APPLICATION OF ASYMMETRIC LAPLACE LAWS IN FINANCIAL RISK MEASURES AND TIME SERIES ANALYSIS

By

YUN ZHU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2007

To Xiaodong, Katie, and my parents

ACKNOWLEDGMENTS

I am grateful to my Ph.D. advisor, Dr. Alex Trindade. This work could not have been written without him who not only served as my supervisor but also encouraged and challenged me throughout my academic program. I am grateful for his immense help at every stage of my research, from initiating the topics, solving problems, to revising numerous drafts. His valuable insights and ideas directly and significantly contributed to my dissertation.

I would like to thank my committee numbers, Dr. Ramon Littell, Dr. Ronald Randles, Dr. Clyde Schoolfield and Dr. Farid AitSahlia, for taking the time to work with me.

Thanks go out to my husband, always offering support and love. Thanks go out to my parents, for taking care of me and my baby. I could not have finished my dissertation without their support. Thank you my dear Katie for providing me happiness and inspiration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

APPLICATION OF ASYMMETRIC LAPLACE LAWS IN FINANCIAL RISK MEASURES AND TIME SERIES ANALYSIS

By

Yun Zhu

May 2007

Chair: Alex Trindade
Major: Statistics

Asymmetric Laplace (AL) laws are applied in financial risk measurement and time series analysis. Traditional methods on financial risk measures and time series analysis are based on the assumption of normality. Recent studies on financial data suggest that the normality assumption is usually violated.

Explicit expressions are derived for maximum likelihood estimators (MLEs) and nonparametric estimators (NPEs) of financial risk measures, Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), under random sampling from the Asymmetric Laplace distribution. Asymptotic distributions are established under very general conditions. Finite sample distributions are investigated by means of saddlepoint approximations. An application of the methodology in modeling currency exchange rates suggests that the AL distribution is successful in capturing the peakedness, leptokurticity and skewness, inherent in such data.

Time series autoregressive moving average (ARMA) models driven by Asymmetric Laplace noise are considered for modeling dependent data. Assuming AL noise, the model marginal distribution is derived analytically. Conditional maximum likelihood estimation is applied to fit ARMA models driven by AL noise and AL general autoregressive conditional heteroscedasticity (GARCH) noise. Daily returns of real estate mutual fund data are fitted by four methods. Models under AL noise have substantially lower

Bias-corrected Akaike Information Criterion (AICc), indicating much better fit for the real financial data.

CHAPTER 1
INTRODUCTION

1.1 Financial Risk Measurement via VaR and CVaR

Value-at-Risk (VaR) has become one of the most important risk measures in modern financial risk management. In general, financial risk comes from three parts, market risk, credit risk, and operational risk. Market risk is defined as the uncertainty due to the changes in financial asset prices such as interest rates, foreign exchange rates, equity prices, and commodity prices; Credit risk comes from the losses associated with the default (or credit downgrade) of an obligor; Operational risk is related with operational failures.

The idea behind Value-at-Risk originated from measuring market risk. VaR is the loss that can occur over a given period, at a given confidence level, due to exposure to market risk. Recently, the idea of VaR has been introduced into measuring credit risk.

The importance of measuring the risk of financial assets has long been realized. Markowitz (1959) first introduced the definition and measurement of risk in *portfolio selection*, where financial risk was measured by the variance and covariance of underlying asset prices.

Since 1990's, Value-at-Risk is widely used by commercial banks, asset management companies, and regulators. For example, the Basel Accord I employs VaR as the measurement for commercial banks' market risk exposure. The Basel Accord is the international capital adequacy standards set up by Basel Committee on Banking Supervision. The 1996 amendment of the Basel Accord extends the capital requirements to include risk-based capital for the market risk in the trading book. Under the supervision of the Basel Committee, banks need to set up their own VaR models to calculate their minimum regulatory capital for market risk. In 1997, the Securities and Exchange Commission in United States began requiring financial institutions to report Value-at-Risk as an important measure of the market risk exposure.

The concept of Value-at-Risk is also widely used by financial institutions and asset managers. For example, in asset management companies, Value-at-Risk is used to set position limits for traders; In commercial banks, Value-at-Risk is used to calculate market risk exposure of their assets and is used in capital allocation.

Loosely speaking, if Y represents "losses" and $0 < \alpha < 1$, given time horizon t, the VaR at confidence level α, $\mathrm{VaR}_\alpha(Y)$, is the lower bound on the worst $(1-\alpha)100\%$ losses during the time horizon.

let Y be a continuous real-valued random variable defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$, with cumulative distribution function (cdf) $F(\cdot)$ and probability density function (pdf) $f(\cdot)$. Let μ and σ^2 denote the mean and variance of Y, respectively, and both are assumed to be finite. With the understanding that Y represents loss, the VaR of Y at probability level α is defined to be the αth quantile of Y.

