An Extreme Value State Space Model with Gumbel Marginals

by

© **Jayde Eustace**

A thesis submitted to the
School of Graduate Studies
in partial fulfillment of the
requirements for the degree of
Master of Science

Department of Mathematics and Statistics
Memorial University

April 2015

St. John's, Newfoundland and Labrador, Canada

Abstract

In a 2009 paper, Fougéres, Nolan and Rootzen published an interesting relationship betweeen the Gumbel and exponential stable distributions. The purpose of this research is to explore this relationship and develop a related state space model that can be used to predict and model time dependent processes with Gumbel marginals. Parameter estimation methods will be discussed, both under a simple AR(1) time series with Gumbel marginals and in the context of our proposed state space model. Since our model has a hidden component, we will then discuss filtering methods as well. Gumbel distributed extreme value processes are often found within natural systems, especially in the field of hydrology and in the study of pollution.

Acknowledgments

I would like to express my thanks to my supervisor, Dr. J.C. Loredo-Osti, for his expertise and guidance during my program. Thank you for your encouragement and for coming in to meet on Sundays when I started working full time. It was a priveledge to be able to work with you over the past few years.

Special thanks as well to Dr. Alwell Oyet, Dr. Brajendra Sutradhar and Dr. Hong Wang, whose courses I attended during my program. Your knowledge and patience were very much appreciated. Thanks also to former department head Dr. Gary Sneddon, who helped me to prepare myself for this statistics program, coming from a pure mathematics background. Thanks to Dr. Shannon Sullivan for making thesis LaTeXtemplates available on his website.

Many thanks to my Mom and Dad for your support and encouragement and for being my very first math teachers. Thanks also to my brother Jeremy for turning down his heavy metal music while I studied, and to my boyfriend Alex for always believing in me.

Finally, thank you to my managers at the Newfoundland and Labrador Statistics Agency, Greg Tucker and Heidi Ryan, for being understanding about my academic commitments and giving me the opportunity to work in my field.

This research was made possible by funding from the School of Graduate Studies and Mathematics and Statistics Department of Memorial University, as well as funding from my supervisor Dr. Loredo-Osti. Thanks as well to Associate Dean of Science Dr. Andy Foster, who helped fund my trip to the Statistical Society of Canada annual meeting in 2012.

Table of contents

List of tables

List of figures

List of abbreviations

γ_E	Euler-Mascheroni constant
CI	Confidence interval
CLS	Conditional Least Squares Method
$\exp \mathcal{S}$	Exponential stable (Exponential-$\mathcal{S}$) distribution
$\mathcal{G}$	Gumbel distribution
GEV	Generalized extreme value distribution
$\mathcal{S}$	α-stable distribution
QFS	Quasi Fisher scoring method
MM	Method of Moments
MSE	Mean squared error
YW	Yule-Walker estimation
SISR	Sequential importance sampling and resampling algorithm

List of appendices

Chapter 1

Introduction

The field of extreme value theory is becoming increasingly important as the world deals with problems where large departures from the norm are becoming more frequent, such as climate change, financial instability and engineering challenges. Often it is the extreme high and low points that we are interested in, since it is the highest waves that can sink a ship and the lowest troughs that can cause panic in the stock market.

One distribution that often arises from extremes in climatology and hydrology is the Gumbel distribution. In addition, we often may have situations with several interacting Gumbel variables, some of which may be difficult to measure. For this reason, we will seek to develop a state space model with Gumbel distributed marginals for time series processes of extremes.

Previously, Nakajima et. al. proposed an AR(1) state space model with Gumbel distributed variables and normal noise [33]. However, we will show in Section 2.8

that in such a model, the assumption of normality is not compatible with extreme value distributed marginals. Naveau and Poncet developed a state space model with Fréchet distributed marginals [34], while Fougéres, Nolan and Rootzn [13] proposed a model in which the observation equation has Gumbel distributed marginals and the state equation marginals have an α-stable distribution.

The model introduced in this thesis is an extension of a theorem due to Fougeres, Nolan and Rootzn [13] that illustrates the relationship between the Gumbel distribution and the distribution of the log of an α-stable random variable, which is called the exponential-$\mathcal{S}$ ($\exp \mathcal{S}$) distribution. That is, the sum of a Gumbel random variable and an exponential-$\mathcal{S}$ random variable with appropriately chosen parameters will also be Gumbel. This allows us to define an additive noise process of a state space model with Gumbel distributed marginals and determine the moments and other properties.

Part way through our research, we discovered that a PhD student of Naveau, Gwladys Toulemonde, had independently proposed a similar model to our own in her thesis with some minor differences [47]. However, her research focus was on the dynamics of the model and paid little attention to parameter estimation and other statistical features. In contrast, the present work will seek to determine the parameter estimation methods that work best under a modest sample size restriction. In a 2013 paper, Toulemonde, Guillou and Naveau [48] proposed another similar model with an application to air pollution data and performed analysis of filtering methods. The models of Toulemonde are detailed in Sections 2.8.4 and 2.8.5.

This report begins with a review of basic concepts, methods and applications of related extreme value theory, with special focus on the Gumbel distribution. We will

then move on to a review of time series analysis and autoregressive models.

The concept of α-stable random variables and the exponential-$\mathcal{S}$ distribution will be introduced, and the relationship between the exponential-$\mathcal{S}$ and Gumbel distributions will be discussed as well. From this relationship, our state space model can be derived, so the next step is to introduce state space models and discuss examples of such models in the existing literature.

Our state space model is comprised of both a state and observation equation (X_t and Y_t) that are defined as a mixture of Gumbel random variables and exponential-$\mathcal{S}$ noise. We will discuss the derivation of our model and its properties.

In the estimation of model parameters, we seek to find a method of estimation that works well under the constraint of small sample sizes. One important issue is that we may not have a closed form of the distribution of the noise, or be able to find an explicit expression for the joint distribution of the X_t or Y_t series. First, in the context of an $AR(1)$ time series, we will discuss the estimation methods of Yule-Walker and conditional least squares. Then we will estimate the parameters in the observation equation using the method of moments and quasi-Fisher's scoring method. Finally we will discuss filtering methods that we might use to determine the value of a hidden state variable from the observed variables.

Chapter 2

Literature review

2.1 Identifying extreme values

Extreme value theory is the study of the distributions of events whose values may deviate far from the median. Our definition of extreme may depend on the context of our application, but we are often interested in sample or time period maxima, or perhaps values over a certain threshold.

Before we can use extreme events in the past to predict what we might see in the future, we must first be able to recognize what makes an event extreme. There are several general methods for identifying extreme events in a data set.

2.1.1 Peaks over threshold (POT) method

One solution is to choose a threshold u, and define any data points greater than u as extreme events. In the early 1970s, Pickands showed that the magnitude of

these events will follow a generalized Pareto distribution, while the timing will follow a Poisson distribution [21] [20]). McNeil and Saladin (1998) extended this method to a two-dimensional point process, to describe both the frequency and severity of insurance lossses due to natural disasters [31]. More recently, Bengtsson and Nilsson (2007) used a similar method to estimate return values for wind damage in Swedish forests [4]. The POT method is often chosen over the block maxima method (discussed in the next section), as we retain more data points and we can easily correct for non-stationarity by employing a time dependent threshold. Unfortunately, extreme events often appear in clusters, which would introduce dependence into our data set. When using the POT method we must employ a declustering method to remove dependent data points (see Section 2.5). Another downside to this method is that the choice of threshold u can be very difficult and depends on the model and the context of our data. An improper choice of threshold may introduce significant error into prediction results. Scarrott and MacDonald (2012) discuss developments in threshold estimation, including methods that account for the additional error resulting from threshold estimation [41].

2.1.2 Block maxima method

Under the block maxima method we can separate our data into time blocks (years, quarters, months etc.) and choose the highest value from each time block as an extreme. The rationale is that as long as the time blocks are large enough (in the context of our application) we can assume that our extremes are independent. Jenkinson (1955) was the first to show that this resulting series of extremes will follow

one of the three generalized extreme value (GEV) distributions, Gumbel, Frèchet or Weibull, each of which has the following general form [22].

$$G(z) = \exp\left(-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right).$$

This method would usually require the assumption of stationarity, but can be extended to non-stationary models (see for example Hanel [17], who models precipitation extremes under a climate change scenerio). However, we must be very careful with our conclusions in such a situation, especially when working with n-year return values. One downfall of the block maxima method is that, since we only take one value per block, we are throwing away much of the data and may require large datasets in order to get enough data for analysis. We can extend the method to the r largest maxima in each block, although we may have to employ decoupling methods (see Section 2.5) to be able to assume independence of the r values (see Soukissian [44]).

2.2 Basic methods

After we have decided whether to use the POT or the block maxima method, we will have an idea of the type of distribution that our identified extreme values will follow. Later in the thesis, we will discuss methods of parameter estimation, but for now we will discuss our motivation. That is, what we can discover once we know the distribution of the extremes. The methods below are largely intuitive, and have been outlined in a 2010 literature review by E. Vanem [49].

2.2.1 n-Year return values

We define the **n-year return value** as the extreme value that we could expect to see only once every n years. This value is often of interest to engineers who would like to build a product that can withstand the elements for n years. For example, if a ship is meant to last for 20 years, the builders would be interested to know the largest wave that the ship might expect to withstand during that time period (the 20 year return value). This process can work in reverse as well, as we can start with an extreme value and determine how often it might occur.

2.2.2 Initial distribution method

We are interested in finding the extreme value h with probability p. That is, h_p such that $F(h_p) = p$, or $P(h \leq h_p) = p$. For an n-year return value, we would set $p = \frac{1}{n}$.

2.2.3 Quantile functions method

Let $Q(p) = x$ if $P(X \leq x) = p$. That is, x is the pth quantile of the data distribution. Imagine that we would like to find an extreme value $Q_T = x$ with return period T. Since a return period of T years is equivalent to a $\frac{1}{T}$ probability of occuring each year, $P(X > x) = \frac{1}{T}$, $P(X \leq x) = 1 - \frac{1}{T}$ and therefore $Q_T = Q\left(1 - \frac{1}{T}\right)$.

2.2.4 Mean number of upcrossings (MENU) method

We can use this method when we are interested in the return period n of a value of size y. Let Y_n be a random variable denoting the number of values in n years that

exceed y in size. Our return period will be n such that $E(Y_n) = 1$.

2.3 Generalized extreme value distribution

Early work in extreme value theory mostly focused on normally distributed data. However, in a 1927 paper, M. Fréchet was the first to show that maxima taken from data with different underlying distributions (which share certain properties) follow the same asymptotic distribution [14]. R.A. Fisher and L.H.C. Tippet published another paper the next year that extended the work of Fréchet, discovering two other asymptotic distributions that would account for initial distributions with different properties [12]. These distributions are now known as the three families of the generalized extreme value theory distribution: the Fréchet, Gumbel and Weibull distributions. The generalized extreme value (GEV) distribution denoted GEV(μ, σ, ξ) has cumulative distribution function (CDF)

$$F(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}.$$

We can take the derivative with respect to x to get the probability density function

$$f(x) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-(1/\xi)-1} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

where μ is a location parameter, $\sigma > 0$ is the scale and ξ is the shape. We require that the term $1 + \xi\left(\frac{x-\mu}{\sigma}\right)$ is non-negative, so that the ξth root will be a real number.

Note that this distribution is useful for analyzing data maxima. If we are interested in data minima (such as low rainfall) we can use a slightly altered form.

2.3.1 Families of GEV

The difference between the three families of the generalized extreme value distribution lies in the parameter ξ, which describes the shape of the distribution. These families differ in the behaviour of the tail and the value of x^+, which is the smallest x value such that $P(X \leq x) = 1$. For an in-depth discussion of these three distributions, see Coles (2001) [8].

Fréchet distribution

This was the first of the three GEV distributions to be discovered, occuring when $\xi > 0$. The Fréchet distribution has a tail that decays polynomically, and the value of x^+ is infinite. Since we know that the fraction $\frac{1}{\xi}$ exists, we can manipulate the cumulative distribution function to become

$$F(x) = \exp\left\{-\left[\frac{(x-(\mu-\sigma\gamma))}{\frac{\sigma}{\xi}}\right]^{-\frac{1}{\xi}}\right\}.$$

This can be re-parametrized to

$$F(x) = \exp\left\{-\left[\frac{(x-m)}{s}\right]^{-\gamma}\right\},$$

where m is the minimum, s is a shape parameter and $\gamma = \frac{1}{\xi}$.

Weibull distribution

The Weibull distribution occurs when $\xi < 0$ and x^+ is finite. This means that we can define an upper bound on maxima. The form of this distribution is closely related to that of the Fréchet distribution.

Gumbel distribution

The Gumbel distribution arises in the case when $\xi = 0$. In this case, the density will decay exponentially, and x^+ will be infinite. To understand how the Gumbel CDF is derived, we should recall that $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \exp(-\lambda)$. Therefore if we let integer $n = [-\frac{1}{\xi}]$, when $\xi \to 0$,

$$F(x) = \exp\left\{-\left[1-\frac{1}{n}\left(\frac{x-\mu}{\sigma}\right)\right]^n\right\} \to \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}.$$

We can then differentiate to get the pdf as well.

$$F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{2.1}$$

$$f(x) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)\exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{2.2}$$

We have arrived at the density of a Gumbel distribution with location μ and scale σ, which will be denoted $\mathcal{G}(\mu, \sigma)$.

2.4 The Gumbel distribution

Our discussion of theory will focus on the Gumbel distribution, since it is relevant to our proposed model.

2.4.1 Moments of the Gumbel distribution

Moment generating function

We will first find the moment generating function for the $\mathcal{G}(0,1)$ distribution, and then generalize.

$$\mathrm{M}_X(t) = \mathrm{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} e^{-x} \exp(-e^{-x}) dx$$

Let $U = e^{-X}$, so that $dU = -e^{-X} dX$ and $X = -\log(U)$. After applying this transformation, by the definition of the gamma function our integral becomes

$$\int e^{t(-\log(u))} e^{-u} du = \int u^{-t} e^{-u} du = \Gamma(1-t),$$

which is finite for all $t < 1$. Now, let $Y \sim \mathcal{G}(\mu, \sigma)$, so that $Y = \sigma X + \mu$.

$$M_Y(t) = \mathrm{E}(e^{tY}) = \mathrm{E}(e^{t\sigma X + t\mu}) = e^{t\mu} E(e^{t\sigma X}) = e^{t\mu}\Gamma(1-\sigma t).$$

Recall that the gamma function is defined as $\Gamma(x) = \int_0^{\infty} x^{t-1} e^{-x} dx$, and $\Gamma(n) = (n-1)!$ if n is a positive integer.

Mean and variance

The mean and variance of the $\mathcal{G}(0,1)$ distribution are known to be the following.

$$\mathrm{E}(X) = \gamma_E \tag{2.3}$$

$$\mathrm{Var}(X) = \frac{\pi^2}{6} \tag{2.4}$$

where γ_E is the Euler-Mascheroni constant, and is approximately equal to 0.57722. In this section we will show how these constants can be derived. Our first thought might be to try taking the derivatives of the moment generating function at 0.

Theorem 2.4.1 *For $x > 0$,*

$$\frac{d^n}{dx^n}\Gamma(x) = \int_0^\infty r^{x-1}e^{-r}(\log(r))^n dr.$$

A discussion and derivation of the theorem above can be found in Bashirov (2014) [3]. Now, since $\mathrm{E}(X) = \mathrm{M}'_X(0)$ and $\mathrm{E}(X^2) = \mathrm{M}''_X(0)$, for a $\mathcal{G}(0,1)$ variable, we will have

$$\begin{aligned}
\frac{d}{dt}\Gamma(1-t) &= -\int_0^\infty r^{-t}e^{-r}(\log(r))dr \\
\frac{d^2}{dt^2}\Gamma(1-t) &= -\int_0^\infty r^{-t}e^{-r}(\log(r))^2 dr.
\end{aligned}$$

These integrals may be difficult to solve, so perhaps it is better to use another method. In fact, we can use the **digamma and trigamma functions** to calculate the mean and variance in an alternate way (See Norman (1970) for a sketch of this proof [23]).

A discussion of the definitions below is available in the CRC Concise Encyclopedia of Mathematics [52].

Definition: The digamma function is defined as

$$\psi(x) = \frac{d}{dx}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}.$$

There is a result due to Gauss that can help us calculate the digamma function for rational numbers.

Theorem 2.4.2 *For rational numbers* $\frac{m}{k}$,

$$\psi\left(\frac{m}{k}\right) = -\gamma_E - \log(2k) - \frac{\pi}{2}\cot\left(\frac{m\pi}{k}\right) + 2\sum_{n=1}^{\lfloor (k-1)/2\rfloor}\cos\left(\frac{2\pi nm}{k}\right)\log\left(\sin\left(\frac{n\pi}{k}\right)\right)$$

Notice that $\psi(1) = -\gamma_E$. Now, in our case where we have $\Gamma(1-t)$, we can let $f(t) = 1 - t$, so that

$$\psi(f(t)) = \frac{d}{dt}\log\Gamma(f(t)) = \frac{\Gamma'(f(t))}{\Gamma(f(t))}f'(t).$$

When $t = 0$,

$$\psi(1) = \frac{d}{dt}\log\Gamma(1) = \frac{\Gamma'(1)}{\Gamma(1)}(-1) = -\Gamma'(1).$$

Therefore we find that $\mathrm{E}(X) = \mathrm{M}'_X(0) = \Gamma'(1) = -\psi(1) = \gamma_E$. To calculate the variance, we need to find the second derivative of the moment generating function. Before we proceed we must first introduce the **trigamma function**.

Definition: The trigamma function is defined as

$$\psi_1(x) = \frac{d^2}{dx^2} \log \Gamma(x) = \frac{d^2}{dx^2} \psi(x).$$

In our case with $\mathrm{M}_X(t) = \Gamma(1-t)$,

$$\begin{aligned} \psi_1(1-t) &= \frac{d}{dt} \frac{(\Gamma'(1-t))(-1)}{\Gamma(1-t)} \\ &= \frac{\Gamma''(1-t)\Gamma(1-t) - (\Gamma'(1-t))^2}{\Gamma(1-t)^2}. \end{aligned}$$

Now, consider the case in which $t = 0$.

$$\begin{aligned} \psi_1(1) &= \frac{\Gamma''(1)\Gamma(1) - (\Gamma'(1))^2}{\Gamma(1)^2} \\ &= \Gamma''(1) - (\Gamma'(1))^2 \\ &= \mathrm{M}''_X(0) - (\mathrm{M}'_X(0))^2 \end{aligned}$$

Therefore $\mathrm{Var}(X) = \psi_1(1)$, which is known to be $\frac{\pi^2}{6}$. If we have a variable $Y \sim \mathcal{G}(\mu, \sigma)$, we can write $Y = \mu + \sigma X$, for some $X \sim \mathcal{G}(0, 1)$. Therefore we can easily derive the moments of a more general Gumbel variable.

$$\mathrm{E}(Y) = \mu + \sigma\gamma_E \tag{2.5}$$

$$\mathrm{Var}(Y) = \frac{\sigma^2\pi^2}{6} \tag{2.6}$$

2.4.2 Relationship between GEV and Gumbel distributions

Since the Gumbel distribution is just a special case of GEV, we are able to transform from one to the other.

Theorem 2.4.3 *If* $X \sim \mathcal{G}(0,1)$, *then* $Y = \mu + \sigma\left(\frac{e^{\xi X}-1}{\xi}\right) \sim \text{GEV}(\mu, \sigma, \xi)$.

Proof Since $X \sim \mathcal{G}(0,1)$, we know that $P(X \leq x) = \exp(-\exp(-x))$.

$$\begin{aligned} P(Y \leq y) &= P\left(\mu + \sigma\left(\frac{e^{\xi X}-1}{\xi}\right) \leq y\right) \\ &= P\left(X \leq \frac{1}{\xi}\log\left(\xi\left(\frac{y-\mu}{\sigma}\right)+1\right)\right) \\ &= \exp\left(-\exp\left(-\frac{1}{\xi}\log\left(\xi\left(\frac{y-\mu}{\sigma}\right)+1\right)\right)\right) \\ &= \exp\left(-\left(1+\xi\left(\frac{y-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right) \end{aligned}$$

We end up with precisely the CDF of a $\text{GEV}(\mu, \sigma, \xi)$ distribution.

2.4.3 Relationship between exponential and Gumbel distributions

Recall that the Gumbel distribution is a member of the generalized extreme value distribution family, and therefore arises as a distrubtion of block maxima. In fact, E.J. Gumbel was able to show that the Gumbel distribution is especially useful when the underlying data is exponentially distributed [16]. The following results illustrate the connections between these two distributions.

Theorem 2.4.4 *Let $X_1, X_2 \ldots X_n$ be independent, identically distributed exponential random variables, and let Y_n be the nth order statistic. The distribution of $Y_n - \log(n)$ will approach a Gumbel distribution as the sample size n approaches infinity.*

Proof Let $f(x)$ and $F(x)$ be the density and cumulative distribution of X_k, $k = 1, 2, \ldots, n$ and recall that the distribution of the nth order statistic is $P(Y_n \leq x) = F(x)^n$, so that $f_{Y_n}(x) = n[1 - F(x)]^{n-1} f(x)$.

Under an Exp(1) distribution, $f(x) = \exp(-x)$ and $F(x) = 1 - \exp(-x)$, and therefore $P(Y_n \leq x) = (1 - \exp(-x))^n$.

$$\begin{aligned} P(Y_n - \log(n) \leq x) &= P(Y_n \leq x + \log(n)) \\ &= F_n(x + \log(n)) \\ &= (1 - \exp(-x - \log(n)))^n \\ &= \left(1 - \frac{\exp(-x)}{n}\right)^n \end{aligned}$$

Taking the limit, we conclude that

$$P(Y_n - \log(n) \leq x) \;\; \rightarrow \;\; \exp(-\exp(-x)) \text{ as } n \rightarrow \infty, \tag{2.7}$$

which leaves us with the cumulative distribution of a $\mathcal{G}(0, 1)$ random variable (see [42] for a sketch of this proof).

Theorem 2.4.5 *If $X \sim \text{Exp}(1)$ then $-\log(X) \sim \mathcal{G}(0, 1)$.*

Proof Define $Y = -\log(X)$, and let $F_X(x) = \exp(-x)$ and $F_Y(y) = f_X(\exp(-y)) =$

$\exp(\exp(-y))$. This is just the density function of a $\mathcal{G}(0,1)$ variable.

Additional exploration of the relationship between the Gumbel distribution and other distributions can be found in a paper by Ojo (2001) [36].

2.4.4 Applications of the Gumbel distribution

As a member of the generalized extreme value distribution family, the Gumbel distribution can be used to model independent block maxima in certain situations. Due to the relationship between the Gumbel and exponential distributions outlined in Section 2.4.3, many of the applications involve exponential processes. One important application is in hydrology: In 1973, Leese used the Gumbel distribution to model annual maximum flooding [28]. Waylen and Woo extended this research to show that when flooding results from different sources (such as rainfall and snow melt), one must estimate the Gumbel distributions for each source separately and then compound them [50].

Another application is in queuing theory. If we have a Poisson process, such as the number of customers arriving at a store, the amount of time between occurrences will be exponentially distributed [9]. Therefore the block maxima of this process can be modelled using a Gumbel distribution (see Asmussen (1998) [2] for a discussion of these methods). Nakajima et al. model extreme returns of daily stock data using a Gumbel distribution [33]. Longin has shown that stock market data sets are more closely modelled using a Fréchet distribution [29], but it is possible that the Gumbel distribution may be useful for some kind of financial application.

2.5 Decoupling methods

When using the peaks over threshold or r-largest maxima models, or when using one of the GEV distributions, we normally require our data to be independent. However, it is common in natural processes to find several extreme values clustered together. In order to be able to use standard techniques developed under the assumption of independence, we can use one of the following techniques, described in Soukissian (2011), to account for clustering [44].

2.5.1 Standard storm length

The standard storm length method, developed by Tawn in 1988 [46], seeks to identify dependent clusters of extreme values and extract the highest value from each as our independent set of extremes. Define a "storm" as a period in which our variable of interest is more active than usual and is producing dependent extremes. To avoid including two dependent maxima, we may require a minimum time period, k, between events. If we require r maxima, we can start with the highest value, H_1, and remove data within $k/2$ time periods. Then we can select the second highest value, H_2, and repeat until we have r data points. Under the POT method, we can perform a similar process but we would stop once we reach a value H_i that drops below u, our chosen threshold. One possible issue with this method is that the assumption that all storms have the same length k is unreasonable. Our chosen value of k might be too small for some storms, making the procedure ineffective in removing dependent values from our data. Otherwise k could be too large, in which case we would lose data, increasing

the variance of our estimates.

2.5.2 Runs declustering

The runs declustering method was proposed by Smith in 1989 [43]. This method has an advantage over the standard storm length method, as it allows for varying storm times. Assume that we have a process H_t, in which values that exceed a threshold of u are considered extreme. We can choose a run-length k, and define a cluster to begin with the first value H_i that exceeds u. The cluster ends only after we observe k consecutive values less than u. We can then take the maximum value from each of the clusters.

Although there is some improvement in the flexibility of this model, as previously we may run into some difficulty with the selection of u and k.

2.5.3 Declustering algorithm (DeCa)

This algorithm was developed by Soukissian and Kalantzi (2009) in the context of ocean waves, but has several other applications as well [45]. The idea behind the declustering algorithm is similar to the runs declustering method, in that we think of our process as a series of storms. The process is as follows:

- Define each storm to end when the underlying state is reduced below a certain threshold u. In the context of waves, this would be the "sea state".
- Perform noise reduction and filtering if necessary.

- Use the threshold to separate the data into independent storms.
- Take the maximum of each storm, to form a set of independent extremes.

This algorithm could be modified to fit other situations as well. For example, if we were measuring snow depth, we might require the snow to melt down completely between independent extremes ($u = 0$ cm).

2.5.4 Choice of parameters

In many of the methods above, we require a parameter such as a threshold u, run length k or number of maxima r. We must be careful with our choices, as a value that is too large or too small can cause us to exclude good data (independent maxima) or include dependent maxima. Since we are usually trying to model a physical situation, we can often use common sense and scientific explanation to choose appropriate values. Another method would be to try several choices for our parameters and check to see which produce stable estimates and low variance. Checking the fit of our assumed underlying model (often generalized extreme value or the generalized Pareto distribution) can also be helpful. When choosing a run-length k value, we can look at the autocorrelation structure of the data, although we may need to remove any trend and seasonality first.

