.7390	2.8722	2.9543	95.2261
		Skw	0.0341	-0.0260	0.0437	0.0057	-0.1047	8.4487
	1500	Kur	3.1024	3.0645	2.9329	3.7487	6.1799	187.9560
		Skw	-0.0299	0.0248	0.0026	0.1707	-0.5935	12.4661

Table 6.2: *Kurtosis and skewness of indirect estimates and blocked Whittle estimates for different sample sizes ($T = 500, 1000, 1500$) assuming known α and β from α-stable tvAR(1) based on $R = 1000$ replications.*

Moreover, Figures 6.1 and 6.2 show the density estimates of each parameter for both scenarios. The density estimates show that the standard error become smaller as T increases. Along with the results from Tables 6.1 and 6.2, we can conclude that indirect estimates behave better than the blocked Whittle estimates in term of mean, standard error, skewness and kurtosis. Therefore, the simulation results show that the indirect inference is satisfactory.

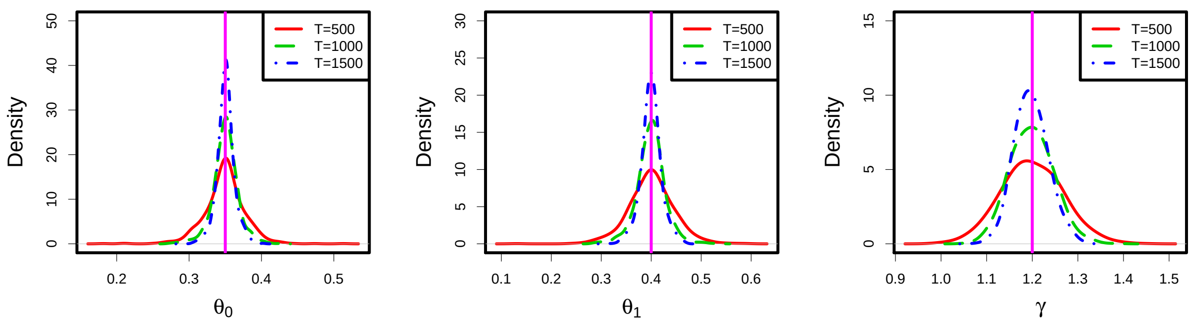


Figure 6.1: *Density estimates of θ_0, θ_1 and γ for different sample sizes based on $R = 1000$ replications from α-stable tvAR(1) with $\alpha = 1.3$, $\beta = 0$, $\theta_0 = 0.35$, $\theta_1 = 0.4$, $\gamma = 1.2$ using indirect inference.*

6.3 Indirect inference for the α-stable tvMA(1)

In this section, we carried out the second simulation with the case of tvMA(q) with $q = 1$ and $\gamma\left(\frac{t}{T}\right) = \gamma$, that is,

$$X_{t,T} = \gamma \left\{ \varepsilon_t + \beta_1 \left(\frac{t}{T} \right) \varepsilon_{t-1} \right\}, \tag{6.4}$$

where $\varepsilon_t \sim S_\alpha \left(1/\sqrt{2}, \beta, 0 \right)$ with known α and β.

cluding those cases, for the first scenario ($\alpha = 1.3$) $R = 945, 898$ and 891 replications are included for $T = 500, 1000$, and 1500, respectively. For the second scenario ($\alpha = 1.9$), all cases converged.

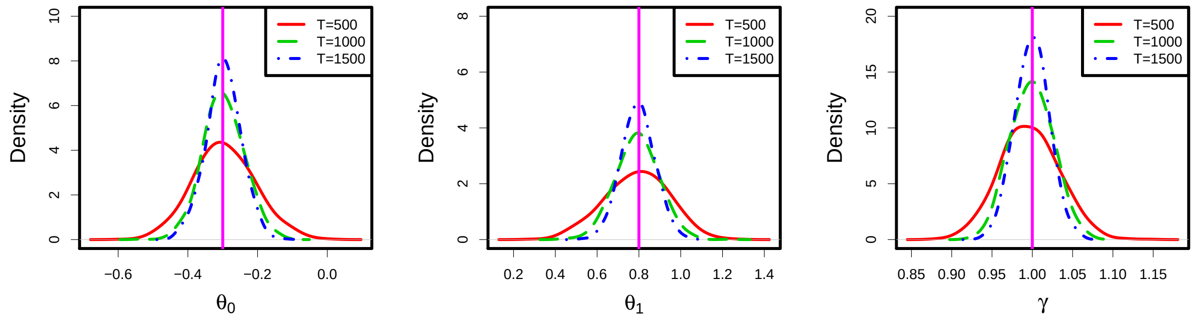


Figure 6.2: *Density estimates of θ_0, θ_1 and γ for different sample sizes based on $R = 1000$ replications from α-stable tvAR(1) with $\alpha = 1.9$, $\beta = 0.9$, $\theta_0 = -0.3$, $\theta_1 = 0.8$, $\gamma = 1$ using indirect inference.*

To continue, we illustrate how the indirect inference can be employed to the tvMA(1) in (6.4) with the linear parametric form of the time-varying coefficient $\beta_1(u) = \theta_0 + \theta_1 u$, and we consider that $\varepsilon_t \sim S_\alpha(1/\sqrt{2}, \beta, 0)$ for known α and β. Therefore, the parameter vector of the model of interest is $\theta = (\theta_0, \theta_1, \gamma)$.

We use the same strategy to estimate this model with the auxiliary model with the parametric form of the process with the scaled t-distribution with $\nu = 3$.

6.3.1 Simulation results

Similarly to the simulation done with tvAR(1), the tvMA(1) simulations were done with two scenarios. The first one with performed by assuming known parameters $\alpha = 1.1$ and $\beta = -0.2$ and unknown $(\theta_0, \theta_1, \gamma) = (0.35, -0.6, 1.2)$. In the second scenario, known parameters $\alpha = 1.8$ and $\beta = 0$ and unknown $(\theta_0, \theta_1, \gamma) = (-0.2, 0.7, 1.1)$ are assumed. Simulations were done for $T = 500$, 1000 and 1500 observations with $S = 100$ based on $R = 1000$ independent replications each scenario. We also performed the blocked Whittle estimation following the same methodology. Excluding those diverged cases, there are $R = 939, 978$ and 978 replications that are included for $T = 500, 1000$, and 1500, respectively, for the first scenario ($\alpha = 1.1$), while for the second scenario ($\alpha = 1.8$), $R = 996, 1000$ and 1000 replications are included for $T = 500, 1000$, and 1500, respectively.

The Monte Carlo mean, standard error, kurtosis and skewness of estimates from the simulation are reported in the Table 6.3 and 6.4 and the density estimates in Figures 6.3, 6.4.

We notice that the results are very similar to the case of tvAR(1). The distribution seem to be less leptokurtic for α close to 2, and the standard error become smaller as T increases. Specifically, for α close to 2, the distribution of indirect estimates is close to the Gaussian distribution (kurtosis are closer to 3 compared with the blocked Whittle estimates and they are more symmetric). On the other hand, for smaller α, the distribution of indirect estimates has heavier tails, and they have similar kurtosis and skewness than the blocked Whittle estimates, except for the parameter γ, indirect estimates behave better. However, in term of standard error and Monte Carlo mean, they still behave better than the blocked Whittle estimates. Therefore, we conclude that the indirect inference is satisfactory.

Scenario		Indirect estimates						Blocked Whittle estimates[2]		
		Model of Interest			Auxiliary model					
$(\alpha,\beta,\theta_0,\theta_1,\gamma)$	T	θ_0	θ_1	γ	$\theta_0^{(A)}$	$\theta_1^{(A)}$	$\gamma^{(A)}$	$\theta_0^{(W)}$	$\theta_1^{(W)}$	$\gamma^{(W)}$
$(1.1,-0.2,0.35,-0.6,1.2)$	500	0.3561	-0.5888	1.1989	0.3562	-0.5887	3.0928	0.3424	-0.5427	18.7932
		(0.0298)	(0.0577)	(0.0600)	(0.0298)	(0.0577)	(0.3093)	(0.1418)	(0.3084)	(38.4343)
	1000	0.3545	-0.5953	1.1986	0.3546	-0.5952	3.0877	0.3386	-0.5532	47.5752
		(0.0186)	(0.0352)	(0.0412)	(0.0185)	(0.0352)	(0.2115)	(0.0870)	(0.1955)	(232.0620)
	1500	0.3536	-0.5982	1.1986	0.3537	-0.5981	3.0859	0.3357	-0.5555	49.4572
		(0.0131)	(0.0244)	(0.0331)	(0.0131)	(0.0244)	(0.1703)	(0.0747)	(0.1690)	(178.3984)
$(1.8,0,-0.2,0.7,1.1)$	500	-0.1981	0.6981	1.0988	-0.1978	0.6976	0.8297	-0.1958	0.6990	1.5694
		(0.0746)	(0.1257)	(0.0450)	(0.0746)	(0.1257)	(0.0678)	(0.1169)	(0.2178)	(0.6809)
	1000	-0.1991	0.6974	1.1009	-0.1991	0.6974	0.8320	-0.1970	0.6951	1.6936
		(0.0517)	(0.0886)	(0.0303)	(0.0515)	(0.0884)	(0.0456)	(0.0785)	(0.1491)	(0.9625)
	1500	-0.1996	0.6996	1.0995	-0.1997	0.6998	0.8298	-0.1976	0.6979	1.6855
		(0.0410)	(0.0677)	(0.0247)	(0.0410)	(0.0677)	(0.0370)	(0.0622)	(0.1135)	(1.0475)

Table 6.3: *Monte Carlo mean and standard error (in parentheses) for different sample sizes ($T = 500, 1000, 1500$) using indirect estimators (both model of interest and auxiliary model) and blocked Whittle estimates assuming known α and β from α-stable tvMA(1) based on $R = 1000$ replications.*

Scenario			Indirect estimates			Blocked Whittle estimates[2]		
$(\alpha,\beta,\theta_0,\theta_1,\gamma)$	T		θ_0	θ_1	γ	$\theta_0^{(W)}$	$\theta_1^{(W)}$	$\gamma^{(W)}$
$(1.1,-0.2,0.35,-0.6,1.2)$	500	Kur	7.9023	6.3460	2.9050	6.9952	7.3434	233.1652
		Skw	1.2950	0.0800	0.2117	0.1961	1.1869	13.5324
	1000	Kur	9.8633	10.4926	2.8616	11.7454	8.6755	385.9194
		Skw	1.6510	0.7121	0.0841	0.8633	1.5096	17.6374
	1500	Kur	8.1466	20.6873	2.8140	9.7011	10.5014	156.1843
		Skw	1.5194	1.4762	0.1493	-0.1837	1.9926	11.4943
$(1.8,0,-0.2,0.7,1.1)$	500	Kur	3.3039	2.9901	2.7817	4.4813	3.1555	66.7817
		Skw	0.0966	-0.0861	-0.0658	0.4187	-0.0800	6.4792
	1000	Kur	3.3611	3.4395	2.8332	3.4253	3.4149	74.8992
		Skw	-0.0234	0.0120	0.1609	0.0556	-0.0736	7.2050
	1500	Kur	3.3234	3.0226	2.7790	3.2663	3.3457	193.5169
		Skw	0.0082	0.0459	-0.0775	0.1853	-0.2094	11.9289

Table 6.4: *Kurtosis and skewness of indirect estimates and blocked Whittle estimates for different sample sizes ($T = 500, 1000, 1500$) assuming known α and β from α-stable tvMA(1) based on $R = 1000$ replications.*

6.4 Indirect inference for the α-stable tvARMA(1,1)

In this section, we carried out the third simulation with the case of tvARMA(p,q) with $p = 1$, $q = 1$ and $\gamma\left(\frac{t}{T}\right) = \gamma$, that is,

$$X_{t,T} + \alpha_1\left(\frac{t}{T}\right)X_{t-1,T} = \gamma\left\{\varepsilon_t + \beta_1\left(\frac{t}{T}\right)\varepsilon_{t-1}\right\}, \tag{6.5}$$

where $\varepsilon_t \sim S_\alpha\left(1/\sqrt{2}, \beta, 0\right)$ with known α and β.

We illustrate how the indirect inference can be employed to the tvARMA(1,1) in (6.5) with the linear parametric form of the time-varying coefficients $\alpha_1(u) = \theta_{a0} + \theta_{a1}u$ and

[2]In tvMA(1) simulations, the blocked Whittle estimates did not converge in some cases. Therefore, excluding those cases, for the first scenario ($\alpha = 1.1$) $R = 939, 978$ and 978 replications are included for $T = 500, 1000$, and 1500, respectively. For the second scenario ($\alpha = 1.8$), $R = 996, 1000$ and 1000 replications are included for $T = 500, 1000$, and 1500, respectively.

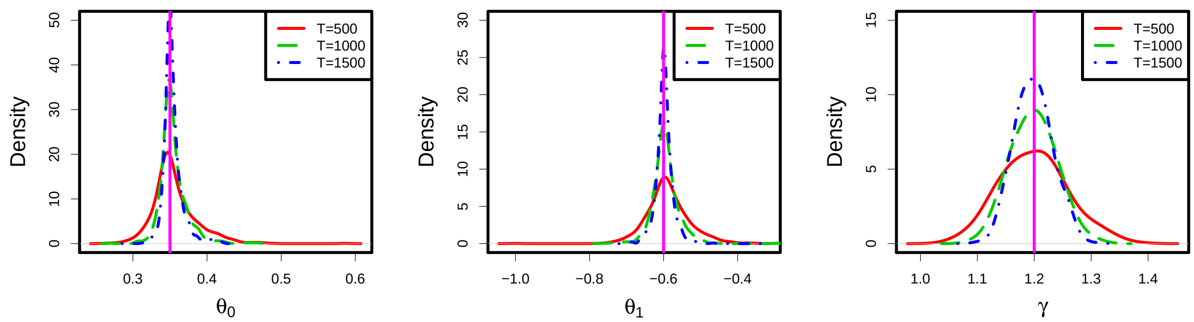


Figure 6.3: *Density estimates of θ_0, θ_1 and γ for different sample sizes based on $R = 1000$ replications from α-stable tvMA(1) with $\alpha = 1.1$, $\beta = -0.2$, $\theta_0 = 0.35$, $\theta_1 = -0.6$, $\gamma = 1.2$ using indirect inference.*

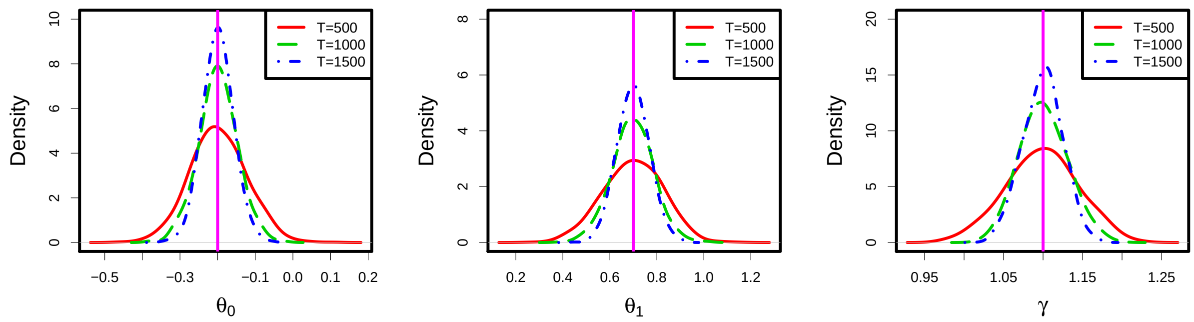


Figure 6.4: *Density estimates of θ_0, θ_1 and γ for different sample sizes based on $R = 1000$ replications from α-stable tvMA(1) with $\alpha = 1.8$, $\beta = 0$, $\theta_0 = -0.2$, $\theta_1 = 0.7$, $\gamma = 1.1$ using indirect inference.*

$\beta_1(u) = \theta_{b0} + \theta_{b1}u$, and we consider that $\varepsilon_t \sim S_\alpha(1/\sqrt{2}, \beta, 0)$ for known α and β. Therefore, the parameter vector of the model of interest is $\theta = (\theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \gamma)$.

We use the same strategy to estimate this model with the auxiliary model with the parametric form of the process with the scaled t-distribution with $\nu = 3$.

6.4.1 Simulation results

In the same way, the tvARMA(1,1) simulations were done with two scenarios. The first one were performed by assuming known parameters $\alpha = 1.3$ and $\beta = 0$ and unknown $(\theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \gamma) = (-0.2, -0.4, 0.2, 0.3, 1.1)$. In the second scenario, known parameters $\alpha = 1.8$ and $\beta = 0.3$ and unknown $(\theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \gamma) = (-0.4, 0.1, 0.1, 0.3, 1.1)$. Again, simulations were done for $T = 500$, 1000 and 1500 observations with $S = 100$ based on $R = 1000$ independent replications each scenario. The blocked Whittle estimation following the same methodology and excluding those cases that diverged, there are $R = 908, 929$ and 926 replications are included for $T = 500, 1000$, and 1500, respectively, for the first scenario ($\alpha = 1.1$), while for the second scenario ($\alpha = 1.8$), $R = 989, 996$ and 994 replications are included for $T = 500, 1000$, and 1500, respectively.

The Monte Carlo mean, standard error, kurtosis and skewness of estimates from the tvARMA(1,1) simulation are reported in the Table 6.5 and 6.6 and the density estimates in Figures 6.5, 6.6.

We again notice similar results compared to tvAR(1) and tvMA(1). Since the standard error become smaller as T increases, simulations suggest that indirect estimates are consistent. In general, the distribution of indirect estimates has heavier tails, and they have similar kurtosis and skewness than the blocked Whittle estimates (except for the parameter γ, indirect estimates behave better). However, in term of standard error and Monte Carlo mean, they behave much better than the blocked Whittle estimates. Therefore, we conclude that the indirect inference is satisfactory for tvARMA(1,1).

6.5 Application

In this section, we illustrate this methodology with two time series: the tree ring and wind power data.

6.5.1 Tree Ring

The tree ring data was collected at Piedra del Aguila, Malleco, Chile from 1242 to 1975 provided by Holmes (2014). Figure 6.7 shows its time series and its (global) sample

[3] In tvARMA(1,1) simulations, the blocked Whittle estimates did not converge in some cases. Therefore, excluding those cases, for the first scenario ($\alpha = 1.1$) $R = 908, 929$ and 926 replications are included for $T = 500, 1000$, and 1500, respectively. For the second scenario ($\alpha = 1.8$), $R = 989, 996$ and 994 replications are included for $T = 500, 1000$, and 1500, respectively.

Scenario	T	Indirect estimates									
		Model of Interest					Auxiliary model				
		θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ	$\theta_{a0}^{(A)}$	$\theta_{a1}^{(A)}$	$\theta_{b0}^{(A)}$	$\theta_{b1}^{(A)}$	$\gamma^{(A)}$
1	500	-0.1977	-0.4028	0.2035	0.2955	1.0977	-0.1977	-0.4029	0.2036	0.2954	1.1843
		(0.0554)	(0.0841)	(0.0571)	(0.0878)	(0.0623)	(0.0554)	(0.0838)	(0.0571)	(0.0878)	(0.1348)
	1000	-0.1995	-0.4003	0.2017	0.2978	1.0987	-0.1994	-0.4005	0.2017	0.2978	1.1845
		(0.0333)	(0.0514)	(0.0347)	(0.0534)	(0.0419)	(0.0333)	(0.0510)	(0.0347)	(0.0534)	(0.0900)
	1500	-0.2003	-0.3986	0.1997	0.3012	1.1009	-0.2003	-0.3990	0.1997	0.3011	1.1888
		(0.0248)	(0.0366)	(0.0258)	(0.0382)	(0.0349)	(0.0247)	(0.0363)	(0.0258)	(0.0381)	(0.0752)
2	500	-0.4000	0.1061	0.0987	0.3097	0.9976	-0.3998	0.1042	0.0991	0.3093	0.6867
		(0.1360)	(0.2222)	(0.1501)	(0.2395)	(0.0386)	(0.1352)	(0.2219)	(0.1496)	(0.2396)	(0.0529)
	1000	-0.3921	0.0881	0.1064	0.2905	0.9982	-0.3913	0.0862	0.1070	0.2891	0.6868
		(0.1001)	(0.1617)	(0.1053)	(0.1652)	(0.0290)	(0.0999)	(0.1613)	(0.1051)	(0.1643)	(0.0400)
	1500	-0.3992	0.1021	0.0988	0.3060	0.9982	-0.3993	0.1015	0.0991	0.3058	0.6865
		(0.0754)	(0.1269)	(0.0793)	(0.1285)	(0.0232)	(0.0748)	(0.1257)	(0.0786)	(0.1279)	(0.0315)
		Blocked Whittle estimates[3]									
		$\theta_{a0}^{(W)}$	$\theta_{a1}^{(W)}$	$\theta_{b0}^{(W)}$	$\theta_{b1}^{(W)}$	$\gamma^{(W)}$					
1	500	-0.1974	-0.4031	0.2185	0.2840	5.6739					
		(0.1632)	(0.2679)	(0.1624)	(0.2724)	(4.7869)					
	1000	-0.2041	-0.4015	0.2039	0.2896	7.0926					
		(0.1266)	(0.2032)	(0.1189)	(0.1941)	(7.4793)					
	1500	-0.2048	-0.3983	0.2019	0.2916	8.1174					
		(0.1052)	(0.1694)	(0.0945)	(0.1520)	(9.0376)					
2	500	-0.3917	0.1021	0.1078	0.3049	1.4151					
		(0.1952)	(0.3522)	(0.2130)	(0.3810)	(0.7603)					
	1000	-0.3850	0.0815	0.1105	0.2880	1.4919					
		(0.1409)	(0.2535)	(0.1470)	(0.2599)	(0.5806)					
	1500	-0.3939	0.0926	0.1055	0.2964	1.5538					
		(0.1085)	(0.1955)	(0.1155)	(0.2040)	(0.8009)					

Table 6.5: *Monte Carlo mean and standard error (in parentheses) for different sample sizes ($T = 500, 1000, 1500$) using indirect estimators (both model of interest and auxiliary model) and blocked Whittle estimates assuming known α and β from α-stable tvARMA(1,1) based on $R = 1000$ replications. Scenario 1 assumes $\alpha = 1.3,\ \beta = 0,\ \theta_{a0} = -0.2,\ \theta_{a1} = -0.4,\ \theta_{b0} = 0.2,\ \theta_{b1} = 0.3,\ \gamma = 1.1$ and Scenario 2 assumes $\alpha = 1.8,\ \beta = 0.3,\ \theta_{a0} = -0.4,\ \theta_{a1} = 0.1,\ \theta_{b0} = 0.1,\ \theta_{b1} = 0.3,\ \gamma = 1$.*