Definition 1.1 (VaR).

$$\mathrm{VaR}_\alpha(Y) \equiv \xi_\alpha(Y) = F^{-1}(\alpha). \tag{1–1}$$

The confidence level α and time horizon t vary among different banks, companies and regulators. A commonly used confidence level is 99%. For example, Basel Accord has set α to be 99% and t to be 10 days in order to measure banks' market risk exposure. Commercial banks typically use overnight Value-at-Risk to measure financial risk exposure for the purpose of internal supervision and risk control, and disclose two-week Value-at-Risk to investors and regulators.

Practically, there is a 'square root of time' rule. That is, if the daily Value-at-Risk is ϕ, then the Value-at-Risk in the time horizon of m days will be $\sqrt{m}\phi$. This result is based on the assumption that daily returns are independent. Under this assumption, Value-at-Risk under different time horizons can easily be transformed by multiplying or dividing by square root of days.

One major criticism of VaR is that it is not a coherent measure, i.e., VaR is not sub-additive. This means that in general the VaR of a portfolio can exceed the sum of the stand-alone VaRs of its components. In addition, VaR provides no information about the loss beyond itself.

As an alternative to VaR, Conditional Value-at-Risk (CVaR) describes the average of the worst $(1-\alpha)\%$ losses. The term of CVaR is drawn from Rochafellar and Uryasev (2000), but synonyms for it also in common usage include: "expected shortfall" (Acerbi and Tasche, 2002), and "tail-conditional expectation" (Artzner *et al.*, 1999).

The CVaR of Y at probability level α, is the mean of the random variable that results by truncating Y at $\mathrm{VaR}_\alpha(Y)$ and discarding its lower tail.

Definition 1.2 (CVaR).

$$\mathrm{CVaR}_\alpha(Y) \equiv \phi_\alpha(Y) = \mathrm{E}(Y|Y \geq \xi_\alpha) = \frac{1}{1-\alpha}\int_{\xi_\alpha}^{\infty} yf(y)dy. \tag{1–2}$$

An equivalent definition of CVaR in terms of the quantile function of Y is

$$\phi_\alpha(Y) = \frac{1}{1-\alpha}\int_{\alpha}^{1} F^{-1}(u)du. \tag{1–3}$$

Methodologies for Estimating VaR. There are various methodologies for estimating Value-at-Risk. Most of them fall into three categories: parametric methods, nonparametric methods and semi-parametric methods.

Parametric methods assume a distribution for the financial data. Under parametric methods, the VaR at confidence level α is just the αth quantile of the distribution. Parametric methods depend on the assumption of a distribution. The normal distribution is the most commonly used family in financial risk measurement. But there are obvious violations of normality in financial data. Financial data are typically skewed and heavy tailed. Another problem with parametric methods is that they are inappropriate when there are discontinuous payoffs in the portfolio.

Nonparametric methods, also called historical methods, use empirical quantiles of the data to estimate VaR. Nonparametric estimation makes no assumptions about the distribution. As a result, it is flexible enough to deal with data with heavy tails. Nonparametric estimation is based on information from the past. If there is a permanent change in major market factors, for example, changes in regulations, nonparametric estimation will underestimate or overestimate VaR.

Semiparametric methods assume the distribution of one tail, which making no assumption about the underlying distribution away from the tail. A typical used tail distribution is pareto tail.

Fat tails and Skewness. Today the presence of heavy tails of financial data is a well-accepted fact. The central limit theorem (CLT) is not valid here because a key assumption behind the central limit theorem is that data that go into a sum are statistically independent. Independence is not a proper description of financial data. And financial data are not symmetric about the mean in most cases. In financial data, one tail represent profit and the other represents loss. Consequently, we can not treat them equally.

Assuming normality on heavy tailed data will cause underestimation of VaR at high confidence level, which will cause major problems in risk control.

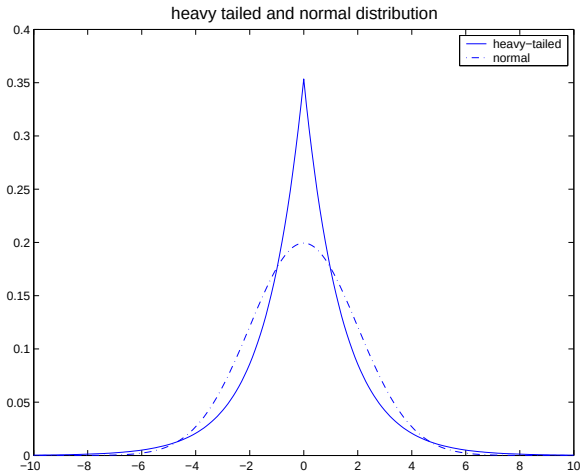

Figure 1-1. Heavy tailed distribution vs. normality.