2.6 Time series processes

The natural alternative to declustering is to try to model the dependence of the extreme process. For this reason, researchers have been interested in studying extremes as time-dependent processes. Many of the extremes that we see in nature, such as ocean waves, rainfall and temperature, arise from dependent time series. Up until the mid 1970s the literature focused on independent extremes, often employing declustering methods to transform data into this format. Early papers include Leadbetter (1986), who examines the properties of extremes in a stationary sequence where declustering has not been performed and concludes that the limiting distribution of the maxima will be unchanged [27]. To deal with this sort of data, new methods that do not require independence will be necessary. Before we discuss these methods we will first touch on the theory of time series processes. A time series is a data set in which each observed value is associated with a point in time. In the analysis of extreme data we are often interested in how a variable changes with time, usually so that we can predict what we should expect in the future. A good discussion of time series concepts is available in Brockwell and Davis [6].

2.6.1 Basic additive auto-correlated time series model

Let X be a process, and $X_1, \ldots, X_T$ be the observations at time $t = 1, \ldots, T$. We call this series of observations a **time series**, often denoted $\{X_t\}$.

Definition: In this work, we will say that a time series $\{X_t\}$ **stationary** if

1. $\mathrm{E}|X_t|^2 < \infty$ for all $t = 0, 1, \ldots$.

2. $\mathrm{E}(X_t) = \mu$ is constant for all t.

3. $\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h)$ for all $t, h \in \mathbb{Z}$, where $\gamma(h)$ is the lag-h autocovariance function.

A time series that does not have these three qualities is said to be **non-stationary**. In particular, a series with a mean that varies with time is said to have a **trend**.

2.6.2 Modelling trend

A discussion of how to model trend can be found in Brockwell and Davis [6]. Trend in the context of extreme value theory has become an important issue due especially to the threat of climate change.

Least squares estimation

Let $m(t)$ be the function that represents our trend. A simple time series model with trend could be written as follows.

$$X_t = m(t) + \epsilon_t \tag{2.8}$$

We can get an idea of what shape this function m might take (linear, exponential, quadratic etc.) by observing a graph of the values. For example, if our trend is linear, we may set m to be

$$m(t) = a_0 + a_1 t.$$

Our goal here will be to minimize

$$\sum_t (x_t - m(t))^2 \tag{2.9}$$

with respect to the parameters of m, which are a_0 and a_1 in our linear case. The benefit of this method is that we are not just removing the trend but also identifying the trend. This may enable us to predict future values.

Smoothing by moving average

Random fluctuations in time series data may obscure the true trend. For this reason, we may want to smooth the data using the moving average method. If we have a time series $X_1, \ldots, X_n$, the k-point moving average at time t is given by

$$\tilde{X}_t \ = \frac{1}{k} \sum_{i=t-j}^{t+j} X_i$$

where k is an odd integer and $j = \frac{k-1}{2}$. In other words, we are taking an average of the j values to either side of X_t.

Differencing

The method of differencing can be used in situations where it is not necessary to develop a model for the trend, rather we would just like to remove trend from the data. Let B be the backward shift operator, $BX_t = X_{t-1}$ and $B^i X_t = X_{t-i}$. Also

note that

$$X_t - X_{t-1} = X_t - BX_t = (1 - B)X_t.$$

The idea behind differencing is that if we have a polynomial trend of degree j, we can apply the operator $(1 - B)^j$ to X_t to remove the trend. For example, if the data has a linear trend, then we would set $j = 1$. Note that we can also use this method to remove seasonality. If we have a seasonal cycle with period d, we can use the operator $(1 - B^d)X_t$ to transform the data. If, for example, we are looking at monthly rainfall data, we would say that $d = 12$. When we perform the transformation, we will obtain $X_t - X_{t-12}$, which just means that we subtract the value from the same month in the previous year. In the following subsection we will detail some other methods of dealing with seasonality in our data.

2.6.3 Modelling seasonality

In many time series datasets, especially those involving natural processes, it is common to see a seasonal pattern emerge. For example, if we were looking at rainfall data for a specific location, we might notice that there is a rainy season and a dry season each year. A time series with seasonality may have the form

$$X_t = m(t) + s(t) + \epsilon_t \tag{2.10}$$

which is similar to Equation (2.8) with an additional seasonal component $s(t)$.

There are several ways to deal with seasonality in our data.

Removing data points

One simple method that we can use to avoid dealing with seasonality is to restrict our data set to only the yearly maxima. In this way we are guaranteed to avoid seasonality, however we lose a significant amount of data. Gilli and Kellezi (2006) use this method to model financial risk, but have a data set that spans several decades [15]. Another option would be to include only the more active months of the year in the data set, or model different seasons separately (see, for example Morton (1997) [32]). We are generally interested in absolute extremes rather than relative extremes; a rainfall during the dry season might be relatively extreme, but could still be less than an average rainfall during the rainy season. This method is not always the best, as we lose data, the remaining months might still have a slight seasonal pattern and we ignore extremes that may occur at unusual times of the year.

Seasonal adjustment

We can estimate seasonality using a sine and cosine model, then subtract it away from the data to remove its effect. The coefficients in the model below can easily be estimated using the *lm* function in **R**.

$$y_t = \alpha_0 + \alpha_1 \sin\left(\frac{2\pi t}{365}\right) + \alpha_2 \cos\left(\frac{2\pi t}{365}\right) + \epsilon_t \tag{2.11}$$

A discussion of this method can be found in Brockwell and Davis [6]. Again, in the context of extreme values, we may not want to remove seasonality because we are usually interested in absolute extremes. After modelling seasonality, it may be useful to set up a state space model, with our observed value of interest modelled as a function of seasonality, noise and other factors.

Season-dependent parameters

Another method of dealing with seasonality is to use season-dependent parameters in our model. For example, Rust et al. (2009) used this method to model monthly maximums of daily precipitation across the United Kingdom [39]. This data set was assumed to have a generalized extreme value distribution with seasonally dependent parameters.

2.6.4 Autoregressive and moving average models

In time series analysis, we are often interested in how current observations can be modelled in terms of the past. Assume that we have some stationary dataset $\{X_t\}$, and we would like to predict future values using our current information. In an **autoregressive** or AR(p) model, the current observation X_t is described in terms of the p previous obervations and a noise term Z_t, where $Z_t \sim \text{IID}(0, \sigma^2)$ (independent and identically distributed).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t \tag{2.12}$$

Conversely, in a **moving average** or MA(q) model, our latest observed value depends on the previous noise terms and some mean μ.

$$X_t = \mu + Z_t - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q} \tag{2.13}$$

In addition, we can have a model in which the observed value depends on a combination of past values and past noise. This is known as an **autoregressive-moving average** or ARMA(p, q) model.

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q} \tag{2.14}$$

How might we know which model is the best for a dataset? Before we answer this, we must first introduce a few concepts.

2.6.5 Autocorrelation function

Since our choice of model seems to depend on how strongly an observation is related to those in the past, it makes sense that we would want to look at the correlation between observations. Let $\gamma(t, t+h) = \text{Cov}(X_t, X_{t+h})$. Since $\{X_t\}$ is stationary this function is independent of time, so we can say that

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) \text{ for all t.} \tag{2.15}$$

This means that the covariance will be the same for any two values of $\{X_t\}$ that are h time units apart. We can use this to calculate our **autocorrelation function** or

ACF, $\rho(h)$, by dividing by the variance, $\gamma(0)$.

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} \tag{2.16}$$

The behaviour of the ACF is as follows.

- In an AR(p) model, X_t has some correlation with all of the past observations. For example, X_{t-1} depends on X_{t-p-1}, so therefore X_t has correlation with X_{t-p-1}. The correlation function should slowly tail off as we move away from time t.
- In the MA(q) model, we should expect X_t to only be correlated with the previous q values through their dependence on $Z_{t-1}, \dots Z_{t-q}$. The plot of $\rho(h)$ should drop off abruptly after lag q.

2.6.6 Partial autocorrelation function

In the section above, we noticed that X_t was correlated with all of the previous observations through cross correlation. The **partial autocorrelation function** (PACF), $p(h)$, measures the correlation between X_t and X_{t-h} after disregarding the cross correlation effect from the observations between the two.

$$p(h) = \text{Corr}(X_t, X_{t-h} | X_{t-1}, \dots, X_{t-h+1}) \tag{2.17}$$

For the PACF function:

- In an AR(p) model, X_t will only be correlated with $X_{t-1} \dots X_{t-p}$ since cross correlation has been discarded. The plot of $p(h)$ should drop off after lag p.
- In an MA(q) model, the plot of $p(h)$ will decay slowly.
- In an ARMA(p, q) model, the ACF will decay exponentially after lag $p-q$, and the PACF after lag $q-p$.

Generally if we have checked the ACF and PACF of our model and can't identify a clear cut-off in either, then it is best to use an ARMA model.

2.6.7 Choice of a model

We now have several guidelines to help us choose the correct model. However, sometimes irregularities in the data can make it difficult to see which p and q values we should use. The **Akaike information criterion** (AIC) can help us choose from several possible models.

$$\mathrm{AIC}(k) = n \log(\hat{\sigma}^2) + 2k \tag{2.18}$$

Note that k is the number of parameters in the model. In choosing between several models, the model with the lowest AIC is often the best choice, as we would like to minimize both variance σ^2 and number of parameters. In **R**, the *arima* function automatically calculates the AIC when fitting an arma type model.

2.6.8 Properties of the AR(1) model

Since it will be important to the model that we will later propose, we will end our time series discussion with a short section on the AR(1) model and its properties. This model has the general form

$$X_t = c + \phi X_{t-1} + Z_t, \tag{2.19}$$

Where ϕ is a parameter, c is constant and $Z_t \sim \text{IID}(0, \sigma^2)$. Notice that this model is an example of a Markov process, since X_t is only dependent on previous values $X_{t-1}, \ldots, X_0$ through X_{t-1}. The AR(1) model is also an example of a random walk. If Z_t has a constant but non-zero mean, we can absorb the mean of the noise into the constant c and continue as usual.

Mean, variance and covariance

Since we know that $\text{E}(Z_t) = 0$, for any $t' < t$ it will hold that

$$\begin{aligned} \text{E}(X_t) &= c + \phi \text{E}(X_{t-1}) \\ &= \phi^{t-t'} \text{E}(X_{t'}) + c \sum_{i=t'}^{t-1} \phi^{i-t'} \\ &= \phi^{t-t'} \text{E}(X_{t'}) + c \left(\frac{1 - \phi^{t-t'}}{1 - \phi} \right). \end{aligned} \tag{2.20}$$

The mean here is not always constant. In fact, AR(1) data is non-stationary unless $\text{E}(X_t) = \frac{c}{(1-\phi)}$, that is, unless the process started long ago in the past. We continue

on to find that the variance is

$$\begin{aligned}\mathrm{Var}(X_t) &= \phi^2 \mathrm{Var}(X_{t-1}) + \sigma^2 \\ &= \phi^{2(t-t')}\mathrm{Var}(X_{t'}) + \sigma^2 \sum_{i=t'}^{t-1} \phi^{2(i-t')} \\ &= \phi^{2(t-t')}\mathrm{Var}(X_{t'}) + \sigma^2 \left(\frac{1-\phi^{2(t-t')}}{1-\phi^2}\right). \end{aligned} \tag{2.21}$$

The variance of this process will only be constant if $\mathrm{Var}(X_t) = \frac{\sigma^2}{(1-\phi^2)}$. Finally, we move on to the covariance. This calculation is simplified due to the fact that the noise terms are independent of the X_t series. Thus, notice that for any positive integer $k < t$ we can write $X_t = c\left(\frac{1-\phi^k}{1-\phi}\right) + \phi^k X_{t-k} + \sum_{i=0}^{k-1} \phi^i Z_{t-i}$. It follows that

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t-k}) &= \mathrm{Cov}(\phi^k X_{t-k}, X_{t-k}) \\ &= \phi^k \mathrm{Var}(X_{t-k}). \end{aligned} \tag{2.22}$$

If the variance is constant, then the covariance will only depend on the lag distance k between X_t and X_{t-k}.

Moment estimation

In the case where the mean and variance are constant and the covariance only depends on the time distance, we can rearrange the equations for $\mathrm{E}(X_t)$, $\mathrm{Var}(X_t)$ and

$\mathrm{Cov}(X_t, X_{t-k}) = \gamma(k)$ to obtain estimates for ϕ, c and σ^2.

$$\phi = \frac{\gamma(1)}{\mathrm{Var}(X_t)}, \tag{2.23}$$

$$c = \mathrm{E}(X_t)(1-\phi), \tag{2.24}$$

$$\sigma^2 = \mathrm{Var}(X_t)(1-\phi^2). \tag{2.25}$$

In the equations above, since the mean and variance are assumed to be constant we can use the sample mean and covariance in place of $\mathrm{E}(X_t)$ and $\mathrm{Var}(X_t)$ respectively, and the lag-1 sample autocovariance in place of $\gamma(1)$.

2.7 The α-stable random variable

A random variable X is said to be **stable** (or have stable distribution) if a linear combination of any two independent random variables that share the distribution of X will also share the same distribution, although possibly with different parameters. For example, the normal distribution is considered to be stable, and therefore the sum of two normal random variables will also be normal. Let X be a stable random variable whose Laplace transform is given by

$$\mathrm{E}(e^{-tX}) = e^{-t^\alpha}. \tag{2.26}$$

We denote the distribution of X by $\mathcal{S}(\alpha)$, and write $X \sim \mathcal{S}(\alpha)$. We will say

that X is α-stable. The standard normal distribution is an example of an α-stable distribution with $\alpha = 2$. In this project we will concentrate on positive random variables whose α values are between 0 and 1. In Figure 2.1 we can see that as α becomes closer to 1, the density of an α-stable random variable will become more mound shaped.

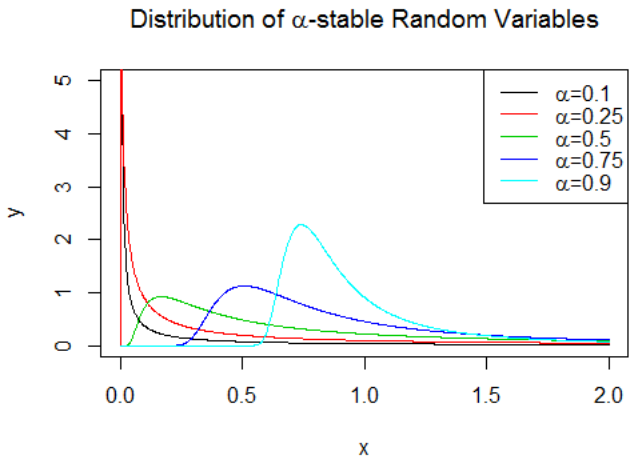

Figure 2.1: Density plots for α-stable random variables at varying levels of α.

2.7.1 Characteristic function of an α-stable random variable

The characteristic function is useful because it always exists and it uniquely determines the distribution of our data. Although we may not know the closed form of the distribution of an α-stable random variable X, we may be able to use the characteristic function $\varphi_X(t)$ to gain useful information.

$$\varphi_X(t) = \mathrm{E}(\exp(itX))$$

We know by Equation (2.26) that $\mathrm{E}(\exp(-uX)) = \exp(-u^\alpha)$ when X is α-stable. If we let $u = -it$, it follows that

$$\varphi_X(t) = \mathrm{E}(\exp(-itX)) = \exp(-(-it)^\alpha) \tag{2.27}$$

Consider a complex number, $z = x + i * y$. In polar coordinates, $z = re^{i\theta}$, where $r = |z|$ and $\theta = arg(z)$. Using De Moivre's formula we can find that

$$\begin{aligned} z^\alpha &= r^\alpha e^{i\theta\alpha} \\ &= r^\alpha(\cos(\theta\alpha) + i\sin(\theta\alpha)). \end{aligned}$$

Now, when $z = -it$, $r = \sqrt{x^2 + y^2} = \sqrt{(-t)^2} = |t|$ and $\theta = \arg(z) = \frac{-\pi}{2}$.

$$\begin{aligned} (-it)^\alpha &= |t|^\alpha e^{i\frac{-\pi}{2}\alpha} \\ &= |t|^\alpha \left(\cos\left(\frac{-\pi}{2}\alpha\right) + i\sin\left(\frac{-\pi}{2}\alpha\right)\right) \\ &= |t|^\alpha \left(\cos\left(\frac{\pi}{2}\alpha\right) - i\sin\left(\frac{\pi}{2}\alpha\right)\right) \\ &= |t|^\alpha \cos\left(\frac{\pi}{2}\alpha\right)\left(1 - i\tan\left(\frac{\pi}{2}\alpha\right)\right). \end{aligned}$$

We can use this to find the characteristic function of an α-stable random variable.

$$\varphi_X(t) = \exp(-(-it)^\alpha) = \exp\left(-|t|^\alpha \cos\left(\frac{\pi}{2}\alpha\right)\left(1 - i\tan\left(\frac{\pi}{2}\alpha\right)\right)\right). \tag{2.28}$$

2.7.2 Four parameter stable random variable

There is a more general parametrization of stable random variables proposed by Samorodnitsky (1994) [40] whose characteristic function is given by

$$\varphi(t;\alpha,\beta,\sigma,\mu) = \exp(-\sigma^\alpha|t^\alpha|(1 - i\beta\tan(\frac{\pi}{2}\alpha)\,\mathrm{sgn}(t) + i\mu t)) \tag{2.29}$$

where $\mathrm{sgn}(t) = 1$ when $t >= 0$ and -1 when $t < 0$.

By comparison of Equations (2.28) and (2.29), we can see that under the Samorodnitsky parametrization for an α-stable random variable X, the parameter α will be preserved, $\beta = \mathrm{sgn}(t)$, $\mu = 0$ and $\sigma = (\cos(\frac{\pi}{2}\alpha))^{\frac{1}{\alpha}}$.

2.7.3 Exponential-$\mathcal{S}$ distribution

Let $S_t \sim \mathcal{S}(\alpha)$. The log of S_t follows an exponential stable (or exponential-$\mathcal{S}$) distribution, denoted by $\exp\mathcal{S}(0,1;\alpha)$, where $0 < \alpha < 1$.

Theorem 2.7.1 *Let S be a standard positive α-stable random variable ($S \sim \mathcal{S}(\alpha)$) . If $M = \mu + \sigma\log(S)$, then $M \sim \exp\mathcal{S}(\mu,\sigma;\alpha)$.*

The exponential-$\mathcal{S}$ and Gumbel distributions have a special relationship. From Fougères et al. [13],

Theorem 2.7.2 *If we have independent $G \sim \mathcal{G}(\mu_1,\sigma)$ and $M \sim \exp\mathcal{S}(\mu_2,\sigma;\alpha)$, then*

$$G + M \sim \mathcal{G}(\mu_1 + \mu_2, \frac{\sigma}{\alpha}).$$

Note that we require that the Gumbel and exponential-$\mathcal{S}$ variables must have the same scale parameter. We can use these ideas to set up a state space model for use with more general Gumbel data.

2.7.4 Gumbel AR(1) model

Theorem 2.7.3 *For $\alpha \in (0,1)$ define the stochastic process*

$$X_{t+1} = \alpha X_t + \alpha \log S_{t+1}$$

where the S_t are independent and identially distributed, and $S_t \sim \mathcal{S}(\alpha)$ for all t in $\mathbb{Z}$. Then $X_t = \sum_{j=1}^{\infty} \alpha^j \log S_{t-j+1}$ and $X_t \sim \mathcal{G}(0,1)$ for all t.

Proof Suppose that $X_{t-1} \sim \mathcal{G}(0,1)$, so that $X_t|S_t$ is Gumbel.

$$\begin{aligned}
P(X_t \leq x) &= \mathrm{E}\left(P(\alpha X_{t-1} + \alpha \log S_t \leq x | S_t)\right) \\
&= \mathrm{E}\left(\exp\left(-\exp\left(-\frac{x - \alpha\log(S_t)}{\alpha}\right)\right)\right) \\
&= \mathrm{E}\left(\exp\left(-\exp\left(-\frac{x}{\alpha}\right)\exp(\log(S_t))\right)\right) \\
&= \mathrm{E}\left(\exp\left(-S_t \exp\left(-\frac{x}{\alpha}\right)\right)\right) \\
&= \exp\left(-\exp\left(-\frac{x}{\alpha}\right)^{\alpha}\right) \\
&= \exp\left(-\exp(-x)\right)
\end{aligned}$$

Note that we end up with the distribution of a $\mathcal{G}(0,1)$ distribution. This result agrees with Theorem 2.7.2.

Now that we have reviewed the necessary background of time series estimation and extreme value methods, we have an understanding of how we might model dependent extremes. We can now move onto the type of model that we will be examining in this thesis, the state space model.

2.8 State space models in extreme value theory

A state space model consists of two components - the state equation and the observation equation. It is useful for modelling processes in which an observed set of values is directly influenced by an underlying and often hidden process. For example, the fluctuation of a single stock is usually related to the behaviour of the entire market, and the values of individual waves are dependent on the sea state. In such contexts it is possible for extremes to result from high spikes in either the state or observed processes, or the additive effect of moderate spikes in both.

One benefit of this type of model is that rather than removing trend or seasonality, we can build those attributes into the state component so that our predictions take that information into account. There have been several authors in the past decade who have explored the use of state space models in an extreme value context. Early work was undertaken by West et al. (1985) who developed a Bayesian model for use in nonlinear, non-normal time series [53]. In 2001, Coles presented an approach to the modelling of time-dependent parameters using the GEV(μ_t, σ, ξ) distribution, expressing the mean μ_t as a deterministic function of time. However, the observations were still assumed to be independent [8]. Huerta and Sansó (2007) explored a variation

on Coles approach, replacing the independence assumption with an assumption of conditional independence given μ_t. The mean μ_t would then be expressed as the observation equation of a linear state space model dependent on an AR(1) state vector θ_t [19]. We will now describe some of the more relevant models in greater detail.

2.8.1 Nakajima state space model (2011)

It is common to model stock market indicators (including the Dow Jones industrial average) with an AR(1) model. Nakajima et al. have proposed the following state space model for use in financial applications, which has an AR(1) state equation [33]. Let $y = \{y_1, \ldots, y_n\}$ be a series of extreme values, and define

$$\begin{aligned} y_t &= \mu + \psi\left(\frac{\exp(\xi\alpha_t) - 1}{\xi}\right) + \epsilon_t, \\ \alpha_{t+1} &= \phi\alpha_t + \eta_t \end{aligned}$$

Where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ and η_t is i.i.d. noise, $\eta_t \sim \mathcal{G}(0, 1)$. In addition, μ, ψ and ξ are parameters such that $\psi > 0$ and $1 + \xi\frac{(y_t - \mu)}{\psi} > 0$, and ϕ is a real number coefficient in the state equation. The state variable α_t is Gumbel distributed. Note that if $\alpha_t \sim \mathcal{G}(0, 1)$ then $y_t|\epsilon_t \sim \text{GEV}(\mu + \epsilon_t, \psi, \xi)$.

It is also assumed that the first value, α_1 is generated by a normal distribution with mean and variance similar to the stationary distribution of α_t. It is clear that $\{\alpha_t\}$ form an AR(1) process, but the authors have also extended their model to the ARMA(p, q) case. There are some theoretical issues with this model, as the authors

claim to be modelling a process with extreme value marginals. If $\alpha_1 \sim \mathcal{N}(\mu, \sigma^2)$ and $\eta_t \sim \mathcal{G}(\theta, \beta)$ with α_1 independent of η_1, what is the distribution of α_t? Consider the moment generating functions,

$$\begin{aligned} \mathrm{M}_{\alpha_1}(r) &= \exp\left(\mu r + \frac{1}{2}\sigma^2 r^2\right) \\ \mathrm{M}_{\eta_t}(r) &= \Gamma(1-\beta r)\exp(\theta r). \end{aligned}$$

Now, since $\alpha_{t+1} = \phi\alpha_t + \eta_t$,

$$\alpha_{t+1} = \phi^t \alpha_1 + \sum_{j=1}^{t} \phi^{j-1}\eta_{t+1-j}$$

$$\begin{aligned} \mathrm{M}_{\alpha_{t+1}}(r) &= \mathrm{E}[\exp(\alpha_{t+1}r)] \\ &= \mathrm{E}[\exp(\phi^t\alpha_1 r + \sum_{j=1}^{t}\phi^{j-1}\eta_{t+1-j})] \\ &= \mathrm{E}[\exp(\phi^t\alpha_1 r)\prod_{j=1}^{t}\exp(\phi^{j-1}\eta_{t+1-j}r)] \\ &= \exp(\mu\phi^t r + \frac{1}{2}\sigma^2\phi^{2t}r^2)\prod_{j=1}^{t}\Gamma(1-\beta\phi^{j-1}r)\exp(\theta\phi^{j-1}r) \end{aligned}$$

This argument shows that it cannot be claimed that the marginals of y_t follow a GEV distribution.

2.8.2 Naveau-Poncet Fréchet model (2007)

Naveau and Poncet [34] proposed the following model

$$Y_t = \max\{F_t X_t, \epsilon_t\}$$
$$X_t = \max\{G_t X_{t-1}, \eta_t\}$$

where $\{\epsilon_t\}$ and $\{\eta_t\}$ are two independent sequences of i.i.d. random variables with a Fréchet distribution, while F_t and G_t are coefficient vectors with $F_t > 0$ and $G_t > 0$.

In this case $\{Y_t\}$ is a Fréchet stochastic process that represents the observations with Fréchet marginals. However, under this model ϵ_t cannot be interpreted as a noise process, so its nature as a source of variation is not quite clear. Although this model could be useful in setting boundaries for extreme values, estimation of parameters may be difficult. Prediction and filtering may be difficult for this model as well.

2.8.3 Fougères et al. model (2009)

Fougères, Nolan and Rootzén proposed a model that arises from the relationship between the Gumbel and exponential-$\mathcal{S}$ distributions [13].

For $t = 0, 1, 2, \ldots, T$

$$Y_t = \psi_t \ \log X_t + \epsilon_t$$
$$X_t = \phi_t X_{t-1} + S_t$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. $\mathcal{G}(\mu_t, \psi_t)$ noise and $\{S_t\}$ is a sequence of i.i.d. α-stable

noise. Thus Y_t is Gumbel and X_t follows an α-stable distribution, in fact,

$$X_t = \sum_{i=0}^{t} c_{t,i} S_i$$

with $c_{t,t} = 1$ and $c_{t,i} = \prod_{k=i+1}^{t} \phi_k$.

The authors applied this model to an engineering analysis of corrosion depth in car aluminum, using the method of maximum likelihood to estimate model parameters. A similar model appears in Naveau and Poncet, 2007 [34].