Scenario	T		Indirect estimates				
			θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ
1	500	Kur	4.7062	5.3065	5.7858	5.9446	3.1403
		Skw	-0.0539	0.1409	-0.1818	0.1172	0.2005
	1000	Kur	4.5786	4.5836	4.3494	4.4781	3.2191
		Skw	0.0217	-0.1284	-0.0273	-0.1239	0.1007
	1500	Kur	6.5317	5.6912	5.4510	4.8657	2.7667
		Skw	-0.4095	0.3837	-0.4359	0.2636	-0.0475
2	500	Kur	3.3650	3.1791	3.3657	3.4297	2.9935
		Skw	0.2754	-0.2426	-0.0746	-0.1274	0.1839
	1500	Kur	3.3964	3.5054	3.4470	3.4690	3.0327
		Skw	0.2002	-0.1558	0.0100	-0.1184	0.2460
	1500	Kur	3.5817	3.1935	3.7052	3.3930	2.9790
		Skw	0.2895	-0.1097	0.0137	-0.0730	0.0718
			Blocked Whittle estimates[3]				
			$\theta_{a0}^{(W)}$	$\theta_{a1}^{(W)}$	$\theta_{b0}^{(W)}$	$\theta_{b1}^{(W)}$	$\gamma^{(W)}$
1	500	Kur	5.4254	4.8197	5.3321	4.3749	139.1506
		Skw	-0.0294	0.2331	-0.0403	-0.0186	8.6437
	1000	Kur	8.1137	5.1403	8.1128	5.0651	210.4645
		Skw	-0.3356	0.1724	-0.4294	0.1899	11.1724
	1500	Kur	10.8431	8.4975	11.3543	7.8360	127.1256
		Skw	0.2951	-0.1960	0.1167	-0.1974	9.2216
2	500	Kur	2.9112	3.0692	3.2168	3.2822	203.2823
		Skw	0.1699	-0.1329	-0.2024	-0.0422	12.0737
	1500	Kur	3.4242	3.2166	3.2601	3.0064	41.6803
		Skw	0.3149	-0.2253	0.0267	-0.1713	4.9608
	1500	Kur	2.9176	2.8801	3.3685	3.3091	96.2273
		Skw	0.0809	0.0268	-0.1083	0.0372	7.7396

Table 6.6: *Kurtosis and skewness of indirect estimates and blocked Whittle estimates for different sample sizes ($T = 500, 1000, 1500$) assuming known α and β from α-stable tvARMA(1,1) based on $R = 1000$ replications. Scenario 1 assumes $\alpha = 1.3,\ \beta = 0,\ \theta_{a0} = -0.2,\ \theta_{a1} = -0.4,\ \theta_{b0} = 0.2,\ \theta_{b1} = 0.3,\ \gamma = 1.1$ and Scenario 2 assumes $\alpha = 1.8,\ \beta = 0.3,\ \theta_{a0} = -0.4,\ \theta_{a1} = 0.1,\ \theta_{b0} = 0.1,\ \theta_{b1} = 0.3,\ \gamma = 1$.*

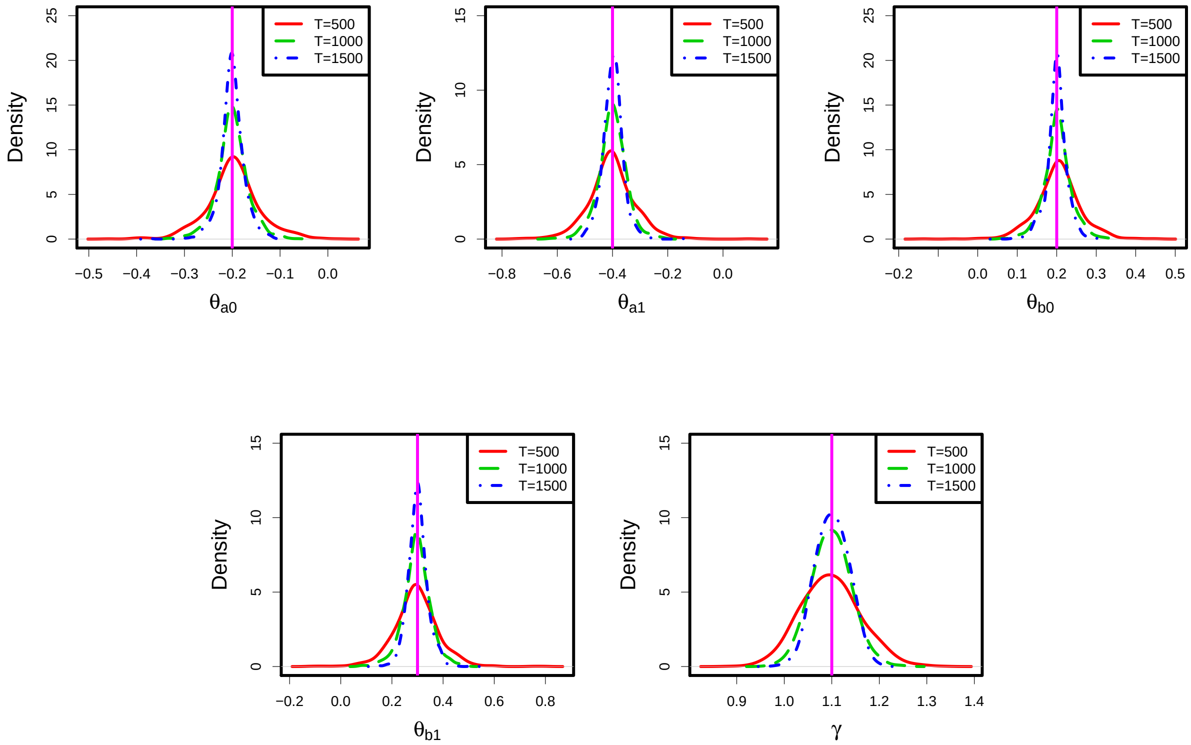


Figure 6.5: *Density estimates of θ_{a0}, θ_{a1}, θ_{b0}, θ_{b1} and γ for different sample sizes based on $R = 1000$ replications from α-stable tvARMA(1,1) with $\alpha = 1.3$, $\beta = 0$, $\theta_{a0} = -0.2$, $\theta_{a1} = -0.4$, $\theta_{b0} = 0.2$, $\theta_{b1} = 0.3$, $\gamma = 1.1$ using indirect inference.*

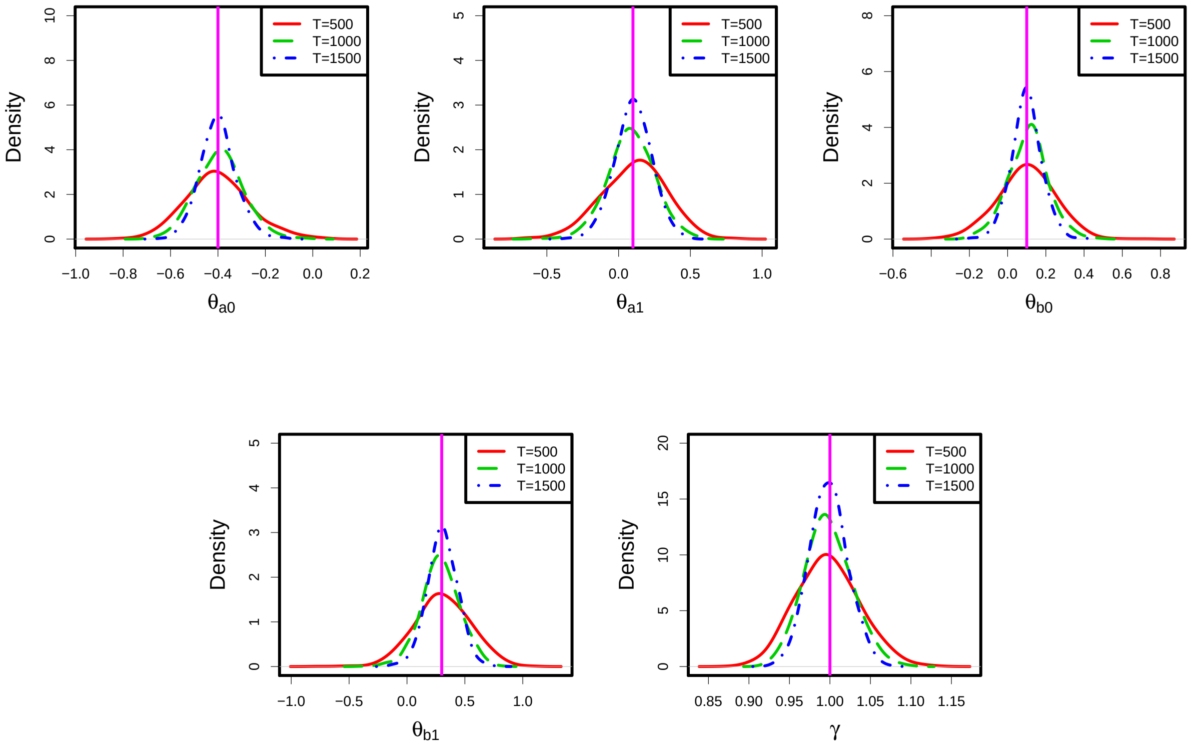
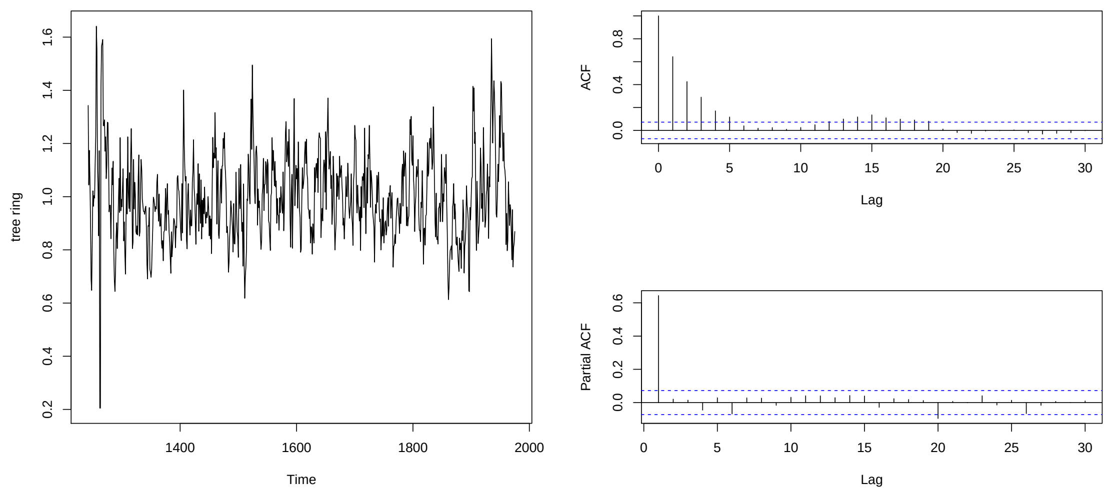


Figure 6.6: *Density estimates of* θ_{a0}, θ_{a1}, θ_{b0}, θ_{b1} *and* γ *for different sample sizes based on* $R = 1000$ *replications from* α*-stable tvARMA(1,1) with* $\alpha = 1.8$, $\beta = 0.3$, $\theta_{a0} = -0.4$, $\theta_{a1} = 0.1$, $\theta_{b0} = 0.1$, $\theta_{b1} = 0.3$, $\gamma = 1$ *using indirect inference.*

autocorrelation function and partial autocorrelation function. An AR(1) model seems to be an appropriate option, but if we analyze for different time periods, they present different structures (Figure 6.8). Moreover, Figure 6.9 shows the blocked smooth periodogram of the series and it presents slowly changed structure throughout the time. For the analysis below, we subtracted the series by its mean in order to have the time series data with mean zero.



(a) Tree ring.

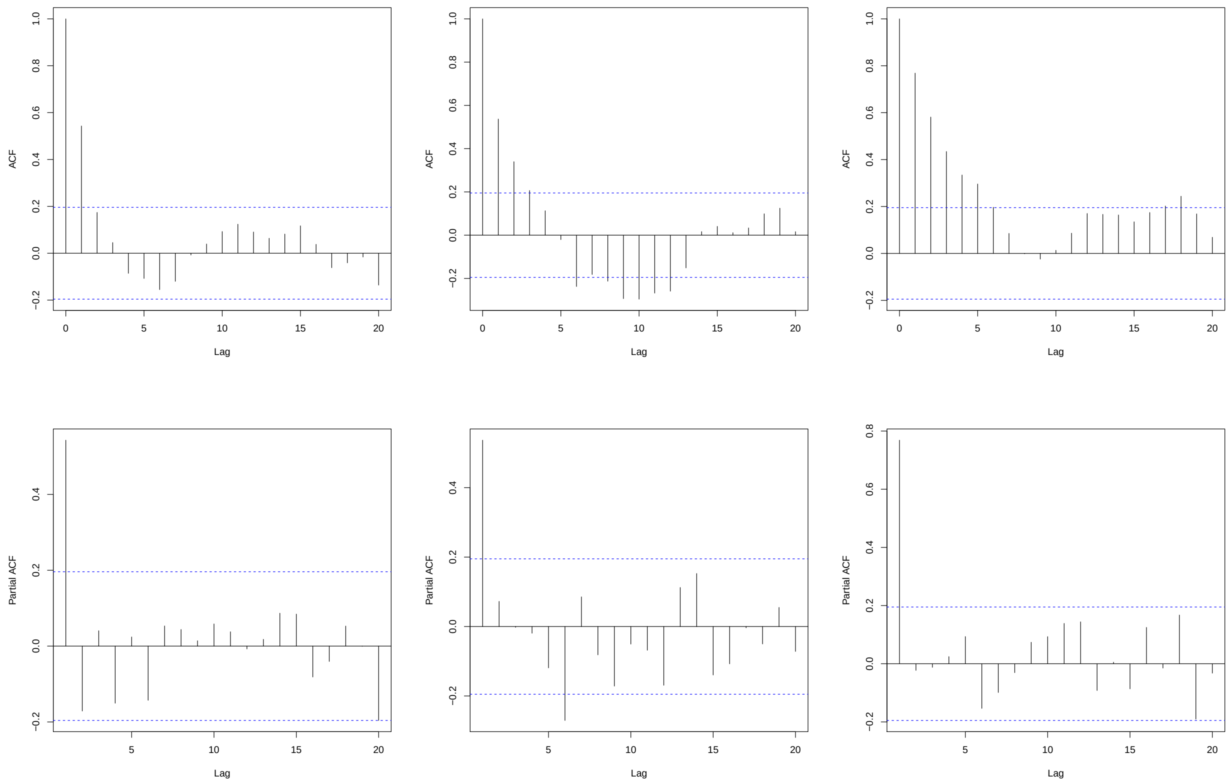
(b) autocorrelation function (ACF) and partial autocorrelation function (partial ACF).

Figure 6.7: *Tree ring data from 1242 to 1975.*

Following the methodology and codes provided by Olea *et al.* (2015), using block size $N = 200$ and shifting each block by $Q = 100$ time units, 6 blocks are obtained. For each block, a stationary AR(1) model is estimated. Then, the smoothed estimated coefficients over time is presented in Figure 6.10. It shows that the coefficients could vary linearly throughout the time. Consequently, a tvAR(1) model with linear coefficients, $\alpha_1(u) = \theta_0 + \theta_1(u)$ and $\gamma(u) = \gamma_0 + \gamma_1(u)$, can be proposed.

One problem that we detected by using the blocked Whittle estimation is that choosing different block size N yields distinct results. To illustrate it, we estimate the Model 1 by using the block size $N = 180$ recommended in Olea *et al.* (2015), then we estimate the Model 2 by considering the suggestion of block size $N = \lfloor T^{0.8} \rfloor = \lfloor 734^{0.8} \rfloor = 196$ in Dahlhaus and Giraitis (1998).

Table 6.7 reports the parameter estimates from both models. For both models, we notice that $\gamma(\cdot)$ presents negative slope, i.e. the process has a decreasing variance throughout the time. On the other hand, if we use 5% of significance level, there could have different interpretations, since the first model implies time-varying autoregressive coefficient and the second model does not. Moreover, the residual analysis shows that although the ACF indicates uncorrelated residuals, the QQ-plot, box-plot and the histogram (Figure 6.11) show that the distribution of error has heavy tail and is positively asymmetric. Additionally, we



(a) Observation 1 to 100 (b) Observation 300 to 400 (c) Observation 600 to 700

Figure 6.8: *Autocorrelation function and partial autocorrelation function of tree ring data for 3 different time windows.*

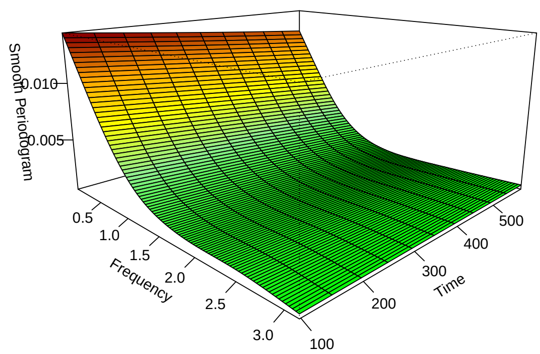
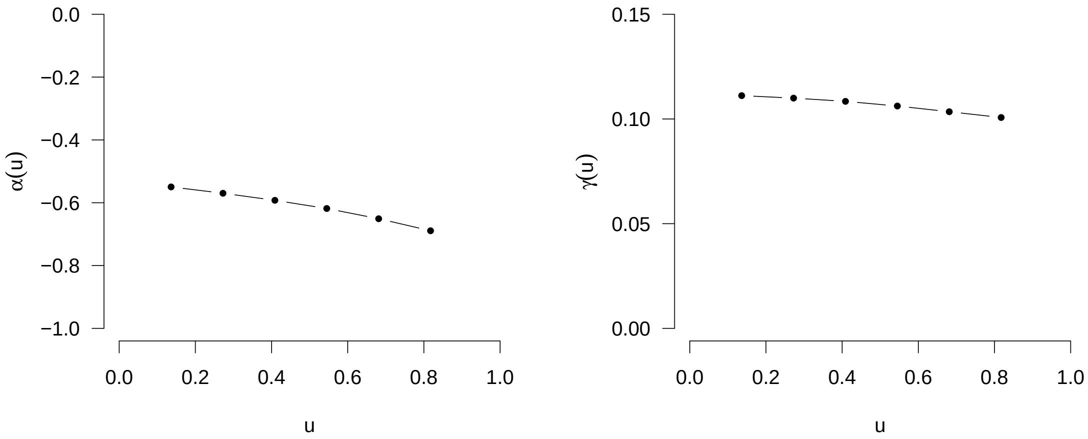


Figure 6.9: *Blocked smooth periodogram of tree ring data.*



(a) $\alpha(u)$ (b) $\gamma(u)$

Figure 6.10: *Estimates of stationary AR(1) model for* 6 *block of size* $N = 200$ *for* $u = \frac{t}{T}$.

also estimated the skewness (0.25) and kurtosis (7.73) and carrying out Shapiro-Wilk test and Jarque-Bera test, both tests rejected the null hypothesis of normality.

Parameter	Model 1				Model 2			
	Estimate	s.e.	z-value	p-value	Estimate	s.e.	z-value	p-value
θ_0	-0.5006	0.0714	-7.0129	0.0000	-0.5280	0.0741	-7.1215	0.0000
θ_1	-0.2542	0.1254	-2.0272	0.0426	-0.1870	0.1369	-1.3661	0.1719
γ_0	0.1130	0.0028	40.4541	0.0000	0.1145	0.0028	41.2186	0.0000
γ_1	-0.0121	0.0051	-2.3684	0.0179	-0.0166	0.0052	-3.1796	0.0015

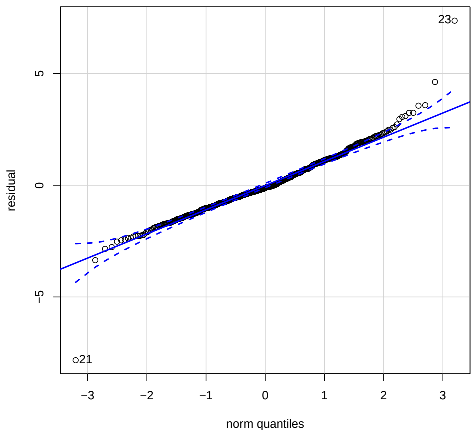
Table 6.7: *Blocked Whittle estimates of tvAR(1) from tree ring time series by considering different block size* $N = 180$ *(Model 1) and* $N = 196$ *(Model 2).*

Next, since the residual analysis shows that the residuals present asymmetry and heavy tail, we propose a more flexible model, $\alpha-$stable tvAR(1), by assuming the parameters of stable innovations as estimated above, i.e. $\alpha = 1.9$, $\beta = 0.98$ and use indirect inference with $S = 30$ to estimate the parameter $(\theta_0, \theta_1, \gamma_0, \gamma_1)$. Table 6.8 presents the indirect estimates along with their (Monte Carlo) standard error calculated based on $R = 200$. Notice that the autoregressive coefficient is not time-varying.

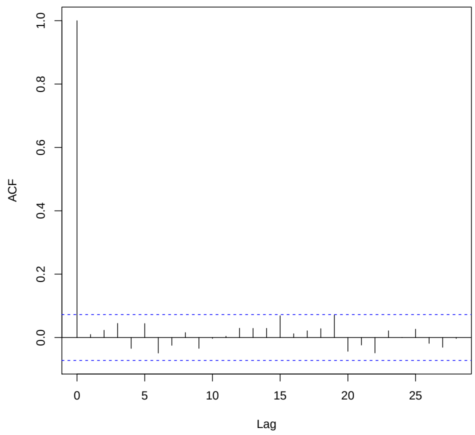
Parameter	θ_0	θ_1	γ_0	γ_1
Indirect estimates	-0.6424	-0.0674	0.1280	-0.0244
Standard error	(0.0499)	(0.0857)	(0.0079)	(0.0136)

Table 6.8: *Indirect estimates of* $\alpha-$*stable tvAR(1) with* $S = 40$.

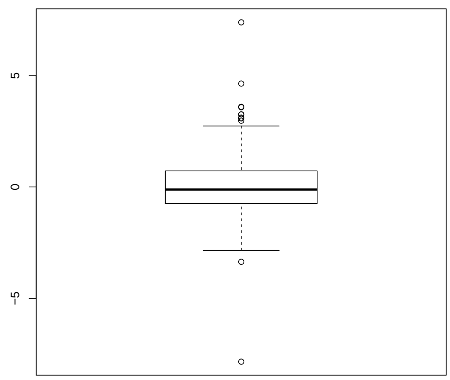
To evaluate the residual distribution with the stable distribution, Nolan (2002) suggested using the stabilized probability plot (stabilized p-p plot), proposed by Michael (1983),



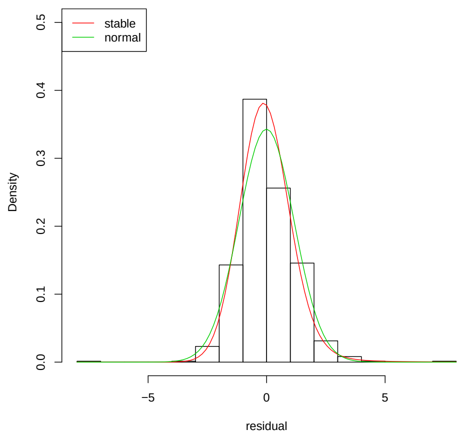
(a) QQ-plot



(b) Autocorreltion function



(c) Box-plot

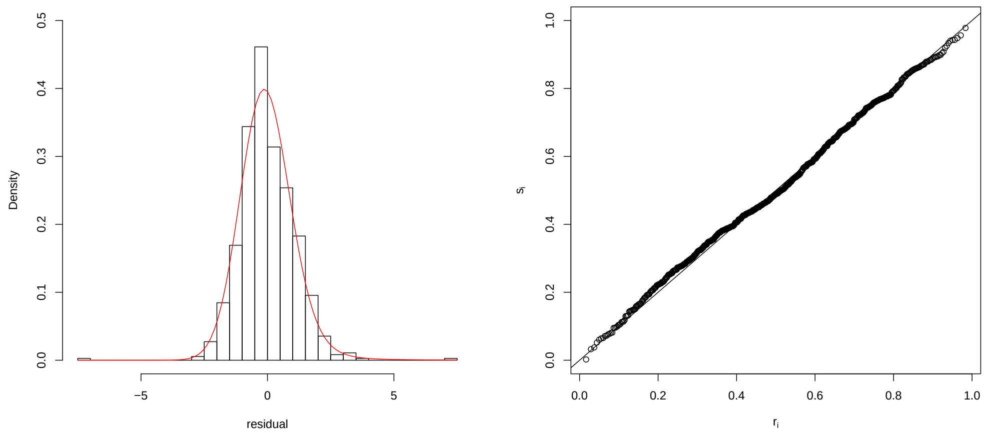


(d) Histogram with estimated stable curve ($\alpha = 1.9$, $\beta = 0.98$) and Gaussian curve.