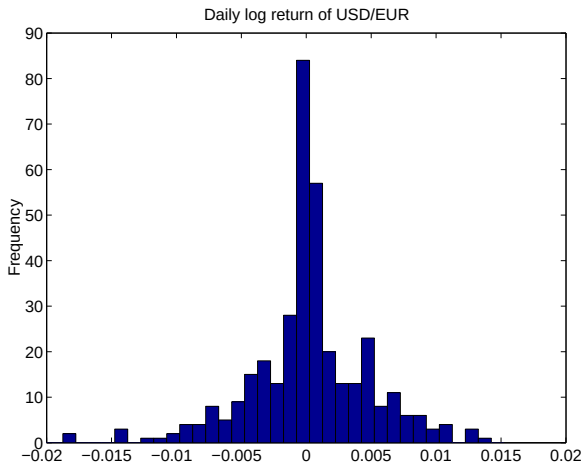

Figure 1-2. Histogram of the daily log returns of USD/EUR exchange rate from Feb 1,2005 to Jan 31, 2006.

Fig. 1-2 is the daily log return of USD/EUR exchange rates. This histogram indicates the typical properties of financial data, heavy tails and skewness. Commonly used measures of heavy tails and skewness are kurtosis and skewness, respectively. For more detailed information about kurtosis and skewness, please refer to Section 1.2. Properly adjusted, the kurtosis of a normal distribution is 0. And when kurtosis is greater than 0, the data are considered heavy tailed. The skewness of any symmetric distribution is 0. The kurtosis and skewness of the data are 2.0119 and -0.3777, respectively. Therefore, we consider them to be heavy tailed and skewed.

Explanation and solution.

Theoretically, there are many explanation about the reasons for fat tails. Two of them are widely accepted: There are some significant discontinuous changes in the financial data due to unexpected changes in market factors, for example, market crash. This is also called 'jumps'. The other explanation is called 'Volatility Clustering', the volatility at time t is highly correlated with volatilities at past time s, $s < t$.

The Mixture normal model and *Jump-diffusion model* are developed to explain data with jumps. The mixture normal model assumes the financial data come from a random mixture of two different normal distributions. One of these normal distributions comes

from an ordinary market situation, and the other comes from a market with higher volatility.

Statistically, let ϕ_1 and ϕ_2 be the density function of two normal distributions with different mean and variance. X is a random variable from a mixture of normals if

$$f_X(x) = P\phi_1(x) + (1 - P)\phi_2(x), \tag{1–4}$$

where P is a Bernoulli random variable with success probability p. Under the mixture normal model, we can choose p and variances of two normal distributions to achieve a given kurtosis, and therefore get a fat tailed X.

The jump-diffusion model was introduced in the context of differential equations. For the jump-diffusion model, financial data is 'jumped' by adding an independent normal random variable.

Auto-Regressive Volatility model (ARV), *Exponentially Weighted Moving Average (EWMA) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models* are developed upon observing the clustering of the volatility. ARV and EWMA are intuitive methods. GARCH models are more flexible to describe time series processes with correlated volatility. Consequently, GARCH models are commonly used in financial industries.

The ARV model assumes

$$\log\sigma_t^2 = \alpha + \beta\log\sigma_{t-1}^2 + \gamma Z_t, \tag{1–5}$$

where α, β and γ are constants. $\{Z_t\}$ is white noise. The volatility at time t is a function of volatility at $t - 1$. As a consequence, the volatilities are correlated. A value of β near zero implies low correlation, while a value of beta near 1 implies high correlation.

The EWMA model was introduced by the RiskMetrics Group. The variance at time t is estimated as

$$\sigma_t^2 = (1-\phi)X_{t-1}^2 + \phi\sigma_{t-1}^2, \tag{1–6}$$

with $\phi > 0$, where ϕ is a 'smoothing constant' and X_{t-1} is the return at time $t-1$. The volatility at time t is considered as a function of the return and volatility at time $t-1$. Since $\phi > 0$, the volatility at time t will be positively correlated with the volatility at $t-1$.