2.8.4 Toulemonde model (2008)

Part way through our research, we came across an online presentation detailing a model quite similar to our own. Further investigation brought us to the 2008 doctoral thesis of Gwladys Toulemonde, a mathematical statistics student of P. Naveau at L'Université Paris [47].

Autoregressive model for the Gumbel distribution

This model is similar to the model that will be presented in this thesis, in that it uses the fact that under certain conditions, the sum of a Gumbel and an $\exp \mathcal{S}$ random variable has a Gumbel distribtuion. The difference lies in the state equation, which is an α-stable random variable in this case.

$$Y_t = \mu + X_t + \sigma \log(\epsilon_t) \tag{2.30}$$

$$\epsilon_t = \rho \epsilon_{t-1} + S_t \tag{2.31}$$

Note that $\epsilon_t = \sum_{i=1}^{t} \rho^i S_{t-i}$, where $\{S_t\}$ is a series of α-stable random variables. By definition, a linear combination of α-stable random variables is also α-stable. The author claims that $Y_t \sim \mathcal{G}(\mu, \sigma)$ and $X_t \sim \mathcal{G}(0, \sigma)$. Recall the following result from Section 2.7.3 - If G and M are independent, with $G \sim \mathcal{G}(\mu_1, \sigma)$ and $M \sim \exp \mathcal{S}(\mu_2, \sigma; \alpha)$, then $G + M \sim \mathcal{G}(\mu_1 + \mu_2, \frac{\sigma}{\alpha})$.

From this result, we can conclude that for Y_t to be $\mathcal{G}(\mu, \sigma)$, $\sigma \log(\epsilon_t)$ is required to follow an $\exp \mathcal{S}(0, \sigma; 1)$ distribution. However, this is in contradiction with the assumption that $\alpha \in (0, 1)$. We would actually need for $\sigma \log(\epsilon_t)$ to follow an $\exp \mathcal{S}(0, \sigma; \alpha)$ distribution, and therefore Y_t would have a $\mathcal{G}(\mu, \frac{\sigma}{\alpha})$ distribution.

Parameter estimation

Since the distribution of Y_t is assumed to be $\mathcal{G}(\mu, \frac{\sigma}{\alpha})$, the distribution is known and the method of moments can be used to find parameter estimates. First recall that the mean, variance and covariance of Y_t are given by

$$\begin{aligned} \mathrm{E}(Y_t) &= \mu + \frac{\sigma}{\alpha}\gamma_E \\ \mathrm{Var}(Y_t) &= \frac{\pi^2}{6}\left(\frac{\sigma}{\alpha}\right)^2 \\ \mathrm{Cov}(Y_t, Y_{t+h}) &= Var(Y_t)\alpha^{|h|} \end{aligned}$$

If we let $\psi = \sigma/\alpha$, then in this case we will have the estimators

$$\begin{aligned}
\hat{\psi} &= \sqrt{\frac{6s_n^2}{\pi^2}} \\
\hat{\mu} &= \bar{Y}_n - \gamma_E \sqrt{\frac{6s_n^2}{\pi^2}} \\
\hat{\alpha} &= \frac{1}{ns^2} \sum_{i=1}^{n} (Y_i - \bar{Y}_n)(Y_{i+1} - \bar{Y}_n) \\
\hat{\sigma} &= \hat{\psi}\hat{\alpha}
\end{aligned}$$

where $\bar{Y}_n$ and s_n^2 are the sample mean and variance respectively, and γ_E is the Euler-Mascheroni constant.

Estimator properties

In her thesis, Toulemonde states the following propositions, relating to the properties of these estimators.

(i) The estimators for α, σ and μ defined above converge almost surely to the true parameter values.

(ii) The vector $\sqrt{n} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \\ \hat{\alpha} - \alpha \end{pmatrix}$ is asymptotically normal with zero mean.

2.8.5 Toulemonde, Guillou and Naveau model (2013)

This model was published in a 2013 paper that we only became aware of recently, and was applied to the problem of predicting levels of small particle air pollution [48].

$$Y_t = v_t + H_t Z_t + \eta_{t,\alpha_2} \tag{2.32}$$

$$Z_t = \alpha_1 Z_{t-1} + \epsilon_{t,\alpha_1} \tag{2.33}$$

where H_t is a positive coefficient and the parameters α_1 and α_2 are between 0 and 1. The authors show that

$$Y_t \sim \mathcal{G}\left(v_t - \frac{H_t \gamma_E \sigma}{\alpha_2}, H_t \frac{\sigma}{\alpha_2}\right)$$

$$Z_t \sim \mathcal{G}(-\gamma_E \sigma, \sigma)$$

whenever $\eta_{t,\alpha_2} \sim \exp \mathcal{S}(H_t \sigma \gamma_E (1/\alpha_2 - 1), H_t \sigma; \alpha_2)$ and $\epsilon_{t,\alpha_1} \sim \exp \mathcal{S}(-\sigma \gamma_E (1-\alpha_1), \alpha_1 \sigma; \alpha_1)$.

This model is very similar to the model that we had developed in the course of this thesis. However, this model also includes a time dependent coefficient H_t and mean v_t. In the paper, the authors assumed that parameters are known and therefore did not discuss estimation methods.

Chapter 3

A State space model with Gumbel marginals

Now that we have completed our literature review and discussion of similar models, we will move on to the model that we shall propose in this thesis. This model depends on the relationship between the Gumbel and exponential stable distributions detailed in Theorem 2.7.2. The model in question is stated in the following theorem.

Theorem 3.0.1 *Let Y_t and X_t be stochastic processes in a state space model that can be written as*

$$\begin{aligned} Y_t &= \mu + \phi(X_t + \log(\epsilon_t)) \\ X_t &= \alpha(X_{t-1} + \log(S_t)) \end{aligned}$$

where $\epsilon_t \sim \mathcal{S}(\beta)$ *and* $S_t \sim \mathcal{S}(\alpha)$. *If* $X_0 \sim \mathcal{G}(0,1)$ *then* $X_t \sim \mathcal{G}(0,1)$ *for all* $t \in \mathbb{Z}$ *and* $Y_t \sim \mathcal{G}(\mu, \frac{\phi}{\beta})$.

Notice that $X_t \sim \mathcal{G}(0,1)$ by Theorem 2.7.3 and $\log(\epsilon_t) \sim \exp \mathcal{S}(0,1;\beta)$ by Theorem 2.7.1. We can then use Theorem 2.7.2 to show that $X_t + \log(\epsilon_t) \sim \mathcal{G}(0, 1/\beta)$. This tells us that $Y_t \sim \mathcal{G}(\mu, \frac{\phi}{\beta})$. The result here is that for any Gumbel response Y_t, we will be able to write it in terms of a $\mathcal{G}(0,1)$ variable and exponential stable noise.

3.1 Moments of the state equation components

Since $X_t \sim \mathcal{G}(0,1)$, we already know the expectation and variance.

$$\mathrm{E}(X_t) = \gamma_E \tag{3.1}$$

$$\mathrm{Var}(X_t) = \frac{\pi^2}{6} \tag{3.2}$$

Note that $\gamma_E \approx 0.5772$, and is called the Euler-Mascheroni constant. Then for all $t \in \mathbb{Z}$ and $s \in \mathbb{N}$,

$$\begin{aligned}
\mathrm{Cov}(X_{t+s}, X_t) &= \mathrm{Cov}(\alpha^s X_t + \sum_{j=1}^{s} \alpha^j \log(S_{t+s-j+1}), X_t) \\
\mathrm{Cov}(X_{t+s}, X_t) &= \alpha^s Var(X_t) \\
\mathrm{Cov}(X_{t+s}, X_t) &= \alpha^s \frac{\pi^2}{6}
\end{aligned}$$

This result holds similarly for $s < 0$, so therefore for all $s, t \in \mathbb{Z}$,

$$\text{Cov}(X_{t+s}, X_t) = \alpha^{|s|}\frac{\pi^2}{6} \tag{3.3}$$

We can use these facts to determine the moments of $\log(S_t)$.

$$\begin{aligned} \text{E}(X_{t+1}) &= \alpha(\text{E}(X_t) + \text{E}(\log(S_t))) \\ \alpha\text{E}(\log(S_t)) &= \text{E}(X_{t+1}) - \alpha\text{E}(X_t) \\ \text{E}(\log(S_t)) &= \frac{\gamma_E(1-\alpha)}{\alpha} \end{aligned} \tag{3.4}$$

Similarly,

$$\begin{aligned} \text{Var}(X_{t+1}) &= \alpha^2(\text{Var}(X_t) + \text{Var}(\log(S_t))) \\ \alpha^2\text{Var}(\log(S_t)) &= \text{Var}(X_{t+1}) - \alpha^2\text{Var}(X_t) \\ \text{Var}(\log(S_t)) &= \frac{(1-\alpha^2)}{\alpha^2}\frac{\pi^2}{6} \end{aligned} \tag{3.5}$$

Note that $\text{Cov}(\log(S_u), \log(S_v)) = 0$ when $u \neq v$ since the noise is independent.

3.2 Moments of the observation equation components

Since we know that $Y_t \sim G(\mu, \frac{\phi}{\beta})$, the mean and variance of Y_t are

$$\mathrm{E}(Y_t) = \mu + \phi\frac{\gamma_E}{\beta} \tag{3.6}$$

$$\mathrm{Var}(Y_t) = \frac{\phi^2\pi^2}{6\beta^2}. \tag{3.7}$$

Now by Equation (3.3) we can find the covariance of the Y_t process. For all $s, t \in \mathbb{Z}$,

$$\begin{aligned}\mathrm{Cov}(Y_t, Y_{t+s}) &= \mathrm{Cov}(\phi X_t + \phi\log(\epsilon_t), \phi X_{t+s} + \phi\log(\epsilon_{t+s})) \\ &= \phi^2\mathrm{Cov}(X_t, X_{t+s}) \\ &= \phi^2\alpha^{|s|}\frac{\pi^2}{6}. \end{aligned} \tag{3.8}$$

Since we know the moments of X_t and Y_t, we can use these to find the moments of $\log(\epsilon_t)$ as well.

$$\begin{aligned}\mathrm{E}(Y_t) &= \mu + \phi(\mathrm{E}(X_t) + \mathrm{E}(\log(\epsilon_t))) \\ \phi\mathrm{E}(\log(\epsilon_t)) &= \mathrm{E}(Y_t) - \mu - \phi\mathrm{E}(X_t) \\ &= \mu + \frac{\phi\gamma_E}{\beta} - \mu - \phi\gamma_E\end{aligned}$$

Then,

$$\mathrm{E}(\log(\epsilon_t)) = \gamma_E\frac{(1-\beta)}{\beta}. \tag{3.9}$$

$$\begin{aligned} \text{Var}(Y_t) &= \phi^2\text{Var}(X_t) + \phi^2\text{Var}(\log(\epsilon_t)) \\ \phi^2\text{Var}(\log(\epsilon_t)) &= \text{Var}(Y_t) - \phi^2\text{Var}(X_t) \\ &= \frac{\phi^2\pi^2}{6\beta^2} - \phi^2\frac{\pi^2}{6} \end{aligned} \tag{3.10}$$

Therefore,

$$\text{Var}(\log(\epsilon_t)) = \frac{\pi^2}{6}\frac{(1-\beta^2)}{\beta^2} \tag{3.11}$$

These moments are similar in form to those of $\log(S_t)$, as we would expect. As before, $\text{Cov}(\log(\epsilon_u), \log(\epsilon_v)) = 0$ when $u \neq v$ due to independence.

Finally we can use Equation (3.3) to calculate the covariance relationship between X_t and Y_{t+s} for $s \geq 0$. For all $s, t \in \mathbb{Z}$,

$$\begin{aligned} \text{Cov}(X_t, Y_{t+s}) &= \text{Cov}(X_t, \mu + \phi X_{t+s} + \phi\log(\epsilon_{t+s})) \\ &= \phi\text{Cov}(X_t, X_{t+s}) \\ &= \alpha^{|s|}\phi\frac{\pi^2}{6} \end{aligned} \tag{3.12}$$

3.3 Simulation of the state space model

We would like to simulate a data set that follows the model

$$Y_t = \phi X_t + \phi \log(\epsilon_t)$$
$$X_t = \alpha X_{t-1} + \alpha \log(S_t)$$

where $\epsilon_t \sim \mathcal{S}(\beta)$ and $S_t \sim \mathcal{S}(\alpha)$. We perform the simulation as follows.

1. Generate the S_t, following an α-stable distribution with parameter α. There is an **R** package called *stabledist* that can generate such variables, we must input the parameters of the Samorodnitsky parametrization that we discussed in Section 2.7.2. Use each S_t to calculate the noise $\alpha \log(S_t)$.

2. Generate X_1 from a $\mathcal{G}(0, 1)$ distribution. The *gumbelSim* function in the *fExtremes* **R** library is useful for this purpose.

3. We can then set up a sequential procedure to calculate each X_{t+1} based on X_t and an independent $\alpha \log(S_{t+1})$.

4. Values for Y_t can be generated using the X_t values and noise $\phi \log(\epsilon_t)$, with the α-stable variable ϵ_t generated in the same manner as S_t.

Chapter 4

Estimation of parameters

In this chapter we will discuss and compare several methods for estimating parameters under a first order auto-regressive model, and apply several estimation methods to the observed component of our model. It is important to keep in mind the nature of the data we will be working with - extremes in datasets are often limited in number. Our emphasis, therefore, is on finding an estimation procedure that can be used with a relatively small sample size.

4.1 Estimation under a first order auto-regressive model

We begin with a simple AR(1) time series model, which can be written

$$X_t = \alpha X_{t-1} + Z_t$$

where $X_t \sim \mathcal{G}(0,1)$ and $Z_t = \alpha \log(S_t) \sim \exp \mathcal{S}(0, \alpha; \alpha)$.

It is usually a good idea to start out with the simplest estimation method, so we will begin our discussion with the Yule-Walker estimators.

4.1.1 Yule Walker estimation

The Yule-Walker equations arose in the 1920s and 30s, appearing in the work of economist George Udny Yule in 1926 and independently in the research of atmospheric scientist Sir Gilbert Walker several years later in 1931 [5]. This method of estimation has become one of the most common, as it is easy to understand and implement. A discussion of this method can be found in the Brockwell and Davis text [6].

We can use the Yule-Walker equations to estimate α and σ_Z^2, the variance of the noise process Z_t, where $Z_t \sim \text{IID}(0, \sigma_Z^2)$. For an AR($p$) model written as $X_t = \psi_1 X_{t-1} + \cdots + \psi_p X_{t-p} + Z_t$, we find that

$$\mathbf{\Gamma}_p \psi = \gamma_p \tag{4.1}$$

and

$$\sigma_Z^2 = \gamma(0) - \psi' \gamma_p, \tag{4.2}$$

where $\mathbf{\Gamma}_p = [\gamma(i-j)]_{i,j}$, a $p \times p$ matrix, $\gamma(h) = \text{Cov}(X_t, X_{t+h})$, $\psi = [\psi_1 \dots \psi_p]'$ is a vector of coefficients and $\gamma_p = [\gamma(1) \dots, \gamma(p)]$.

In the AR(1) case these equations become simpler:

$$\gamma(0)\alpha = \gamma(1) \tag{4.3}$$

$$\sigma_Z^2 = \gamma(0) - \alpha\gamma(1) \tag{4.4}$$

Since the autocovariances $\gamma(0)$ and $\gamma(1)$ are unknown, we can substitute the sample covariances $\hat{\gamma}(0)$ and $\hat{\gamma}(1)$, where

$$\hat{\gamma}(j) = \frac{1}{n-j-1}\sum_{t=1}^{n-j}(X_t - \bar{X})(X_{t+j} - \bar{X}).$$

In our Gumbel AR(1) equation the noise Z_t does not have a zero mean, so we will have to be extra careful. Let $\delta_t = Z_t - C$, where $C = \gamma_E(1-\alpha)$, the mean of Z_t. Recall that the Yule-Walker equations are derived by multiplying the $AR(p)$ equation by $X_{t-1}, \ldots, X_{t-p}$ on both sides and taking the expectation. Therefore in our $AR(1)$ case, we can find the following result.

$$
\begin{aligned}
X_t - \alpha X_{t-1} - C &= \delta_t \\
X_{t-1}X_t - \alpha X_{t-1}X_{t-1} - X_{t-1}C &= X_{t-1}\delta_t \\
\mathrm{E}(X_{t-1}X_t) - \mathrm{E}(\alpha X_{t-1}X_{t-1}) - \mathrm{E}(X_{t-1}C) &= \mathrm{E}(X_{t-1}\delta_t) \\
\gamma(1) + \mathrm{E}(X)^2 - \alpha(\gamma(0) + \mathrm{E}(X)^2) - C\mathrm{E}(X) &= \mathrm{E}(\alpha X)\mathrm{E}(\delta_t) \\
\gamma(1) - \alpha\gamma(0) + \gamma_E^2(1-\alpha) - \gamma_E^2(1-\alpha) &= 0 \\
\gamma(1) - \alpha\gamma(0) &= 0
\end{aligned}
$$

In this derivation, we used the fact that $X_t \sim \mathcal{G}(0,1)$, so that $\mathrm{E}(X)$ is known to be γ_E, the Euler-Mascheroni constant. Even though our noise process is not zero-mean, we are left with the same Yule-Walker estimate,

$$\hat{\alpha} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}. \tag{4.5}$$

Note that by (4.5) and (4.4),

$$\hat{\sigma}_Z^2 \quad = \hat{\gamma}(0) - \tfrac{\hat{\gamma}(1)}{\hat{\gamma}(0)}\hat{\gamma}(1)$$

therefore

$$\hat{\sigma_Z^2} = \frac{\hat{\gamma}(0)^2 - \hat{\gamma}(1)^2}{\hat{\gamma}(0)}. \tag{4.6}$$

Results

How does this estimation method perform for different values of α? Figure 4.1 shows estimates of α plotted against the true values of α, with 95% bootstrap confidence intervals at different sample sizes. The width of the confidence intervals decreases as α increases towards 1. The method seems to slightly underestimate α at lower sample sizes, but generally works quite well. A table of mean values, confidence intervals and mean squared error estimates is available in Appendix B.

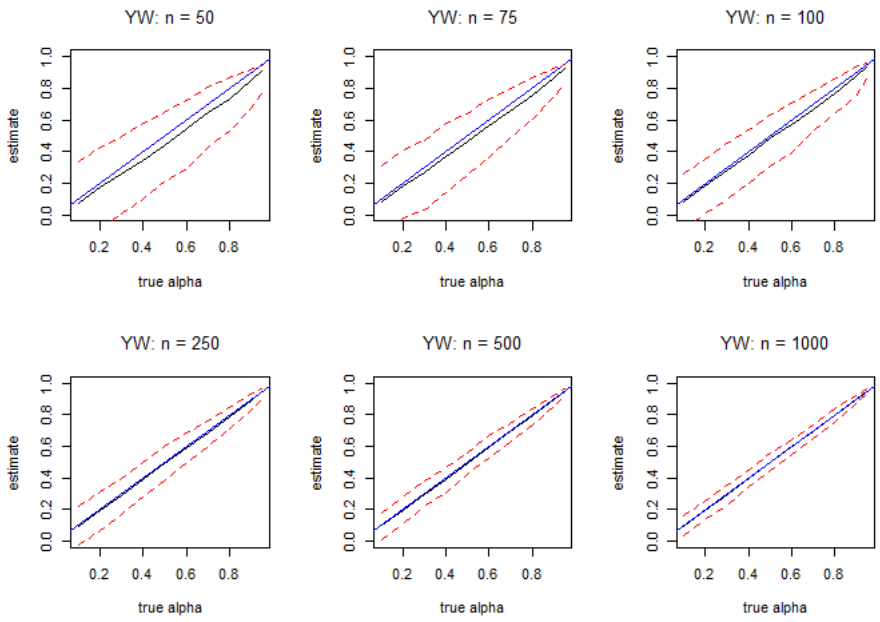

Figure 4.1: Yule-Walker α estimates with 95% bootstrap confidence intervals (red dotted lines). The blue line represents a perfect fit.

Properties of Yule Walker estimators

To determine the properties of this estimator, we refer to a theorem described in Brockwell and Davis [6].

Theorem 4.1.1 *Let $\hat{\psi}$ denote the Yule-Walker estimate of ψ. If $\{X_t\}$ is an AR(p) process and $\{Z_n\} \sim$ IID$(0, \sigma^2)$, then*

$$\sqrt{n}(\hat{\psi} - \psi) \to N(0, \sigma^2 \mathbf{\Gamma}_p^{-1})$$

where $\mathbf{\Gamma}_p$ is the covariance matrix of $\{X_t\}$, $[\gamma(i-j)]_{i,j=1}^{p}$. It is also known that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

From this theorem, we conclude that $\sqrt{n}(\hat{\alpha} - \alpha) \to \mathrm{N}(0, \sigma_Z^2 \gamma(1))$, thus,

$$\mathrm{Var}(\hat{\alpha}) = \frac{\sigma_Z^2 \gamma(1)}{n} \tag{4.7}$$

which goes to 0 as $n \to \infty$. The theorem also tells us that $\hat{\sigma}_Z^2$ is a consistent estimator of σ_Z^2.

4.1.2 Conditional least squares estimation

The method of conditional least squares estimation, introduced by Kilmko and Nelson in 1978 [26], is quite useful for situations in which our process is dependent but can be broken into independent conditional variables.

Let $X_0, X_1, \ldots, X_T$ be a stochastic process and θ be a vector of parameters that

we would like to estimate. Let F_t be a sub-sigma field generated by any subset of $X_0, X_1, \ldots, X_{t-1}$. Our goal is to minimize the conditional sum of squares

$$\mathcal{Q}_n(\theta) = \sum_{t=0}^{n-1} [X_{t+1} - g_t(\theta)]^2 \tag{4.8}$$

where $g_t(\theta) = \mathrm{E}_\theta(X_{t+1}|F_t)$, which would be minimized by solving the least squares equation,

$$\frac{\partial \mathcal{Q}_n(\theta)}{\partial \theta} = 0.$$

Let $\mathbf{X}_0 = [X_0, \ldots, X_{N-1}]'$ and $\mathbf{X}_1 = [X_1, \ldots, X_N]$. In the case of a linear model of the form $g_t(\theta) = \theta \mathbf{X}_0$, we can obtain a closed form for $\hat{\theta}$,

$$\hat{\theta} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}_1. \tag{4.9}$$

Application to our AR(1) model

Consider the model $X_t = \alpha X_{t-1} + \alpha \log(S_t)$.

Since $\alpha \log(S_t) \sim \exp \mathcal{S}(0, \alpha; \alpha)$, and $\mathrm{E}(\log(S_t) = \gamma_E \frac{(1-\alpha)}{\alpha}$, we find that

$$\mathrm{E}_\alpha(X_t|X_{t-1}) = \alpha X_{t-1} + \gamma_E(1-\alpha).$$

Therefore we can write the conditional sum of squares as

$$\begin{aligned}\mathcal{Q}_n(\theta) &= \sum_{t=0}^{n-1} \left(X_{t+1} - \mathrm{E}_\alpha(X_{t+1}|X_t)\right)^2 \\ &= \sum_{t=0}^{n-1} \left(X_{t+1} - \alpha X_{t-1} - \gamma_E(1-\alpha)\right)^2 \\ &= \sum_{t=0}^{n-1} \left((X_{t+1} - \gamma_E) - \alpha(X_{t-1} - \gamma_E)\right)^2.\end{aligned}$$

It follows from Equation (4.9) that the closed form of $\hat{\alpha}$ will be

$$\hat{\alpha} = \left[(\mathbf{X_0} - \gamma_E\mathbf{1})'(\mathbf{X_0} - \gamma_E\mathbf{1})\right]^{-1} (\mathbf{X_0} - \gamma_E\mathbf{1})'(\mathbf{X_1} - \gamma_E\mathbf{1}). \tag{4.10}$$

Results

As before we perform our simulation 500 times for various values of α and N (see Figure 4.2). Conditional least squares estimates have a slight advantage over the Yule-Walker estimates discussed in Section 4.1.1, which tend to suffer from downward bias at low sample sizes. A table of mean values, confidence intervals and mean squared error estimates is available in Appendix B.

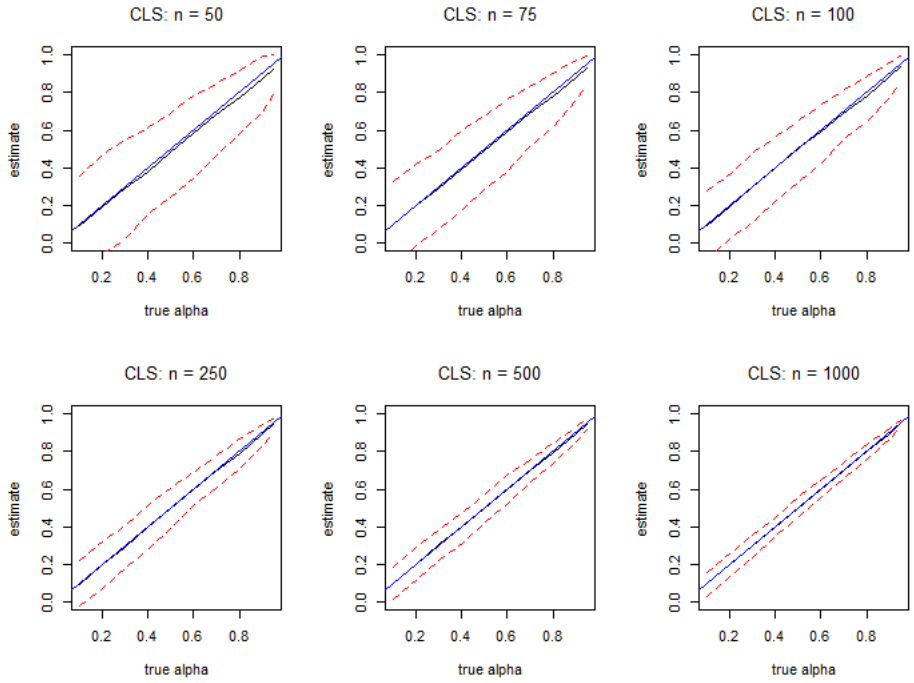

Figure 4.2: Conditional least squares α estimates with 95% bootstrap confidence intervals (red dotted lines). The blue line represents a perfect fit.