Figure 6.11: *Residual analysis of the model 1 (standadized residual).*

instead of the QQ-plot due to the fact that the last one is not suitable to evaluate stable distributions. In case of the QQ-plot, large fluctuation for the extreme values in case of the heavy-tailed distribution will produce very large standard errors in the tails. Also, the standard p-p plot tend to emphasize behavior around the mode of the distribution and the plotted points near the tails have more variation. By using the transformation below, a stabilized p-p plot is defined to eliminates this non-uniformity so that the variance of the plotted points are approximately equal. Let $y_1 \leq \cdots \leq y_n$ be an ordered random sample of size n from the distribution F. The stabilized p-p plot is defined as the plot of $s_i = \left(\frac{2}{\pi}\right) \arcsin(F^{\frac{1}{2}}(y_i))$ versus $r_i = \left(\frac{2}{\pi}\right) \arcsin\left(\left[\left(i - \frac{1}{2}\right)/n\right]^{\frac{1}{2}}\right)$.

In this way, the histogram and the stabilized p-p plot in figure 6.12 show that the stable distribution fits well the residuals.



(a) Histogram with the stable curve (b) Stabilized p-p plot

Figure 6.12: *Residual analysis from the $\alpha-$stable tvAR(1) model assuming that the innovation distribution is stable with $\alpha = 1.42$ and $\beta = 0$.*

In order to compare the method, we compare the Mean square error (MSE), Root mean square error (RMSE) and Mean absolute error (MAE) from the two methods. This is important to notice that MSE and RMSE do not make sense theoretically since the $\alpha-$stable tvAR(1) does not have the second moment. On the other hand, MAE does make sense since the process assumes the first moments finite. Nevertheless, since the time series data is available, we calculated them to compare the two methods. From the Table 6.9, we observe that using blocked Whittle estimation (assuming finite second moment), MSE and RMSE are slightly lower, while using the indirect inference presents lower MAE.

As conclusion, since the residual analysis indicates heavy-tailed and skewed error, it is reasonable to consider the $\alpha-$stable tvAR(1) instead of tvAR(1) with Gaussian innovations. In addition, by assuming $\alpha = 1.9$, which is close to 2 (Gaussian case), the blocked Whittle estimation seems to be well fitted compared to the indirect inference. However, by assuming stable innovations, MSE and RMSE do not make sense. Based on MAE, the indirect inference performs slightly better, and thus, the interpretation of estimated coefficients of the model also changed.

Accuracy	Blocked Whittle estimates	Indirect inference
MSE	0.015704	0.015762
RMSE	0.125315	0.125560
MAE	0.093919	0.093823

Table 6.9: *Goodness of fit of two estimation methods for the tree ring data*

6.5.2 Wind Power

In this subsection, we illustrate an application for the total wind power generated in offshore wind farms in Germany from 16/06/2015 at 00:00 to 27/07/2015 at 24:00 ($T = 1008$ hours), obtained from the EMHIRES (European Meteorological High resolution RES time series) datasets (Gonzales-Aparicio *et al.*, 2016). The reason of selecting just a small segment of the data is due to the fact that the whole time series has more complex structure, such as seasonality, thus a non-parametric approach could be more appropriate. For daily wind power time series, the Gaussian assumption of the innovations seems to be appropriate, but the hourly time series present heavy tails and Gaussian assumption is inadequate as we present below. The Figure 6.13 shows the original time series (y_t) and its first difference (Δy_t). The original time series seems to be non-stationary and difficult to analyze. We focus on the differencing time series, which shows heavy-tailed behavior.

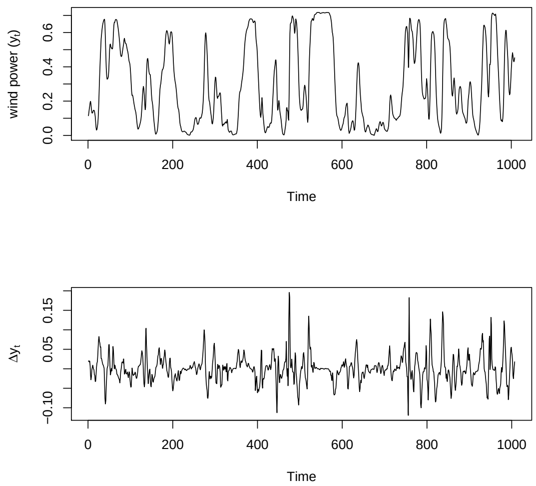
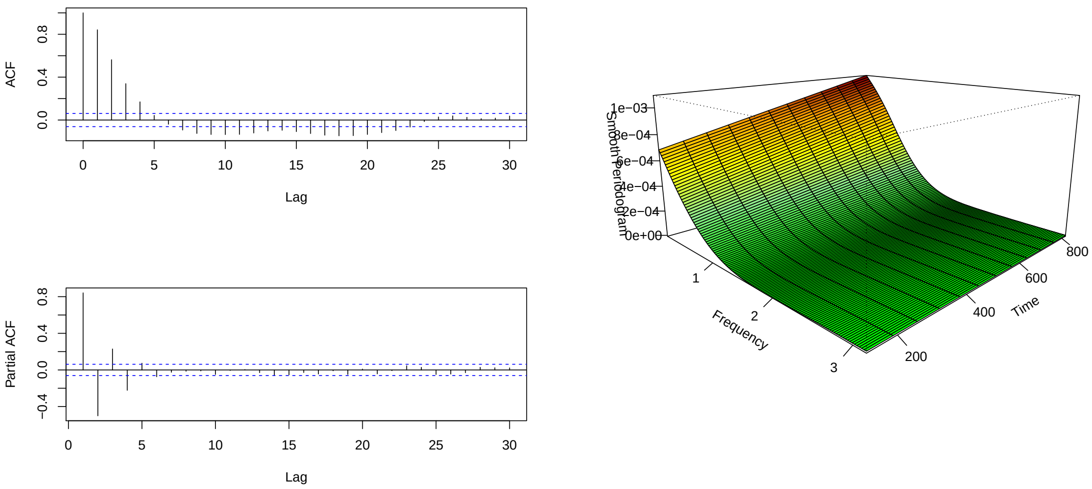


Figure 6.13: *Wind power generated hourly from 16/06/2015 at 00:00 to 27/07/2015 at 24:00.*

Figure 6.14 shows its (global) sample autocorrelation function, and partial autocorrelation function. Traditional models: ARMA(1,1) and AR(4) seem to be appropriate, but the blocked smooth periodogram shows its slowly changed structure over the time.



(a) ACF and partial ACF. (b) Blocked smooth periodogram.

Figure 6.14: *ACF, partial ACF and blocked smooth periodogram of wind power data.*

In order to explore its local structure, we first estimate ARMA(1,1) model for 9 time blocks following the same methodology in the previous subsection (block size $N = 200$ and shifting each block by $Q = 100$ time units). The smoothed estimated coefficients over time is presented in Figure 6.15. Then, we also estimate AR(4) model for 9 time blocks. The smoothed estimated coefficients over time is presented in Figure 6.16. Both cases show that the coefficients are approximately linear throughout the time. Consequently, two models are proposed:

- tvARMA(1,1) model with linear coefficients, $\alpha_1(u) = \theta_{a0} + \theta_{a1}(u)$, $\beta_1(u) = \theta_{b0} + \theta_{b1}(u)$ and $\gamma(u) = \gamma_0 + \gamma_1(u)$.
- tvAR(4) model with linear coefficients, $\alpha_1(u) = \theta_{a0} + \theta_{a1}(u)$, $\alpha_2(u) = \theta_{b0} + \theta_{b1}(u)$, $\alpha_3(u) = \theta_{c0} + \theta_{c1}(u)$, $\alpha_4(u) = \theta_{d0} + \theta_{d1}(u)$ and $\gamma(u) = \gamma_0 + \gamma_1(u)$.

After estimating both models, the residuals of tvARMA(1,1) are correlated, and thus, we focus only on the other model. The residuals of the tvAR(4) model are approximately white noise, and the parameter estimates of this model are reported in Table 6.10.

Figure 6.17 presents the residual analysis for the blocked Whittle estimates (assuming finite variance) and the QQ-plot, box-plot and the histogram show that the distribution of error has heavy tail. Additionally, we also estimated the skewness (0.31) and kurtosis (12.86) and carrying out Shapiro-Wilk test and Jarque-Bera test, both tests rejected the null hypothesis of normality. Moreover, Figure 6.18 presents the variogram of the first difference of the wind data and the residuals from the tvAR(4) model. It is clear to observe that both of the variograms do not converge.

Next, since the residual analysis shows that the residuals present heavy tail, we propose a more flexible model, $\alpha-$stable tvAR(4), by assuming the parameters of symmetric stable

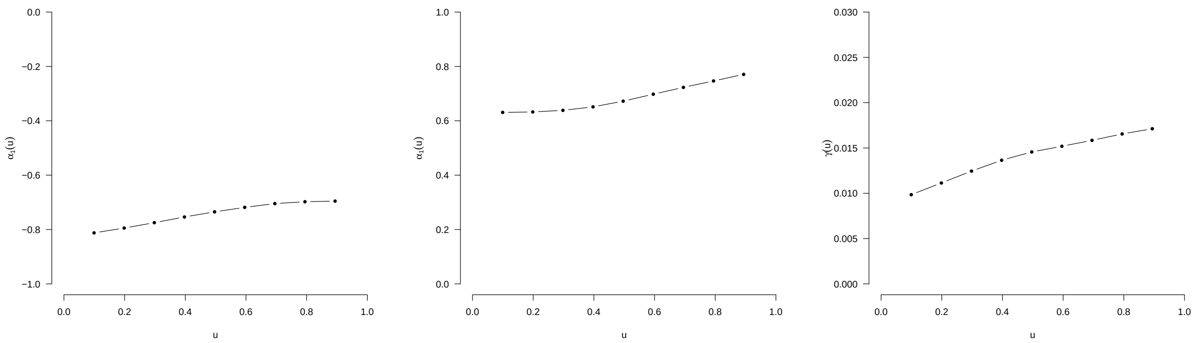


Figure 6.15: *(Smoothed) $\alpha(u)$, $\beta(u)$ and $\gamma(u)$ estimates of stationary ARMA(1,1) model for* 9 *block of size $N = 200$ with $u = t/T$ center point of each block.*

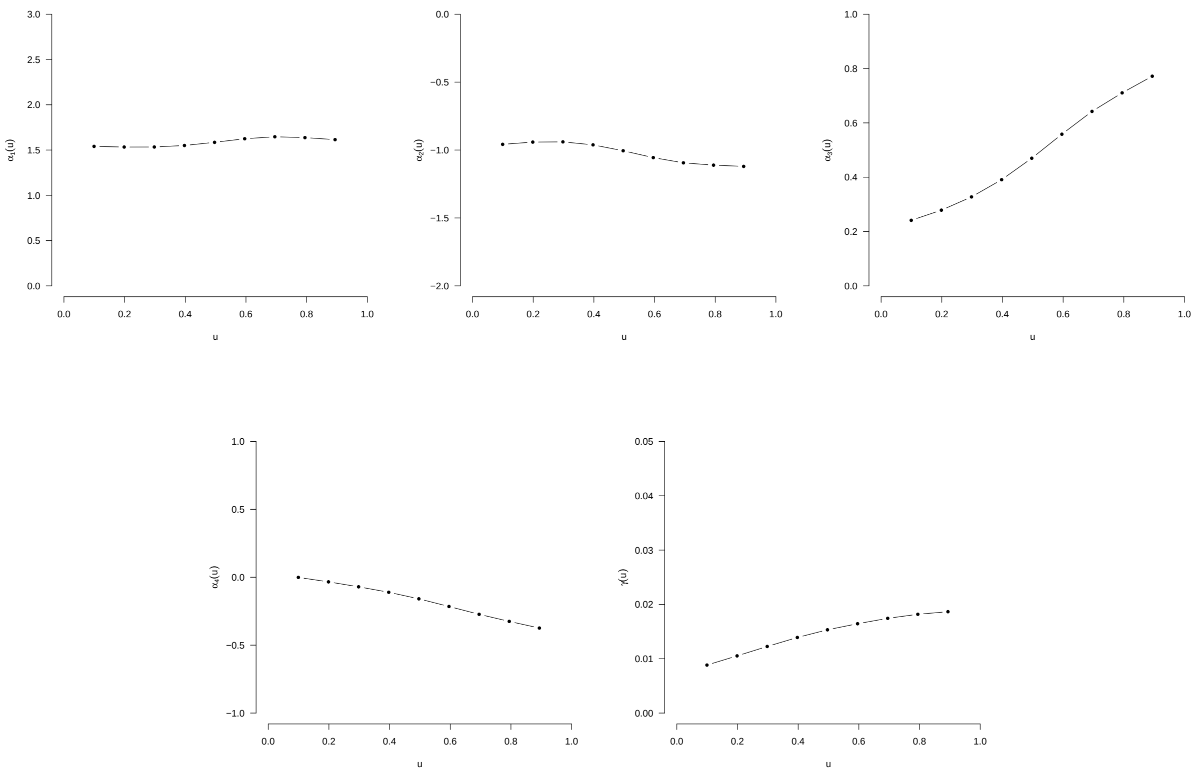
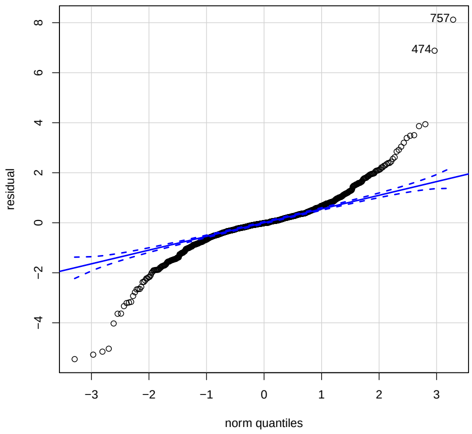
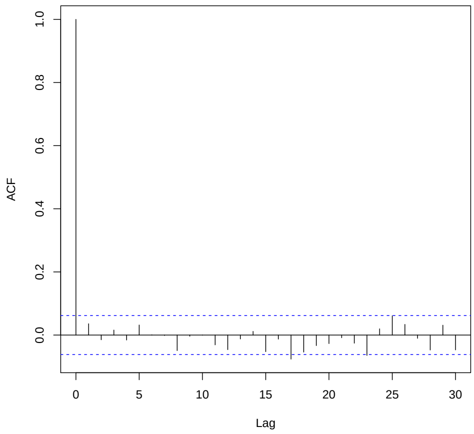


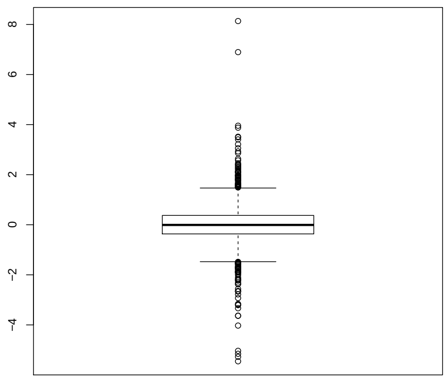
Figure 6.16: *(Smoothed) $\alpha_1(u)$, $\alpha_2(u)$, $\alpha_3(u)$, $\alpha_4(u)$ and $\gamma(u)$ estimates of stationary AR(4) model for* 9 *block of size $N = 200$ with $u = t/T$ center point of each block.*



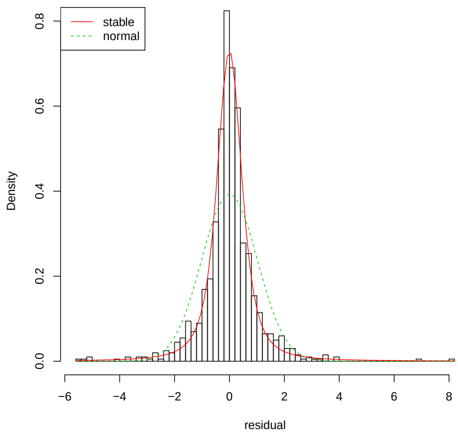
(a) QQ-plot



(b) Autocorreltion function

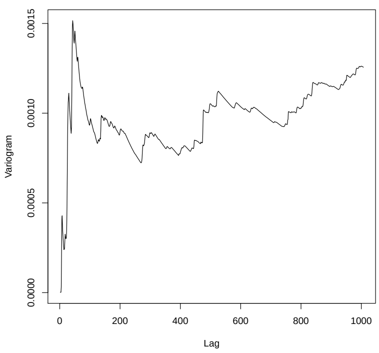


(c) Box-plot

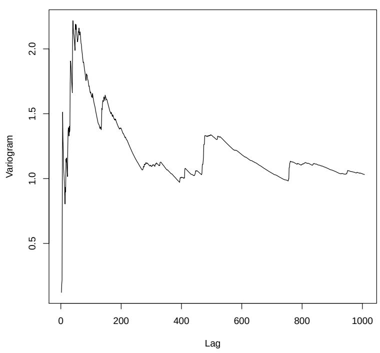


(d) Histogram with estimated stable curve ($\alpha = 1.34$, $\beta = 0$) and Gaussian curve.

Figure 6.17: *Residual analysis using the blocked Whittle estimation (standadized residual).*



(a) Δy_t



(b) Residuals

Figure 6.18: *Variogram of the first differenced wind data Δy_t and the residuals of the tvAR(4) model.*

Parameter	Blocked Whittle estimates			
	Estimate	s.e.	z-value	p-value
θ_{a0}	-1.5985	0.0768	-20.8171	0.0000
θ_{a1}	0.3305	0.1406	2.3508	0.0187
θ_{b0}	0.9135	0.1373	6.6536	0.0000
θ_{b1}	0.0207	0.2447	0.0847	0.9325
θ_{c0}	-0.0585	0.1372	-0.4266	0.6697
θ_{c1}	-0.7153	0.2445	-2.9254	0.0034
θ_{d0}	-0.1316	0.0767	-1.7158	0.0862
θ_{d1}	0.5454	0.1405	3.8831	0.0001
γ_0	0.0077	0.0003	24.5452	0.0000
γ_1	0.0152	0.0007	21.6400	0.0000

Table 6.10: *Blocked Whittle estimates of tvAR(4) from wind power time series.*

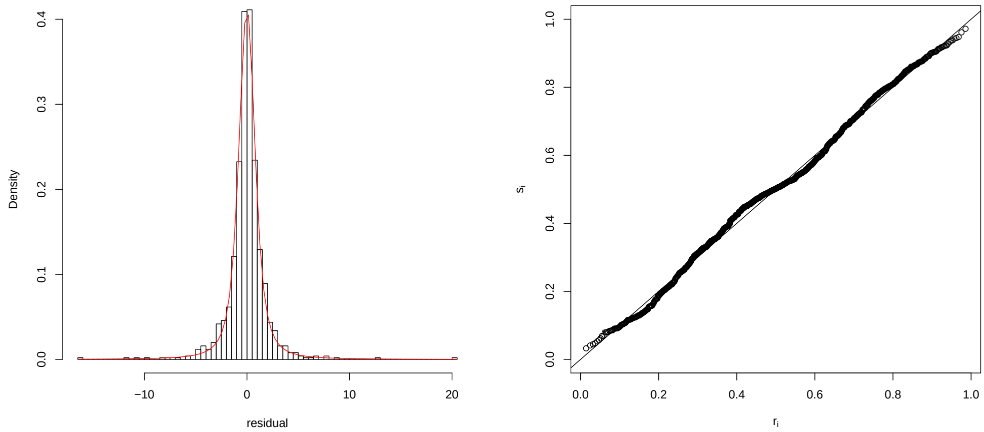
innovations as estimated above, i.e. $\alpha = 1.34$, and use indirect inference with $S = 40$ to estimate the parameter $(\theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \theta_{c0}, \theta_{c1}, \theta_{d0}, \theta_{d1}, \gamma_0, \gamma_1)$. Table 6.11 presents the indirect estimates along with their (Monte Carlo) standard error calculated based on $R = 1000$. Similar to the previous example, the histogram and the stabilized p-p plot in figure 6.19 show that the stable distribution fits well the residuals. Moreover, we notice that $\alpha_1(u)$ and $\alpha_2(u)$ are not time-varying, while $\alpha_3(u)$, $\alpha_4(u)$ and $\gamma(u)$ vary linearly in time.

Parameter	Indirect estimate	Standard error
θ_{a0}	-1.5549	0.0230
θ_{a1}	0.0102	0.0397
θ_{b0}	0.9230	0.0415
θ_{b1}	0.0370	0.0715
θ_{c0}	-0.2754	0.0412
θ_{c1}	-0.2086	0.0736
θ_{d0}	0.0436	0.0215
θ_{d1}	0.1794	0.0389
γ_0	0.0067	0.0006
γ_1	0.0023	0.0011

Table 6.11: *Indirect estimates of $\alpha-$stable tvAR(4) with $S = 40$.*

We compare the Mean square error (MSE), Root mean square error (RMSE) and Mean absolute error (MAE) of tvARMA(1,1) and tvAR(4) using two estimation methods (blocked Whittle estimates and indirect estimates). As before, it is important to notice that MSE and RMSE do not make sense theoretically since the $\alpha-$stable tvAR(4) does not have the second moment. In Table 6.12, we observe that using blocked Whittle estimation (assuming finite second moment), MSE and RMSE are slightly lower, while using the indirect inference presents lower MAE.

Since the residual analysis indicates heavy-tailed, $\alpha-$stable tvAR(4) is a better model to describe the data. In this case, by assuming $\alpha = 1.34$, which is farther from 2 (Gaussian



(a) Histogram with the stable curve (b) Stabilized p-p plot

Figure 6.19: *Residual analysis from the $\alpha-$stable tvAR(4) model assuming that the innovation distribution is stable with $\alpha = 1.34$ and $\beta = 0$.*

case), the simulation done in the previous section shows that the blocked Whittle estimation is not appropriate. Even though MSE and RMSE are lower for blocked Whittle estimation, they are not appropriate for $\alpha-$stable process since they cannot be theoretically handled. Based on MAE, the indirect inference performs slightly better. Moreover, the interpretation of estimated coefficients of the model also changed.

Model	MSE	RMSE	MAE
tvARMA(1,1)	0.000248	0.015739	0.009675
α-stable tvARMA(1,1)	0.000257	0.016028	0.009469
tvAR(4)	0.000242	0.015542	0.009468
α-stable tvAR(4)	0.000256	0.015993	0.009094

Table 6.12: *Goodness of fit of different models for the wind power data.*

Chapter 7

Indirect inference for α-stable tvARMA process with unknown α.