Properties of conditional least squares estimates

Conditional least squares estimates are strongly consistent and unbiased. Klinko and Nelson show that under certain conditions, these estimates are asymptotically multivariate normal [26]. In fact, if $\hat{\alpha}_n$ is the least squares estimate of α at sample size n,

$$n^{1/2}(\hat{\alpha}_n - \alpha) \to \text{MVN}(\mathbf{0}^{p\times 1}, \mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}) \text{ as } n \to \infty$$

where $\mathbf{V}$ and $\mathbf{W}$ are matrices such that if $\mathbf{V}_n$ is the matrix of second partial derivatives

of $\mathcal{Q}_n(\alpha)$ then $(2n)^{-1}\mathbf{V}_n \to \mathbf{V}$ and

$$\frac{1}{2}n^{-1/2}\frac{\partial \mathcal{Q}_n(\alpha)}{\partial \alpha} \to \text{MVN}(\mathbf{0}^{p\times 1}, \mathbf{W}) \text{ as } n \to \infty.$$

4.1.3 Comparison of AR(1) model parameter estimators

We have illustrated two different methods for estimating the parameter α of our AR(1) model - Yule-Walker estimation (YW) and conditional least squares (CLS). Both of these methods perform well when n is large. However, in extreme value applications, we are often limited to a relatively small sample size.

In Figure 4.3 we can see that for the smaller sample sizes, both the Yule-Walker and conditional least squares estimates are biased slightly downward. However, the conditional least squares estimates are closer to the true values. The mean squared error (MSE) is similar between the two methods, although conditional least squares performs slightly better (see Figure 4.4). Notice also that for both methods, the MSE gets smaller as α gets closer to 1.

If there are several methods that are close in performance, it is advisable to use the method that is the least computationally expensive. Table 4.1 includes program run-times for the two estimation methods. The user time gives the CPU time spent by the current R session, while the system time would include CPU time spent by the operating system on behalf of the R session. For example, in our programs this would include saving the results to CSV. Elapsed time is the true (real world) time that has passed since the program started. Each program was run 500 times over each of the 10 parameter values and 6 sample sizes. Yule-Walker was significantly

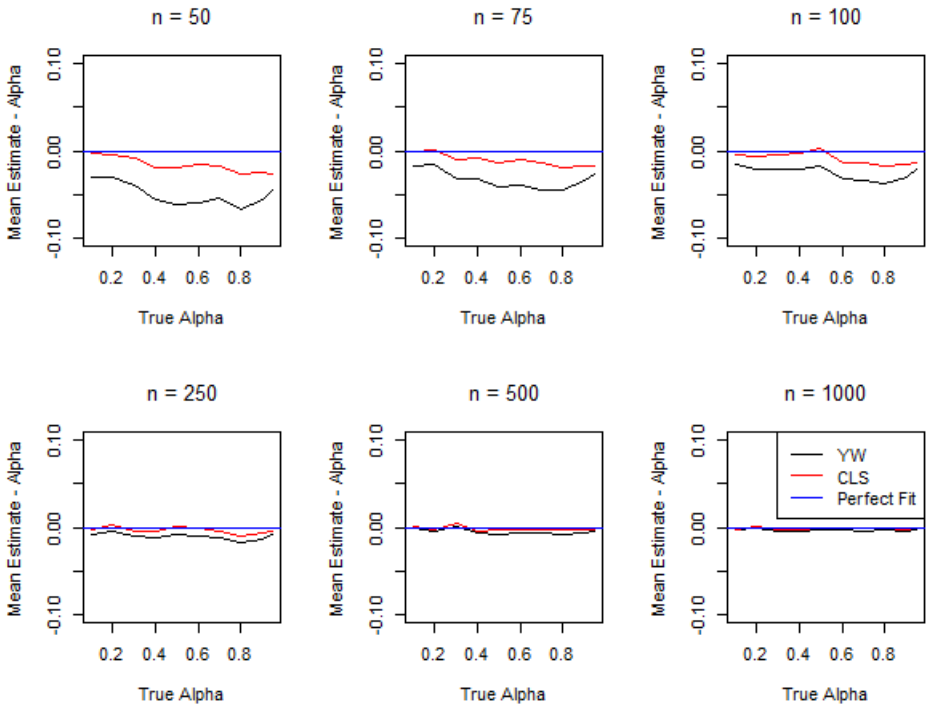

Figure 4.3: Comparison of mean parameter estimates - AR(1) Model

faster than conditional least squares.

4.2 Estimation of parameters in the state space model

For the purposes of our parameter estimation, we will assume that $\mu = 0$ and that the state equation is hidden, i.e. not observed. However, for several of these methods, we assume that values from the state process X have been determined using a filtering method (see Chapter 5 for an in-depth discussion).

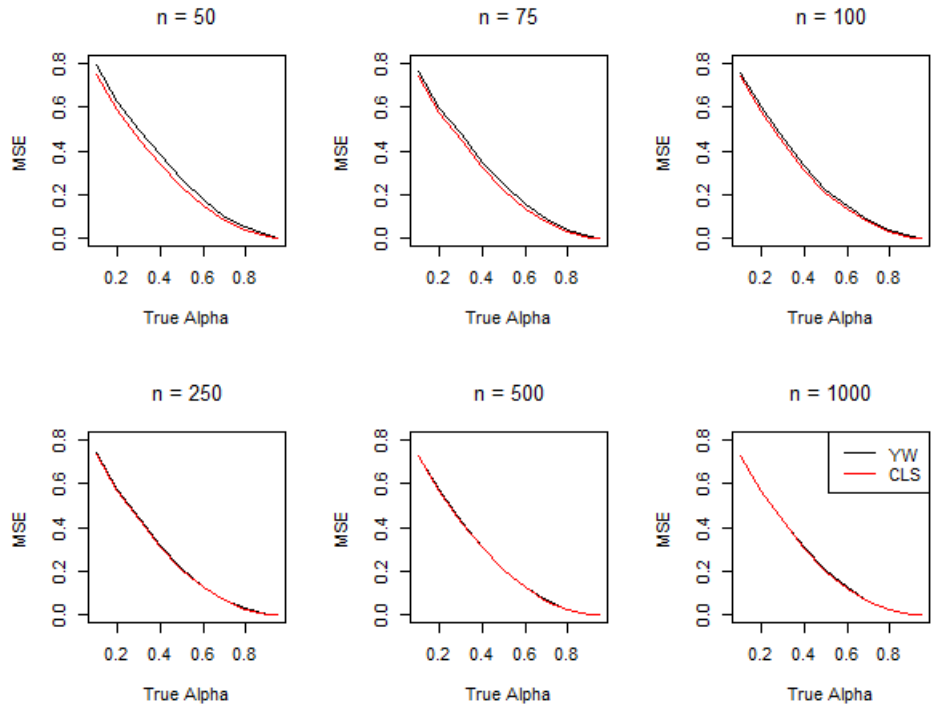

Figure 4.4: Comparison of mean squared errors of parameter estimates - AR(1) Model

Recall that our observation equation can be written

$$Y_t \quad = \phi(X_t + \log(\epsilon_t)) \tag{4.11}$$

where $\log(\epsilon_t) \sim \exp \mathcal{S}(0, 1; \beta)$, $X_t \sim \mathcal{G}(0, 1)$ and $Y_t \sim \mathcal{G}(0, \frac{\phi}{\beta})$. As with the state equation, we will test several parameter estimation methods starting with the simplest, the method of moments.

Table 4.1: Comparison of running time (seconds) - AR(1) model

Method	Parameter	User	System	Elapsed
Yule-Walker	α	78.22	0.55	79.03
Conditional least squares	α	262.38	1.74	279.24

4.2.1 Method of moments estimation

Method of Moments estimation is a simple parameter estimation method first proposed by Karl Pearson in 1894. A discussion of this method is available in a 1970 paper by Robertson and Fryer, in which they also discuss the limitations and issues that we may find with this method [38].

The method of moments estimators are derived by equating the first k sample moments to the average of Y_i^k where k is the number of parameters in our model.

$$E(Y_t) = \sum_{i=1}^{T} y_i$$
$$E(Y_t^2) = \sum_{i=1}^{T} y_i^2$$
$$\dots$$
$$E(Y_t^k) = \sum_{i=1}^{T} y_i^k$$

Recall that $Y_t \sim \mathcal{G}(0, \phi/\beta)$, so these theoretical moments are known.

$$\mathrm{E}(Y_t) = \frac{\phi}{\beta}\gamma_E \tag{4.12}$$

$$\begin{aligned}\mathrm{Var}(Y_t) &= \frac{\pi^2}{6}\frac{\phi^2}{\beta^2} \\ &= \frac{\pi^2}{6\gamma_E}\mathrm{E}(Y_t)\end{aligned} \tag{4.13}$$

Notice that $\mathrm{Var}(Y_t)$ is a function of $\mathrm{E}(Y_t)$. This means that the system of those two equations is not linearly independent and cannot be solved to find ϕ and β. We will need to use the covariance structure (Yule-Walker method) to gain some extra information about the parameters.

$$\mathrm{Cov}(Y_t, Y_{t+h}) = \phi^2\frac{\pi^2}{6}\alpha^{|h|} \tag{4.14}$$

$$\frac{\mathrm{Cov}(Y_t, Y_{t+2})}{\mathrm{Cov}(Y_t, Y_{t+1})} = \alpha \tag{4.15}$$

Now the moments can be rearranged to find the other parameter estimates.

$$\hat{\phi} = \sqrt{\frac{6\hat{\mathrm{Cov}}(Y_t, Y_{t+1})}{\pi^2\hat{\alpha}}} \tag{4.16}$$

$$\hat{\beta} = \frac{\hat{\phi}\gamma_E}{\bar{Y}_n} \tag{4.17}$$

Performance

For each of the simulations below, default values of $\alpha = 0.75$, $\beta = 0.6$ and $\phi = 1$ were used, varying one value as necessary. For each set of values, we looked at n values of 50, 75, 100, 250, 500 and 1000, and ran the simulations for 500 trials. For the estimation of ϕ and β it was assumed that α was known, to aid in comparison with the methods that we will see in subsequent sections. Tables of mean values, confidence intervals and mean squared error estimates are available in Appendix B.

Estimation of α

Estimates of α do not behave well until our sample size is quite large (see Figure 4.5). In the estimation of β and ϕ here, we will assume that α is known. However, in practice it would be preferable to use a method that either does not depend on α or is robust to inaccurate α estimates.

Estimation of β

The method of moments does not work well for the estimation of β when we have small sample sizes. The confidence intervals are quite large and the method is not robust to outliers. This estimation method tends to overestimate β values that are close to 1, but the tendency is reduced as the sample size increases.

Note that although β can only be between 0 and 1, we did not restrict the values so that we could see the true distribution of the estimates (see Figure 4.6).

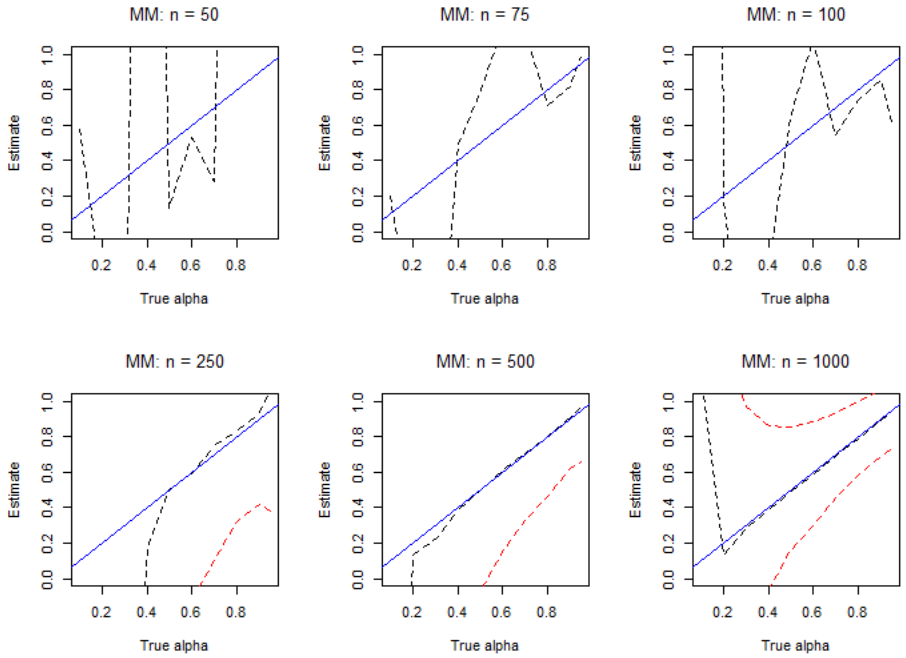

Figure 4.5: Method of moments estimates of α with 95% bootstrap confidence intervals (red dashed lines). The blue line represents a perfect fit.

Estimation of ϕ

At low sample sizes, this method tended to underestimate ϕ slightly and had very wide confidence intervals, especially as ϕ became larger (see Figure 4.7). However, for ϕ we do not see the same wild variation at lower n values as we saw with our β estimates.

Properties of the method of moments estimators

Generalized method of moments estimators are consistent due to the law of large numbers, but may often be biased. Another issue that may occur is that parameter

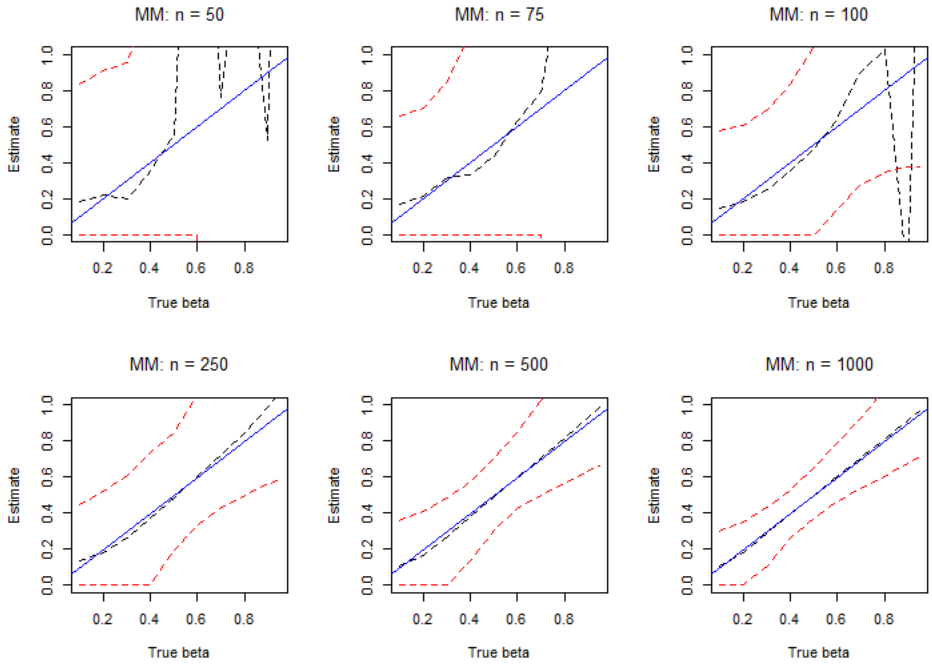

Figure 4.6: Method of moments estimates of β with 95% bootstrap confidence intervals (red dashed lines). The blue line represents a perfect fit.

estimates may fall outside of the allowed range [38]. Method of moments estimates are a special case of generalized method of moments estimates (GMM) which have been shown to be asymptotically normal and efficient [30].

4.2.2 Quasi Fisher's scoring method

Fisher's Scoring method, developed by R.A. Fisher, is a method of refining parameter estimates. Usually this method would require a score function and information matrix, which would require knowledge of the likelihood function and the probability density function respectively. In a quasi-likelihood context, we can substitute the

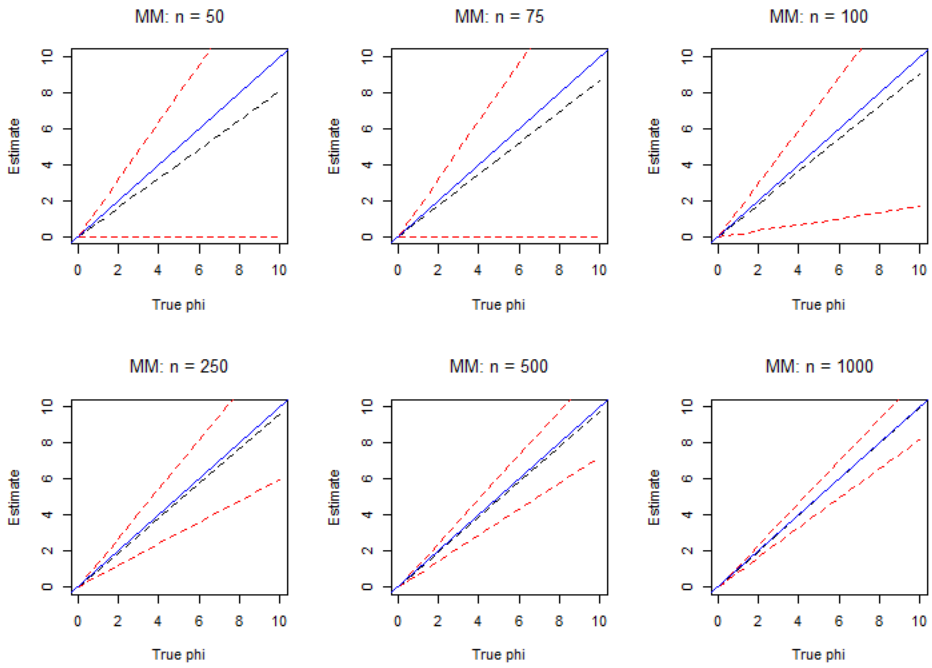

Figure 4.7: Method of moments estimates of ϕ with 95% bootstrap confidence intervals (red dashed lines). The blue line represents a perfect fit.

quasi-likelihood function and use a similar method for parameter estimation. We will refer to this method as quasi Fisher's scoring (QFS).

Heyde (1997) [18] tells us that for any estimating equation $G(\hat{\Theta}) = 0$ for parameters Θ, the scoring iterations will be

$$\Theta_{m+1} = \Theta_m - (\mathrm{E}(\dot{G}(\Theta_m)))^{-1} G(\Theta_m)$$

Note that when $G()$ is a standardized estimating function, $-\mathrm{E}(\dot{G}(\Theta))$ is the covariance matrix. In the context of quasi Fisher's scoring, $G(\Theta)$ will be the quasi-likelihood score function and $-\mathrm{E}(\dot{G}(\Theta))$ is the quasi-information matrix.

Following the method of Wedderburn [51] we first define the quasi-likelihood function to be

$$\mathcal{K}_\Theta(y_i) = \int^{\mu} \frac{y_i - \mu}{V(\mu)} d\mu \tag{4.18}$$

where μ is the mean function and $V(\mu)$ is the variance expressed as a function of the mean. Since $V(\mu) = \frac{\pi^2}{6\gamma_E^2}\mu^2$ (where γ_E is the Euler-Mascheroni constant), it follows that

$$\begin{aligned} \mathcal{K}_\Theta(y_i) &= \int \tfrac{y_i-\mu}{C_1\mu^2} d\mu \\ &= \tfrac{1}{C_1} \int \left(\tfrac{y_i}{\mu^2} - \tfrac{1}{\mu} \right) d\mu \\ &= \tfrac{-1}{C_1} \left(\tfrac{y_i}{\mu} + \log(\mu) + C_0 \right) \end{aligned} \tag{4.19}$$

where $C_1 = \frac{\pi^2}{6\gamma_E^2}$ and C_0 is an unknown constant. The quasi-likelihood score function is given as

$$G(\hat{\Theta}) = \sum_{i=1}^{n} \frac{\partial}{\partial \Theta} \mathcal{K}_\Theta(y_i). \tag{4.20}$$

The derivatives of $\mathcal{K}$ with respect to the parameters β and ϕ are

$$\begin{aligned} \frac{\partial}{\partial \phi}\mathcal{K}_\Theta(y_i) &= \tfrac{\partial}{\partial \phi} \tfrac{-1}{C_1} \left(\tfrac{y_i \beta}{\phi \gamma_E} + \log\left(\tfrac{\phi \gamma_E}{\beta} \right) + C_0 \right) \\ &= \tfrac{-1}{C_1} \left(\tfrac{-y_i \beta}{\phi^2 \gamma_E} + \tfrac{1}{\phi} \right) \qquad (4.21) \\ \frac{\partial}{\partial \beta}\mathcal{K}_\Theta(y_i) &= \tfrac{-1}{C_1} \left(\tfrac{y_i}{\phi \gamma_E} - \tfrac{1}{\beta} \right), \qquad (4.22) \end{aligned}$$

which leaves us with the result

$$G(\hat{\Theta}) = \sum_{i=1}^{n} \begin{bmatrix} \frac{\partial}{\partial\phi}\mathcal{K}_\Theta(y_i) \\ \frac{\partial}{\partial\beta}\mathcal{K}_\Theta(y_i) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} \frac{-1}{C_1}(\frac{-y_i\beta}{\phi^2\gamma_E} + \frac{1}{\phi}) \\ \frac{-1}{C_1}(\frac{y_i}{\phi\gamma_E} - \frac{1}{\beta}) \end{bmatrix} \tag{4.23}$$

with $C_1 = \frac{\pi^2}{6\gamma_E^2}$ as before.

Now, we can move on to the quasi-information matrix, $n\mathcal{I}_q(\Theta)$. In general, for $\Theta = [\theta_1, \ldots, \theta_N]'$, the diagonal entries of $\mathcal{I}_q(\Theta)$ will be given by

$$E\left(\frac{\partial\mathcal{K}}{\partial\mu}\right)^2 = \frac{1}{V(\mu)} \tag{4.24}$$

while the off-diagonal entry in row i and column j will be

$$E\left(\frac{\partial\mathcal{K}}{\partial\theta_i}\frac{\partial\mathcal{K}}{\partial\theta_j}\right) = \frac{1}{V(\mu)}\frac{\partial\mu}{\partial\theta_i}\frac{\partial\mu}{\partial\theta_j} \tag{4.25}$$

When we apply these equations to our context with $\Theta = [\beta, \phi]'$, we find that since $\mu = \frac{\phi}{\beta}\gamma_E$,

$$\frac{\partial\mu}{\partial\beta} = \tfrac{-\phi\gamma_E}{\beta^2} \tag{4.26}$$

$$\frac{\partial\mu}{\partial\phi} = \tfrac{\gamma_E}{\beta} \tag{4.27}$$

so that

$$\mathcal{I}_q(\Theta) = \frac{6\beta^2}{\pi^2\phi^2}\begin{pmatrix} 1 & \frac{-\phi\gamma_E^2}{\beta^3} \\ \frac{-\phi\gamma_E^2}{\beta^3} & 1 \end{pmatrix}. \tag{4.28}$$

We require the inverse of this matrix, which is

$$\mathcal{I}_q^{-1}(\Theta) = \frac{\pi^2\phi^2}{6\beta^2}\left(\frac{1}{1-\frac{\phi^2\gamma_E^4}{\beta^6}}\right)\begin{pmatrix} 1 & \frac{\phi\gamma_E^2}{\beta^3} \\ \frac{\phi\gamma_E^2}{\beta^3} & 1 \end{pmatrix}. \tag{4.29}$$

The parameter vector $\Theta = [\beta, \phi]'$ can therefore be estimated using the following iteration equations:

$$\Theta_{m+1} = \Theta_m + \frac{\mathcal{I}_q^{-1}(\Theta_m)}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\Theta_m}\mathcal{K}_{\Theta_m}(y_i). \tag{4.30}$$

where $\sum_{i=1}^{n}\frac{\partial}{\partial\Theta}\mathcal{K}_\Theta(y_i)$ is the quasi-likelihood score function and $n\mathcal{I}_q(\Theta)$ is the quasi-information matrix.

We have now defined all of the components and can proceed with the method as follows.

1. Use the method of moments to get initial values for ϕ and β, which will be our Θ_0. Note that our esimation only depends on α through these initial values.

2. For $m = 0, 1, 2, \ldots,$ use Equation (4.30) to compute Θ_{m+1}.

3. We stop when we reach convergence, that is $|\Theta_{m+1} - \Theta_m| < \epsilon$ for some pre-chosen small $\epsilon > 0$. In our simulations we used $\epsilon = 0.001$.

Correction Factor

This optimization problem should be solved using more robust numerical methods, but this would make its implementation impractical. As an ad-hoc alternative we can

make sure that our parameter estimates of ϕ and β do not go outside their respective bounds, $(0, \infty)$ and $(0, 1)$ respectively, by multiplying by a correction factor. With this method of quasi Fisher scoring it is often the case that the direction of the parameter adjustment is correct but the magnitude may be too high.

Therefore we choose a random variable $k \sim U(\frac{1}{3}, \frac{1}{2})$ that will act as a correction to keep the estimates within the proper bounds. We may need to multiply estimates by k several times for estimates that are further out of bounds. In addition to making our estimates more accurate, this will keep the program from producing errors due to negatives being placed inside a log or square root function.

We must also make sure that our initial values (found using the method of moments) are within the proper bounds so that our program does not produce errors.

Estimation of β

Simulations were run 500 times, with default values of $\alpha = 0.75$ and $\phi = 1$, over varying levels of β and n, assuming α is known and ϕ is unknown. This method tends to underestimate higher values of β, but is overall an improvement upon the method of moments, especially at low sample sizes (see Figure 4.8).

Estimation of ϕ

Simulations were run 500 times, with default values of $\alpha = 0.75$ and $\beta = 0.6$, over varying levels of ϕ and n amd assuming that α is known and β is unknown. As with other methods, the quasi Fisher scoring estimates of ϕ have wide confidence intervals when the sample size is low and for larger values of ϕ (see Figure 4.9). We also have

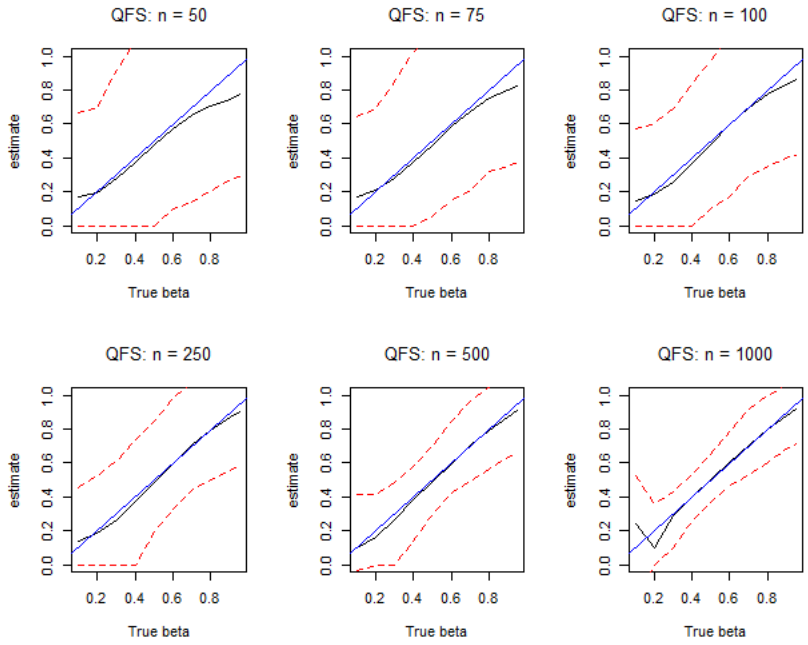

Figure 4.8: Quasi Fisher's scoring estimates of β with 95% bootstrap confidence intervals (red dashed lines). The blue line represents a perfect fit.

a downward bias that becomes less evident as the sample size increases.

4.2.3 Comparison of state space model parameter estimators

We have now estimated the parameters of the observation equation with both the method of moments and the quasi Fisher scoring method. Quasi Fisher scoring is the recommended method of estimation for the parameter β, as the estimates are much more stable than the method of moments estimators at low sample sizes (see Figure 4.10). Quasi fisher scoring also results in a lower mean squared error (see Figure 4.11). However, when $N = 1000$, we can see some irregularity in the QFS estimates

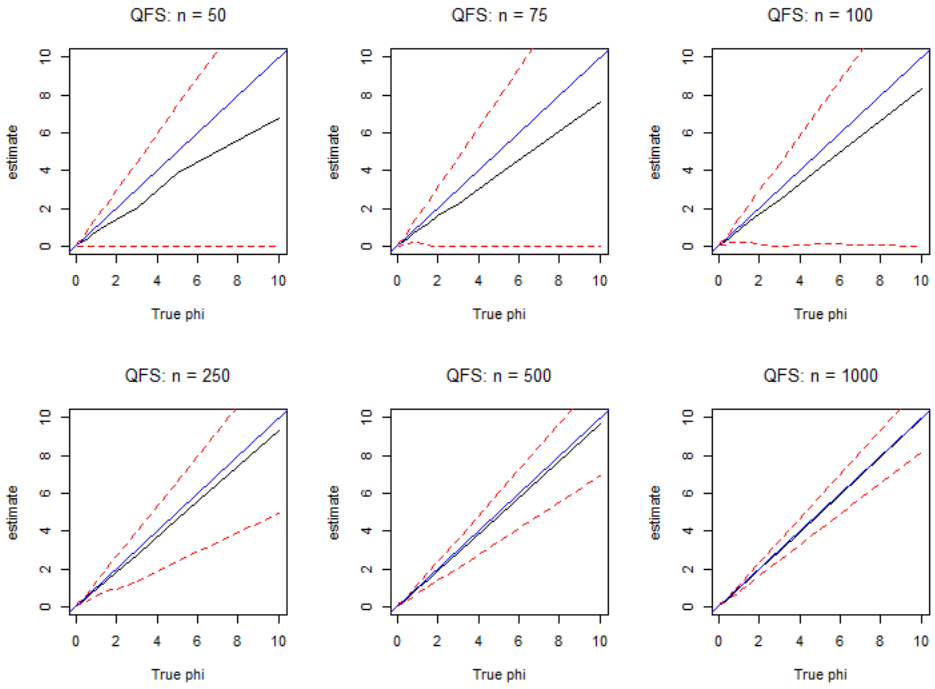

Figure 4.9: Quasi Fisher's scoring estimates of ϕ with 95% bootstrap confidence intervals (red dashed lines). The blue line represents a perfect fit.

but not the MM estimates. This tells us that quasi Fisher scoring may not be as robust to large extreme values as the method of moments.

Among the ϕ estimation methods, both methods tend to underestimate ϕ but the method of moments method is closer to the true value (see Figure 4.12). The quasi Fisher scoring method has a lower mean squared error however, and therefore may be the better choice for parameter estimation (see Figure 4.13).

In Table 4.2, we can see the run times for the two estimation methods. Recall that the user time gives the CPU time spent by the current R session, while the system time would include CPU time spent by the operating system on behalf of the

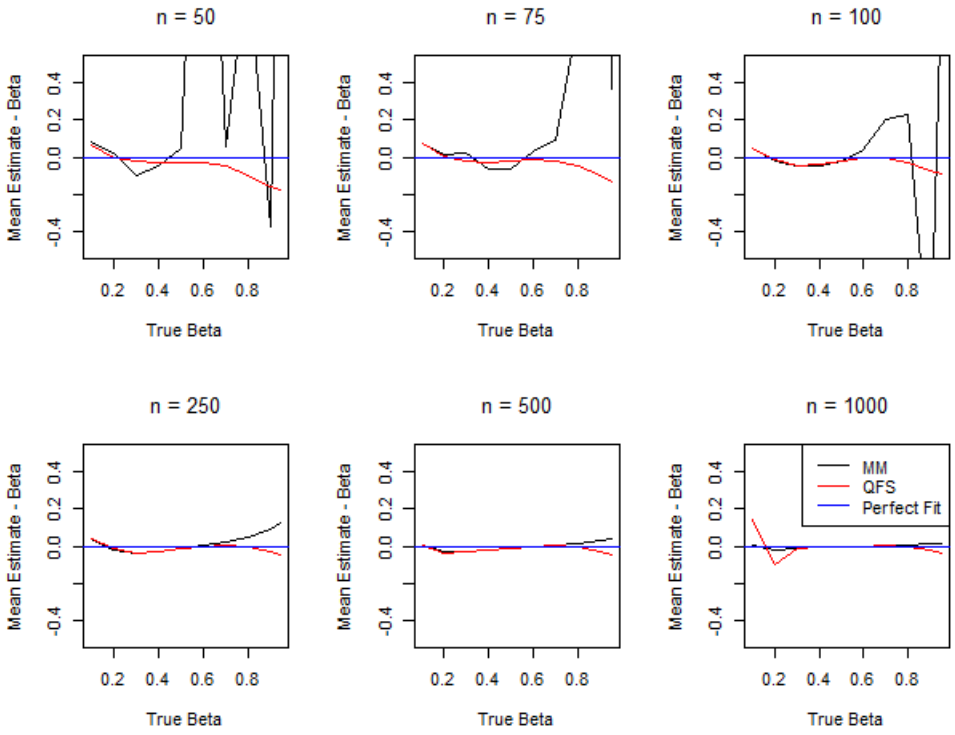

Figure 4.10: Comparison of mean parameter estimates of β - State space model

R session and elapsed time is the true (real world) time that has passed since the program started. Each program was run 500 times over each of the 10 parameter values and 6 sample sizes, varying either ϕ or β. The method of moments was the fastest, followed by quasi Fisher scoring.

When we defined our model, we assumed that the $\{X_t\}$ process was unknown. Howeever, in some applications we may be intererested in using a filtering method to predict each X_t value from Y_t and the previously predicted value of X_{t-1}. In the next chapter we will discuss such methods.

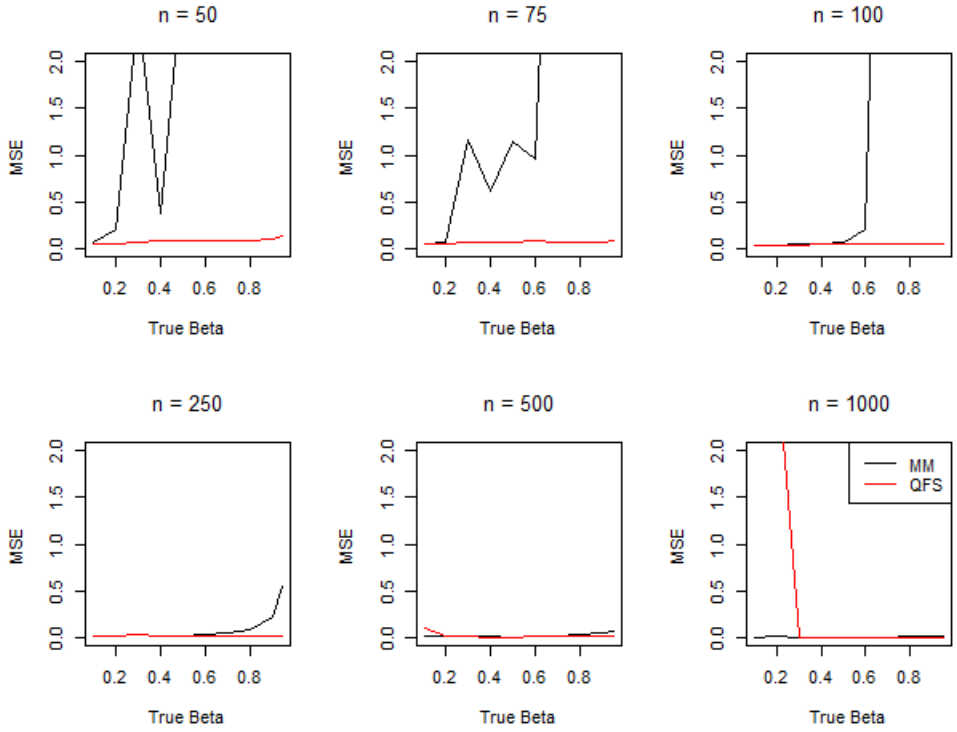

Figure 4.11: Comparison of MSE of β estimates - State space model

Table 4.2: Comparison of running time (seconds) - State space model

Method	Parameter	User	System	Elapsed
Method of Moments	β	84.99	0.38	85.66
Quasi Fisher Scoring	β	196.00	0.55	197.78
Method of Moments	ϕ	82.89	0.27	83.45
Quasi Fisher Scoring	ϕ	192.16	0.72	193.72

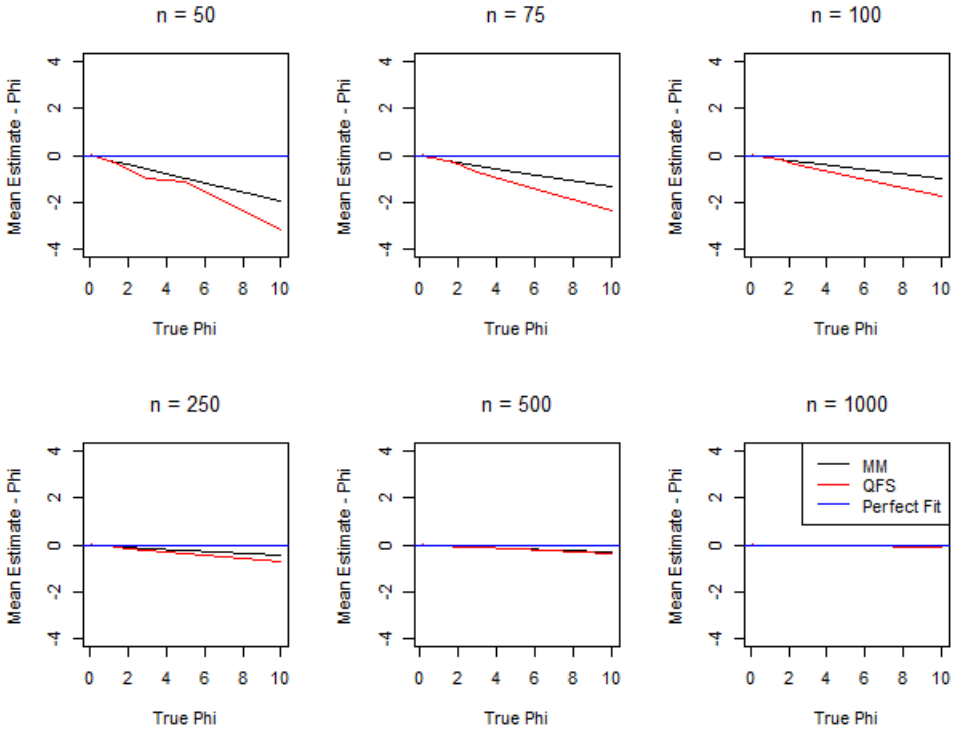

Figure 4.12: Comparison of mean parameter estimates of ϕ - State space model

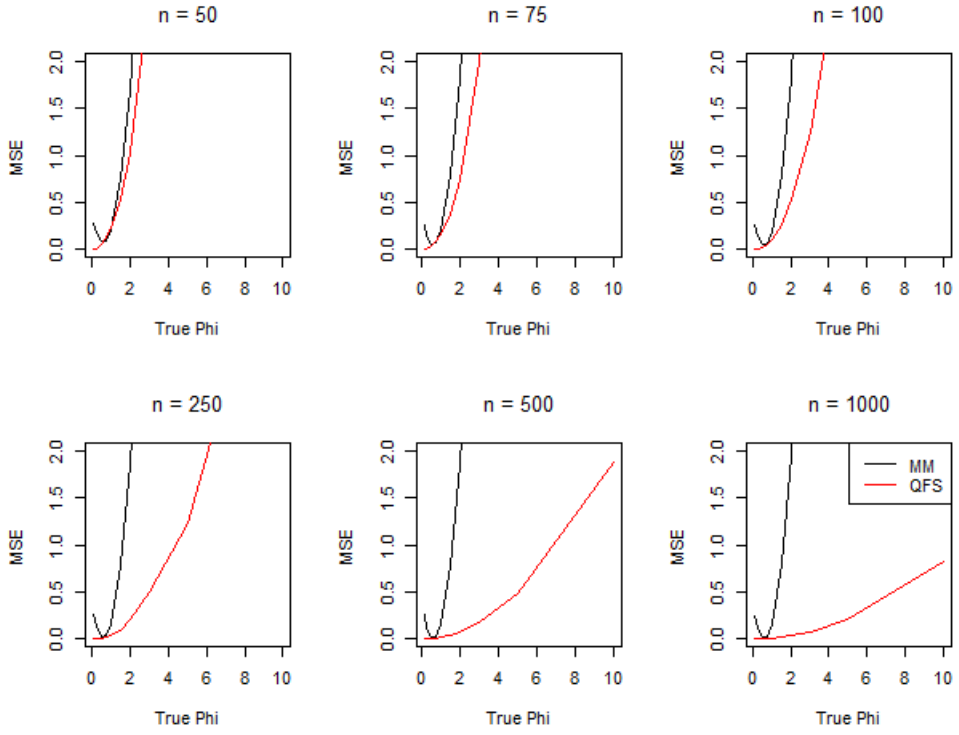

Figure 4.13: Comparison of MSE of ϕ estimates - State space model

Chapter 5

Filtering and the SISR algorithm

5.1 Importance sampling

Imagine that we have a hidden process X with distribution f, and an observed dataset Y. We are interested in the value of $I(h_t)$, an integral representing the expectation of the function h_t evaluated at $x_{0:t}$, given the observed data.

$$I(h_t) = E_{f(x_{0:t}|y_{1:t})}(h_t(x_{0:t})).$$

Since this integral is often too complicated to evaluate directly, we often must turn to numerical methods. The most obvious choice would be to generate N vectors $\{x_{0:t}^{(i)}\}_{i=1}^{N}$ from the distribution of $x_{0:t}|y_{1:t}$. Then we can estimate $I(h_t)$ as

$$I_N(h_t) = \frac{1}{N}\sum_{i=1}^{N} h_t(x_{0:t}^{(i)}).$$

Basically we have evaluated h_t at each of our generated x particles, and taken the average. One issue here is that sometimes the distribution of $x_{0:t}|y_{1:t}$ is difficult to determine or hard to sample from. For this reason we will require the concept of **importance sampling** (see Doucet [11] for an in depth discussion). We can write $I(h_t)$ as

$$I(h_t) = \int h_t(x_{0:t}) f(x_{0:t}|y_{1:t}) dx_{0:t} \tag{5.1}$$

$$= \frac{\int h_t(x_{0:t}) w(x_{0:t}) g(x_{0:t}|y_{1:t}) dx_{0:t}}{\int w(x_{0:t}) g(x_{0:t}|y_{1:t}) dx_{0:t}}. \tag{5.2}$$

where $w(x_{0:t}) = \frac{f(x_{0:t}|y_{1:t})}{g(x_{0:t}|y_{1:t})}$. We refer to $g()$ as the **importance distribution**, ideally a distribution that is somehow similar to $f()$ but easier to sample. Our choice of $g()$ is somewhat arbitrary, but we require that $g()$ include the support of $f()$. The closer $g()$ is to $f()$, the less likely it is that our method will break down.

5.2 Sequential importance sampling and resampling algorithm

The importance sampling method described in the previous section has some areas that can be improved upon, the first of which is the possibility that the method can become very computationally expensive. Since we are working with an AR(1) model, if we are sampling values at time t, we can simply retain the values from time $t-1$

instead of creating new samples each time. That way, we will end up with a set of **particles**,

$$x_t^{(j)} = (x_{t-1}^{(j)}, \tilde{x}_t^{(j)}).$$

One problem with this approach is that, as t increases, the sample weights may become almost zero for some particle values. This means that these particles will not be useful for estimation, so should be removed. The **sequential importance sampling and resampling algorithm (SISR)** solves this problem by resampling from $\{x_t^{(j)}\}_{j=1}^N$ with replacement using the sample weights $\{W_t^{(j)}\}$ [10].

We begin by initializing the values $x_0 \sim \mathcal{G}(0,1)$ and $W_0 = \frac{1}{N}$. For each value t from 1 to T, we perform the following steps.

Step 1: Sample from the importance distribution

Sample N values $\tilde{x}_t$ from $g(x_t|x_{t-1}, y_t)$, the importance distribution, and store the particles $\{x_{t-1}^{(j)}, \tilde{x}_t^{(j)}\}$. Some possible options for $g()$ include

- The unscented transformation.
- $g(x_t^{(j)}|x_{t-1}^{(j)}) \sim \exp \mathcal{S}(\alpha x_{t-1}^{(j)}, \alpha; \alpha)$.
- $g(x_t^{(j)}|x_{t-1}^{(j)}) \sim \mathcal{N}(\alpha x_{t-1} + \gamma_E(1-\alpha), (1-\alpha^2)\frac{\pi^2}{6})$

Exponential-$\mathcal{S}$ is the true distribution of $x_t|x_{t-1}$ so that is the option we will choose. How can we generate this data? Recall that the state component of our model is given by

$$X_t = \alpha X_{t-1} + \alpha \log(S_t)\ ,\ S_t \sim \mathcal{S}(\alpha).$$

In our notation, $p(X = x)$ denotes the density of X evaluated at the point x. The conditional density is given by

$$\begin{aligned} p(X_t^{(j)} = x_t | X_{t-1}^{(j)} = x_{t-1}) &= p(\alpha \log(S_t) = x_t - \alpha x_{t-1}) \\ &= \left| \exp\left(\frac{1}{\alpha} x_t - x_{t-1}\right) \right| p\left(S_t = \exp\left(\frac{1}{\alpha} x_t - x_{t-1}\right)\right) \end{aligned}$$

where the vertical brackets denote absolute value. We can generate the distribution of S_t using the *dstable* function of the *stabledist* R package, and then calculate $x_t^{(j)}$ using $S_t^{(j)}$ and $x_{t-1}^{(j)}$ Note that we have left off the effect of Y_t. However, our choice of $g()$ should theoretically not affect our results if it is close enough to the true distribution so we do not expect this to be an issue.

Step 2: Calculate weights

Generally the weights will have the form

$$\tilde{W}_t^{(j)} = W_{t-1}^{(j)} \frac{p(\tilde{x}_t^{(j)} | x_{t-1}^{(j)}) p(y_t | \tilde{x}_t^{(j)})}{g(x_t | x_{t-1}, y_t)}$$

However, since $g(x_t | x_{t-1}, y_t) = p(\tilde{x}_t^{(j)} | x_{t-1}^{(j)})$ in our case, we will only need to calculate $p(y_t | \tilde{x}_t^{(j)})$. Again we can use the *dstable* function and the distribution of ϵ_t to calculate this result.

$$Y_t = \phi X_t + \phi \log(\epsilon_t) \ , \ \epsilon_t \sim \mathcal{S}(\beta)$$

$$p(Y_t^{(j)} = y_t | \tilde{X}_t^{(j)} = x_t) = p(\phi \log(\epsilon_t) = y_t - \phi x_t)$$
$$= \left| \exp\left(\frac{1}{\phi} y_t - x_t\right) \right| p\left(\epsilon_t = \exp\left(\frac{1}{\phi} y_t - x_t.\right)\right)$$

Once the weights have been calculated, we should then proceeed to normalize the weights.

$$W_t^{(j)} = \frac{\tilde{W}_t^{(j)}}{\sum_{j=1}^{N} \tilde{W}_t^{(j)}}$$

Step 3: Compute a function of X_t

If $h(X_t)$ is our function of interest, then we can calculate an estimator

$$\hat{h}(X_t) = \sum_{j=1}^{N} W_t^{(j)} h_t(\tilde{x}_t^{(j)}).$$

In our case, we are interested in the expectation and variance,

$$\hat{\mathrm{E}}(X_t) = \sum_{j=1}^{N} W_t^{(j)} \tilde{x}_t^{(j)} \tag{5.3}$$

$$\hat{\mathrm{Var}}(X_t) = \sum_{j=1}^{N} W_t^{(j)} (\tilde{x}_t^{(j)} - \hat{\mathrm{E}}(X_t))^2. \tag{5.4}$$

Before moving on to the next step, we should resample N values with replacement from $\{\tilde{x}_t^{(j)}\}$ using the probabilities $\{W_t^{(j)}\}$, and save these along with the corresponding weights to use in the next round.

5.2.1 Results

Figure 5.1 shows predicted values for X_t for selected values of α, β and ϕ, using a fixed time series of $T = 10$, and 100 repetitions in which $n = 1000$ particles were generated each time.

This method works quite well, unless β is low and X is relatively high (greater than 1). The effectiveness of the bootstrap method is more sensitive to changes in β than in the other two parameters.

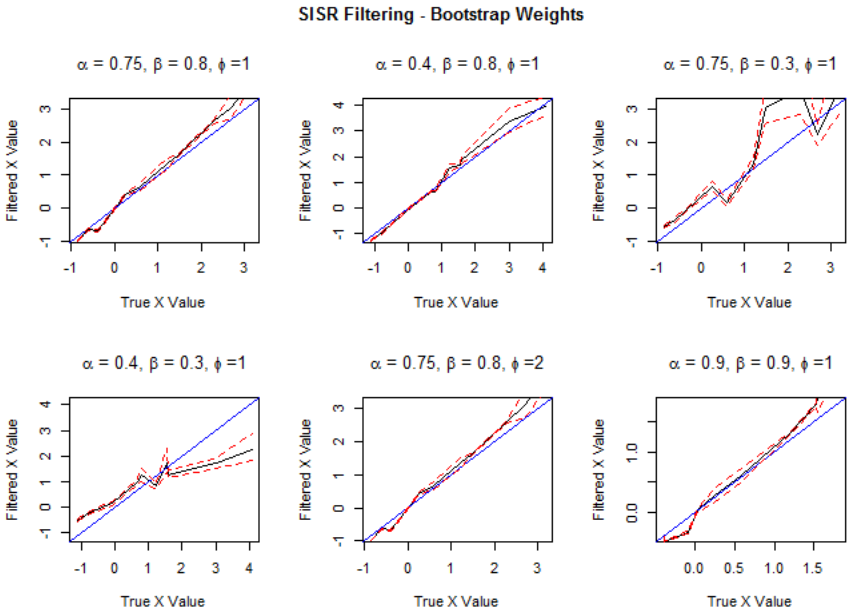

Figure 5.1: SISR Estimates for X with 95% confidence intervals (red dashed lines) - Bootstrap importance distribution. The blue line represents a perfect fit.

5.3 Alternative weights

Previously, we used the bootstrap importance distribution because it is the simplest to implement. However, it is possible that another distribution may give us better results. We will discuss several alternatives.

5.3.1 Optimal importance distribution

The optimal importance distribution is given by

$$g() = p(X_k^{(i)}|X_{k-1}^{(i)}, Y_k) \tag{5.5}$$

This is often used when our state space variable has only a finite number of discrete states, and may also be used in cases where we can determine the distribution of $X_k^{(i)}|X_{k-1}^{(i)}, Y_k$. In this case, the weights will simplify to $W_t^{(j)} = W_{t-1}^{(j)} p(Y_t|X_{t-1}^{(j)})$.

$$\begin{aligned} W_t^{(j)} &= W_{t-1}^{(j)} \frac{p(X_t^{(j)}|X_{t-1}^{(j)}) p(Y_t|X_t^{(j)})}{P(X_t^{(i)}|X_{t-1}^{(i)}, Y_t)} \\ &= W_{t-1}^{(j)} \frac{p(X_{t-1}^{(j)}, Y_t)}{p(Y_t|X_{t-1}^{(j)})} \frac{p(X_t^{(j)}, X_{t-1}^{(j)}) p(Y_t, X_t^{(j)})}{p(X_t^{(j)}, X_{t-1}^{(j)}, Y_t) p(X_t^{(j)})} \\ &= W_{t-1}^{(j)} p(Y_t|X_{t-1}^{(j)}) \frac{p(X_{t-1}^{(j)}|X_t^{(j)})}{p(X_{t-1}^{(j)}|X_t^{(j)}, Y_t)} \\ &= W_{t-1}^{(j)} p(Y_t|X_{t-1}^{(j)}) \end{aligned} \tag{5.6}$$

Note that $p(X_{t-1}^{(j)}|X_t^{(j)}) = p(X_{t-1}^{(j)}|X_t^{(j)}, Y_t)$ because $X_{t-1}^{(j)}$ only depends on Y_t through $X_t^{(j)}$.

What is the distribution of $p(Y_t|X_{t-1}^{(j)})$ for our model? Recall that our state and observation equations are

$$Y_t = \phi X_t + \phi \log(\epsilon_t)$$
$$X_t = \alpha X_{t-1} + \alpha \log(S_t).$$

Rearranging these equations, we find that

$$Y_t = \phi\alpha X_{t-1} + \alpha\phi \log(S_t) + \phi \log(\epsilon_t)$$

and therefore

$$p(y_t|x_{t-1}^{(j)}) = p(\alpha\phi x_{t-1}^{(j)} + \alpha\phi \log(S_t) + \phi \log(\epsilon_t)|x_{t-1}^{(j)}). \tag{5.7}$$

Since the terms on the right hand side are given, we can find this probability if we know the distribution of $\alpha \log(S_t) + \log(\epsilon_t)$. This distribution is a mixture of two exponential stable distributed variables. It will be difficult to find probabilities from this exact distribution, but we can determine its moments:

$$\mathrm{E}(y_t|x_{t-1}^{(j)}) = \alpha\phi x_{t-1}^{(j)} + \phi\gamma_E\left(1-\alpha+\frac{(1-\beta)}{\beta}\right), \tag{5.8}$$

$$\mathrm{Var}(y_t|x_{t-1}^{(j)}) = \phi^2\frac{\pi^2}{6}\left(1-\alpha^2+\frac{(1-\beta^2)}{\beta^2}\right). \tag{5.9}$$

5.3.2 Normal importance distribution

In the previous section, we found that we were unable to use the optimal importance distribution in our case because we could not determine probabilities associated with the mixture distribution of $\alpha\log(S_t)+\log(\epsilon_t)$.