In Section 6.1, we introduced the possibility of using the skew-t distribution and its advantage of having four parameters that are similar to the α-stable distribution and it has the likelihood function available. Lombardi and Calzolari (2008) studied the binding function by simulations in case of independent samples and it seems to behave well. However, the α-stable distributions have the characteristic of the asymmetry parameter β being unidentified when α approximates 2. Hence, in this chapter, we study the parameter estimation of a parametric tvARMA model with α-stable innovations with unknown stability index α but with known β. In this case, the standardized t distribution with unknown ν is used for the auxiliary model. It is important to notice that we report results for $\alpha = 0.8, 0.85$ and 0.9 and for small values of α, the convergence is difficult. We suspect that it is because of highly heavy tail and the auxliary model (student-t) cannot capture well this behavior.

7.1 Indirect inference for the α-stable tvAR(1)

Similar to the Section 6.2, we consider the tvAR(1) model

$$X_{t,T} + \alpha_1\left(\frac{t}{T}\right)X_{t-1,T} = \gamma\left(\frac{t}{T}\right)\varepsilon_t, \tag{7.1}$$

where $\varepsilon_t \sim S_\alpha\left(1/\sqrt{2}, \beta, 0\right)$ with known β.

We illustrate how the indirect inference can be employed to the tvAR(1) in (6.2) with the linear parametric form of the time-varying coefficient $\alpha_1(u) = \theta_0 + \theta_1 u$, and $\gamma(u) = \gamma_0 + \gamma_1 u$. Therefore, the parameter vector of the model of interest is $\theta = (\theta_0, \theta_1, \alpha, \gamma_0, \gamma_1)$. On the other hand, we use the same parametric form of the process with the scaled t-distribution with unknown ν as the auxiliary model with the likelihood function defined in (2.33), that is, $\lambda = (\theta_0^{(A)}, \theta_1^{(A)}, \nu, \gamma_0^{(A)}, \gamma_1^{(A)})$. We report two scenarios assuming $\alpha = 0.8$ and 1.4 letting other parameters the same $(\beta, \theta_0, \theta_1, \gamma_0, \gamma_1) = (0, 0.35, -0.6, 0.5, 0.1)$. We carried out simulations for $T = 500$, 1000 and 1500 observations based on $R = 1000$ independent replications each scenario. The indirect inference was carried out using $S = 100$.

Table 7.1 reports the Monte Carlo mean and standard error of the estimates from both model of interest and auxiliary model. We notice that the Monte Carlo mean from the indirect estimates seems to be consistent, that is, they approximate to the real parameters and present lower standard errors as T increases. Table 7.2 presents the kurtosis and skewness of indirect estimates. For both scenarios, indirect estimates related to autoregressive part (θ_0 and θ_1) have high kurtosis and it is specially noticeable for $\alpha = 0.8$. In general, all other indirect estimates behave similar to the case when α is known.

We also compare two scenarios all parameters are set the same except α. We notice that θ_0 and θ_1 are highly affected by the value of α, that is, they have less standard error and high kurtosis.

α	T	Indirect estimates									
		Model of Interest					Auxiliary model				
		θ_0	θ_1	α	γ_0	γ_1	$\theta_0^{(A)}$	$\theta_1^{(A)}$	ν	$\gamma_0^{(A)}$	$\gamma_1^{(A)}$
0.8	500	0.3489	-0.5988	0.8002	0.4898	0.1156	0.3489	-0.5988	0.7225	0.3024	0.0708
		(0.0064)	(0.0111)	(0.0383)	(0.0799)	(0.1436)	(0.0064)	(0.0111)	(0.0479)	(0.0517)	(0.0886)
	1000	0.3495	-0.5996	0.8000	0.4947	0.1085	0.3495	-0.5996	0.7220	0.3056	0.0669
		(0.0026)	(0.0048)	(0.0278)	(0.0563)	(0.1042)	(0.0026)	(0.0048)	(0.0348)	(0.0354)	(0.0644)
	1500	0.3496	-0.5996	0.8013	0.4941	0.1087	0.3496	-0.5996	0.7232	0.3058	0.0669
		(0.0016)	(0.0027)	(0.0232)	(0.0472)	(0.0831)	(0.0016)	(0.0027)	(0.0293)	(0.0300)	(0.0514)
1.4	500	0.3482	-0.5980	1.4083	0.4922	0.1111	0.3482	-0.5980	1.8853	0.3994	0.0897
		(0.0406)	(0.0715)	(0.0737)	(0.0527)	(0.0960)	(0.0407)	(0.0716)	(0.2351)	(0.0446)	(0.0778)
	1000	0.3492	-0.5986	1.4037	0.4974	0.1033	0.3492	-0.5986	1.8622	0.4033	0.0834
		(0.0244)	(0.0430)	(0.0520)	(0.0370)	(0.0661)	(0.0244)	(0.0429)	(0.1570)	(0.0311)	(0.0533)
	1500	0.3498	-0.5988	1.4000	0.4976	0.1011	0.3499	-0.5988	1.8478	0.4030	0.0818
		(0.0187)	(0.0323)	(0.0417)	(0.0305)	(0.0546)	(0.0187)	(0.0323)	(0.1244)	(0.0255)	(0.0441)

Table 7.1: *Monte Carlo mean and standard error (in parentheses) for different sample sizes ($T = 500, 1000, 1500$) using indirect estimators (model of interest and auxiliary model) assuming $\alpha = 0.8, 1.4$, and $\beta = 0$, $\theta_0 = 0.35$, $\theta_1 = -0.6$, $\gamma_0 = 0.5$, $\gamma_1 = 0.1$ with known β from α-stable tvAR(1) based on $R = 1000$ replications.*

α	T		Indirect estimates				
			θ_0	θ_1	α	γ_0	γ_1
0.8	500	kur	26.8926	17.5308	2.8292	2.9540	3.0422
		skw	-1.1025	0.4991	0.2203	0.2658	0.0305
	1000	kur	50.1470	71.1822	3.3781	3.2657	3.1101
		skw	3.9116	-4.4294	0.0729	0.2901	-0.0323
	1500	kur	19.0859	16.3044	6.8858	2.9669	2.9937
		skw	1.1142	-0.7635	0.5434	0.1051	0.0270
1.4	500	kur	3.7767	3.6800	3.1078	2.8924	3.0343
		skw	-0.1369	0.1213	0.2730	0.1583	0.0156
	1000	kur	4.7209	3.8513	2.7680	3.1397	3.0710
		skw	-0.1548	0.1008	0.0889	0.0672	-0.0654
	1500	kur	4.2029	3.8664	2.7385	3.0881	3.0266
		skw	0.1274	-0.0192	0.0967	0.0973	0.0108

Table 7.2: *Kurtosis and skewness of indirect estimates for different sample sizes ($T = 500, 1000, 1500$) assuming $\alpha = 0.8, 1.4$ and $\beta = 0$, $\theta_0 = 0.35$, $\theta_1 = -0.6$, $\gamma_0 = 0.5$, $\gamma_1 = 0.1$ with known β from α-stable tvAR(1) based on $R = 1000$ replications.*

Finally, Figures 7.1 and 7.2 show the density estimates of each parameter. Similarly, $\alpha = 0.8$ scenario shows high kurtosis for autoregressive part and other indirect estimates seem to behave well. In general, the density estimates show that the standard error become smaller as T increases. Along with the results from Tables 7.1 and 7.2, we can conclude that the distribution of indirect estimates is leptokurtic and suggest consistency.

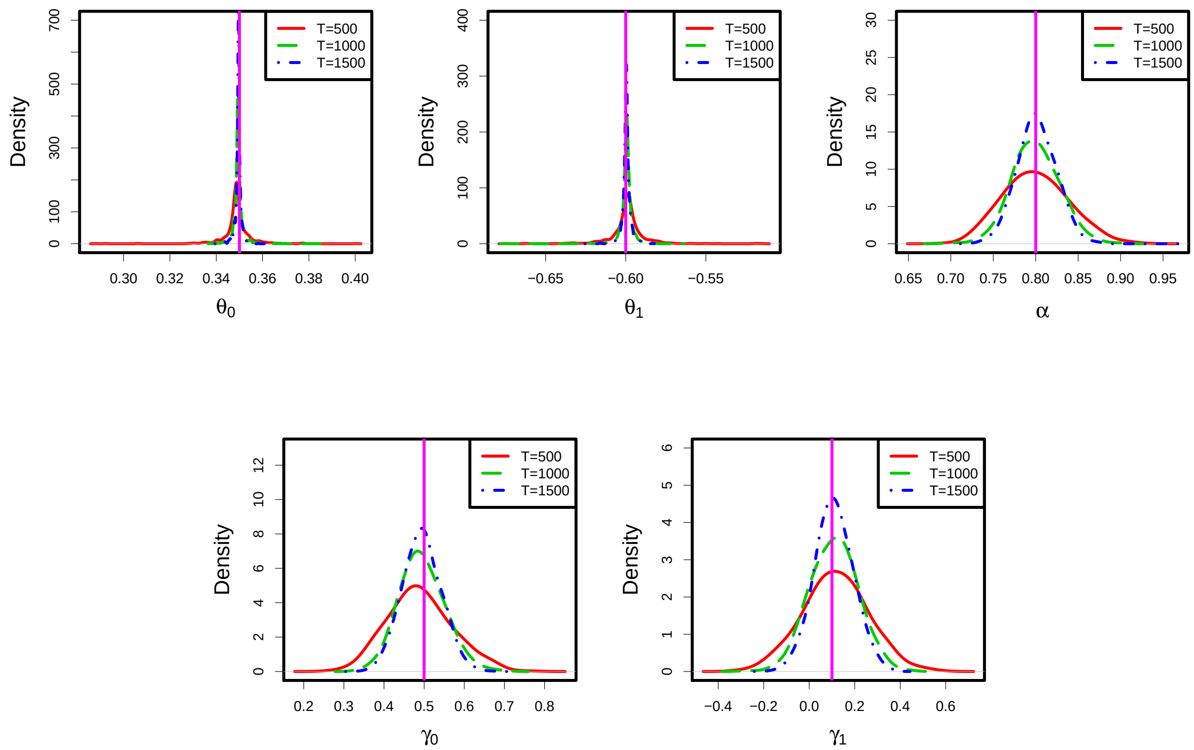


Figure 7.1: *Density estimates of* θ_0*,* θ_1*,* α*,* γ_0 *and* γ_1 *for different sample sizes based on* $R = 1000$ *replications from* α*-stable tvAR(1) with* $\alpha = 0.8,\ \beta = 0,\ \theta_0 = 0.35,\ \theta_1 = -0.6,\ \gamma_0 = 0.5,\ \gamma_1 = 0.1$ *using indirect inference.*

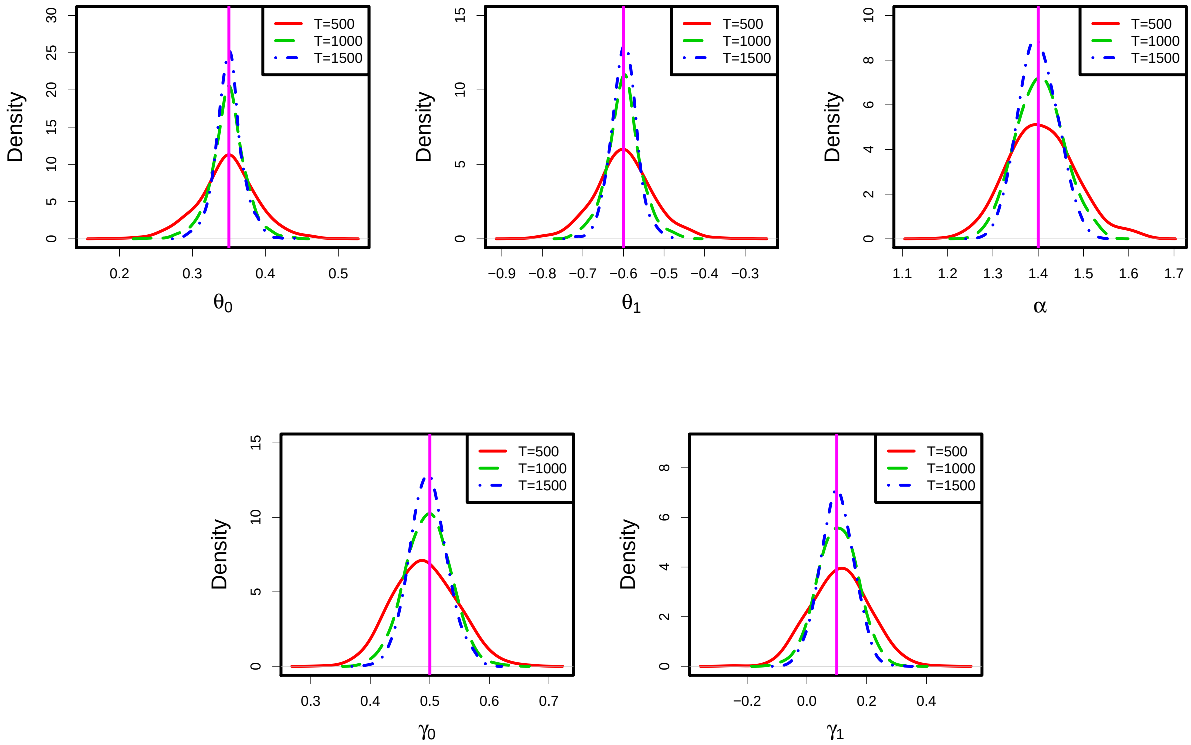


Figure 7.2: *Density estimates of* θ_0*,* θ_1*,* α*,* γ_0 *and* γ_1 *for different sample sizes based on* $R = 1000$ *replications from* α*-stable tvAR(1) with* $\alpha = 1.4, \beta = 0, \theta_0 = 0.35, \theta_1 = -0.6, \gamma_0 = 0.5, \gamma_1 = 0.1$ *using indirect inference.*

7.2 Indirect inference for the α-stable tvMA(1)

In this section, we illustrate the indirect inference for the model (6.4) with unknown α. This implies that the parameter of model of interest is $\theta = (\theta_0, \theta_1, \alpha, \gamma)$ and the parameter of auxiliary model is $\lambda = (\theta_0^{(A)}, \theta_1^{(A)}, \nu, \gamma^{(A)})$. The simulation was performed by assuming two scenarios with $\alpha = 0.85, 1.75$ and $(\beta, \theta_0, \theta_1, \gamma) = (0.2, -0.35, 0.4, 0.7)$ for three different period of time ($T = 500$, 1000 and 1500) based on $R = 1000$ independent replications. The indirect inference was carried out using $S = 100$.

Similarly to the tvAR(1) case, the Monte Carlo mean and standard error of the estimates from both model of interest and auxiliary model are reported in Table 7.3, and kurtosis and skewness are presented in Table 7.4. Along with the density estimates showed in Figures 7.3 and 7.4, the indirect estimates have similar behavior than the tvAR(1) case. They all seem to be consistent with these sample path length.

Comparing two scenarios that have all parameters set the same except α. We notice again that θ_0 and θ_1 are highly affected by the value of α, that is, they have less standard error and high kurtosis. However, they are less noticeable than the tvAR(1) case.

One interesting result is that for the scenario 2, while $\alpha < 2$ implies the process has infinite variance, the auxiliary model was estimated with $\nu > 3$, that is, it has finite variance.

α	T	Indirect estimates							
		Model of Interest				Auxiliary model			
		θ_0	θ_1	α	γ	$\theta_0^{(A)}$	$\theta_1^{(A)}$	ν	$\gamma^{(A)}$
0.85	500	-0.3450	0.3975	0.8499	0.7007	-0.3450	0.3975	0.9550	0.4631
		(0.0126)	0.0220)	(0.0430)	(0.0634)	(0.0125)	(0.0220)	(0.1185)	(0.0628)
	1000	-0.3473	0.3988	0.8498	0.7024	-0.3473	0.3988	0.9555	0.4676
		(0.0065)	0.0107)	(0.0192)	(0.0588)	(0.0065)	(0.0107)	(0.0826)	(0.0451)
	1500	-0.3486	0.3995	0.8522	0.7056	-0.3486	0.3995	0.9568	0.4712
		(0.0037)	0.0065)	(0.0385)	(0.0409)	(0.0037)	(0.0065)	(0.1002)	(0.0426)
1.75	500	-0.3518	0.4016	1.7566	0.7008	-0.3518	0.4016	3.9795	0.3810
		(0.0699)	(0.1245)	(0.0739)	(0.0296)	(0.0694)	(0.1237)	(1.0183)	(0.0390)
	1000	-0.3487	0.3987	1.7527	0.6999	-0.3486	0.3987	3.8307	0.3776
		(0.0446)	(0.0787)	(0.0559)	(0.0229)	(0.0445)	(0.0788)	(0.6414)	(0.0299)
	1500	-0.3504	0.4009	1.7525	0.7003	-0.3502	0.4007	3.7874	0.3785
		(0.0375)	(0.0663)	(0.0457)	(0.0187)	(0.0373)	(0.0661)	(0.4852)	(0.0242)

Table 7.3: *Monte Carlo mean and standard error (in parentheses) for different sample sizes ($T = 500, 1000, 1500$) using indirect estimators (model of interest and auxiliary model) assuming $\alpha = 0.85, 1.75$ and $\beta = 0.2$, $\theta_0 = -0.35$, $\theta_1 = 0.4$, $\gamma = 0.7$ with known β from α-stable tvMA(1) based on $R = 1000$ replications.*

7.3 Indirect inference for the α-stable tvARMA(1,1)

Finally, as in Section 6.4 we study the indirect inference with simulations for the case of tvARMA(1,1) in (6.5), but α is assumed to be unknown. The time-varying coefficients are assumed to be linear, that is, $\alpha_1(u) = \theta_{a0} + \theta_{a1}u$ and $\beta_1(u) = \theta_{b0} + \theta_{b1}u$, and we consider that $\varepsilon_t \sim S_\alpha(1/\sqrt{2}, \beta, 0)$ for known β. Therefore, the parameter vector of the model of interest is $\theta = (\theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \alpha, \gamma)$, while the auxiliary model has the parameter $\lambda = \left(\theta_{a0}^{(A)}, \theta_{a1}^{(A)}, \theta_{b0}^{(A)}, \theta_{b1}^{(A)}, \nu, \gamma^{(A)}\right)$.

α	T		Indirect estimates			
			θ_0	θ_1	α	γ_0
0.85	500	kur	13.1427	11.6976	266.5685	8.5182
		skw	2.1758	-0.2905	14.0235	1.0395
	1000	kur	16.6546	15.9780	150.5931	212.2336
		skw	2.6774	0.2625	-7.7802	9.8820
	1500	kur	17.9085	12.2223	395.1609	36.4959
		skw	2.7648	-0.1490	18.8091	3.0698
1.75	500	kur	3.8667	3.2718	2.8731	3.3445
		skw	-0.0413	-0.0030	-0.1140	0.0003
	1000	kur	3.7260	3.4565	2.8049	2.9412
		skw	0.0436	0.0582	-0.0081	0.1446
	1500	kur	3.6876	3.4211	3.0489	3.0002
		skw	-0.0043	0.0187	-0.2133	0.0557

Table 7.4: *Kurtosis and skewness of indirect estimates for different sample sizes* ($T = 500, 1000, 1500$) *assuming* $\alpha = 0.85, 1.75$ *and* $\beta = 0.2,\ \theta_0 = -0.35,\ \theta_1 = 0.4,\ \gamma = 0.7$ *with known* β *from* α*-stable tvMA(1) based on* $R = 1000$ *replications.*

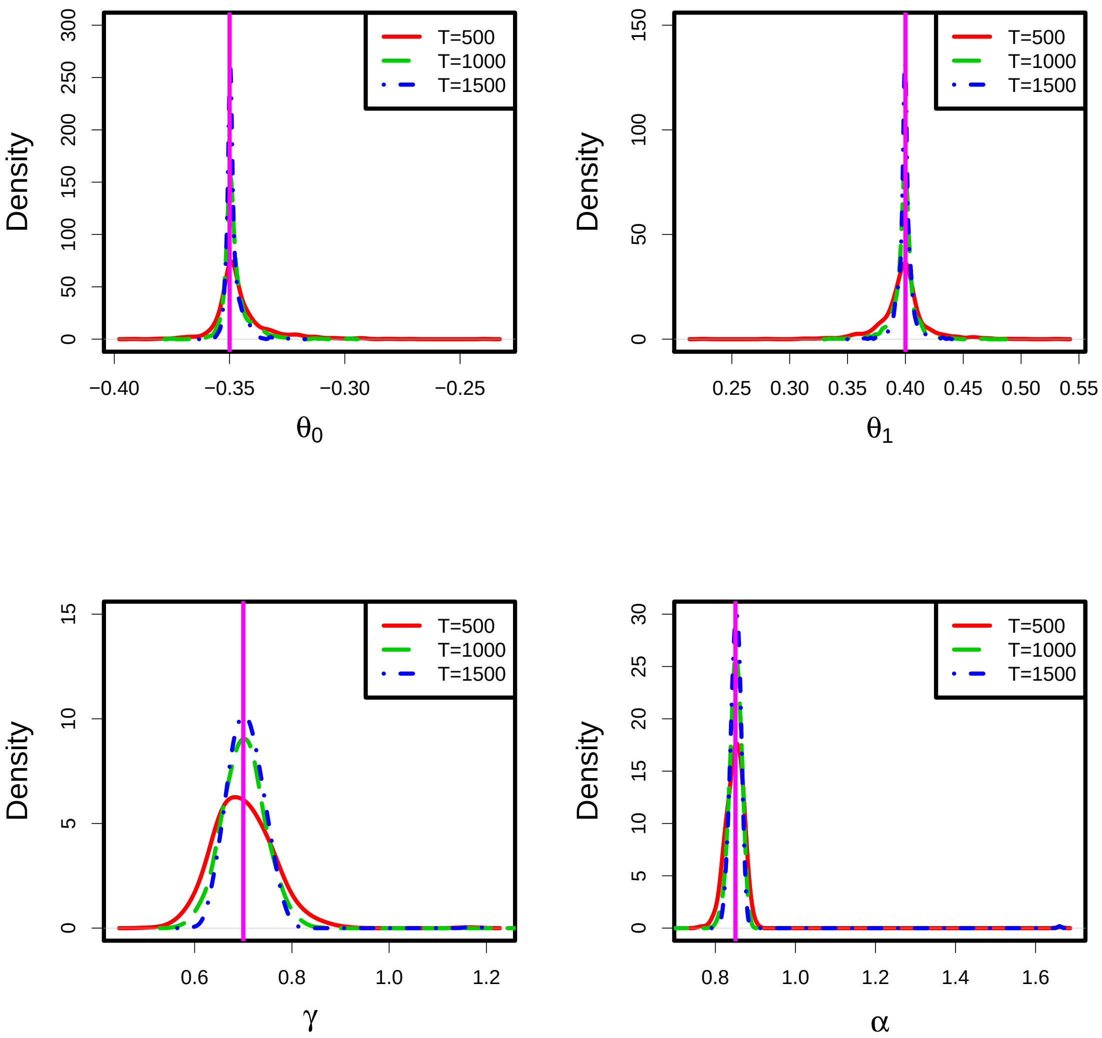


Figure 7.3: *Density estimates of* θ_0*,* θ_1*,* α *and* γ *for different sample sizes based on* $R = 1000$ *replications from* α*-stable tvMA(1) with* $\alpha = 0.85,\ \beta = 0.2,\ \theta_0 = -0.35,\ \theta_1 = 0.4,\ \gamma = 0.7$ *using indirect inference.*