It is possible to approximate this distribution using a normal distribution with the mean and variance found in Equations (5.8) and (5.9). In Figure 5.2 we can see that the mixed distribution is mound shaped and slightly skewed right. Using the normal distribution will result in a biased estimate.

5.3.3 Approximation using cumulants

The normal importance distribution was not a good choice due to the skewness of the data. For this reason we will need to construct a distribution that more accurately resembles $p(y_t|x_{t-1}^{(j)})$.

We will use an approximation described in Abramowitz and Stegun (1964) [1]. Let $X_1,\ldots,X_n$ be our random variables of interest with mean μ and variance σ^2. An

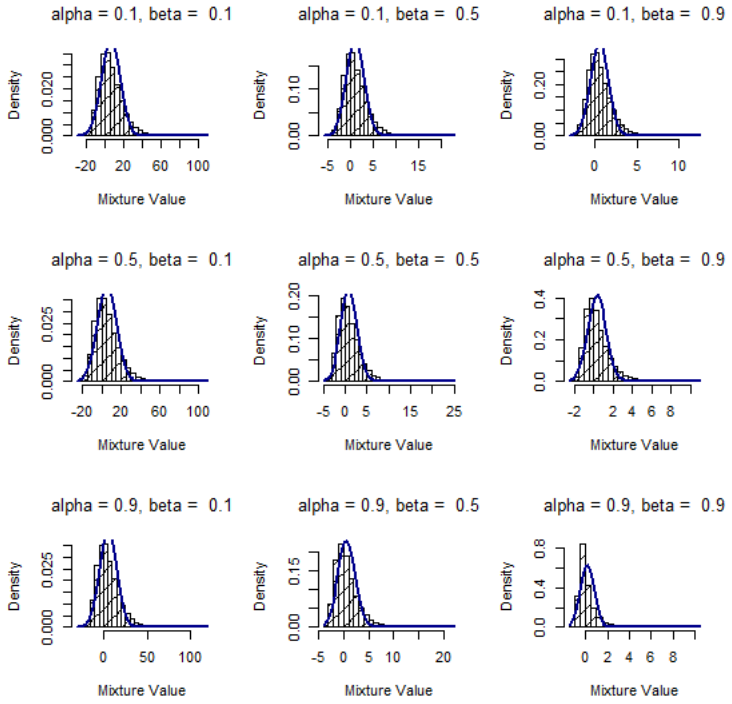

Figure 5.2: Distribution histogram of $\alpha \log(S_t) + \log(\epsilon_t)$, as compared to a normal distribution with the same mean and variance.

approximation for the distribution of X is given by

$$\begin{aligned} f(x) \sim Z(x) - \left[\frac{\gamma_1}{6} Z^{(3)}(x)\right] + \left[\frac{\gamma_2}{24} Z^{(4)}(x) + \frac{\gamma_1^2}{72} Z^{(6)}(x)\right] \\ - \left[\frac{\gamma_3}{120} Z^{(5)}(x) + \frac{\gamma_1 \gamma_2}{144} Z^{(7)}(x) + \frac{\gamma_1^3}{1296} Z^{(9)}(x)\right] + \dots \end{aligned} \tag{5.10}$$

where $\gamma_{r-2} = \frac{\kappa_r}{\kappa_2^{r/2}}$, $Z^{(r)}(x)$ is the rth derivative of the standard normal density and κ_r is the rth cumulant of Y. These standard normal derivatives can be calculated in **R**, and the cumulants of $Y_t | X_{t-1}^{(j)}$ are given by the following theorem (see Appendix

A for derivation).

Theorem 5.3.1 *The ith cumulant of* $Y_t|X_{t-1}^{(j)}$ *is equal to*

$$\kappa_i(Y_t|X_{t-1}^{(j)}) = \alpha\phi x_{t-1}^{(j)} + \phi^i(1-\alpha^i)\kappa_i(X) + \phi^i\frac{(1-\beta^i)}{\beta^i}\kappa_i(X) \text{ for } i \geq 1,$$

where $\kappa_1(X) = \mu + \sigma\gamma_E$, $\kappa_i(X) = (k-1)!\sigma^k\zeta(k)$ *for* $i > 1$, $\zeta_k = \sum_{n=1}^{\infty}\frac{1}{n^k}$ *and* $\gamma_E \approx 0.57721$ *is the Euler-Mascheroni constant.*

Many of the pieces of this formula, such as the standard normal (Z) derivatives and the cumulants of the $\mathcal{G}(0,1)$ distribution only needed to be calculated once at the beginning of the **R** program. This program will run more quickly than the bootstrap program, since we do not need to use the computationally expensive *dstable* function.

Note that sometimes this method results in negative weights, which we set to 0 to prevent errors in the program.

Results of this program are available in Figure 5.3 and the table in Appendix C. Notice that this method does not perform well for low values of α and β. To understand why, we can plot the distribution of $Y_t|X_{t-1} = 1$ and overlay a plot of the cumulant expansion with 4 terms (see Figure 5.4). When β is low, the distribution of $Y_t|X_{t-1}$ is not mound shaped, and is in fact quite uniform. Even if we were to replicate the distribution precisely, the normalized weights would be close to $\frac{1}{n}$, and therefore would be of limited use.

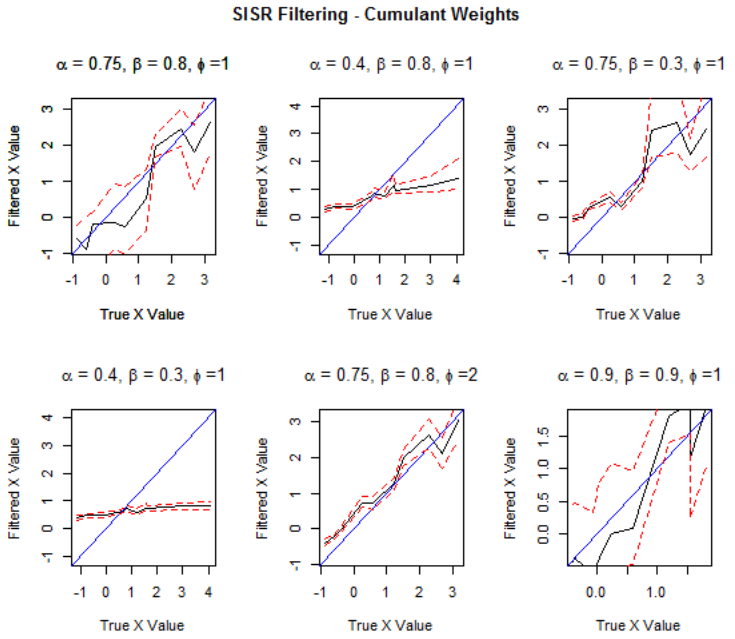

Figure 5.3: SISR estimation with cumulant method weights for various parameter values. The blue line represents a perfect fit and the red dashed lines visualize the 95% quantile confidence interval.

5.3.4 Saddlepoint approximation

The method of **saddlepoint approximation** is another method by which we can use cumulants to approximate the density of a random variable. Saddlepoint approximation was introduced by Daniels in 1954, and is also known as the method of steepest descent. We will follow the discussion and notation of Broda et al. (2012) [7].

Let X be a random variable, with density $f_X(x)$. In the previous section, we tried to estimate $f_X(x)$ using an expansion about the normal distribution, but found that the method tended to work well in the area of the mean and overestimate the tails. The basic idea of saddlepoint approximation is that we will express $f_X(x)$ in

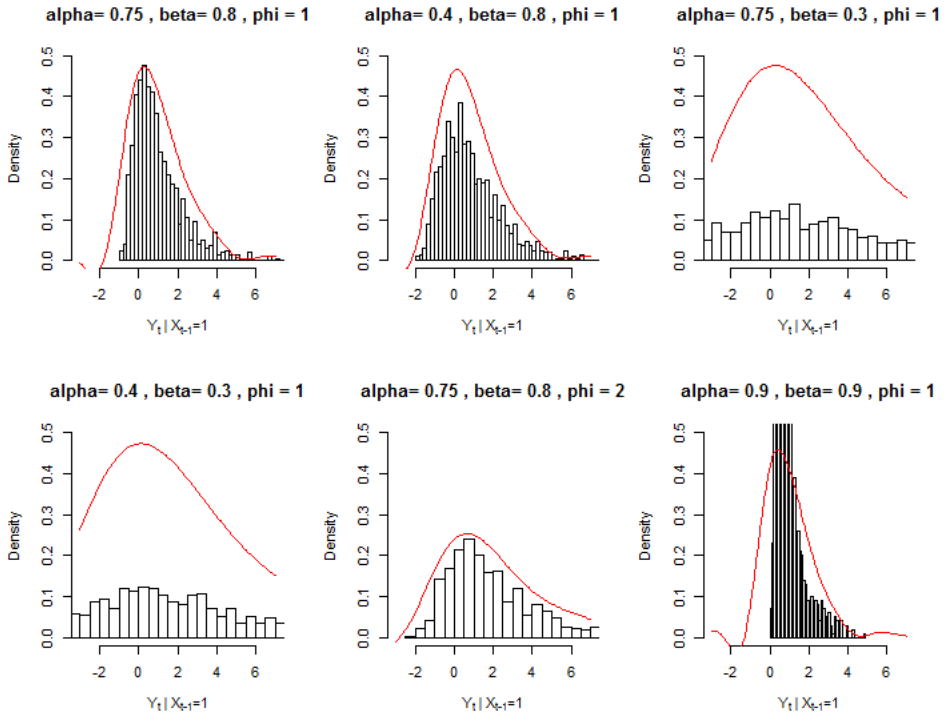

Figure 5.4: Histogram of $Y_t|X_{t-1} = 1$ with cumulant based expansion to 4 terms

terms of a normal density where x is near the mean, taking advantage of the optimal behaviour in that area.

Let T_s be a random variable with density

$$f_{T_s}(x; s) = \frac{\exp(xs) f_X(x)}{\mathbb{M}_X(s)}$$

where $\mathbb{M}_X(t)$ is the moment generating function of X. We say that T_s is exponentially tilted, and note that $\mathrm{E}(T_s) = h'_X(s)$ and $\mathrm{Var}(T_s) = h''_X(s)$, where $h_X(s)$ is the cumulant generating function of X evaluated at s. Note that when $s = 0$, the mean

and variance of T_s will be the same as for X.

Rearranging the equation, we can find an expression for $f_X(x)$ in terms of the moment generating function of X and density of T_s,

$$f_X(x) = \exp(-xs) f_{T_s}(x; s) \mathbb{M}_X(s). \tag{5.11}$$

Now, in order to fulfill our requirement that x is near the mean of the distribution, we will set $s = \check{s}$ such that $\mathrm{E}(T_s) = h_X(\check{s}) = x$. This value $\check{s}$ is calld the **saddle point**.

The first and second order approximations of the density of X will be

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi h''_X(\check{s})}} \exp(h_X(\check{s} - xs)) \tag{5.12}$$

and

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi h''_X(\check{s})}} \exp(h_X(\check{s}) - xs) \left(1 + \frac{\gamma_4}{8} - \frac{5}{24}\gamma_3^2\right) \tag{5.13}$$

where $\gamma_i = \frac{h_X^{(i)}(\check{s})}{(h''_X(\check{s}))^{i/2}}$.

In the context of our model, we will require the cumulant generating function of $Y_t|X_{t-1}$, $h_{Y_t|X_{t-1}}(t)$. It is difficult to get a closed form for this expression, but we can use the cumulants to get an approximation using the definition of the cumulant

generating function.

$$h_{Y_t|X_{t-1}}(t) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!} \tag{5.14}$$

We can see from Figures 5.5 and 5.6 that the method of saddlepoint approximation should be an improvement over the cumulant expansion method. The second order approximation works quite a bit better for high values of α and β than the first order approximation.

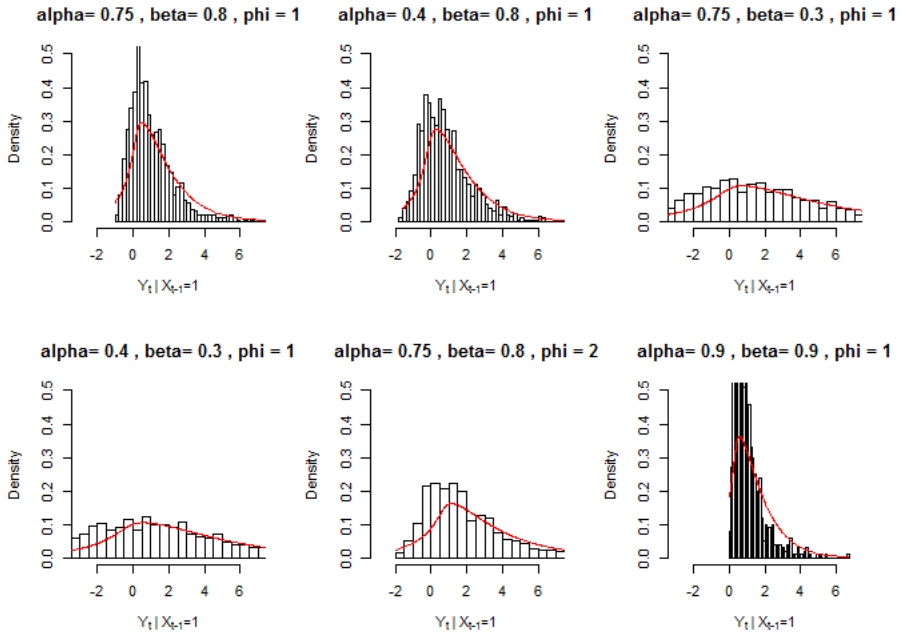

Figure 5.5: First order Saddle point approximation for various parameter values, plotted against a histogram of generateed values

The resulting first and second order estimates can be found in Figures 5.7 and 5.8 respectively, with a table of values available in Appendix C. The saddlepoint method

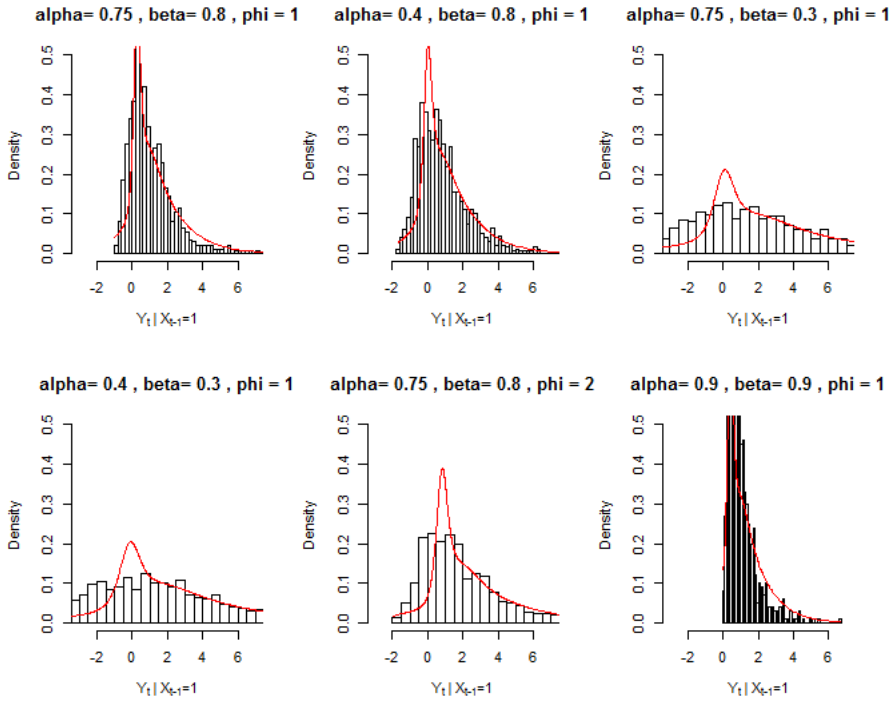

Figure 5.6: Second order Saddle point approximation for various parameter values, plotted against a histogram of generateed values

is an improvement over the previous cumulant method. However, this still does not work very well for low values of α and β. Even though the method does a better job of approximating the distribution for such parameter values, the nearly-uniform nature of the distribution will result in weights that are close to $\frac{1}{n}$. This means that we do not gain much information from the weights.

If X has a uniform distribution, then the moment generating function is not very well-behaved, and is in fact equal to 1 when $t = 0$. If we look at Figures 5.5 and 5.6, we can see that the first order approximation actually works better than the second order approximation when α or β is low. This is due to the almost-uniform nature of

the distribution - the program is trying to approximate something that is not there. If we compare Figures 5.7 and 5.8, we can see that the first order approximation works slightly better for small values of α and β, but neither method works well. One option would be to take a mixture of the two, as the first order approximation tends to underestimate the peak of each distribution and the second order approximation tends to overestimate. However, it is unlikely that even the mixture method would work as well as the bootstrap method.

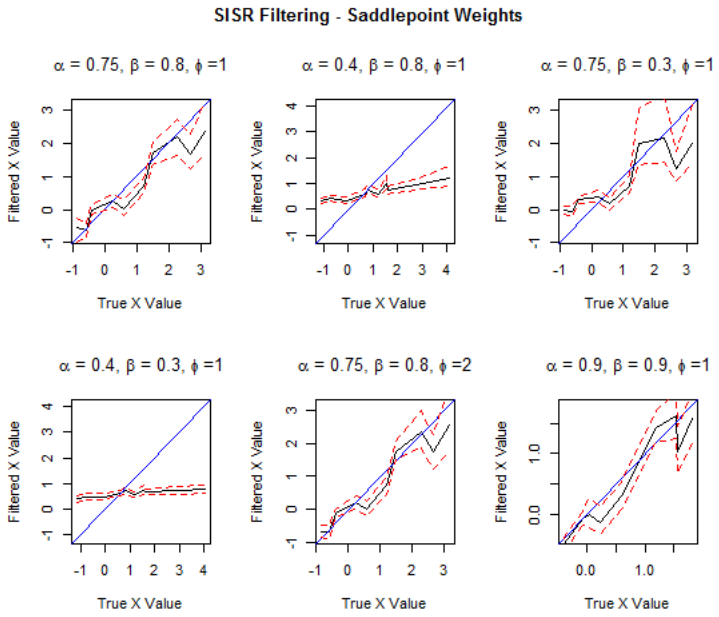

Figure 5.7: SISR estimation with first order saddle point weights for various parameter values. The blue line represents a perfect fit and the red dashed lines visualize the 95% quantile confidence interval.

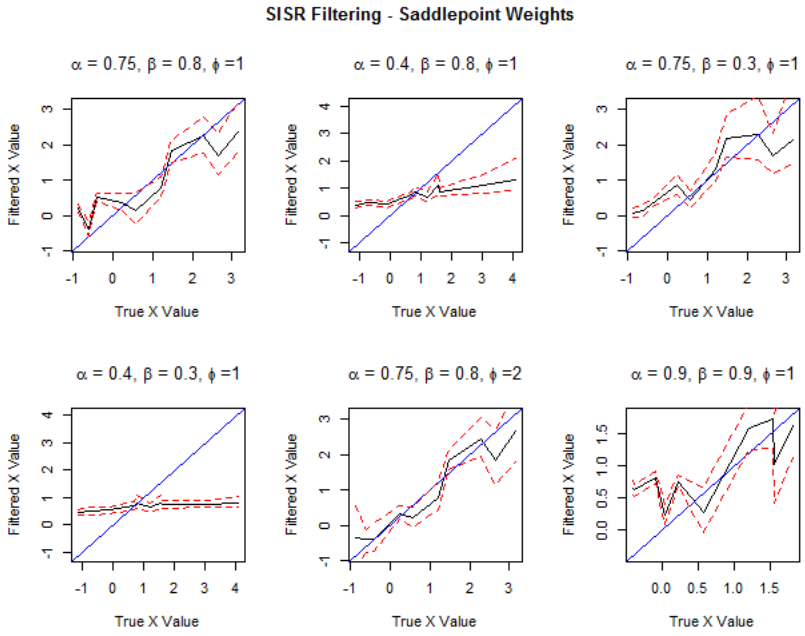

Figure 5.8: SISR estimation with second order saddle point weights for various parameter values. The blue line represents a perfect fit and the red dashed lines visualize the 95% quantile confidence interval.

Chapter 6

Applications and Concluding Remarks

6.1 Applications

Since the Gumbel distribution is part of the Generalized Extreme Value (GEV) family, it occurs naturally as the distribution of block maxima or minima of certain underlying distributions. One example would be the exponential distribution, which is used to model the time between Poisson events, or the distribution of daily and annual rainfall and river discharge volumes.

Nakajima et. al. propose a state space model with an AR(1) Gumbel distributed state step and a GEV observed step, which they use to model minimum daily stock return prices from the Tokyo Stock Price Index (TOPIX) [33]. Although Longin [29] has shown that the maxima and minima of stock prices can be best approximated

using a Fréchet distribution, there may be some industries in which a model similar to ours might be useful.

In the field of hydrology, the Gumbel distribution is often used to model block maxima of the previously mentioned exponential processes - daily rainfall totals and river discharge volumes. Perhaps a model similar to ours could be used to predict river discharge or flood levels from rainfall, although in winter those values would also be dependent on snow melt. Such a mixture of Gumbel variables would be better estimated by a GEV distribution (see [25] for a review of the methods used for extremes in hydrology).

Inter-arrival times for climate events such as floods, droughts, heat and cold waves are Gumbel (see, for example [24]). Therefore if we can identify these events as dependent on another underlying Gumbel process, we may be able to use a state space model for prediction. As of our knowledge, the model in current literature that is the most similar to ours is that of Toulemonde et. al., which is used to estimate concentrations of greenhouse gases [48].

6.1.1 Toulemonde, Guillou and Naveau 2013

This paper is an application of a state space model with Gumbel marginals to prediction of air pollution levels [48]. Methane and nitrous oxide are two greenhouse gases with correlated concentrations in the atmosphere. The maxima and minima of the concentration of these gases are both Gumbel distributed. The model that they used is as follows:

$$Y_t = v_t + H_t Z_t + \eta_{t,\alpha_2}$$

$$Z_t = \alpha_1 Z_{t-1} + \epsilon_{t,\alpha_1}$$

where $\eta_{t,\alpha_2} \sim \exp \mathcal{S}(\alpha_2, -H_t\sigma\gamma_E(\frac{1}{\alpha_2-1}), H_t\sigma)$ and $\epsilon_{t,\alpha_1} \sim \exp \mathcal{S}(\alpha_1, -\sigma\gamma_E(1-\alpha_1), \alpha_1\sigma)$ and are independent of Y_t and Z_t respectively. The distributions of the state and observed variables are $Z_t \sim \mathcal{G}(\gamma_E\sigma, \sigma)$ and $Y_t \sim \mathcal{G}(v_t - \frac{H_t\gamma_E\sigma}{\alpha_2}, H_t\frac{\sigma}{\alpha_2})$

The authors assumed that parameters were known, so did not proceed with estimation. Instead they focused on particle filtering, comparing four different filtering methods, including the Kalman filter, bootstrap filter and two versions of an auxillery particle filtering approach developed by Pitt and Shephard (1999) [37].

6.2 Conclusion

Throughout the course of this thesis, we have examined the properties of Gumbel and Exponential-$\mathcal{S}$ random variables and the relationship between them, and proposed a state space model that exploits this relationship.

$$Y_t = \phi X_t + \phi \log(\epsilon_t)$$

$$X_t = \alpha X_{t-1} + \alpha \log(S_t)$$

where $Y_t \sim \mathcal{G}(0, \frac{\phi}{\beta}$, $X_t \sim \mathcal{G}(0,1)$, $\log(\epsilon_t) \sim \exp \mathcal{S}(0,1;\beta)$ and $\log(\S_t) \sim \exp \mathcal{S}(0,1;\alpha)$. The parameters α and β are between 0 and 1, and ϕ is a positive coefficient.

Although similar models have been published in the recent past by Toulemonde, Guillou and Naveau (2013) and the PhD Thesis of Toulemonde (2008), we expand upon the research by testing different methods for parameter estimation and variable filtering [48] [47].

First we discussed how the methods of Yule-Walker and conditional least squares could be used to estimate parameters in an AR(1) time series model with $\mathcal{G}$(0,1) marginals. The method of moments and quasi Fisher scoring were then used to estimate the parameters in our proposed state space model. For each method, we tested the performance of the estimators at varying sample sizes and true parameter values, computing the mean estimate, the 95% quantile confidence interval and the mean squared error. Quasi Fisher scoring tended to work very well for this type of model.

In defining our model, the state component was assumed to be hidden. There may be some applications where it would be useful to employ filtering methods to determine the values of $X_1, X_2, \ldots, X_T$. The sequential importance sampling and re-sampling (SISR) algorithm was chosen for this purpose, with a bootstrap importance distribution and several different weighting methods. First the bootstrap weights were tested, and then several methods of estimating the optimal weights using cumulant expansions of the distribution (including the saddlepoint method). The bootstrap method worked quite well in predicting the hidden values of X_t, although the method breaks down when the parameter β is close to 0.

In the future, it would be interesting to test out estimation methods for time-dependent parameters. Gumbel distributed random variables are often encountered in the context of natural processes, so with the threat of climate change it may be increasingly necessary to include time dependence in our models. Another useful direction would be to examine how we might use our model to predict future values from current data.

Bibliography

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions*. U.S. Government Printing Office, Washington D.C., 1964.

[2] Soren Asmussen. Extreme value theory for queues vis cycle maxima. *Extremes*, 1(2):137--168, 1998.

[3] Agamirza Bashirov. *Mathematical analysis fundamentals*. Elsevier Insights, 2014.

[4] A. Bengtsson and C. Nilsson. Extreme value modelling of storm damage in swedish forests. *Natural Hazards and Earth System Sciences*, 7:515--521, 2007.

[5] N.H. Bingham and John M. Fry. *Regression: Linear Models in Statistics*. SpringerVerlag, London, 2010.