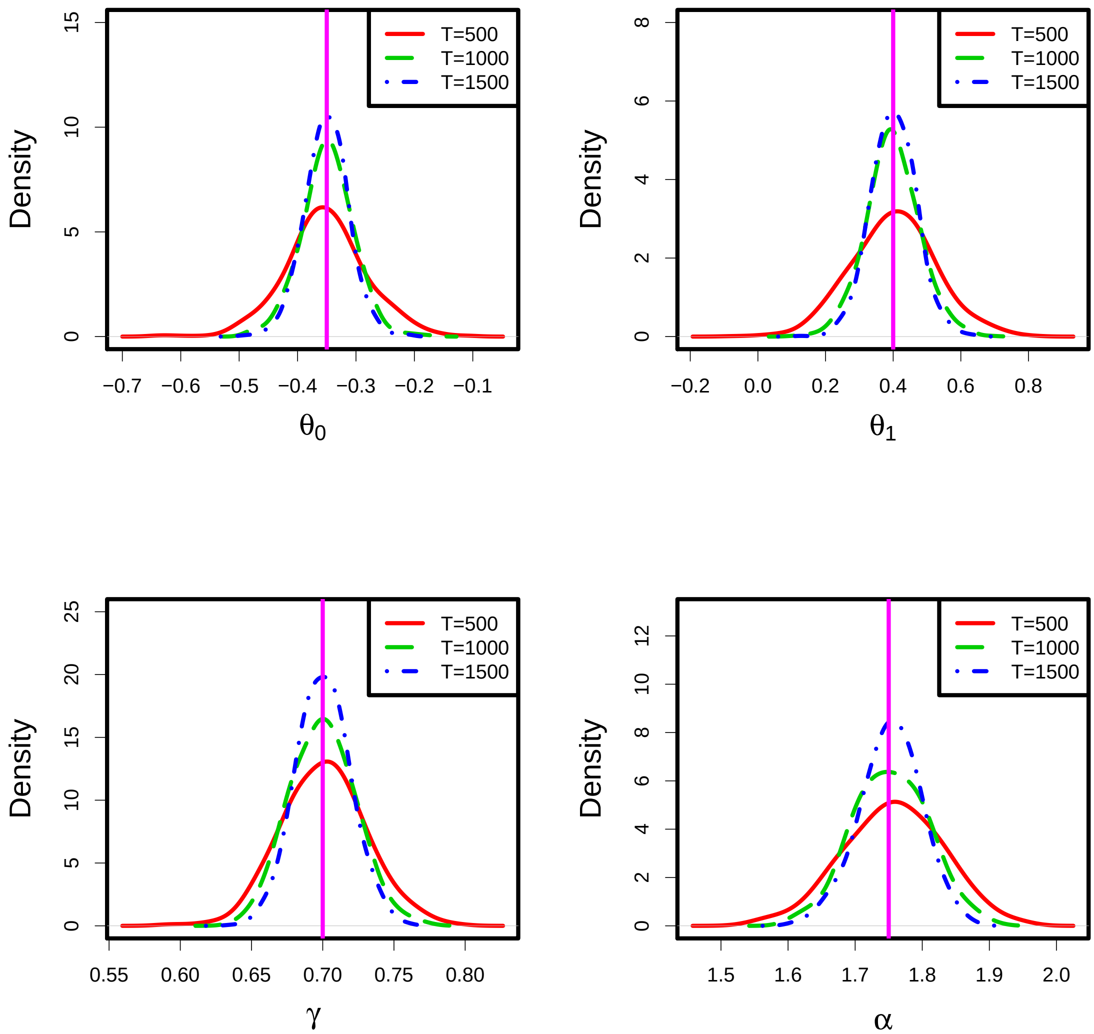


Figure 7.4: *Density estimates of* θ_0, θ_1, α *and* γ *for different sample sizes based on* $R = 1000$ *replications from* α*-stable tvMA(1) with* $\alpha = 1.75$, $\beta = 0.2$, $\theta_0 = -0.35$, $\theta_1 = 0.4$, $\gamma = 0.7$ *using indirect inference.*

The simulation was performed by assuming the same values as in the first scenario in Section 6.4, that is, $(\alpha, \beta, \theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \alpha, \gamma) = (1.3, 0, -0.2, -0.4, 0.2, 0.3, 1.1)$ for three different period of time ($T = 500$, 1000 and 1500) based on $R = 1000$ independent replications. The indirect inference was carried out using $S = 100$. We also add an extra scenario with the same parameter except $\alpha = 0.9$.

The Monte Carlo mean and standard error of the estimates from both model of interest and auxiliary model are reported in Table 7.5, and kurtosis and skewness are presented in Table 7.6. Along with density estimates showed in Figures 7.5 and 7.6, they present similar results than tvAR(1) and tvMA(1), i.e. they seem to be consistent. For lower α, $\theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}$ present high kurtosis and but lower standard error.

Finally, if we compare with simulation results from the case when α is known, they present similar standard error, kurtosis and asymmetry.

α	T	Model of Interest (indirect estimates)					
		θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	α	γ
0.9	500	-0.2000 (0.0159)	-0.4002 (0.0248)	0.2007 (0.0176)	0.2991 (0.0275)	0.9005 (0.0456)	1.0974 (0.0806)
	1000	-0.2002 (0.0080)	-0.4000 (0.0123)	0.2004 (0.0083)	0.2995 (0.0129)	0.9005 (0.0337)	1.0974 (0.0588)
	1500	-0.2003 (0.0041)	-0.3998 (0.0065)	0.2002 (0.0043)	0.2997 (0.0069)	0.9019 (0.0283)	1.0982 (0.0469)
1.3	500	-0.2036 (0.0585)	-0.3932 (0.0869)	0.1971 (0.0587)	0.3064 (0.0891)	1.3018 (0.0698)	1.0923 (0.0587)
	1000	-0.2005 (0.0319)	-0.3986 (0.0489)	0.2003 (0.0329)	0.3004 (0.0504)	1.3045 (0.0471)	1.0976 (0.0433)
	1500	-0.1998 (0.0233)	-0.3999 (0.0359)	0.2012 (0.0250)	0.2983 (0.0374)	1.2998 (0.0390)	1.0953 (0.0347)
		Auxiliary model					
		$\theta_{a0}^{(A)}$	$\theta_{a1}^{(A)}$	$\theta_{b0}^{(A)}$	$\theta_{b1}^{(A)}$	ν	$\gamma^{(A)}$
0.9	500	-0.2001 (0.0158)	-0.4001 (0.0248)	0.2007 (0.0175)	0.2991 (0.0274)	0.8548 (0.0635)	0.5384 (0.0906)
	1000	-0.2002 (0.0080)	-0.4000 (0.0124)	0.2005 (0.0082)	0.2995 (0.0128)	0.8546 (0.0452)	0.5380 (0.0630)
	1500	-0.2003 (0.0041)	-0.3998 (0.0065)	0.2002 (0.0043)	0.2997 (0.0069)	0.8558 (0.0375)	0.5386 (0.0516)
1.3	500	-0.2036 (0.0584)	-0.3936 (0.0864)	0.1971 (0.0586)	0.3062 (0.0889)	1.5904 (0.1731)	0.7465 (0.0940)
	1000	-0.2006 (0.0319)	-0.3986 (0.0483)	0.2003 (0.0329)	0.3006 (0.0505)	1.5917 (0.1160)	0.7542 (0.0668)
	1500	-0.1998 (0.0232)	-0.4009 (0.0344)	0.2012 (0.0250)	0.2983 (0.0374)	1.5772 (0.0947)	0.7487 (0.0540)

Table 7.5: *Monte Carlo mean and standard error (in parentheses) for different sample sizes ($T = 500, 1000, 1500$) using indirect estimators (model of interest and auxiliary model) assuming $\alpha = 0.9, 1.3$ and $\beta = 0$, $\theta_{a0} = -0.2$, $\theta_{a1} = -0.4$, $\theta_{b0} = 0.2$, $\theta_{b1} = 0.3$, $\gamma = 1.1$ with known β from α-stable tvARMA(1,1) based on $R = 1000$ replications.*

α	T		Indirect estimates					
			θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	α	γ
0.9	500	kur	7.6335	11.2571	26.5658	21.7708	3.8779	3.4606
		skw	0.5099	-0.1177	-1.2743	-0.1933	0.3693	0.3314
	1000	kur	24.0671	18.3498	24.5761	16.6943	3.2849	3.2163
		skw	1.2605	-0.4986	1.8389	-0.9324	0.1080	0.0874
	1500	kur	6.3449	7.1998	6.3046	7.3008	3.0136	3.2006
		skw	-0.0312	-0.0450	0.1387	-0.0387	0.2148	0.1162
1.3	500	kur	5.2893	4.5345	6.0807	5.5860	3.0705	3.3815
		skw	0.2593	-0.1945	-0.1951	0.1297	0.1406	0.2709
	1000	kur	4.6288	4.1073	4.5984	4.1306	3.4796	2.9077
		skw	-0.1406	0.1144	0.0360	-0.0328	0.1167	-0.0713
	1500	kur	4.9301	4.0790	5.1471	4.5091	3.1301	3.0701
		skw	0.0236	-0.1964	0.0964	-0.2378	0.1004	0.0833

Table 7.6: *Kurtosis and skewness of indirect estimates for different sample sizes ($T = 500, 1000, 1500$) assuming $\alpha = 0.9, 1.3$ and $\beta = 0$, $\theta_{a0} = -0.2$, $\theta_{a1} = -0.4$, $\theta_{b0} = 0.2$, $\theta_{b1} = 0.3$, $\gamma = 1.1$ with known β from α-stable tvARMA(1,1) based on $R = 1000$ replications.*

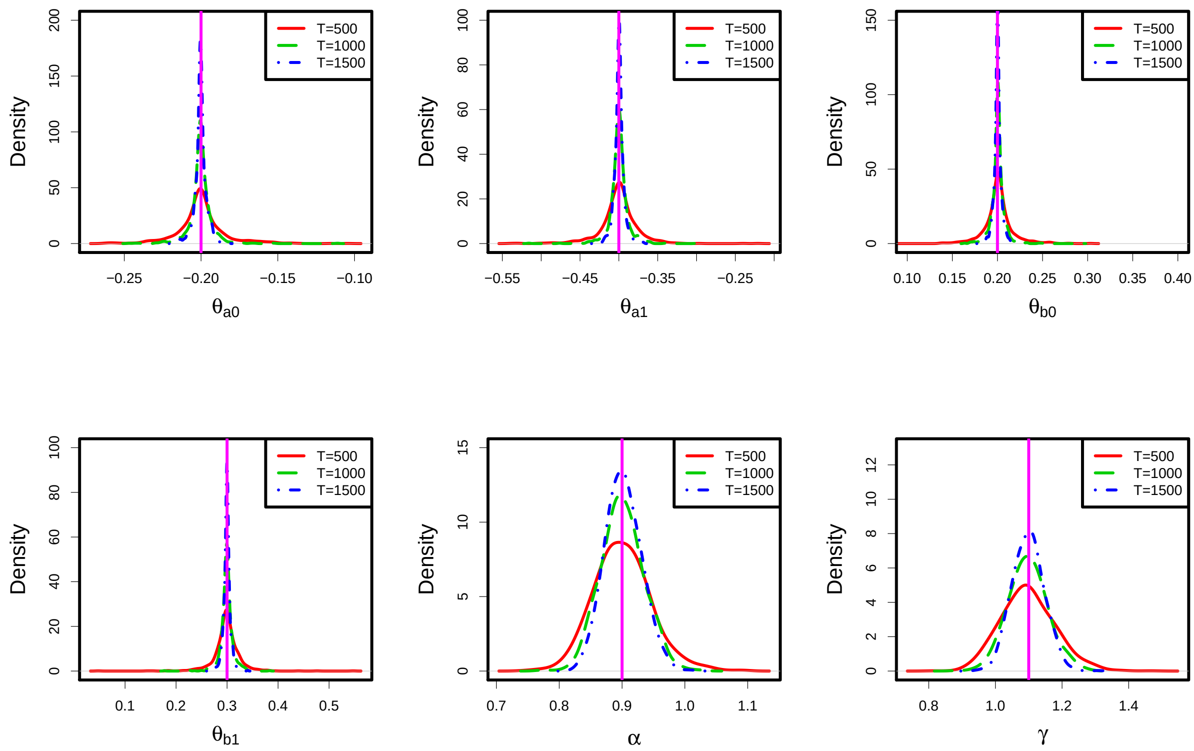


Figure 7.5: *Density estimates of θ_{a0}, θ_{a1}, θ_{b0}, θ_{b1}, α and γ for different sample sizes based on $R = 1000$ replications from α-stable tvARMA(1,1) with $\alpha = 0.9$, $\beta = 0$, $\theta_{a0} = -0.2$, $\theta_{a1} = -0.4$, $\theta_{b0} = 0.2$, $\theta_{b1} = 0.3$, $\gamma = 1.1$ using indirect inference.*

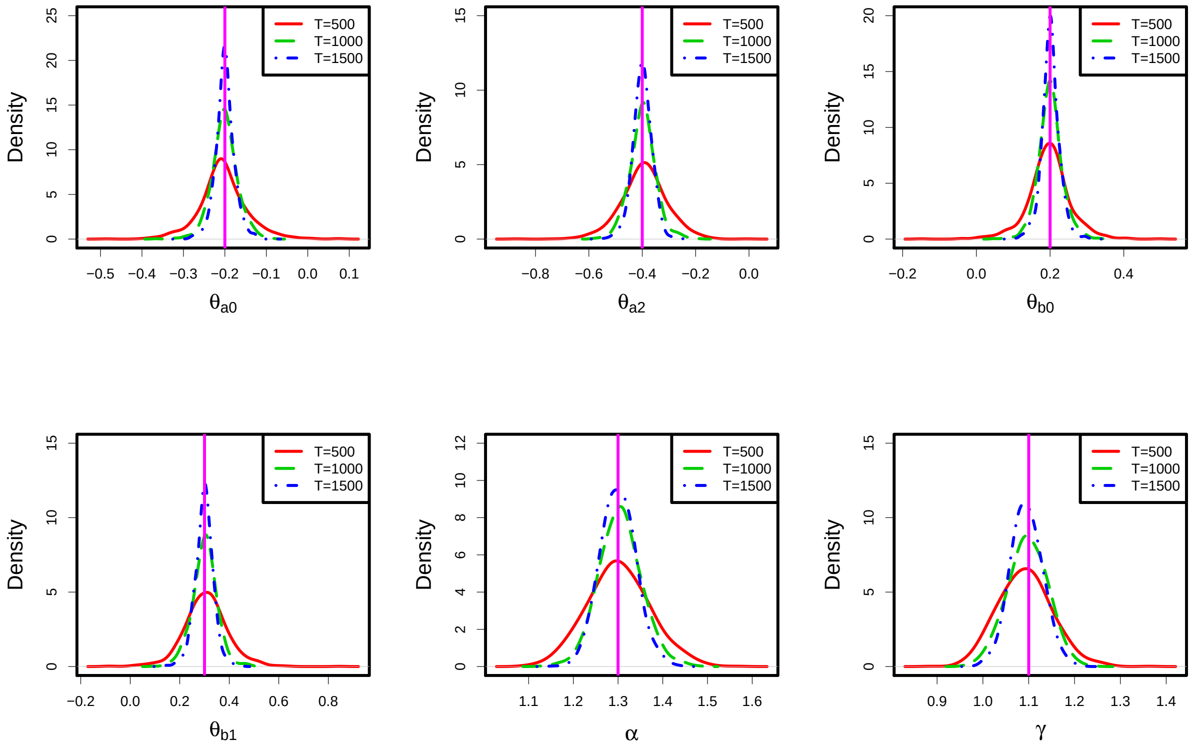


Figure 7.6: *Density estimates of θ_{a0}, θ_{a1}, θ_{b0}, θ_{b1}, α and γ for different sample sizes based on $R = 1000$ replications from α-stable tvARMA(1,1) with $\alpha = 1.3$, $\beta = 0$, $\theta_{a0} = -0.2$, $\theta_{a1} = -0.4$, $\theta_{b0} = 0.2$, $\theta_{b1} = 0.3$, $\gamma = 1.1$ using indirect inference.*

7.4 Application

In this section, we illustrate this methodology with the same time series presented in Section 6.5. For the tree ring data presented in Section 6.5.1, we use indirect inference to estimate the parameter $(\theta_0, \theta_1, \gamma_0, \alpha, \gamma_1)$ of the same structure of the tvAR(1) but letting α unknown. We obtained two different indirect estimates assuming $\beta = 0.98$ and $\beta = 0$. However, we report the case assuming $\beta = 0$ since β becomes irrelevant when α is close to 2 and both estimates are very similar. In Table 7.7, indirect estimates assuming $\beta = 0$ are reported along with their Monte Carlo standard error estimated with $R = 100$ replications. Similar results are obtained as in known α case.

Parameter	θ_0	θ_1	α	γ_0	γ_1
Indirect estimates	-0.6364	-0.0851	1.9208	0.1241	-0.0228
Standard error	(0.0507)	(0.0859)	(0.0391)	(0.0074)	(0.0119)

Table 7.7: *Indirect estimates and Monte Carlo standard error of $\alpha-$stable tvAR(1) with $S = 100$ from tree ring data.*

On the other hand, for the wind data (Section 6.5.2) we use indirect inference to estimate the parameter $(\theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \theta_{c0}, \theta_{c1}, \theta_{d0}, \theta_{d1}, \alpha, \gamma_0, \gamma_1)$ of the same structure of the tvAR(4) but letting α unknown. The indirect inference was done by assuming symmetric α-stable innovations, that is, $\beta = 0$. In Table 7.8, indirect estimates are reported along with their Monte Carlo standard error estimated with $R = 1000$ replications. Similar results are obtained as in known α case, i.e. $\alpha_1(u)$ and $\alpha_2(u)$ are constant, while $\alpha_3(u)$, $\alpha_4(u)$ and $\gamma(u)$ vary linearly in time.

Parameter	Indirect estimate	Standard error
θ_{a0}	-1.5434	0.0251
θ_{a1}	-0.0316	0.0426
θ_{b0}	0.9036	0.0442
θ_{b1}	0.1083	0.0764
θ_{c0}	-0.2818	0.0437
θ_{c1}	-0.2235	0.0752
θ_{d0}	0.0639	0.0246
θ_{d1}	0.1496	0.0412
α	1.3875	0.0528
γ_0	0.0065	0.0005
γ_1	0.0033	0.0010

Table 7.8: *Indirect estimates of $\alpha-$stable tvAR(4) with $S = 40$ from wind data.*

In both applications, we can observe that the indirect estimates and their associated MC standard error are very similar with the case assuming known α. Moreover, the indirect estimation of α is very close to the case when α is estimated using residuals from the Whittle blocked estimation (assuming finite variance). We also did the residual analysis of this case, but they are not presented here because the histogram and the stabilized p-p plot are almost the same as the case when α is known. Moreover, the MAE indicates that the α-stable assumption is better.

Chapter 8

tvARMA process with tempered stable innovations

In the previous chapters, we studied locally stationary processes with stable innovations and implemented the indirect estimation to this type of processes. This allows the possibility of modeling time series that present locally stationarity with infinite variance behavior. Although stable distribution presents attractive theoretical properties, such as the extremely heavy tails and stability under linear combinations, the fact that moments of order greater than two do not exist is a restrictive assumption in real-world applications. In this Chapter, we study the tvARMA process with tempered stable innovations.

8.1 tvARMA process with tempered stable innovations

Similarly to the $\alpha-$stable tvARMA presented in Chapter 5, we study the tvARMA process from (2.18) with i.i.d. tempered stable innovations. The system of difference equations is defined by

$$\sum_{j=0}^{p} \alpha_j \left(\frac{t}{T}\right) X_{t-j,T} = \sum_{k=0}^{q} \beta_k \left(\frac{t}{T}\right) \gamma \left(\frac{t-k}{T}\right) \varepsilon_{t-k}. \tag{8.1}$$

Here, we will assume that ε_t are i.i.d. and $\varepsilon_t \sim stdCTS(\alpha, \lambda_+, \lambda_-)$ with $\alpha \in (0, 2)$ and $\lambda_+, \lambda_- > 0$. Moreover, the regularity conditions in proposition 2.1 are assumed. In this way, there exists a solution of the form

$$X_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j)\varepsilon_{t-j}, \tag{8.2}$$

which fullfills (2.6), (2.7) and (2.8) of Assumption 2.1.

The reason we implement standardized classical tempered stable innovations is that all moments are finite and specifically it has zero mean and unit variance.

8.2 Two-step estimation

In the case of $\alpha-$stable tvARMA, the absence of second moment causes the difficulty of estimating parameters using blocked Whittle estimators. In addition, simulation study (Chapter 6 and 7) showed that this estimation method diverges in some cases. However, by assuming tempered stable innovations, the innovations $\{\varepsilon_t\}$ has zero mean and unit variance. Consequently, the blocked Whittle estimates proposed by Dahlhaus (1997) can be used.

Formally, the parameter space can be separated in two sets, i.e. $\theta = (\theta_1, \theta_2)$ where $\theta_1 = (\alpha, \lambda_+, \lambda_-)$ is the parameters related to the innovations and θ_2 is the parameter vector related to the locally stationary process. The natural candidate for estimating θ is using (2.33). However, the difficulty arises because of the absence of tempered stable density function. It involves numerical computation of the Fourier transform of the characteristic function.

We propose a two-step parametric estimation in the following manner. Suppose that we are interested in estimating θ by maximizing a likelihood function $\mathcal{L}_T(\theta_1, \theta_2)$ such as (2.33). In the first step, we obtain the blocked Whittle estimates $\hat{\theta}_2$, which does not depends on θ_1, since for all different values of θ_1, ε_t has zero mean and unit variance. In the second step, we estimate θ_1 by maximizing $\mathcal{L}_T(\theta_1, \hat{\theta}_2)$. Note that if θ_2 is known, we can recursively obtain ε_t and by assuming $\{\varepsilon_t\} \overset{iid}{\sim} stdCTS(\alpha, \lambda_+, \lambda_-)$, consistent maximum likelihood estimates for θ_1 are obtained (see Appendix A). This estimation procedure was also implemented in GARCH model with tempered stable innovations by Kim *et al.* (2008).

Although $\hat{\theta}_2$ is asymptotically consistent and normal, it has a bias $\hat{\theta}_2 - \theta_2$. Here, we will study the parametric estimation of θ using this method.

8.3 Simulation results

In this section, we carried out simulations in order to study the parameter estimation of the model for $\alpha \in (0, 1)$ since this case the tempered stable distribution can be simulated exactly (see Section 3.3.2).

The estimation procedure is done as follows. First, we performed the blocked Whittle estimation presented in the Section 2.2.1 considering the suggestion of block size $N = \lfloor T^{0.8} \rfloor$ and shifting each block by $Q = \lfloor 0.2N \rfloor$ time units from Dahlhaus and Giraitis (1998). After $\hat{\theta}_2$ are obtained, we fix $\theta_2 = \hat{\theta}_2$ and then we estimate the parameters of stdCTS distribution θ_1. In the second stage, it is necessary to alter the original parameter space $\Theta_1 = (0, 2) \times (0, \infty)^2$ into $\Theta_1^* = (\epsilon, 2-\epsilon) \times (\epsilon, M)^2$ in order to guarantee the strong consistency (see Appendix A). We set $\epsilon = 0.01$ and $M = 3$.