[6] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods, Second Edition*. Springer Science+Business Media LLC, New York, 2006.

[7] Simon A. Broda and Marc S. Paolella. Saddlepoint approximations: A review and some new applications. In James E. Gentle, Wolfgang Karl Härdle, and Yuichi Mori, editors, *Handbook of computational statistics*. SpringerVerlag, BerlinHeidelberg, 2012.

[8] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. SpringerVerlag, London, 2001.

[9] J.C.B. Cooper. The poisson and exponential distributions. *Mathematical Spectrum*, 37(3):123--125, 2005.

[10] Shaojie Deng, Kay Giesecke, and Tze Leung Lai. Sequential importance sampling and resampling for dynamic portfolio credit risk. *Operations Research*, 60(1):78--91, JanuaryFebruary 2012.

[11] Arnaud Doucet, Nando de Freitas, and Neil Gordon. *Sequential Monte Carlo Methods in Practice*. Springer Science+Business Media, New York, 2001.

[12] R.A. Fisher and L.H.C. Tippett. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180--190, 1928.

[13] A.L. Fougères, J.P. Nolan, and H. Rootzén. Models for dependant extremes using stable mixtures. *Scandanavian Journal of Statistics*, 36(1):42--59, March 2009.

[14] Maurice Fréchet. Sur la loi de probabilit de l'cart maximum. *Annales de la Socit Polonaise de Mathematique*, 6:93--116, 1927.

[15] Manfred Gilli and Evis Kellezi. An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(1):1--23, 2006.

[16] E.J. Gumbel. *Statistics of Extremes*. Columbia University Press, London, 1958.

[17] Martin Hanel, T. Adri Buishand, and Christopher A.T. Ferro. A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *Journal of Geophysical Research: Atmospheres*, 114(D15), August 2009.

[18] Christopher C. Heyde. *QuasiLikelihood And Its Application*. SpringerVerlag, New York, 1997.

[19] Gabriel Huerta and Bruno Sansó. Timevarying models for extreme values. *Environmental and Ecological Statistics*, 14:285--299, August 2007.

[20] James Pickands III. The twodimensional poisson process and extremal processes. *Journal of Applied Probability*, 8:745--756, 1971.

[21] James Pickands III. Statistics inference using extreme order statisticss. *The Annals of Statistics*, 3(1):119--131, 1975.

[22] A.F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158--171, 1955.

[23] Norman I. Johnson and Samuel Kotz. *Continuous Univariate Distributions 1*. Houghton Mifflin Company, Boston, 1970.

[24] K.K. Jose and Bindu Abraham. A counting process with gumbel interarrival times with applications in modeling climate data. *Journal of Environmental Statistics*, 4(5):1287--1304, February 2013.

[25] Richard W. Katz, Marc B. Parlange, and Philippe Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25:1287--1304, 2002.

[26] Lawrence A. Klimko and Paul I. Nelson. On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6(3):629--642, 1978.

[27] M.R. Leadbetter. Extremes and local dependence in stationary sequences. *Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete*, 65:291--306, 1983.

[28] Morven N. Leese. Use of censored data in the estimation of gumbel distribution parameters for annual maximum flood series. *Water Resources Research*, 9(6):1534--1542, December 1973.

[29] François M. Longin. The asympotic distribution of extreme stock market returns. *Journal of Business*, 69(3):383--408, 1996.

[30] László Mátyás. *Generalized Method of Moments*. Cambridge University Press, Cambridge, UK, 1999.

[31] Alexander J. McNeil and Thomas Saladin. Developing scenarios for future extreme losses using the pot model. In Paul Embrechts, editor, *Extremes and Integrated Risk Management*. RISK books, London, 2000.

[32] I.D. Morton, J. Bowers, and G. Mould. Estimating return period wave heights and wind speeds using a seasonal point process model. *Coastal Engineering*, 31(14):305--326, July 1997.

[33] Jouchi Nakajima, Tsuyoshi Kunihama, Yasuhiro Omori, and Sylvia Fruhwirth-Schnatter. Generalized extreme value distribution with timedependence using the ar and ma models in state space form. *Computational Statistics and Data Analysis*, 56(11):3241--3259, November 2012.

[34] P. Naveau and P. Poncet. Statespace models for precipitation maxima. *Journal de la Societe Francaise de Statistique et Revue de Statistique Appliquee*, 148:107--120, 2007.

[35] J.P. Nolan. *Stable Distributions: Models for Heavy Tailed Data*. Birkhauser, Boston, 2012.

[36] M.O. Ojo. Some relationships between the generalized gumbel and other distributions. *Kragujevac Journal of Mathematics*, 23:101--106, 2001.

[37] M.K. Pitt and N. Shephard. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, 94:590--599, 1999.

[38] C.A. Robertson and J.G. Fryer. The bias and accuracy of moment estimator. *Biometrika*, 57(1):57--65, April 1970.

[39] H.W. Rust, D. Mauran, and T.J. Osborn. Modelling seasonality in extreme precipitation. *The European Physical Journal Special Topics*, 174(1):99--111, July 2009.

[40] Gennady Samorodnitsky and M.S. Taqqu. *Stable NonGaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York, 1994.

[41] Carl Scarrott and Anna MacDonald. A review of extreme value theory threshold estimation and uncertainty quantification. *REVSTAT Statistical Journal*, 10(1):33--60, March 2012.

[42] Kyle Siegrist. Virtual laboratories in probability and statistics, 19972013.

[43] R.L. Smith. Extreme value analysis of environmental time series: an application to trend detection in groundlevel ozone. *Statistical Science*, 4(4):367--377, 1989.

[44] Takvor H. Soukissian and Panagiota M Arapi. The effect of declustering in the rlargest maxima model for the estimation of sh design values. *The Open Ocean Engineering Journal*, 4:34--43, 2011.

[45] Takvor H. Soukissian and G. Kalantzi. A new method for applying the rlargest maxima model for design seastate prediction. *International Journal of Offshore Polar Engineering*, 19(3):176--182, 2009.

[46] J.A. Tawn. An extremevalue theory model for dependent observations. *Journal of Hydrology*, 101(14):227--250, 1988.

[47] G. Toulemonde. *Estimation et tests en théorie des valeurs extrêmes*. PhD thesis, L'Université Pierre et Marie Curie, 2008.

[48] G. Toulemonde, A. Guillou, and P. Naveau. Particle filtering for gumbel-distributed daily maxima of methane and nitrous oxide. *Environmetrics*, 24:51--62, 2013.

[49] E. Vanem. Longterm timedependent stochastic modelling of extreme waves. *Stochastic Environmental Research and Risk Assessment*, 25:185--209, 2011.

[50] Peter Waylen and Ming ko Woo. Prediction of annual floods generated by mixed processes. *Water Resources Research*, 18(4):1283--1286, 1982.

[51] R.W.M. Wedderburn. Quasilikelihood functions, generalized linear models, and the gaussnewton method. *Biometrika*, 61(3):439--447, 1974.

[52] Eric W. Weisstein. *CRC Concise Encyclopedia of Mathematics, Second Edition*. CRC Press, Boca Raton, FL, 2002.

[53] Mike West, P. Jeff Harrison, and Helio S. Mignon. Dynamic generalized linear models and baysian forecasting. *Journal of the American Statistical Association*, 80(389):73--83, 1985.

Appendix A

Cumulants of observations conditional on the previous state variable

We are interested in calculating the cumulants of $Y_t|X_{t-1}$. Recall that the moment generating function of the distribution of X is given by

$$M_X(t) = \mathrm{E}(\exp(tX)).$$

If we take the log of this function, we end up with

$$h_X(t) = \log(\mathrm{E}(\exp(tX))),$$

which is referred to as the **cumulant generating function**. Cumulants follow these

rules:

- If X and Y are independent random variables, then $\kappa_i(X+Y) = \kappa_i(X) + \kappa_i(Y)$
- If c is a constant, then $\kappa_i(cX) = c^i\kappa_i(X)$ and $\kappa_i(c+X) = c + \kappa_i(X)$.

Therefore, if we want to find the cumulants of $\alpha\phi x_{t-1}^{(j)} + \alpha\phi\log(S_t) + \phi\log(\epsilon_t)|x_{t-1}^{(j)}$, we can start by calculating the cumulants of $\log(S_t)$ and $\log(\epsilon_t)$. Note that $x_{t-1}^{(j)}$ will be treated as a constant here.

We do not know the cumulant generating function of $\log(S_i)$. If we try to calculate this in the usual way, we will find that

$$\begin{aligned} h_{\log(S_i)}(t) &= \log(\mathrm{E}(\exp(t\log(S_i)))) \\ &= \log(\mathrm{E}(S_i^t)). \end{aligned} \tag{A.1}$$

Unfortunately, since S_i is α-stable with $\alpha < 1$, the second and higher moments of S_i will not exist, so $\mathrm{E}(S_i^t)$ will not be defined (see Nolan (2009) [35]). If we try to obtain the moment generating function we will run into the same problem as well. However, there is a way of getting around this issue.

In our model, $X_t = \alpha X_{t-1} + \alpha\log(S_t)$, and we know that X_{t-1} and $\log(S_t)$ are independent. Therefore we can use our rules to determine that

$$\begin{aligned} \kappa_i(X_t) &= \alpha^i\kappa_i(X_{t-1}) + \alpha^i\kappa_i(\log(S_t)) \\ \kappa_i(\log(S_t)) &= \frac{(1-\alpha^i)}{\alpha^i}\kappa_i(X) \end{aligned} \tag{A.2}$$

We have set $\kappa_i(X_t) = \kappa_i(X_{t-1}) = \kappa_i(X)$ because X_t and X_{t-1} are identically distributed and would therefore have the same moments and cumulants. So therefore, if we can determine the cumulant generating function of a Gumbel random variable, we can use that to find the cumulant generating function of $\log(S_t)$.

In Section 2.4.1 we found that the moment generating function of $X \sim \mathcal{G}(\mu, \sigma)$ is given by $M_X(t) = \exp(\mu t)\Gamma(1 - \sigma t)$, where $\Gamma()$ is the gamma function. We can take the log of this expression to find that the cumulant generating function of X is

$$\begin{aligned} h_X(t) &= \log(\exp(\mu t)\Gamma(1 - \sigma t)) \\ &= \mu t + \log(\Gamma(1 - \sigma t)). \end{aligned} \tag{A.3}$$

From this cumulant generating function, we can take the ith derivative and evaluate at 0 to get the ith cumulant.

$$\kappa_i(X) = \frac{\partial^i}{\partial t^i} h_X(0) \tag{A.4}$$

Our next step is to find the derivative of $h_X(t)$. In Section 2.4.1, we derived the mean and variance of the Gumbel distribution using the digamma and trigamma functions. Since we will be looking at higher moments, we will need to use the more general **polygamma function**, given by

$$\varphi^(n)(Z) = \frac{\partial^n}{\partial Z^n} \log \Gamma(Z) \tag{A.5}$$

$$\begin{aligned} h'_X(t) &= \mu - \sigma \frac{\Gamma'(1-\sigma t)}{\Gamma(1-\sigma t)} \\ &= \mu - \sigma\varphi(1-\sigma t) \end{aligned} \tag{A.6}$$

$$\begin{aligned} h^{(2)}_X(t) &= \sigma^2\varphi^{(1)}(1-\sigma t) \\ h^{(3)}_X(t) &= -\sigma^3\varphi^{(2)}(1-\sigma t) \\ &\cdots \\ h^{(k)}_X(t) &= (-1)^k\sigma^k\varphi^{(k-1)}(1-\sigma t). \end{aligned} \tag{A.7}$$

When $t = 0$, $1 - \sigma t = 1$, so we are interested in calculating the first few derivatives of φ evaluated at $Z = 1$. Fortunately, this has a known form (Abramowitz and Stegun (1964) [1]):

$$\varphi^{(n)}(1) = (-1)^{n+1} n! \zeta(n+1) \tag{A.8}$$

where $\zeta(n) = \sum_{k=1}^{\infty} \frac{1}{k^n}$, the Riemann zeta function, which converges for $n > 1$. These values can be easily found using a short **R** program.

$$\begin{aligned} \zeta(2) &= \frac{\pi^2}{6} \\ \zeta(3) &\approx 1.202 \\ \zeta(4) &\approx 1.0823 \end{aligned}$$

We are now ready to start putting all of these pieces together to find the cumulants

of $X \sim \mathcal{G}(\mu, \sigma)$.

$$\begin{aligned}\kappa_1(X) &= h'_X(0) \\ &= \mu - \sigma\varphi(1) \\ &= \mu + \sigma\gamma_E \end{aligned} \tag{A.9}$$

where $\gamma_E \approx 0.57721$ is the Euler-Mascheroni constant, which is known to be the value of $-\varphi(1)$. In general for $i > 1$, the cumulants of X are given by

$$\begin{aligned}\kappa_i(X) &= h_X^{(i)}(0) \\ &= (-1)^k \sigma^k \varphi^{(k-1)}(1) \\ &= (-1)^k \sigma^k (-1)^k (k-1)! \zeta(k) \\ &= (k-1)! \sigma^k \zeta(k) \end{aligned} \tag{A.10}$$

For example, the 2nd, 3rd and 4th cumulants are:

$$\begin{aligned}\kappa_2(X) &= \sigma^2 \frac{\pi^2}{6}, \\ \kappa_3(X) &= 2\sigma^3 \zeta(2), \\ \kappa_4(X) &= 6\sigma^4 \zeta(3). \end{aligned}$$

$\kappa_1(X)$ and $\kappa_2(X)$ are equal to the mean and variance respectively of a Gumbel distribution with parameters μ and σ, as we would expect.

For $X_0 \sim \mathcal{G}(0, 1)$, these cumulants will be

$$\begin{aligned}\kappa_1(X_0) &= \gamma_E \\ \kappa_2(X_0) &= \frac{\pi^2}{6}, \\ \kappa_3(X_0) &= 2\zeta(2), \\ \kappa_4(X_0) &= 6\zeta(3).\end{aligned}$$

The cumulants of $\log S_t$ will be

$$\kappa_i(\log(S_t)) = \frac{(1-\alpha^i)}{\alpha^i}\kappa_i(X_0)$$

so that the first four cumulants will be

$$\begin{aligned}\kappa_1(\log(S_t)) &= \frac{(1-\alpha)}{\alpha}\gamma_E, \\ \kappa_2(\log(S_t)) &= \frac{(1-\alpha^2)}{\alpha^2}\frac{\pi^2}{6}, \\ \kappa_3(\log(S_t)) &= 2\frac{(1-\alpha^3)}{\alpha^3}\zeta(2), \\ \kappa_4(\log(S_t)) &= 6\frac{(1-\alpha^4)}{\alpha^4}\zeta(3).\end{aligned}$$

Notice that again, the first two cumulants match up to the mean and variance of $\log(S_t)$ that we had previously calculated in Section 3.1. Since $\log(S_t) \sim \exp \mathcal{S}(0, 1; \alpha)$ and $\log(\epsilon_t) \sim \exp \mathcal{S}(0, 1; \beta)$, it can be shown that

$$\kappa_i(\log(\epsilon_t)) = \frac{(1-\beta^i)}{\beta^i}\kappa_i(X_0).$$

Finally we find that the ith cumulant of $Y_t|X_{t-1}^{(j)} = \alpha\phi x_{t-1}^{(j)} + \alpha\phi\log(S_t) + \phi\log(\epsilon_t)|x_{t-1}^{(j)}$ will be

$$\kappa_i(Y_t|X_{t-1}^{(j)}) = \alpha\phi x_{t-1}^{(j)} + \phi^i(1-\alpha^i)\kappa_i(X_0) + \phi^i\frac{(1-\beta^i)}{\beta^i}\kappa_i(X_0) \text{ for } i \geq 1,$$

where $\kappa_1(X_0) = \mu + \sigma\gamma_E$, $\kappa_i(X_0) = (k-1)!\sigma^k\zeta(k)$ for $i > 1$ and $\zeta_k = \sum_{n=1}^{\infty}\frac{1}{n^k}$.

Appendix B

Tables of estimation results

The tables in this section include parameter estimation results under the first order auto-regressive model and also the observed component of our state space model. Each table includes the mean estimate over 500 samples, the 95% quantile bootstrap confidence interval and the mean squared error for each of the methods, by sample size N and true parameter value. Default values of $\alpha = 0.75$, $\beta = 0.6$ and $\phi = 1$ were used, with only one parameter varying at a time. When estimating β and ϕ, we assumed that both were unknown but α was known.

Table B.1: Comparison of estimation methods - First order auto-regressive model.

		Yule-Walker			**Conditional least squares**		
N	α	Mean	95 % CI	MSE	Mean	95 % CI	MSE
50	0.1	0.070	(-0.20, 0.33)	0.7929	0.097	(-0.18, 0.36)	0.7474
50	0.2	0.171	(-0.10, 0.43)	0.6251	0.196	(-0.06, 0.47)	0.5882

Continued on next page

Table B.1 – continued from previous page

		Yule-Walker			Conditional least squares		
N	α	Mean	95 % CI	MSE	Mean	95 % CI	MSE
50	0.3	0.260	(0.00, 0.50)	0.4922	0.292	(0.02, 0.55)	0.4503
50	0.4	0.344	(0.10, 0.58)	0.3812	0.380	(0.15, 0.61)	0.3394
50	0.5	0.440	(0.21, 0.64)	0.2738	0.482	(0.24, 0.69)	0.2334
50	0.6	0.542	(0.29, 0.72)	0.1792	0.584	(0.34, 0.78)	0.1469
50	0.7	0.646	(0.43, 0.81)	0.1015	0.684	(0.46, 0.85)	0.0810
50	0.8	0.733	(0.53, 0.87)	0.0542	0.775	(0.58, 0.91)	0.0385
50	0.9	0.846	(0.67, 0.92)	0.0153	0.875	(0.70, 0.99)	0.0111
50	0.95	0.906	(0.77, 0.95)	0.0037	0.924	(0.79, 1.00)	0.0039
75	0.1	0.082	(-0.13, 0.31)	0.7646	0.099	(-0.12, 0.32)	0.7372
75	0.2	0.185	(-0.02, 0.41)	0.5981	0.202	(-0.01, 0.42)	0.5724
75	0.3	0.268	(0.03, 0.47)	0.4760	0.291	(0.07, 0.49)	0.4462
75	0.4	0.369	(0.14, 0.58)	0.3498	0.393	(0.17, 0.60)	0.3221
75	0.5	0.460	(0.25, 0.63)	0.2499	0.487	(0.28, 0.67)	0.2241
75	0.6	0.562	(0.36, 0.73)	0.1598	0.591	(0.38, 0.76)	0.1378
75	0.7	0.655	(0.49, 0.80)	0.0933	0.686	(0.51, 0.83)	0.0764
75	0.8	0.756	(0.61, 0.87)	0.0424	0.781	(0.62, 0.90)	0.0339
75	0.9	0.867	(0.75, 0.93)	0.0093	0.883	(0.76, 0.97)	0.0071
75	0.95	0.925	(0.84, 0.95)	0.0016	0.933	(0.84, 0.99)	0.0020
100	0.1	0.085	(-0.10, 0.26)	0.7567	0.096	(-0.09, 0.28)	0.7377
100	0.2	0.180	(0.01, 0.36)	0.6018	0.193	(0.02, 0.37)	0.5811
100	0.3	0.280	(0.10, 0.45)	0.4571	0.296	(0.11, 0.47)	0.4360
100	0.4	0.379	(0.20, 0.54)	0.3341	0.397	(0.22, 0.56)	0.3136
100	0.5	0.483	(0.31, 0.63)	0.2257	0.502	(0.32, 0.65)	0.2078
100	0.6	0.568	(0.39, 0.71)	0.1521	0.587	(0.42, 0.73)	0.1379
100	0.7	0.667	(0.53, 0.78)	0.0842	0.686	(0.54, 0.81)	0.0744
100	0.8	0.762	(0.64, 0.86)	0.0385	0.783	(0.65, 0.89)	0.0317
100	0.9	0.871	(0.75, 0.93)	0.0083	0.886	(0.77, 0.96)	0.0065
100	0.95	0.930	(0.86, 0.96)	0.0012	0.936	(0.85, 0.99)	0.0015
250	0.1	0.092	(-0.03, 0.22)	0.7398	0.097	(-0.02, 0.22)	0.7315
250	0.2	0.196	(0.07, 0.31)	0.5723	0.202	(0.07, 0.32)	0.5632

Continued on next page

Table B.1 – continued from previous page

		Yule-Walker			Conditional least squares		
N	α	Mean	95 % CI	MSE	Mean	95 % CI	MSE
250	0.3	0.290	(0.17, 0.40)	0.4396	0.295	(0.18, 0.40)	0.4323
250	0.4	0.388	(0.28, 0.50)	0.3197	0.395	(0.28, 0.51)	0.3111
250	0.5	0.493	(0.38, 0.60)	0.2119	0.501	(0.39, 0.60)	0.2047
250	0.6	0.591	(0.50, 0.69)	0.1310	0.600	(0.51, 0.69)	0.1250
250	0.7	0.688	(0.60, 0.76)	0.0705	0.696	(0.60, 0.78)	0.0665
250	0.8	0.782	(0.71, 0.85)	0.0293	0.790	(0.71, 0.86)	0.0269
250	0.9	0.887	(0.83, 0.93)	0.0045	0.895	(0.83, 0.94)	0.0039
250	0.95	0.943	(0.90, 0.96)	0.0003	0.945	(0.90, 0.98)	0.0004
500	0.1	0.099	(0.01, 0.18)	0.7265	0.101	(0.01, 0.18)	0.7223
500	0.2	0.195	(0.11, 0.28)	0.5716	0.198	(0.12, 0.29)	0.5677
500	0.3	0.301	(0.22, 0.38)	0.4233	0.304	(0.22, 0.39)	0.4191
500	0.4	0.393	(0.31, 0.47)	0.3119	0.396	(0.31, 0.47)	0.3081
500	0.5	0.493	(0.42, 0.56)	0.2103	0.497	(0.42, 0.57)	0.2067
500	0.6	0.594	(0.52, 0.67)	0.1280	0.598	(0.52, 0.67)	0.1252
500	0.7	0.693	(0.63, 0.75)	0.0669	0.698	(0.63, 0.76)	0.0646
500	0.8	0.793	(0.74, 0.84)	0.0255	0.797	(0.74, 0.85)	0.0240
500	0.9	0.894	(0.86, 0.92)	0.0035	0.897	(0.85, 0.93)	0.0032
500	0.95	0.945	(0.92, 0.96)	0.0001	0.947	(0.91, 0.97)	0.0002
1000	0.1	0.097	(0.03, 0.16)	0.7281	0.098	(0.03, 0.16)	0.7261
1000	0.2	0.199	(0.14, 0.26)	0.5648	0.200	(0.14, 0.26)	0.5628
1000	0.3	0.296	(0.23, 0.35)	0.4287	0.297	(0.23, 0.35)	0.4267
1000	0.4	0.396	(0.34, 0.45)	0.3082	0.397	(0.35, 0.45)	0.3061
1000	0.5	0.498	(0.45, 0.55)	0.2053	0.500	(0.45, 0.56)	0.2036
1000	0.6	0.597	(0.55, 0.65)	0.1249	0.600	(0.55, 0.65)	0.1234
1000	0.7	0.696	(0.65, 0.74)	0.0649	0.699	(0.65, 0.74)	0.0636
1000	0.8	0.798	(0.76, 0.83)	0.0236	0.800	(0.76, 0.84)	0.0229
1000	0.9	0.896	(0.87, 0.92)	0.0031	0.898	(0.87, 0.92)	0.0029
1000	0.95	0.948	(0.93, 0.96)	0.0001	0.949	(0.93, 0.97)	0.0001

Table B.2: Comparison of estimation methods - Observed step, estimation of α.

		Method of Moments		
N	α	Mean	95 % CI	MSE
50	0.1	0.578	(-10.40, 12.64)	178.96
50	0.2	-0.319	(-11.91, 11.08)	7640.4
50	0.3	-1.250	(-14.23, 7.45)	232.80
50	0.4	7.270	(-5.71, 9.36)	23605
50	0.5	0.132	(-5.15, 6.92)	103.95
50	0.6	0.531	(-4.80, 7.61)	29.673
50	0.7	0.280	(-6.30, 5.13)	21.075
50	0.8	7.835	(-3.58, 5.66)	23785
50	0.9	2.160	(-5.41, 7.02)	1351.6
50	0.95	1.148	(-6.48, 5.56)	578.44
75	0.1	0.199	(-9.29, 13.57)	120.49
75	0.2	-0.544	(-14.72, 12.25)	129.64
75	0.3	-1.158	(-13.84, 9.89)	533.17
75	0.4	0.485	(-6.26, 7.80)	27.186
75	0.5	0.777	(-3.81, 7.19)	26.836
75	0.6	1.142	(-2.65, 6.24)	108.91
75	0.7	1.139	(-2.52, 3.82)	133.33
75	0.8	0.713	(-1.10, 3.62)	58.022
75	0.9	0.815	(-1.64, 4.92)	10.311
75	0.95	0.982	(-3.05, 7.52)	55.782
100	0.1	40.609	(-15.01, 13.35)	788238
100	0.2	0.175	(-13.23, 15.17)	3450.1
100	0.3	-0.946	(-7.29, 7.19)	268.41
100	0.4	-0.186	(-8.77, 7.05)	171.19
100	0.5	0.636	(-2.72, 4.06)	11.3931
100	0.6	1.083	(-1.85, 3.13)	82.003
100	0.7	0.546	(-1.57, 2.82)	8.7567

Continued on next page

Table B.2 – continued from previous page

		Method of Moments		
N	α	Mean	95 % CI	MSE
100	0.8	0.741	(-0.60, 3.21)	10.707
100	0.9	0.855	(-1.29, 4.00)	8.1131
100	0.95	0.624	(-3.83, 4.88)	28.329
250	0.1	-1.355	(-20.44, 12.26)	247.65
250	0.2	-0.222	(-7.84, 6.16)	66.982
250	0.3	-2.028	(-2.99, 2.96)	2704.8
250	0.4	0.180	(-1.30, 2.01)	28.8659
250	0.5	0.498	(-0.46, 1.82)	0.9826
250	0.6	0.591	(-0.13, 1.59)	0.8956
250	0.7	0.748	(0.10, 1.47)	0.4128
250	0.8	0.836	(0.32, 1.51)	0.2511
250	0.9	0.932	(0.42, 1.74)	0.1309
250	0.95	1.049	(0.38, 2.32)	0.9913
500	0.1	-4.093	(-5.80, 7.85)	13119.4
500	0.2	0.129	(-3.71, 3.75)	33.511
500	0.3	0.221	(-1.05, 1.91)	6.1852
500	0.4	0.390	(-0.39, 1.36)	0.2348
500	0.5	0.492	(-0.08, 1.14)	0.1124
500	0.6	0.608	(0.15, 1.15)	0.0902
500	0.7	0.704	(0.33, 1.15)	0.0432
500	0.8	0.805	(0.47, 1.19)	0.0307
500	0.9	0.908	(0.62, 1.28)	0.0265
500	0.95	0.965	(0.66, 1.34)	0.0337
1000	0.1	1.148	(-4.81, 8.29)	633.51
1000	0.2	0.133	(-0.97, 1.38)	2.0203
1000	0.3	0.283	(-0.33, 0.97)	0.1127
1000	0.4	0.386	(-0.07, 0.86)	0.0526
1000	0.5	0.487	(0.15, 0.85)	0.0311
1000	0.6	0.588	(0.30, 0.88)	0.0206
1000	0.7	0.689	(0.45, 0.93)	0.0147

Continued on next page

Table B.2 – continued from previous page

		Method of Moments		
N	α	Mean	95 % CI	MSE
1000	0.8	0.789	(0.59, 1.00)	0.0110
1000	0.9	0.890	(0.69, 1.07)	0.0090
1000	0.95	0.940	(0.74, 1.13)	0.0095

Table B.3: Comparison of estimation methods - Observed step, estimation of β.