We performed simulations for three scenarios of the tvARMA(1,1) model with stdCTS innovations where the coefficients are linear functions $\alpha_1(u) = \theta_{a0} + \theta_{a1} u$, $\beta_1(u) = \theta_{b0} + \theta_{b1} u$, and $\gamma(u) = \gamma$. We carried out simulations for $T = 500, 1000$, 1500, 2000 and 3000 observations based on $R = 1000$ independent replications each scenario.

The first scenario assumes $(\alpha, \lambda_+, \lambda_-, \theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \gamma) = (0.2, 1, 1, 0.3, -0.3, -0.5, 0.4, 1.2)$. In this case, the innovation distribution is symmetric and leptokurtic. The Monte Carlo mean, standard error, kurtosis and skewness of estimates from the tvARMA(1,1) simulation are reported in the Table 8.1 and 8.2 and the density estimates in Figures 8.1. All blocked Whittle estimates behave as expected and seem to be Gaussian. On the other hand, λ_+ and λ_- estimates behave appropriately, but their sample distribution is bimodal when $T = 500$. Nevertheless, its behavior disappears when T increases. α estimator is biased for small sample path, but its Monte Carlo mean approaches to the real value when T increases.

T	α	λ_+	λ_-	θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ
500	0.2989	0.9663	0.9698	0.3025	-0.3235	-0.4987	0.3768	1.1913
	(0.2625)	(0.2446)	(0.2444)	(0.2285)	(0.5855)	(0.2260)	(0.5911)	(0.0794)
1000	0.2693	0.9678	0.9673	0.3068	-0.3330	-0.4950	0.3686	1.1966
	(0.1829)	(0.1736)	(0.1743)	(0.1546)	(0.3992)	(0.1521)	(0.3953)	(0.0551)
1500	0.2355	0.9876	0.9857	0.2967	-0.3055	-0.5014	0.3952	1.1976
	(0.1466)	(0.1403)	(0.1416)	(0.1234)	(0.3169)	(0.1198)	(0.3131)	(0.0452)
2000	0.2340	0.9853	0.9851	0.2983	-0.2956	-0.5003	0.4005	1.1983
	(0.1260)	(0.1221)	(0.1224)	(0.1050)	(0.2674)	(0.1008)	(0.2600)	(0.0376)
3000	0.2207	0.9902	0.9901	0.2962	-0.2987	-0.5050	0.4062	1.1995
	(0.1041)	(0.1017)	(0.1012)	(0.0822)	(0.2114)	(0.0777)	(0.2038)	(0.0325)

Table 8.1: *Monte Carlo mean and standard error (in parentheses) of two-step estimators for different T sample sizes assuming $\alpha = 0.2, \lambda_+ = 1, \lambda_- = 1, \theta_{a0} = 0.3, \theta_{a1} = -0.3, \theta_{b0} = -0.5, \theta_{b1} = 0.4, \gamma = 1.2$ from tvARMA(1,1) with stdCTS innovations based on $R = 1000$ replications.*

T		α	λ_+	λ_-	θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ
500	kur	2.5468	2.3023	2.2717	3.2051	3.0259	3.2725	3.0547	2.9928
	skw	0.6314	-0.2488	-0.2226	-0.0605	-0.0410	0.2343	0.1258	0.2158
1000	kur	2.5576	2.5213	2.4212	3.0045	3.3690	3.1830	3.2570	3.1912
	skw	0.4032	-0.0530	-0.0642	-0.0286	-0.0834	0.0935	0.0511	0.1031
1500	kur	2.7362	2.6395	2.6459	2.9303	2.9114	2.8817	3.0034	3.0287
	skw	0.3650	-0.1349	-0.1544	-0.0655	0.0483	0.1590	0.1257	0.0134
2000	kur	2.8761	2.7006	2.7478	2.9323	3.0834	3.1401	3.1718	3.1502
	skw	0.2672	0.0307	-0.0320	0.0287	-0.0029	0.1037	0.1469	0.1542
3000	kur	2.7710	2.8416	2.7187	3.0465	3.2093	2.8993	3.2595	2.9748
	skw	0.0477	0.2253	0.1790	-0.1514	0.0557	-0.0116	0.1697	0.1015

Table 8.2: *Kurtosis and skewness of two-step estimators for different T sample sizes assuming $\alpha = 0.2, \lambda_+ = 1, \lambda_- = 1, \theta_{a0} = 0.3, \theta_{a1} = -0.3, \theta_{b0} = -0.5, \theta_{b1} = 0.4, \gamma = 1.2$ from tvARMA(1,1) with stdCTS innovations based on $R = 1000$ replications.*

The second scenario assumes $(\alpha, \lambda_+, \lambda_-, \theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \gamma)$=$(0.3, 0.5, 1, -0.3, 0.8, 0.5, -0.1, 1)$ and its innovation distribution is positively asymmetric and leptokurtic. The Monte Carlo mean, standard error, kurtosis and skewness of estimates from the tvARMA(1,1) simulation are reported in the Table 8.3 and 8.4 and the density estimates in Figures 8.2. First, we

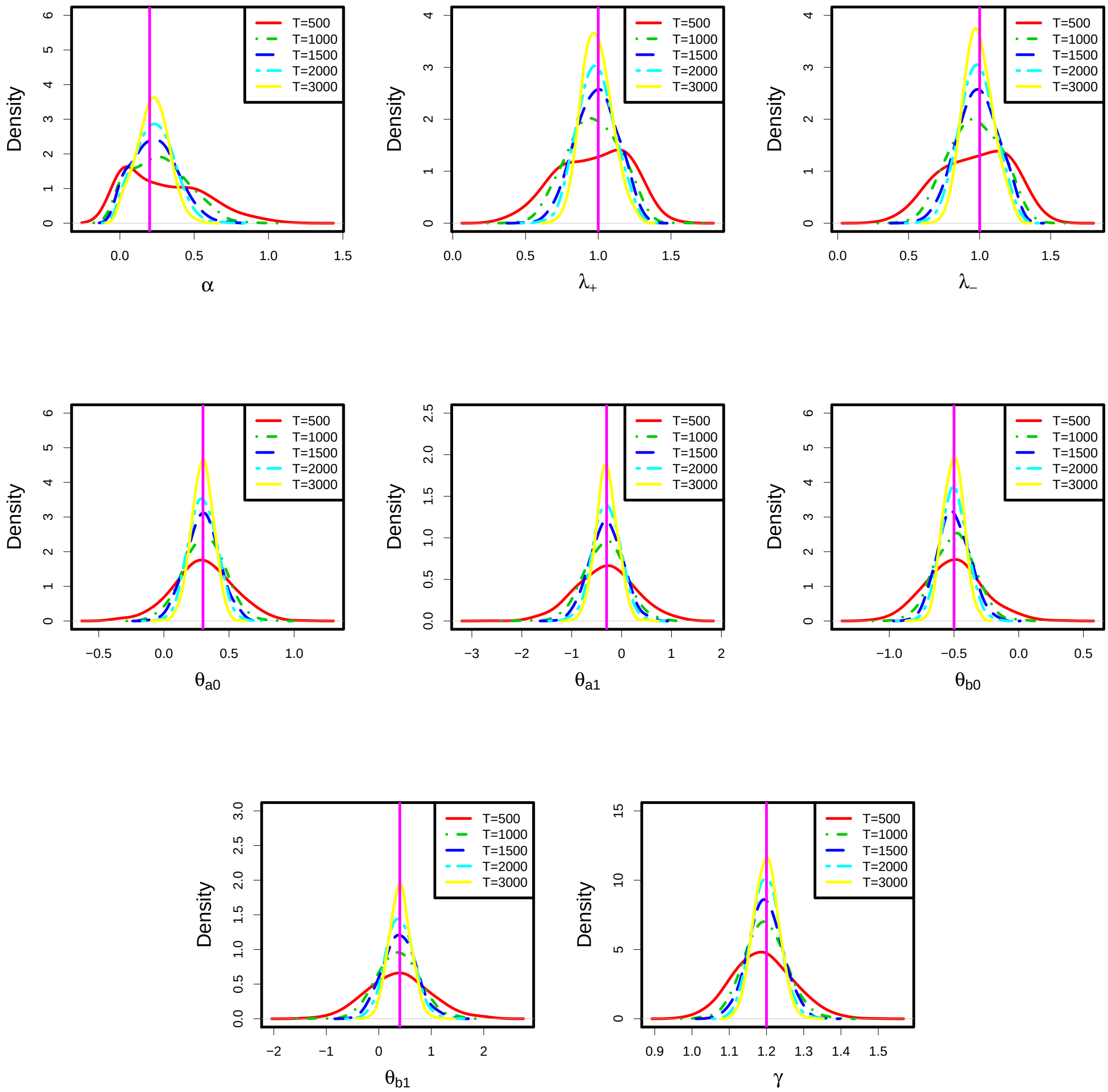


Figure 8.1: *Density estimates of estimators for different T sample sizes based on $R = 1000$ replications from tvARMA with stdCTS innovations with* $\alpha = 0.2, \lambda_+ = 1, \lambda_- = 1, \theta_{a0} = 0.3, \theta_{a1} = -0.3, \theta_{b0} = -0.5, \theta_{b1} = 0.4, \gamma = 1.2.$

notice that all blocked Whittle estimates behave appropriately, as the previous scenario. For θ_1, similar Monte Carlos mean and standard error are presented. However, the kurtosis and skewness are far from the Gaussian case. We conclude that the simulation agree to the theoretical result (strong consistency), but we cannot guarantee the asymptotic normality of the estimators.

T	α	λ_+	λ_-	θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ
500	0.3732	0.5160	1.0100	-0.2686	0.7138	0.5245	-0.1753	0.9891
	(0.1679)	(0.1647)	(0.1796)	(0.2711)	(0.7640)	(0.2520)	(0.7377)	(0.1043)
1000	0.3612	0.4948	0.9888	-0.2931	0.7927	0.5012	-0.1017	0.9983
	(0.1077)	(0.1266)	(0.1226)	(0.1751)	(0.5234)	(0.1652)	(0.5056)	(0.0725)
1500	0.3464	0.4867	0.9908	-0.2908	0.7709	0.5074	-0.1267	0.9977
	(0.0842)	(0.0732)	(0.0887)	(0.1429)	(0.4266)	(0.1368)	(0.4171)	(0.0565)
2000	0.3318	0.4930	0.9944	-0.2934	0.7843	0.5048	-0.1152	1.0004
	(0.0769)	(0.0678)	(0.0814)	(0.1153)	(0.3437)	(0.1076)	(0.3295)	(0.0520)
3000	0.3193	0.4951	0.9982	-0.2931	0.7854	0.5053	-0.1141	0.9995
	(0.0721)	(0.0594)	(0.0726)	(0.0928)	(0.2755)	(0.0850)	(0.2618)	(0.0431)

Table 8.3: *Monte Carlo mean and standard error (in parentheses) of two-step estimators for different T sample sizes assuming $\alpha = 0.3, \lambda_+ = 0.5, \lambda_- = 1, \theta_{a0} = -0.3, \theta_{a1} = 0.8, \theta_{b0} = 0.5, \theta_{b1} = -0.1, \gamma = 1$ from tvARMA(1,1) with stdCTS innovations based on $R = 1000$ replications.*

T		α	λ_+	λ_-	θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ
500	kur	6.6391	9.5239	6.1957	3.1675	3.2482	3.4967	3.5037	3.9955
	skw	0.4994	1.5074	-0.0480	0.0972	-0.3716	-0.2039	-0.4640	0.5446
1000	kur	3.1408	27.3207	11.0789	3.0195	3.2693	3.2417	3.3428	3.3931
	skw	-0.1212	3.5038	-0.7241	0.0318	-0.2118	-0.2340	-0.2320	0.3512
1500	kur	3.5276	17.6460	3.1597	3.4821	3.9615	3.8693	4.2494	3.0668
	skw	-0.1706	1.8607	0.1045	-0.0142	-0.4793	-0.2568	-0.5146	0.2863
2000	kur	4.7877	16.2723	3.7344	3.2455	3.3039	3.2883	3.3466	3.3949
	skw	-0.5666	1.8473	0.4418	0.0362	-0.0351	-0.0549	-0.0675	0.2757
3000	kur	5.9832	7.1008	4.8855	3.0839	3.2296	3.0412	3.3390	3.2774
	skw	-0.9616	1.2562	0.6662	0.0384	-0.2878	-0.0336	-0.3403	0.2919

Table 8.4: *Kurtosis and skewness of two-step estimators for different T sample sizes assuming $\alpha = 0.3, \lambda_+ = 0.5, \lambda_- = 1, \theta_{a0} = -0.3, \theta_{a1} = 0.8, \theta_{b0} = 0.5, \theta_{b1} = -0.1, \gamma = 1$ from tvARMA(1,1) with stdCTS innovations based on $R = 1000$ replications.*

Finally, the third scenario assumes $(\alpha, \lambda_+, \lambda_-, \theta_{a0}, \theta_{a1}, \theta_{b0}, \theta_{b1}, \gamma) = (0.7, 1, 0.5, 0.3, -0.3, -0.5, 0.4, 1.2)$. This process has innovation with $\alpha = 0.7$ which is less leptokurtic and asymmetric. The Monte Carlo mean, standard error, kurtosis and skewness of estimates from the tvARMA(1,1) simulation are reported in the Table 8.5 and 8.6 and the density estimates in Figures 8.3. In this case, we notice again that the blocked Whittle estimates perform as expected. For θ_1, we notice that the estimates behave closer to the Gaussian case although the innovation distribution is asymmetric. We conjecture that it is because α is higher than the previous scenario.

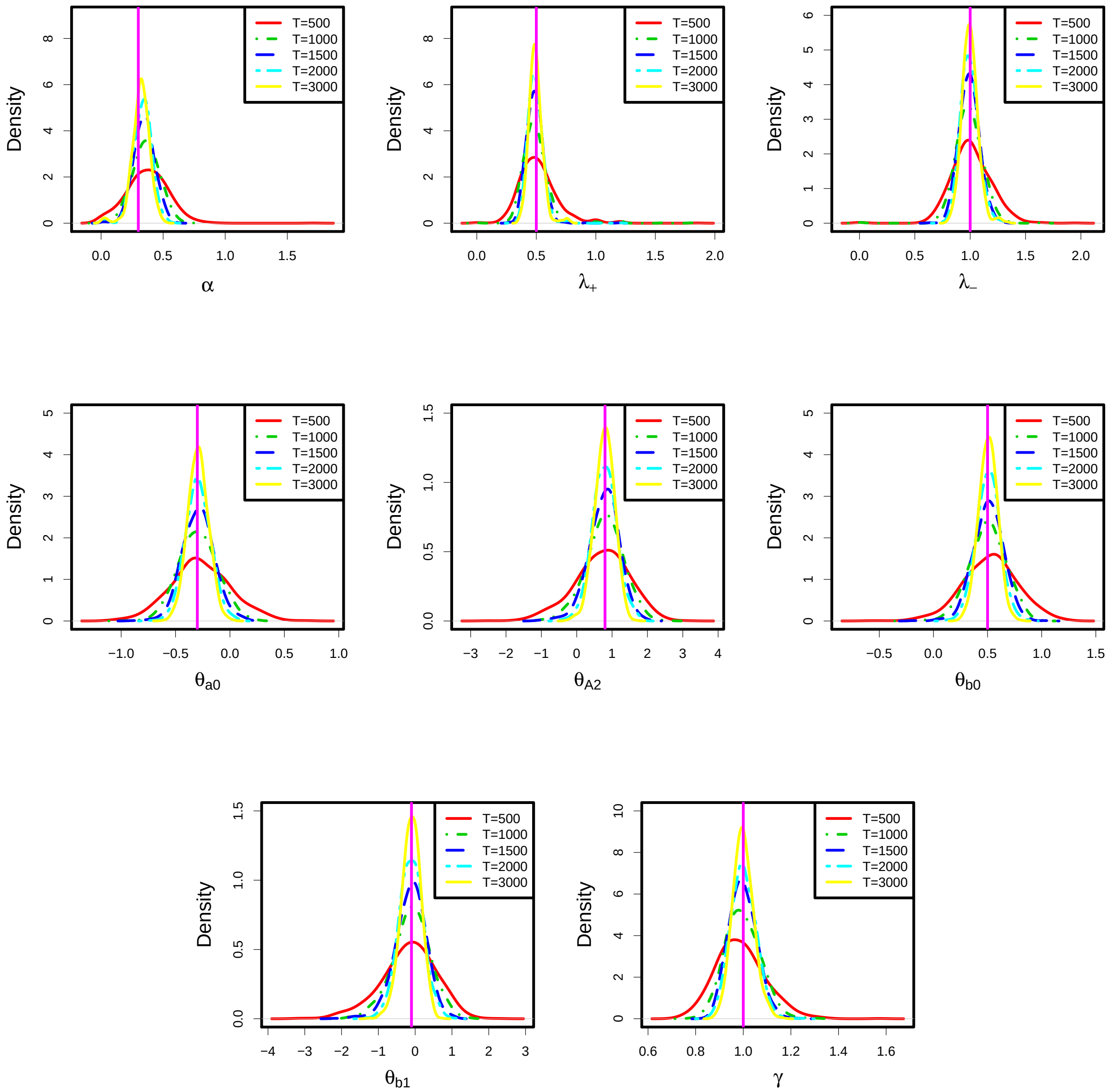


Figure 8.2: *Density estimates of estimators for different T sample sizes based on $R = 1000$ replications from tvARMA with stdCTS innovations with* $\alpha = 0.3, \lambda_{+} = 0.5, \lambda_{-} = 1, \theta_{a0} = -0.3, \theta_{a1} = 0.8, \theta_{b0} = 0.5, \theta_{b1} = -0.1, \gamma = 1$.

T	α	λ_+	λ_-	θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ
500	0.6762	1.0782	0.5675	0.3054	-0.3306	-0.4984	0.3784	1.1857
	(0.2930)	(0.2935)	(0.2463)	(0.2269)	(0.5777)	(0.2198)	(0.5724)	(0.1020)
1000	0.6935	1.0298	0.5247	0.3025	-0.3058	-0.4954	0.3863	1.1953
	(0.1983)	(0.1947)	(0.1604)	(0.1491)	(0.3818)	(0.1476)	(0.3797)	(0.0700)
1500	0.7040	1.0094	0.5106	0.2972	-0.2978	-0.5049	0.4026	1.1996
	(0.1598)	(0.1554)	(0.1302)	(0.1187)	(0.3049)	(0.1177)	(0.3015)	(0.0561)
2000	0.7007	1.0091	0.5089	0.3006	-0.3103	-0.4962	0.3843	1.1989
	(0.1439)	(0.1377)	(0.1160)	(0.1083)	(0.2739)	(0.1008)	(0.2594)	(0.0517)
3000	0.6996	1.0071	0.5064	0.2979	-0.2976	-0.5010	0.4005	1.1972
	(0.1144)	(0.1104)	(0.0922)	(0.0770)	(0.2053)	(0.0758)	(0.2035)	(0.0402)

Table 8.5: *Monte Carlo mean and standard error (in parentheses) of two-step estimators for different T sample sizes assuming $\alpha = 0.7, \lambda_+ = 1, \lambda_- = 0.5, \theta_{a0} = 0.3, \theta_{a1} = -0.3, \theta_{b0} = -0.5, \theta_{b1} = 0.4, \gamma = 1.2$ from tvARMA(1,1) with stdCTS innovations based on $R = 1000$ replications.*

T		α	λ_+	λ_-	θ_{a0}	θ_{a1}	θ_{b0}	θ_{b1}	γ
500	kur	2.6833	3.0095	3.1353	2.9216	3.1001	3.1375	3.4460	3.7251
	skw	-0.5157	0.3563	0.7647	0.0513	-0.1086	0.2629	0.0768	0.6667
1000	kur	3.2161	3.0973	3.4723	2.9338	3.0667	2.9823	2.9215	2.8907
	skw	-0.3283	0.3894	0.6380	-0.1271	-0.0667	0.1997	-0.0237	0.3053
1500	kur	3.1654	3.1584	3.4874	3.0103	3.1361	2.9630	3.0521	3.0971
	skw	-0.3203	0.4250	0.5738	-0.0419	-0.0312	0.0873	0.0880	0.1313
2000	kur	3.1776	2.9805	3.3796	2.8149	2.9761	3.0181	3.0578	3.2647
	skw	-0.2914	0.2486	0.4563	-0.0226	0.0241	0.1353	0.0365	0.2892
3000	kur	3.4588	3.2988	3.5399	3.0361	3.2010	2.9095	2.8987	3.2337
	skw	-0.3294	0.2599	0.4583	0.0719	-0.0601	0.0808	0.0618	0.0516

Table 8.6: *Kurtosis and skewness of two-step estimators for different T sample sizes assuming $\alpha = 0.7, \lambda_+ = 1, \lambda_- = 0.5, \theta_{a0} = 0.3, \theta_{a1} = -0.3, \theta_{b0} = -0.5, \theta_{b1} = 0.4, \gamma = 1.2$ from tvARMA(1,1) with stdCTS innovations based on $R = 1000$ replications.*

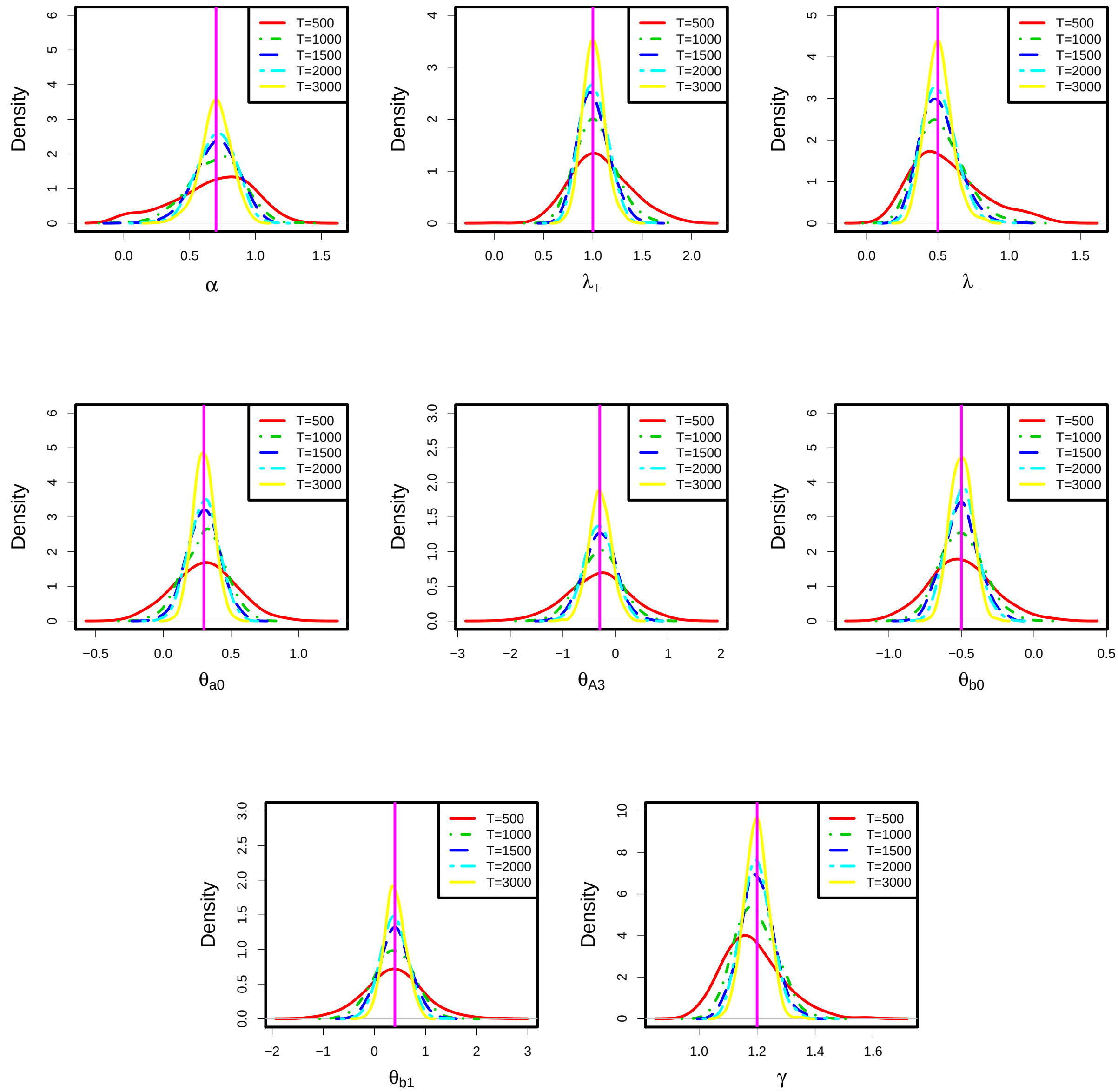


Figure 8.3: *Density estimates of estimators for different T sample size based on $R = 1000$ replications from tvARMA with stdCTS innovations with* $\alpha = 0.7, \lambda_+ = 1, \lambda_- = 0.5, \theta_{a0} = 0.3, \theta_{a1} = -0.3, \theta_{b0} = -0.5, \theta_{b1} = 0.4, \gamma = 1.2$.

8.4 Application

In this section, we continue with the application of wind data (Section 6.5.2). By assuming tempered stable innovations, the first step estimation procedure to estimate the time-varying structure of the model is using the Blocked Whittle estimation and it is presented in Table 6.10. The second step of tempered stable parameter estimation is done by maximum likelihood estimation and we obtain $\hat{\theta}_1 = (\hat{\alpha}, \hat{\lambda}_+, \hat{\lambda}_-) = (0.5017, 0.6702, 0.6798)$. The disadvantage of this method is the difficulty of compute standard error of estimates. The only property from independent and identically distributed sample from tempered stable distribution is the strong consistency. Hence, we performed parametric Bootstrapping with $R = 1000$ replications to recover the standard error of estimates. The results are presented in Table 8.7. It is interesting to see that the bootstrapping standard errors and the asymptotic standard error estimated using Blocked Whittle estimator (see Table 6.10) are very similar. Finally, Figure 8.4 shows the histogram of the residuals from the model with 3 different assumptions (normal, stable and tempered stable innovations). It is clear to see that the tempered stable innovations assumption is slightly better.

Parameter	Estimate	Standard error
α	0.5017	0.1600
λ_+	0.6702	0.1411
λ_-	0.6798	0.1408
θ_{a0}	-1.5985	0.0771
θ_{a1}	0.3305	0.1356
θ_{b0}	0.9135	0.1379
θ_{b1}	0.0207	0.2351
θ_{c0}	-0.0585	0.1394
θ_{c1}	-0.7153	0.2383
θ_{d0}	-0.1316	0.0770
θ_{d1}	0.5454	0.1346
γ_0	0.0077	0.0015
γ_1	0.0152	0.0032

Table 8.7: *two-step estimates of tvAR(4) with tempered stable innovations from wind power time series.*

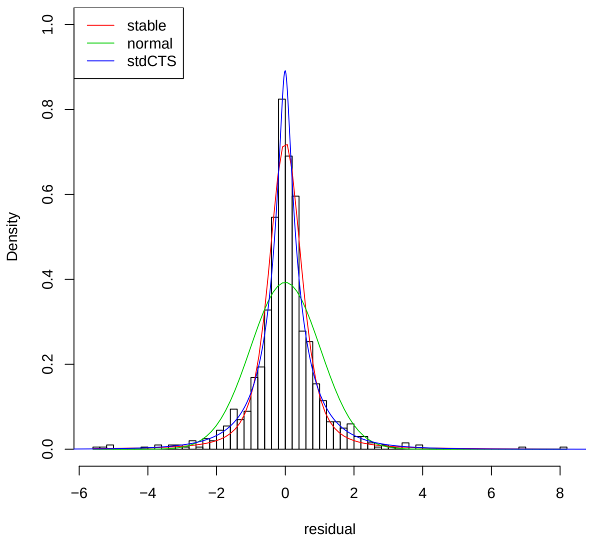


Figure 8.4: *Standardized residual histogram with estimated stable curve* $(\alpha = 1.34,\ \beta = 0)$*, tempered stable curve* $(\alpha = 0.5017,\ \lambda_+ = 0.6702,\ \lambda_- = 0.6798)$ *and Gaussian curve.*

Chapter 9

Conclusions

In this thesis, we focused on two specific approaches on the locally stationary processes with heavy-tailed innovation. First, we studied $\alpha-$stable locally stationary ARMA processes and presented their properties. In contrast to the locally stationary processes with finite second moments, this type of processes involves the infinite variance phenomenon observed in different fields of study. Since the α-stable family of distributions, as a generalization of the Gaussian distribution, is closed under linear combinations which includes the possibility of handling asymmetry and thicker tails, the proposed model presents the same tail behavior, which is characterized by the index of stability, throughout the time. We also proposed an indirect inference method for the process with parametric time-varying coefficients. Specifically, we performed simulations for some basic models with linear parametric coefficients for known and unkown α. The results show that indirect inference is unbiased and suggest consistency. Thus, we conclude that the estimation methodology is satisfactory.

Then, we studied the locally stationary process with tempered stable innovations. In this case, the process presents less attractive properties because it is not closed under linear combinations. However, since its moments of all orders are finite, time series models involving tempered stable innovations can be estimated using traditional methods with weakly stationary assumption. We concentrate in the standardized tempered stable innovation, so that a two-step estimation can be performed. In the first step, the blocked Whittle estimation can be used to estimate the time-varying structure of the process. In the second step, by assuming independent innovations, recovering from residuals of the model, consistent estimation related to the tempered stable distribution can be obtained by maximum likelihood estimation. Simulations were done to study the properties of the estimators. As expected, blocked Whittle estimates (first step) behave approximately Gaussian. In the second step, nevertheless, we notice that for small α, estimates seem to be biased for small time series length, but they approximate to the real parameter when time series length increases. That is, theoretical result (strong consistency) is satisfied, but we cannot guarantee the asymptotic normality of the estimators.

There are some limitations that still need to be solved and they remain as future research. In the stable innovation case, since the time-varying spectral representation does not exist, identifying the local structure using traditional methods (autocorrelation and partial autocorrelation) or using blocked smooth periodogram are informal ways to identify the time-varying structure. One possibility is to explore the local version of the dependence measure called autocovariation (Kokoszka and Taqqu, 1994). Next, the indirect inference simulation is performed by assuming known time-varying linear structure throughout the time. Second, the simulation study does not reach to more complex structure and the possibility of non-parametric estimation. In this case, it is reasonable to consider periodic time structure. Third, the indirect inference is more time-consuming since it involves simulation-based estimation, but they are appropriate when heavy-tailed innovations are present. Fourth, asymptotic properties of the indirect estimates are unknown. Simulations suggest that when α is close to 2, Gaussian innovations can be assumed and thus, blocked Whittle estimation can be used. Issues related to model selection are still an open question. Also, there is few work about the prediction of locally stationary process.

In the tempered stable innovation case, all the traditional methods with weakly stationary condition can be used (identification of time varying structure using autocorrelation, partial autocorrelation and blocked smooth periodogram) since the second moments exist. However, the properties of the process are unknown and alternative estimation methods are still unexplored. Finally, it is important to explore methodologies in order to distinguish between heavy-tailed and semi-heavy-tailed residuals.

Appendix A

Maximum likelihood estimation of standardized classical tempered stable distribution

In this section, we present the strong consistency of the maximum likelihood estimator (MLE) from independent and identically distributed standardized classical tempered stable random sample. It is well known that the traditional conditions for the consistency are based on the probability density function. However, Grabchak (2016b) proposed conditions based on properties of the Lévy triplet. First, we briefly review background about infinitely divisible distribution and selfdecomposable distribution (for more detail see Cont and Tankov , 2015; Grabchak, 2016a). Second, we check the conditions for the consistency of MLE from standardized classical tempered stable distribution. Finally, some simulations are performed.

A.1 Background

An infinitely divisible distribution ν has a characteristic function of the form

$$\hat{\nu}(z) = \exp\left\{-\frac{1}{2}a^2z^2 + ibz + \int_{\mathbb{R}} \left(e^{izx} - 1 - izx\mathbb{1}_{|x|\le 1}\right) M(dx)\right\}, z \in \mathbb{R}, \tag{A.1}$$

where $a \geq 0$ is called the Gaussian part, $b \in \mathbb{R}$ is the shift, and M is the Lévy measure satisfying

$$M(\{0\}) = 0 \text{ and } \int_{\mathbb{R}} (1 \wedge |x|^2) M(dx) < \infty. \tag{A.2}$$

The Lévy triplet (a, b, M) uniquely determines the distribution and it is denoted by $\nu \sim ID(a, b, M)$.

A probability measure μ is selfdecomposable if and only if $\mu = ID(a, b, M)$, where

$$M(dx) = \frac{g(x)}{|x|}dx, \tag{A.3}$$

for a function that is increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$. In this case, the triplet (a, b, g) uniquely determines the distribution μ and it is denoted $\mu \sim SD(a, b, g)$. In general, the parametric class is considered, that is, let $\Theta \in \mathbb{R}^d$ the parameter space and $\{\mu_\theta : \theta \in \Theta\}$ be the family of selfdecomposable distributions with $\mu_\theta \sim SD(a_\theta, b_\theta, g_\theta)$.

If $X \sim CTS(\alpha, \lambda_+, \lambda_-, C_+, C_-, \mu)$ with characteristic function defined in (3.24), then $\theta = (\alpha, \lambda_+, \lambda_-, C_+, C_-, \mu)$ and $X \sim SD(0, b_\theta, g_\theta)$ with

$$g_\theta(x) = C_-|x|^{-\alpha} e^{-|x|\lambda_-} \mathbb{1}_{x<0} + C_+|x|^{-\alpha} e^{-|x|\lambda_+} \mathbb{1}_{x>0}, \tag{A.4}$$

and

$$b_\theta = \mu - \int_1^\infty [g_\theta(x) - g_\theta(-x)]\, dx. \tag{A.5}$$

Finally, the stdCTS distribution is SD since it is a special case of CTS distribution.

A.2 Maximum likelihood estimation

Let $X_1, X_2, ..., \overset{iid}{\sim} \mu_{\theta_0}$ for some $\theta_0 \in \Theta$. The MLE based on the observations $X_1, X_2, ..., X_n$ is given by

$$\hat{\theta}_n^{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \prod_{i=1}^n f_\theta(X_i), \tag{A.6}$$

where f_θ is the density function of μ_θ and $\prod_{i=1}^n f_\theta(X_i)$ is the likelihood function. In the following, the conditions for the strong consistency of MLE are stated.

(A1) The parameter space Θ is a closed set.

(A2) If $\theta, \theta' \in \Theta$, $a_\theta = a_{\theta'}$, $b_\theta = b_{\theta'}$ and $g_\theta(x) = a_{\theta'}(x)$ for Lebesgue almost every x then $\theta = \theta'$.

(A3) If $\lim_{i\to\infty} \theta_i = \theta$ then $\mu_{\theta_i} \xrightarrow{d} \mu_\theta$.

(A4) For every $\theta \in \Theta$

$$\int_{|x|>1} g_\theta(x) \frac{\log |x|}{|x|} dx < \infty. \tag{A.7}$$

(A5) We can write $\Theta = \Theta_1 \cup \Theta_2 \cup \Theta_3$ such that

$$\inf_{\theta \in \Theta_1} a_\theta > 0, \tag{A.8}$$

$$\lim_{x\to 0} \inf_{\theta \in \Theta_2} [g_\theta(x) + g_\theta(-x)] > 1, \tag{A.9}$$

and there exists a $\beta \in [0, 2)$ and $c > 0$ such that for any $\delta \in [0, 1]$

$$\inf_{\theta \in \Theta_3} \int_0^\delta x\, [g_\theta(x) + g_\theta(-x)]\, dx \geq c\delta^\beta. \tag{A.10}$$

(A6) If $\{\theta_i'\}$ is a sequence in Θ with $|\theta_i| \to \infty$ then every subsequence has a further subsequence $\{\theta_i\}$ such that either there exists an $x_0 > 0$ with

$$\lim_{i\to\infty} [a_{\theta_i} + g_{\theta_i}(x_0) + g_{\theta_i}(-x_0)] = \infty \tag{A.11}$$

or there exists a $\beta \in [0,2)$ and a sequence $\{c_i\}$ such that $c_i \to \infty$ and for any $\delta \in [0,1]$

$$\int_0^\delta x\,[g_{\theta_i}(x) + g_{\theta_i}(-x)] \geq c_i \delta^\beta. \tag{A.12}$$

Theorem A.1. *If (A1)- (A6) hold then $\hat{\theta}_n^{MLE} \to \theta_0$ with probability 1.*

Grabchak (2016b) showed the MLE is strongly consistent for several classes of stable, tempered stable and others distributions with some modification in the parameter space due to the fact that it has to be close. In the following, we prove that the stdCTS distribution satisfies these conditions.

Let $X \sim stdCTS(\alpha, \lambda_+, \lambda_-)$ with $\theta = (\alpha, \lambda_+, \lambda_-)$. It is easy to see from (A.4) and (A.5) that $X \sim SD(0, b_\theta, g_\theta)$ with

$$g_\theta(x) = C\left[|x|^{-\alpha} e^{-|x|\lambda_-} \mathbb{1}_{x<0} + |x|^{-\alpha} e^{-|x|\lambda_+} \mathbb{1}_{x>0}\right], \tag{A.13}$$

and

$$b_\theta = \mu - \int_1^\infty [g_\theta(x) - g_\theta(-x)]\,dx, \tag{A.14}$$

where $C = C_{\alpha,\lambda_+,\lambda_-} = \frac{1}{\Gamma(2-\alpha)\left(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}\right)}$ from (3.29).

We consider the restricted parameter space $\Theta = (\epsilon, 2-\epsilon) \times (\epsilon, M)^2$ where $\epsilon \in (0, 0.5)$ and $M > 1$ in order to guarantee the conditions for strongly consistent MLE. Since the Θ is a compact set, then (A1) and (A6) hold (see Grabchak, 2016b). (A2) can be easily verified and (A3) holds from the Proposition 3.1 in Küchler and Tappe (2013). To see that (A4) holds, let $\theta \in \Theta$,

$$\begin{aligned}
\int_{|x|>1} g_\theta(x) \frac{\log|x|}{|x|} dx &= C\left[\int_1^\infty \frac{e^{-x\lambda_-}\log x}{x^{1+\alpha}} dx + \int_1^\infty \frac{e^{-x\lambda_+}\log x}{x^{1+\alpha}} dx\right] \\
&\leq C\left[e^{-\lambda_-}\int_1^\infty \frac{\log x}{x^{1+\alpha}} dx + e^{-\lambda_+}\int_1^\infty \frac{\log x}{x^{1+\alpha}} dx\right] \\
&\leq \frac{C}{\alpha^2}\left(e^{-\lambda_-} + e^{-\lambda_+}\right) \leq \frac{C}{\epsilon^2}\left(e^{-\epsilon} + e^{-\epsilon}\right) < \infty.
\end{aligned} \tag{A.15}$$

Finally, to show (A5) notice that

$$g_\theta(x) - g_\theta(-x) = C\left[(-x)^{-\alpha}\left(e^{x\lambda_-} + e^{x\lambda_+}\right)\mathbb{1}_{x<0} + (x)^{-\alpha}\left(e^{-x\lambda_-} + e^{-x\lambda_+}\right)\mathbb{1}_{x>0}\right]. \tag{A.16}$$

Consider $0 < x < 1$,

$$\inf_{\theta\in\Theta} g_\theta(x) - g_\theta(-x) = \inf_{\theta\in\Theta} \frac{1}{\Gamma(2-\alpha)\left(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}\right)} x^{-\alpha}\left(e^{-x\lambda_-} + e^{-x\lambda_+}\right)$$
$$= \frac{1}{2\Gamma(\epsilon)M^{2-\epsilon}} x^{-\epsilon}\left(e^{-xM} + e^{-xM}\right) = \frac{1}{\Gamma(\epsilon)M^{2-\epsilon}x^\epsilon e^{xM}}. \quad (A.17)$$

Then,

$$\lim_{x\to 0^+} \inf_{\theta\in\Theta} g_\theta(x) - g_\theta(-x) = \frac{1}{\Gamma(\epsilon)M^{2-\epsilon}} \lim_{x\to 0^+} \frac{1}{x^\epsilon e^{xM}} = \infty > 1. \quad (A.18)$$

Similarly for $x < 0$, the same result is obtained.

$$\lim_{x\to 0^-} \inf_{\theta\in\Theta} g_\theta(x) - g_\theta(-x) = \infty > 1. \quad (A.19)$$

A.3 Simulation

This section presents the Monte Carlo simulation in order to investigate the properties of MLE from finite sample. We concentrate the case when $\alpha \in (0,1)$ due to the fact that the CTS random variable can be generated exactly through acceptance-rejection sampling (see Kawai and Masuda, 2011). Here, we present results of the scenario with $\theta = (\alpha, \lambda_+, \lambda_-) = (0.4, 0.5, 1.5)$ for $T = 500, 1000$ and 1500 independent observations based on $R = 1000$ replications. Some other scenario were carried out and similar results were obtained.

Table A.1 reports the Monte Carlo mean, standard error, kurtosis and skewness of MLE. The estimates are likely to be biased for small sample, but it seems to be asymptotically unbiased since they are strongly consistent (see Section A.2). Moreover, kurtosis and skewness are similar to the Gaussian distribution.

T		α	λ_+	λ_-
500	Mean	0.3853	0.5128	1.5112
	Std. Dev.	0.1540	0.1054	0.1289
	kurtosis	2.8294	3.0112	3.7146
	skewness	-0.0153	0.3819	0.1773
1000	Mean	0.3898	0.5069	1.5083
	Std. Dev.	0.1007	0.0706	0.0893
	kurtosis	3.0356	3.7443	3.6659
	skewness	-0.2467	0.5931	0.1177
1500	Mean	0.3932	0.5041	1.5077
	Std. Dev.	0.0830	0.0568	0.0732
	kurtosis	3.0767	3.0755	3.1715
	skewness	0.0438	0.1893	0.1301

Table A.1: *Monte Carlo mean, standard error, kurtosis and skewness for maximum likelihood estimates of independent stdCTS sample with* $\alpha = 0.4,\ \lambda_+ = 0.5,\ \lambda_- = 1.5$.

Finally, Figures A.1, A.2, A.3 and A.4 show the density estimates, histogram and QQ-plot of each estimated parameter. They show satisfactory results and close to Gaussian distribution.

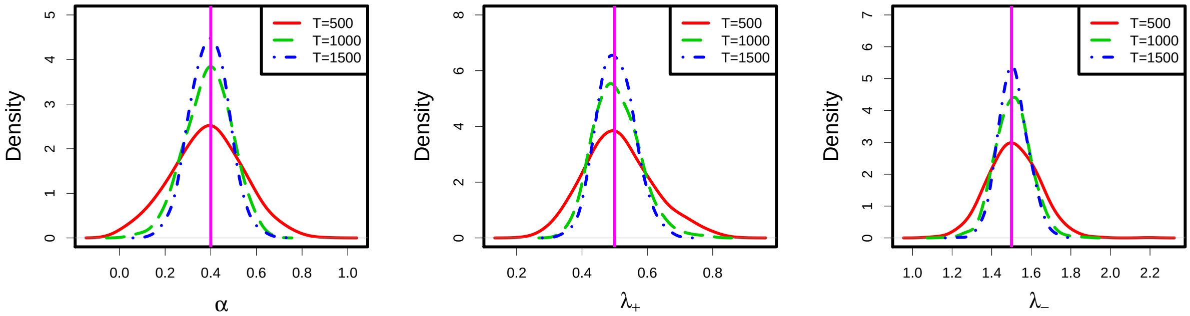


Figure A.1: *Density estimates of maximum likelihood estimates for different sample size based on* $R = 1000$ *replications from stdCTS distribution with* $\alpha = 0.4,\ \lambda_+ = 0.5,\ \lambda_- = 1.5.$

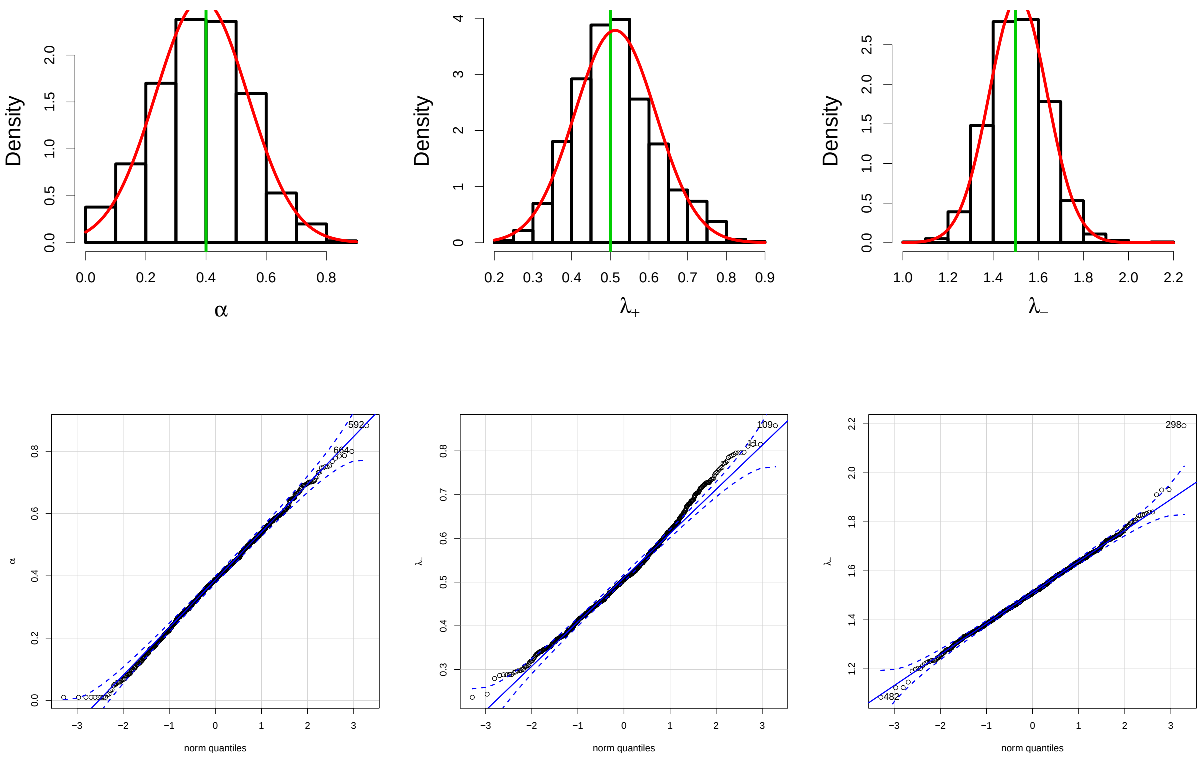


Figure A.2: *Histogram and QQ-plot of maximum likelihood estimates for* $T = 500$ *based on* $R = 1000$ *replications from stdCTS distribution with* $\alpha = 0.4,\ \lambda_+ = 0.5,\ \lambda_- = 1.5.$