		Method of Moments			Quasi Fisher Scoring		
N	β	Mean	95 % CI	MSE	Mean	95 % CI	MSE
50	0.1	0.181	(0.00, 0.83)	0.0799	0.169	(0.00, 0.67)	0.0524
50	0.2	0.224	(0.00, 0.91)	0.2132	0.198	(0.00, 0.69)	0.0568
50	0.3	0.203	(0.00, 0.95)	2.5294	0.279	(0.00, 0.92)	0.0747
50	0.4	0.354	(0.00, 1.35)	0.3783	0.375	(0.00, 1.09)	0.0907
50	0.5	0.552	(0.00, 2.11)	2.9838	0.476	(0.00, 1.20)	0.0938
50	0.6	3.305	(0.00, 2.58)	3324.9	0.570	(0.10, 1.18)	0.0927
50	0.7	0.760	(-1.74, 4.37)	11.038	0.652	(0.14, 1.23)	0.0943
50	0.8	1.817	(-5.14, 4.97)	1247.3	0.707	(0.20, 1.25)	0.096
50	0.9	0.527	(-8.34, 6.28)	19.991	0.747	(0.27, 1.26)	0.1098
50	0.95	3.655	(-6.23, 8.84)	2688.1	0.777	(0.29, 1.27)	0.1409
75	0.1	0.171	(0.00, 0.66)	0.049	0.171	(0.00, 0.64)	0.0473
75	0.2	0.214	(0.00, 0.70)	0.0717	0.210	(0.00, 0.69)	0.0509
75	0.3	0.320	(0.00, 0.85)	1.1725	0.281	(0.00, 0.85)	0.0679
75	0.4	0.336	(0.00, 1.11)	0.6179	0.373	(0.00, 1.02)	0.0755
75	0.5	0.437	(0.00, 1.52)	1.1415	0.481	(0.06, 1.15)	0.0791
75	0.6	0.632	(0.00, 2.29)	0.9646	0.586	(0.15, 1.20)	0.0803
75	0.7	0.789	(0.00, 3.67)	7.2381	0.677	(0.21, 1.23)	0.0769
75	0.8	1.544	(-0.78, 4.75)	93.774	0.754	(0.32, 1.24)	0.0714

Continued on next page

Table B.3 – continued from previous page

		Method of Moments			Quasi Fisher Scoring		
N	β	Mean	95 % CI	MSE	Mean	95 % CI	MSE
75	0.9	6.478	(-6.83, 7.08)	12951	0.803	(0.35, 1.27)	0.0769
75	0.95	1.318	(-4.49, 7.17)	96.165	0.822	(0.37, 1.26)	0.0819
100	0.1	0.145	(0.00, 0.57)	0.0333	0.146	(0.00, 0.57)	0.0331
100	0.2	0.183	(0.00, 0.61)	0.0377	0.185	(0.00, 0.61)	0.0372
100	0.3	0.253	(0.00, 0.69)	0.048	0.258	(0.00, 0.69)	0.0452
100	0.4	0.353	(0.00, 0.84)	0.0581	0.364	(0.00, 0.84)	0.0492
100	0.5	0.479	(0.00, 1.04)	0.0757	0.480	(0.10, 0.97)	0.0543
100	0.6	0.639	(0.14, 1.47)	0.2002	0.596	(0.17, 1.11)	0.0594
100	0.7	0.906	(0.28, 2.15)	8.969	0.694	(0.30, 1.18)	0.0601
100	0.8	1.029	(0.34, 2.72)	6.6407	0.773	(0.35, 1.20)	0.0585
100	0.9	-0.251	(0.38, 4.21)	1401.8	0.833	(0.40, 1.25)	0.0601
100	0.95	1.891	(0.38, 5.91)	99.707	0.860	(0.41, 1.26)	0.0619
250	0.1	0.138	(0.00, 0.45)	0.0234	0.140	(0.00, 0.45)	0.0242
250	0.2	0.184	(0.00, 0.52)	0.0273	0.184	(0.00, 0.52)	0.027
250	0.3	0.264	(0.00, 0.61)	0.0316	0.263	(0.00, 0.61)	0.0324
250	0.4	0.372	(0.00, 0.74)	0.0288	0.374	(0.00, 0.74)	0.0273
250	0.5	0.488	(0.20, 0.85)	0.0257	0.488	(0.20, 0.85)	0.0245
250	0.6	0.602	(0.33, 1.05)	0.0307	0.598	(0.33, 0.98)	0.0256
250	0.7	0.720	(0.43, 1.23)	0.0472	0.703	(0.43, 1.07)	0.0282
250	0.8	0.847	(0.50, 1.52)	0.088	0.794	(0.50, 1.11)	0.0285
250	0.9	0.990	(0.56, 2.03)	0.2242	0.871	(0.56, 1.15)	0.0276
250	0.95	1.078	(0.59, 2.34)	0.5432	0.903	(0.59, 1.18)	0.0279
500	0.1	0.105	(0.00, 0.36)	0.015	0.102	(-0.03, 0.41)	0.0988
500	0.2	0.168	(0.00, 0.41)	0.0181	0.167	(0.00, 0.41)	0.0183
500	0.3	0.269	(0.00, 0.49)	0.0168	0.268	(0.00, 0.49)	0.0172
500	0.4	0.382	(0.14, 0.58)	0.0119	0.382	(0.14, 0.58)	0.0115
500	0.5	0.491	(0.31, 0.71)	0.0104	0.491	(0.31, 0.71)	0.0104
500	0.6	0.598	(0.42, 0.85)	0.0127	0.598	(0.42, 0.85)	0.0124
500	0.7	0.706	(0.50, 1.03)	0.0189	0.702	(0.50, 0.97)	0.0158
500	0.8	0.817	(0.56, 1.26)	0.0304	0.796	(0.56, 1.05)	0.0173

Continued on next page

Table B.3 – continued from previous page

		Method of Moments			**Quasi Fisher Scoring**		
N	β	Mean	95 % CI	MSE	Mean	95 % CI	MSE
500	0.9	0.931	(0.63, 1.55)	0.0508	0.874	(0.63, 1.10)	0.0169
500	0.95	0.990	(0.66, 1.70)	0.0665	0.907	(0.66, 1.12)	0.0172
1000	0.1	0.106	(0.00, 0.30)	0.0109	0.244	(-0.25, 0.52)	9.5684
1000	0.2	0.182	(0.00, 0.35)	0.0119	0.100	(0.00, 0.36)	2.9406
1000	0.3	0.291	(0.10, 0.43)	0.0072	0.291	(0.10, 0.43)	0.0075
1000	0.4	0.397	(0.26, 0.53)	0.0047	0.397	(0.26, 0.53)	0.0047
1000	0.5	0.499	(0.37, 0.65)	0.0049	0.499	(0.37, 0.65)	0.0049
1000	0.6	0.601	(0.46, 0.79)	0.0065	0.601	(0.46, 0.79)	0.0065
1000	0.7	0.703	(0.54, 0.93)	0.0095	0.702	(0.54, 0.92)	0.009
1000	0.8	0.807	(0.60, 1.09)	0.0143	0.798	(0.60, 1.00)	0.0103
1000	0.9	0.913	(0.68, 1.28)	0.0217	0.880	(0.68, 1.06)	0.0099
1000	0.95	0.967	(0.71, 1.36)	0.0267	0.914	(0.71, 1.08)	0.0099

Table B.4: Comparison of estimation methods - Observed step, estimation of ϕ.

		Method of Moments			**Quasi Fisher Scoring**		
N	ϕ	Mean	95 % CI	MSE	Mean	95 % CI	MSE
50	0.1	0.080	(0.00, 0.16)	0.2717	0.143	(0.01, 0.20)	0.0048
50	0.25	0.201	(0.00, 0.40)	0.17	0.218	(0.01, 0.37)	0.0071
50	0.5	0.402	(0.00, 0.79)	0.0822	0.391	(0.02, 0.76)	0.046
50	0.75	0.603	(0.00, 1.19)	0.0967	0.577	(0.01, 1.16)	0.1163
50	1	0.804	(0.00, 1.58)	0.2134	0.768	(0.01, 1.55)	0.2197
50	1.5	1.206	(0.00, 2.38)	0.7535	1.139	(0.00, 2.23)	0.5265
50	2	1.607	(0.00, 3.17)	1.7024	1.452	(0.00, 2.99)	1.0325
50	3	2.411	(0.00, 4.75)	4.8271	2.038	(0.00, 4.46)	2.7608
50	5	4.018	(0.00, 7.92)	15.983	3.871	(0.00, 7.47)	123.8

Continued on next page

Table B.4 – continued from previous page

		Method of Moments			Quasi Fisher Scoring		
N	ϕ	Mean	95 % CI	MSE	Mean	95 % CI	MSE
50	10	8.037	(0.00, 15.84)	72.497	6.800	(0.00, 14.87)	30.349
75	0.1	0.087	(0.00, 0.16)	0.265	0.148	(0.05, 0.20)	0.0045
75	0.25	0.216	(0.00, 0.40)	0.1558	0.230	(0.08, 0.40)	0.0054
75	0.5	0.433	(0.00, 0.80)	0.0624	0.420	(0.15, 0.80)	0.0369
75	0.75	0.649	(0.00, 1.20)	0.0799	0.625	(0.20, 1.20)	0.0899
75	1	0.865	(0.00, 1.59)	0.2082	0.836	(0.20, 1.60)	0.1638
75	1.5	1.298	(0.00, 2.39)	0.7973	1.244	(0.11, 2.35)	0.3761
75	2	1.731	(0.00, 3.19)	1.8296	1.619	(0.01, 3.14)	0.7461
75	3	2.596	(0.00, 4.78)	5.2243	2.292	(0.00, 4.70)	2.0012
75	5	4.327	(0.00, 7.97)	17.333	3.810	(0.00, 7.84)	5.6705
75	10	8.654	(0.00, 15.94)	78.638	7.646	(0.00, 15.68)	22.444
100	0.1	0.090	(0.02, 0.15)	0.2611	0.152	(0.06, 0.20)	0.0044
100	0.25	0.225	(0.04, 0.37)	0.1468	0.237	(0.11, 0.37)	0.0035
100	0.5	0.450	(0.08, 0.74)	0.0473	0.443	(0.20, 0.74)	0.0245
100	0.75	0.675	(0.13, 1.10)	0.0615	0.664	(0.20, 1.10)	0.0592
100	1	0.900	(0.17, 1.47)	0.1894	0.887	(0.20, 1.47)	0.1086
100	1.5	1.350	(0.25, 2.21)	0.7864	1.337	(0.25, 2.21)	0.2533
100	2	1.800	(0.33, 2.94)	1.8383	1.687	(0.05, 2.95)	0.5348
100	3	2.701	(0.50, 4.41)	5.3065	2.493	(0.01, 4.41)	1.2537
100	5	4.501	(0.83, 7.36)	17.701	4.135	(0.12, 7.35)	3.6403
100	10	9.002	(1.67, 14.71)	80.526	8.293	(0.03, 14.71)	14.353
250	0.1	0.095	(0.06, 0.14)	0.2549	0.158	(0.10, 0.20)	0.0042
250	0.25	0.239	(0.15, 0.34)	0.133	0.243	(0.20, 0.33)	0.0017
250	0.5	0.477	(0.30, 0.68)	0.0249	0.474	(0.30, 0.67)	0.0102
250	0.75	0.716	(0.45, 1.02)	0.0358	0.712	(0.45, 1.00)	0.0235
250	1	0.955	(0.59, 1.36)	0.1655	0.951	(0.59, 1.35)	0.0419
250	1.5	1.432	(0.89, 2.03)	0.7816	1.430	(0.89, 2.04)	0.0958
250	2	1.910	(1.19, 2.71)	1.8733	1.865	(0.92, 2.66)	0.2046
250	3	2.864	(1.78, 4.07)	5.4836	2.775	(1.35, 3.99)	0.4937
250	5	4.774	(2.97, 6.78)	18.411	4.648	(2.43, 6.66)	1.2336

Continued on next page

Table B.4 – continued from previous page

		Method of Moments			Quasi Fisher Scoring		
N	ϕ	Mean	95 % CI	MSE	Mean	95 % CI	MSE
250	10	9.548	(5.95, 13.56)	84.022	9.307	(4.94, 13.31)	4.8257
500	0.1	0.097	(0.07, 0.12)	0.2531	0.160	(0.11, 0.20)	0.0041
500	0.25	0.243	(0.18, 0.30)	0.1288	0.244	(0.20, 0.30)	0.0009
500	0.5	0.485	(0.36, 0.61)	0.0174	0.485	(0.36, 0.60)	0.0045
500	0.75	0.728	(0.54, 0.91)	0.026	0.727	(0.54, 0.91)	0.0101
500	1	0.971	(0.72, 1.21)	0.1544	0.970	(0.72, 1.21)	0.0179
500	1.5	1.456	(1.08, 1.82)	0.7709	1.456	(1.08, 1.82)	0.0403
500	2	1.941	(1.44, 2.43)	1.867	1.941	(1.44, 2.42)	0.0717
500	3	2.912	(2.16, 3.64)	5.4978	2.898	(2.07, 3.63)	0.1736
500	5	4.853	(3.60, 6.07)	18.514	4.831	(3.46, 6.05)	0.4741
500	10	9.706	(7.19, 12.15)	84.623	9.664	(6.91, 12.10)	1.889
1000	0.1	0.099	(0.08, 0.12)	0.2508	0.161	(0.12, 0.20)	0.0041
1000	0.25	0.248	(0.20, 0.29)	0.1243	0.248	(0.20, 0.29)	0.0005
1000	0.5	0.496	(0.41, 0.58)	0.0128	0.496	(0.41, 0.58)	0.0021
1000	0.75	0.744	(0.61, 0.88)	0.0254	0.744	(0.61, 0.88)	0.0046
1000	1	0.993	(0.81, 1.17)	0.1622	0.993	(0.81, 1.17)	0.0082
1000	1.5	1.489	(1.22, 1.75)	0.8083	1.489	(1.22, 1.75)	0.0185
1000	2	1.985	(1.63, 2.33)	1.951	1.985	(1.63, 2.33)	0.0329
1000	3	2.978	(2.44, 3.50)	5.7263	2.978	(2.44, 3.50)	0.074
1000	5	4.963	(4.07, 5.84)	19.236	4.963	(4.07, 5.84)	0.2054
1000	10	9.925	(8.15, 11.67)	87.776	9.925	(8.15, 11.67)	0.8217

Appendix C

Table of SISR filtering results

The following table contains filtering results from the four weighting methods described in chapter 5:

- Bootstrap weights
- Cumulant estimation of optimal importance distribution
- First order saddlepoint weights
- Second order saddlepoint weights

For each method, we generated the same time series of length $T = 10$, and then performed filtering over 1000 particles, 100 times. The mean and 95% quantile bootstrap confidence intervals of the X_t predictions are included in this table.

Table C.1: Comparison of SISR filtering weights

	Bootstrap		**Cumulant**		**1st Order SP**		**2nd Order SP**	
X	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
α =0.75, β = 0.8, ϕ = 1								
-0.86	-0.99	(-1.00, -0.97)	-0.59	(-1.37, -0.23)	-0.59	(-1.15, -0.32)	0.24	(0.13, 0.35)
-0.60	-0.64	(-0.67, -0.61)	-0.91	(-1.39, 0.00)	-0.65	(-0.89, -0.42)	-0.36	(-0.56, -0.16)
-0.40	-0.68	(-0.71, -0.64)	-0.17	(-1.66, 0.13)	-0.02	(-0.16, 0.12)	0.51	(0.42, 0.62)
0.25	0.41	(0.36, 0.47)	-0.16	(-0.89, 0.93)	0.26	(0.05, 0.47)	0.38	(0.15, 0.65)
0.58	0.65	(0.51, 0.79)	-0.27	(-1.00, 0.86)	0.03	(-0.23, 0.33)	0.16	(-0.23, 0.63)
1.21	1.34	(1.16, 1.48)	0.52	(-0.39, 1.33)	0.72	(0.52, 0.95)	0.78	(0.52, 1.10)
1.49	1.62	(1.60, 1.65)	1.99	(1.67, 2.32)	1.68	(1.34, 2.03)	1.84	(1.51, 2.16)
2.27	2.61	(2.55, 2.66)	2.47	(1.99, 3.02)	2.14	(1.65, 2.69)	2.25	(1.75, 2.78)
2.67	3.04	(2.70, 3.42)	1.84	(0.79, 2.56)	1.60	(1.20, 2.10)	1.70	(1.17, 2.32)
3.16	3.89	(3.69, 4.04)	2.66	(1.75, 3.59)	2.28	(1.73, 3.01)	2.39	(1.81, 3.21)
α =0.4, β = 0.8, ϕ = 1								
-1.13	-1.26	(-1.29, -1.23)	0.27	(0.16, 0.37)	0.30	(0.17, 0.44)	0.38	(0.27, 0.50)
-0.75	-0.95	(-0.98, -0.92)	0.39	(0.32, 0.48)	0.41	(0.28, 0.56)	0.46	(0.35, 0.55)
-0.08	-0.15	(-0.18, -0.11)	0.36	(0.26, 0.47)	0.32	(0.17, 0.45)	0.40	(0.29, 0.52)
0.66	0.65	(0.60, 0.69)	0.75	(0.62, 0.89)	0.62	(0.47, 0.78)	0.74	(0.63, 0.85)
0.78	0.63	(0.60, 0.67)	0.86	(0.72, 1.04)	0.76	(0.60, 0.94)	0.86	(0.69, 1.05)
1.21	1.57	(1.43, 1.70)	0.77	(0.66, 0.90)	0.60	(0.48, 0.75)	0.69	(0.53, 0.86)
1.55	1.68	(1.63, 1.75)	1.15	(0.89, 1.51)	1.04	(0.80, 1.42)	1.12	(0.77, 1.53)
1.60	1.89	(1.73, 2.04)	0.97	(0.85, 1.14)	0.79	(0.63, 0.99)	0.86	(0.70, 1.01)
3.01	3.40	(2.96, 3.91)	1.15	(0.92, 1.48)	1.07	(0.85, 1.34)	1.10	(0.82, 1.48)
4.07	3.94	(3.59, 4.32)	1.36	(1.02, 2.09)	1.26	(0.91, 1.75)	1.31	(0.94, 2.07)
α =0.75, β = 0.3, ϕ = 1								
-0.86	-0.55	(-0.61, -0.51)	-0.03	(-0.13, 0.05)	-0.02	(-0.14, 0.12)	0.08	(-0.05, 0.21)
-0.60	-0.37	(-0.45, -0.29)	0.00	(-0.09, 0.10)	-0.04	(-0.18, 0.11)	0.15	(0.00, 0.33)
-0.40	-0.12	(-0.16, -0.08)	0.28	(0.21, 0.35)	0.28	(0.19, 0.37)	0.34	(0.24, 0.45)
0.25	0.66	(0.53, 0.83)	0.57	(0.43, 0.71)	0.39	(0.19, 0.59)	0.86	(0.59, 1.17)
0.58	0.15	(0.04, 0.27)	0.29	(0.17, 0.42)	0.17	(-0.02, 0.34)	0.46	(0.24, 0.70)
1.21	1.34	(1.14, 1.53)	0.99	(0.81, 1.20)	0.73	(0.51, 0.99)	1.30	(0.99, 1.78)

Continued on next page

Table C.1 – continued from previous page

	Bootstrap		Cumulant		1st Order SP		2nd Order SP	
X	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
1.49	3.01	(2.55, 3.74)	2.42	(1.68, 3.54)	2.03	(1.41, 3.03)	2.18	(1.65, 2.87)
2.27	3.56	(2.84, 4.74)	2.64	(1.77, 4.06)	2.21	(1.44, 3.46)	2.28	(1.59, 3.39)
2.67	2.22	(1.91, 2.56)	1.74	(1.28, 2.20)	1.26	(0.87, 1.77)	1.68	(1.22, 2.34)
3.16	3.61	(2.82, 5.02)	2.45	(1.67, 3.78)	2.05	(1.34, 3.41)	2.15	(1.48, 3.90)
α =0.4, β = 0.3, ϕ = 1								
-1.13	-0.52	(-0.57, -0.46)	0.40	(0.31, 0.48)	0.43	(0.33, 0.56)	0.46	(0.36, 0.57)
-0.75	-0.19	(-0.26, -0.15)	0.47	(0.39, 0.54)	0.49	(0.37, 0.60)	0.51	(0.38, 0.63)
-0.08	0.16	(0.08, 0.24)	0.49	(0.40, 0.58)	0.49	(0.38, 0.64)	0.55	(0.42, 0.66)
0.66	1.00	(0.91, 1.12)	0.64	(0.56, 0.75)	0.64	(0.54, 0.77)	0.69	(0.56, 0.85)
0.78	1.24	(1.02, 1.51)	0.71	(0.59, 0.84)	0.70	(0.54, 0.88)	0.79	(0.63, 1.07)
1.21	0.84	(0.72, 0.95)	0.60	(0.50, 0.71)	0.60	(0.50, 0.72)	0.66	(0.50, 0.80)
1.55	1.72	(1.34, 2.32)	0.75	(0.62, 0.90)	0.74	(0.56, 0.90)	0.82	(0.61, 1.11)
1.60	1.29	(1.14, 1.44)	0.72	(0.64, 0.82)	0.68	(0.53, 0.81)	0.73	(0.60, 0.90)
3.01	1.72	(1.53, 1.94)	0.80	(0.69, 0.93)	0.74	(0.60, 0.91)	0.77	(0.64, 0.92)
4.07	2.23	(1.82, 2.86)	0.80	(0.70, 0.95)	0.79	(0.61, 0.95)	0.83	(0.66, 1.05)
α =0.75, β = 0.8, ϕ = 2								
-0.86	-0.99	(-1.01, -0.97)	-0.39	(-0.46, -0.30)	-0.64	(-0.87, -0.45)	-0.37	(-1.22, 0.56)
-0.60	-0.64	(-0.68, -0.61)	-0.22	(-0.29, -0.16)	-0.66	(-0.85, -0.49)	-0.40	(-0.77, -0.11)
-0.40	-0.68	(-0.71, -0.65)	-0.02	(-0.08, 0.06)	-0.11	(-0.24, 0.02)	-0.39	(-0.73, 0.08)
0.25	0.42	(0.36, 0.47)	0.73	(0.59, 0.89)	0.19	(-0.02, 0.43)	0.34	(0.18, 0.55)
0.58	0.66	(0.54, 0.78)	0.72	(0.55, 0.91)	-0.03	(-0.27, 0.24)	0.23	(-0.04, 0.52)
1.21	1.34	(1.19, 1.52)	1.27	(1.09, 1.41)	0.70	(0.48, 0.94)	0.79	(0.44, 1.25)
1.49	1.62	(1.60, 1.65)	1.99	(1.76, 2.21)	1.71	(1.42, 2.09)	1.85	(1.58, 2.15)
2.27	2.61	(2.54, 2.65)	2.63	(2.25, 3.06)	2.29	(1.82, 2.91)	2.45	(1.96, 3.06)
2.67	3.09	(2.67, 3.50)	2.10	(1.68, 2.53)	1.71	(1.13, 2.36)	1.82	(1.16, 2.69)
3.16	3.90	(3.58, 4.07)	3.03	(2.47, 3.62)	2.50	(1.79, 3.56)	2.67	(1.80, 3.67)
α =0.9, β = 0.9, ϕ = 1								
-0.40	-0.43	(-0.44, -0.41)	-0.43	(-0.87, 0.44)	-0.53	(-0.65, -0.42)	0.64	(0.55, 0.76)
-0.39	-0.48	(-0.49, -0.47)	-0.36	(-0.70, 0.46)	-0.49	(-0.63, -0.36)	0.62	(0.52, 0.72)
-0.09	-0.34	(-0.36, -0.32)	-0.61	(-1.01, 0.33)	-0.09	(-0.17, 0.00)	0.81	(0.72, 0.92)

Continued on next page

Table C.1 – continued from previous page

	Bootstrap		**Cumulant**		**1st Order SP**		**2nd Order SP**	
X	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
0.03	0.04	(0.00, 0.06)	-0.36	(-0.72, 0.80)	0.01	(-0.27, 0.41)	0.23	(0.08, 0.40)
0.22	0.27	(0.15, 0.38)	-0.01	(-0.58, 1.09)	-0.14	(-0.39, 0.16)	0.76	(0.68, 0.86)
0.58	0.63	(0.52, 0.73)	0.08	(-0.48, 0.96)	0.33	(0.03, 0.80)	0.27	(-0.03, 0.67)
1.20	1.31	(1.29, 1.33)	1.80	(1.40, 2.36)	1.46	(1.17, 1.75)	1.58	(1.26, 1.96)
1.54	1.79	(1.75, 1.82)	2.00	(1.53, 2.72)	1.67	(1.30, 2.11)	1.72	(1.27, 2.18)
1.55	1.89	(1.67, 2.10)	1.17	(0.25, 2.01)	1.07	(0.73, 1.42)	1.03	(0.43, 1.81)
1.81	2.41	(2.30, 2.49)	1.94	(1.02, 2.77)	1.64	(1.23, 2.14)	1.63	(1.13, 2.31)