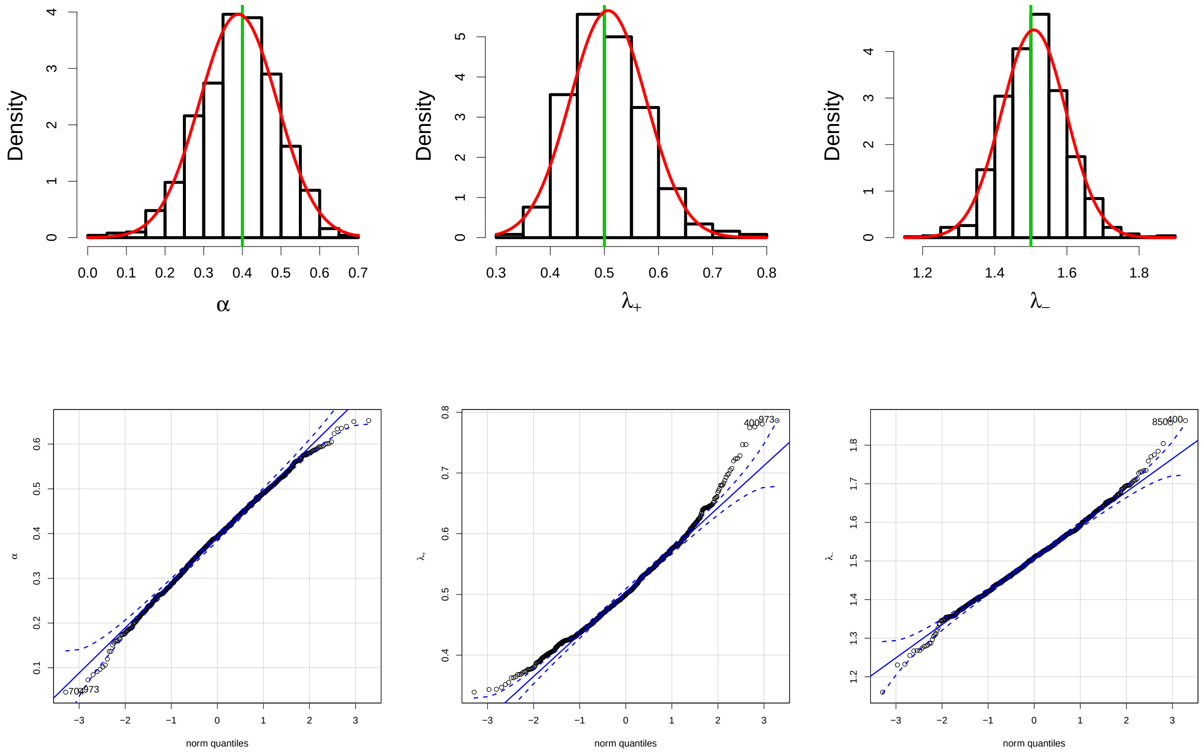


Figure A.3: *Histogram and QQ-plot of maximum likelihood estimates for* $T = 1000$ *based on* $R = 1000$ *replications from stdCTS distribution with* $\alpha = 0.4,\ \lambda_+ = 0.5,\ \lambda_- = 1.5.$

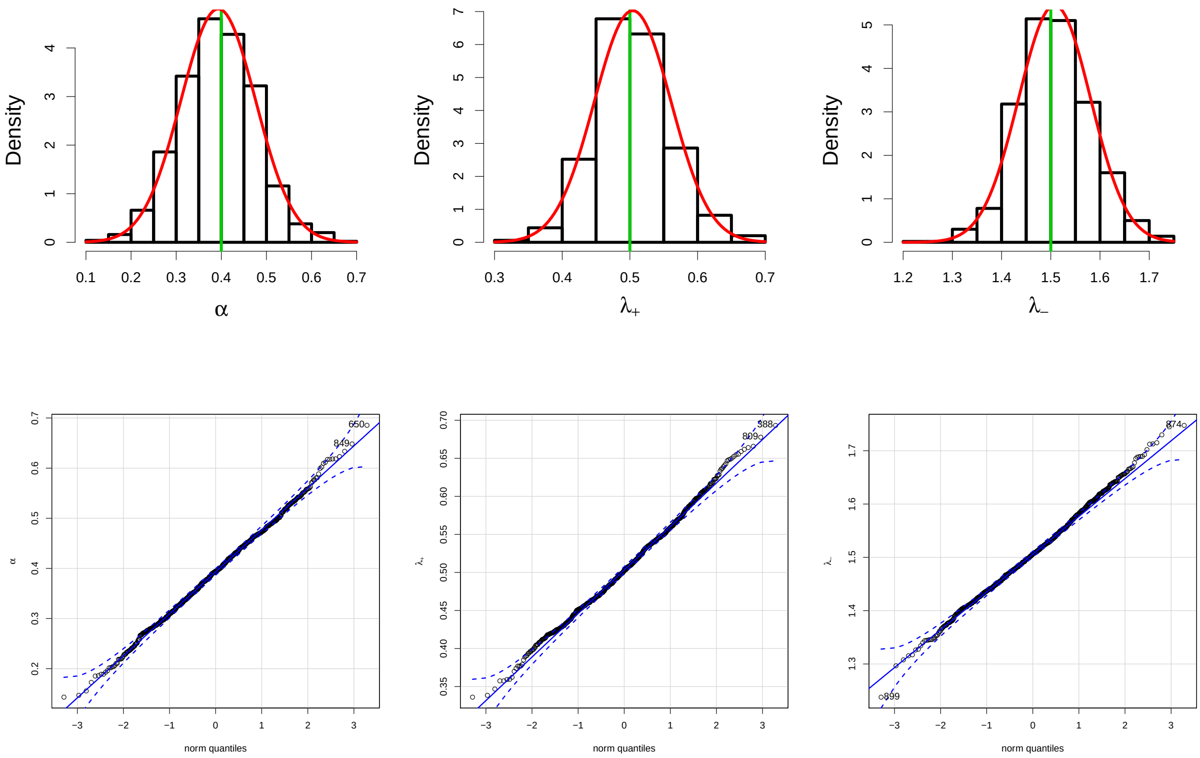


Figure A.4: *Histogram and QQ-plot of maximum likelihood estimates for* $T = 1500$ *based on* $R = 1000$ *replications from stdCTS distribution with* $\alpha = 0.4,\ \lambda_+ = 0.5,\ \lambda_- = 1.5.$

Bibliography

Azzalini and Capitanio (2003) Adelchi Azzalini and Antonella Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(2):367–389. ISSN 13697412, 14679868. URL http://www.jstor.org/stable/3647510. From page 61

Baeumer and Meerschaert (2010) Boris Baeumer and Mark M. Meerschaert. Tempered stable lévy motion and transient super-diffusion. *Journal of Computational and Applied Mathematics*, 233(10):2438 – 2448. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2009.10.027. URL http://www.sciencedirect.com/science/article/pii/S0377042709007250. From page 37

Bardet and Doukhan (2017) Jean-Marc Bardet and Paul Doukhan. Non-parametric estimation of time varying AR(1)–processes with local stationarity and periodicity. From page 19

Boyarchenko and Levendorskiĭ (2000) Svetlana I. Boyarchenko and Sergei Z. Levendorskiĭ. Option pricing for truncated lÉvy processes. *International Journal of Theoretical and Applied Finance*, 03(03):549–552. doi: 10.1142/S0219024900000541. URL https://doi.org/10.1142/S0219024900000541. From page 31

Brockwell and Davis (1991) Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer-Verlag New York. From page 49, 50

Calzolari and Halbleib (2018) Giorgio Calzolari and Roxana Halbleib. Estimating stable latent factor models by indirect inference. *Journal of Econometrics*, 205(1):280 – 301. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2018.03.014. From page 2

Calzolari *et al.* (2004) Giorgio Calzolari, Gabriele Fiorentini and Enrique Sentana. Constrained indirect estimation. *The Review of Economic Studies*, 71(4):945–973. doi: 10.1111/0034-6527.00310. URL +http://dx.doi.org/10.1111/0034-6527.00310. From page 46, 47

Calzolari *et al.* (2014) Giorgio Calzolari, Roxana Halbleib and Alessandro Parrini. Estimating garch-type models with symmetric stable innovations: Indirect inference versus maximum likelihood. *Computational Statistics & Data Analysis*, 76:158 – 171. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2013.07.028. From page 2

Carr and Geman (2002) Peter Carr and Helyette Geman. The fine structure of asset returns: An empirical investigation. *The Journal of Business*, 75(2):305–332. URL https://EconPapers.repec.org/RePEc:ucp:jnlbus:v:75:y:2002:i:2:p:305-332. From page 31

Chan and Palma (2019) Ngai Hang Chan and Wilfredo Palma. On the estimation of locally stationary long-memory processes. *Statistica Sinica*. doi: 10.5705/ss.202017.0472. URL https://app.dimensions.ai/details/publication/pub.1101621210andhttps://doi.org/10.5705/ss.202017.0472. From page 1

Chow and Teicher (2003) Yuan Shih Chow and Henry Teicher. *Probability Theory: Independence, Interchangeability, Martingales (Springer Texts in Statistics)*. Springer, 3rd ed. ISBN 0387406077,9780387406077. From page 51

Cont and Tankov (2015) R. Cont and P. Tankov. *Financial Modelling with Jump Processes, Second Edition*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis. ISBN 9781420082197. URL https://books.google.com.br/books?id=-fZtKgAACAAJ. From page 30, 32, 107

Contreras-Cristán *et al.* (2006) Alberto Contreras-Cristán, Eduardo Gutiérrez-Peña and Stephen G. Walker. A note on whittle's likelihood. *Communications in Statistics - Simulation and Computation*, 35(4):857–875. doi: 10.1080/03610910600880203. URL https://doi.org/10.1080/03610910600880203. From page 61

Dahlhaus (1996a) R. Dahlhaus. Maximum likelihood estimation and model selection for locally stationary processes. *Journal of Nonparametric Statistics*, 6(2-3):171–191. doi: 10.1080/10485259608832670. URL http://dx.doi.org/10.1080/10485259608832670. From page 1, 8

Dahlhaus (1996b) R. Dahlhaus. On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, 62(1):139 – 168. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(95)00090-9. URL http://www.sciencedirect.com/science/article/pii/0304414995000909. From page 15, 18

Dahlhaus (1996c) R. Dahlhaus. *Asymptotic statistical inference for nonstationary processes with evolutionary spectra*, pages 145–159. Springer New York, New York, NY. ISBN 978-1-4612-2412-9. doi: 10.1007/978-1-4612-2412-9_11. URL https://doi.org/10.1007/978-1-4612-2412-9{_}11. From page 15

Dahlhaus (1997) R. Dahlhaus. Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37. ISSN 00905364. URL http://www.jstor.org/stable/2242711. From page 1, 2, 10, 17, 18, 19, 96

Dahlhaus (2000) Rainer Dahlhaus. A likelihood approximation for locally stationary processes. *Ann. Statist.*, 28(6):1762–1794. doi: 10.1214/aos/1015957480. URL http://dx.doi.org/10.1214/aos/1015957480. From page 17

Dahlhaus (2012) Rainer Dahlhaus. Locally stationary processes. Em Suhasini Subba Rao Tata Subba Rao and C.R. Rao, editors, *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*, pages 351 – 413. Elsevier. doi: https://doi.org/10.1016/B978-0-444-53858-1.00013-2. URL http://www.sciencedirect.com/science/article/pii/B9780444538581000132. From page 5, 6, 8, 9, 16

Dahlhaus and Giraitis (1998) Rainer Dahlhaus and Liudas Giraitis. On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis*, 19(6):629–655. ISSN 1467-9892. doi: 10.1111/1467-9892.00114. URL http://dx.doi.org/10.1111/1467-9892.00114. From page 63, 72, 96

Dahlhaus and Polonik (2009) Rainer Dahlhaus and Wolfgang Polonik. Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39. ISSN 13507265. URL http://www.jstor.org/stable/20680139. From page 8, 54, 56

Dzhaparidze and Kotz (2012) K. Dzhaparidze and S. Kotz. *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer Series in Statistics. Springer New York. ISBN 9781461248422. URL https://books.google.com.br/books?id=lyfjBwAAQBAJ. From page 16

Embrechts *et al.* (1997) Paul Embrechts, Claudia Klüppelberg and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag Berlin Heidelberg. From page 49

Feng and Shi (2017) L. Feng and Y. Shi. Fractionally integrated garch model with tempered stable distribution: a simulation study. *Journal of Applied Statistics*, 44(16):2837–2857. ISSN 0266-4763. doi: 10.1080/02664763.2016.1266310. From page 2

Gallant and Tauchen (1996) A. Ronald Gallant and George Tauchen. Which moments to match? *Econometric Theory*, 12(4):657–681. ISSN 02664666, 14694360. URL http://www.jstor.org/stable/3532789. From page 2, 39, 41

Gonzales-Aparicio *et al.* (2016) Iratxe Gonzales-Aparicio, Andreas Zucker, Francesco Carerri, Fabio Monforti, Thomas Huld and Jake Badger. EMHIRES dataset. part I: Wind power generation European Meteorological derived high resolution res generation time series for present and future scenarios. Technical report, EUR 28171 EN; 10.2790/831549. From page 77

Gouriéroux and Monfort (1997) C. Gouriéroux and A. Monfort. *Simulation-based Econometric Methods*. Core Lectures. OUP Oxford. ISBN 9780191525094. URL https://books.google.com.br/books?id=k5jF-7VRMnEC. From page 39

Gourieroux *et al.* **(1993)** C. Gourieroux, A. Monfort and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118. ISSN 1099-1255. doi: 10.1002/jae.3950080507. URL http://dx.doi.org/10.1002/jae.3950080507. From page 2, 39, 42, 46

Grabchak (2016a) M. Grabchak. *Tempered Stable Distributions: Stochastic Models for Multiscale Processes*. SpringerBriefs in Mathematics. Springer International Publishing. ISBN 9783319249278. URL https://books.google.com.br/books?id=UWB6CwAAQBAJ. From page 1, 2, 21, 107

Grabchak (2016b) Michael Grabchak. On the consistency of the mle for ornstein-uhlenbeck and other selfdecomposable processes. *Statistical Inference for Stochastic Processes*, 19(1):29–50. ISSN 1572-9311. doi: 10.1007/s11203-015-9118-9. URL https://doi.org/10.1007/s11203-015-9118-9. From page 107, 109

Hitaj *et al.* **(2018)** Asmerilda Hitaj, Friedrich Hubalek, Lorenzo Mercuri and Edit Rroji. On properties of the mixedts distribution and its multivariate extension. *International Statistical Review*, 86(3):512–540. doi: 10.1111/insr.12265. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12265. From page 2

Holmes (2014) R. L. Holmes. Tree: PIEDRA DEL AGUILA CHILE, CLIMATIC, ARAR, 1536, Location: MALLECO, ARAR, 1300m -3750-07302, years: 1242-1975, 2014. data retrieved from https://datamarket.com/data/set/22tn. From page 68

Kawai and Masuda (2011) Reiichiro Kawai and Hiroki Masuda. On simulation of tempered stable random variates. *Journal of Computational and Applied Mathematics*, 235(8):2873 – 2887. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2010.12.014. URL http://www.sciencedirect.com/science/article/pii/S0377042710006643. From page 37, 110

Kim (2001) Woocheol Kim. Nonparametric kernel estimation of evolutionary autoregressive processes. SFB 373 Discussion Papers 2001,103, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes. URL https://ideas.repec.org/p/zbw/sfb373/2001103.html. From page 11

Kim *et al.* **(2006)** Young Shin Kim, Svetlozar Rachev, Myung Dong and Dong Chung. The modified tempered stable distribution, garch models and option pricing. *Probability and Mathematical Statistics*, 29. From page 2

Kim *et al.* **(2008)** Young Shin Kim, Svetlozar T. Rachev, Michele Leonardo Bianchi and Frank Fabozzi. Financial market models with lévy processes and time-varying volatility. *Journal of Banking & Finance*, 32(7):1363–1378. URL https://EconPapers.repec.org/RePEc:eee:jbfina:v:32:y:2008:i:7:p:1363-1378. From page 2, 32, 96

Kokoszka and Taqqu (1994) Piotr S. Kokoszka and Murad S. Taqqu. Infinite variance stable ARMA processes. *Journal of Time Series Analysis*, 15(2):203–220. doi:

10.1111/j.1467-9892.1994.tb00185.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1994.tb00185.x. From page 28, 49, 106

Kokoszka and Taqqu (1995) Piotr S. Kokoszka and Murad S. Taqqu. Fractional ARIMA with stable innovations. *Stochastic Processes and their Applications*, 60(1): 19 – 47. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(95)00034-8. URL http://www.sciencedirect.com/science/article/pii/0304414995000348. From page 28, 29, 49

Koponen (1995) Ismo Koponen. Analytic approach to the problem of convergence of truncated lévy flights towards the gaussian stochastic process. *Phys. Rev. E*, 52:1197–1199. doi: 10.1103/PhysRevE.52.1197. URL https://link.aps.org/doi/10.1103/PhysRevE.52.1197. From page 29, 31

Koutrouvelis (1981) Ioannis A. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics - Simulation and Computation*, 10(1):17–28. doi: 10.1080/03610918108812189. URL https://doi.org/10.1080/03610918108812189. From page 2, 39

Küchler and Tappe (2008) Uwe Küchler and Stefan Tappe. Bilateral gamma distributions and processes in financial mathematics. *Stochastic Processes and their Applications*, 118 (2):261 – 283. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2007.04.006. URL http://www.sciencedirect.com/science/article/pii/S0304414907000609. From page 31

Küchler and Tappe (2013) Uwe Küchler and Stefan Tappe. Tempered stable distributions and processes. *Stochastic Processes and their Applications*, 123(12):4256 – 4293. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2013.06.012. URL http://www.sciencedirect.com/science/article/pii/S0304414913001828. From page 2, 33, 109

Lombardi and Calzolari (2008) Marco J. Lombardi and Giorgio Calzolari. Indirect estimation of α-stable distributions and processes. *Econometrics Journal*, 11(1):193–208. ISSN 1368-423X. doi: 10.1111/j.1368-423X.2008.00234.x. URL http://dx.doi.org/10.1111/j.1368-423X.2008.00234.x. From page 2, 39, 61, 83

Madan and Seneta (1990) Dilip B. Madan and Eugene Seneta. The variance gamma (v.g.) model for share market returns. *The Journal of Business*, 63(4):511–524. ISSN 00219398, 15375374. URL http://www.jstor.org/stable/2353303. From page 32

Martin and Flandrin (1985) W. Martin and P. Flandrin. Wigner-ville spectral analysis of nonstationary processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(6):1461–1470. ISSN 0096-3518. doi: 10.1109/TASSP.1985.1164760. From page 8

McCulloch (1986) J. Huston McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulation and Computation*, 15(4):1109–1136. doi: 10.1080/03610918608812563. URL http://www.tandfonline.com/doi/abs/10.1080/03610918608812563. From page 2, 39

Michael (1983) John R. Michael. The stabilized probability plot. *Biometrika*, 70(1):11–17. ISSN 00063444. URL http://www.jstor.org/stable/2335939. From page 74

Mikosch *et al.* (1995) Thomas Mikosch, Tamar Gadrich, Claudia Kluppelberg and Robert J. Adler. Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics*, 23(1):305–326. ISSN 00905364. URL http://www.jstor.org/stable/2242413. From page 49

Miller (1968) K.S. Miller. *Linear Difference Equations*. Mathematics monograph series. W. A. Benjamin. URL https://books.google.com.br/books?id=1_-oAAAAIAAJ. From page 52, 54

Nolan (2002) John P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. From page 74

Olea *et al.* (2015) Ricardo Olea, Wilfredo Palma and Pilar Rubio. *LSTS: Locally Stationary Time Series*, 2015. URL https://CRAN.R-project.org/package=LSTS. R package version 1.0. From page 72

Palma *et al.* (2013) Wilfredo Palma, Ricardo Olea and Guillermo Ferreira. Estimation and forecasting of locally stationary processes. *Journal of Forecasting*, 32(1):86–96. ISSN 1099-131X. doi: 10.1002/for.1259. URL http://dx.doi.org/10.1002/for.1259. From page 19

Priestley (1965) M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2):204–237. ISSN 00359246. URL http://www.jstor.org/stable/2984191. From page 5

R Core Team (2019) R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/. From page 3

Rroji and Mercuri (2015) Edit Rroji and Lorenzo Mercuri. Mixed tempered stable distribution. *Quantitative Finance*, 15(9):1559–1569. doi: 10.1080/14697688.2014.969763. From page 2, 32

Samorodnitsky and Taqqu (1994) G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Stochastic Modeling Series. Taylor & Francis. ISBN 9780412051715. URL https://books.google.com.br/books?id=wTTUfYwjksAC. From page 22, 23

Sampaio and Morettin (2015) Jhames M. Sampaio and Pedro A. Morettin. Indirect estimation of randomized generalized autoregressive conditional heteroskedastic models. *Journal of Statistical Computation and Simulation*, 85(13):2702–2717. doi: 10.1080/00949655.2014.934244. URL https://doi.org/10.1080/00949655.2014.934244. From page 2, 39

Sampaio and Morettin (2018) Jhames M. Sampaio and Pedro A. Morettin. Stable randomized generalized autoregressive conditional heteroskedastic models. *Econometrics and Statistics*. ISSN 2452-3062. doi: https://doi.org/10.1016/j.ecosta.2018.11.002. URL http://www.sciencedirect.com/science/article/pii/S2452306218300947. From page 2, 39

Shelton Peiris and Thavaneswaran (2001a) M. Shelton Peiris and A. Thavaneswaran. On the properties of some nonstationary arma processes with infinite variance. *International Journal of Modelling and Simulation*, 21(4):301–304. doi: 10.1080/02286203.2001.11442217. URL https://doi.org/10.1080/02286203.2001.11442217. From page 49, 50, 57

Shelton Peiris and Thavaneswaran (2001b) M. Shelton Peiris and A. Thavaneswaran. Multivariate stable arma processes with time dependent coefficients. *Metrika*, 54(2): 131–138. ISSN 1435-926X. doi: 10.1007/s001840100127. URL https://doi.org/10.1007/s001840100127. From page 49, 50, 52, 58

Van Bellegem and von Sachs (2004) Sebastien Van Bellegem and Rainer von Sachs. Forecasting economic time series with unconditional time-varying variance. *International Journal of Forecasting*, 20(4):611–627. URL https://EconPapers.repec.org/RePEc:eee:intfor:v:20:y:2004:i:4:p:611-627. From page 19, 57, 58

Weron and Weron (1995) Aleksander Weron and Rafal Weron. *Computer simulation of Lévy α-stable variables and processes*, pages 379–392. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-540-44722-1. doi: 10.1007/3-540-60188-0_67. URL https://doi.org/10.1007/3-540-60188-0_67. From page 36

Whittle (1953) P. Whittle. Estimation and information in stationary time series. *Ark. Mat.*, 2(5):423–434. doi: 10.1007/BF02590998. URL https://doi.org/10.1007/BF02590998. From page 10