GARCH is a more flexible and general model describing the clustering of volatility. Bollerslev (1986) introduced the GARCH process. The volatility is estimated as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_j Z_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2, \tag{1–7}$$

where $\{Z_t\}$ is white noise. $\alpha_0 \geq 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$, $j = 1, 2, \ldots$. Under a GARCH model, the volatility at time t depends on the volatility of the past. For more detailed information about GARCH model, please refer to Chapter 4.

Some potentially more flexible models include: Exponential GARCH model, EGARCH , Cross-Market GARCH, etc.

Distributions for fat tails and skewness. Various families of distribution have been introduced to describe the two characteristics of financial data: fat tails and skewness. A very intuitive one is the t distribution. But t distributions can not capture the skewness of financial data. Fernandez and Steel (1998), Kuester and Mittnik (2006) and Patton (2004) introduced a generalized t-distribution and a skewed t-distribution. In general, those densities are not log-concave.

Exponential Power Distribution (EPD), also called Generalized Power Distribution (GPD), has also been used in financial risk measure. The density of EPD is

$$f(x) = \frac{1}{2a\Gamma(1+1/b)} \exp(1 - |x/a|^b). \tag{1–8}$$

For $b = 1$ this reduces to the Laplace Distribution. For $b = 2$, it has the same form as a normal distribution with $a = \sqrt{2}\alpha$. EPD is a flexible distribution but does not allow for any asymmetry in the data. Komunjer (2006) introduces the Asymmetric Power Distribution (APD) for estimating Expected Shortfall. Some of the properties of the APD distribution have been studied by Fernandez, Osiewalski and Steel (1995), Ayebo and Kozubowski (2003). Komunjer (2006) gives no result about the distribution of the estimators of risk measures nor confidence intervals.

The Symmetric Laplace Distribution (SLD), also called Double Exponential Distribution, has been used for modeling data with heavy tails. See Balakrishnan and Basu (1995), Bain and Engelhardt (1973), Kotz , Kozubowski and Podgoriski (2001). SLD distribution has the same problem as other symmetric distributions which do not allow for any asymmetry. Therefore, we consider the asymmetric form of the Laplace distribution, the Asymmetric Laplace distribution.

1.2 Asymmetric Laplace Distribution

One of the intents of this work is to apply Asymmetric Laplace law to financial data. The *Asymmetric Laplace* (AL) distribution, introduced by Kotz *el al.* (2001), is a generalization of the Symmetric Laplace distribution (Double Exponential distribution).The AL distribution demonstrates flexibility in fitting data with heavy tails and skewness, which make it a promising candidate for financial data modeling.

Definition 1.3 (The Asymmetric Laplace distribution). Random variable Y is said to be distributed as Asymmetric Laplace distribution with location parameter θ, scale parameter $\tau > 0$, and skewness parameter $\kappa > 0$, $Y \sim \mathcal{AL}(\theta, \kappa, \tau)$, if its pdf is of the form

$$f(y) = \frac{\kappa\sqrt{2}}{\tau(1+\kappa^2)} \begin{cases} \exp\left(-\frac{\kappa\sqrt{2}}{\tau}|y-\theta|\right), & \text{if } y \geq \theta, \\ \exp\left(-\frac{\sqrt{2}}{\kappa\tau}|y-\theta|\right), & \text{if } y < \theta, \end{cases} \tag{1–9}$$

or, the distribution function of Y is the form

$$F(y) = \begin{cases} 1 - \frac{1}{1+\kappa^2}\exp\left(-\frac{\sqrt{2}\kappa}{\tau}|y-\theta|\right), & \text{if } y \geq \theta, \\ \frac{\kappa^2}{1+\kappa^2}\exp\left(-\frac{\sqrt{2}}{\tau\kappa}|y-\theta|\right), & \text{if } y < \theta. \end{cases} \tag{1–10}$$

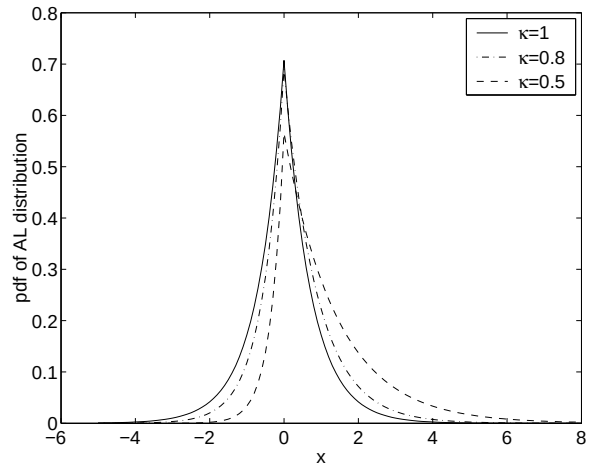

Figure 1-3. Asymmetric Laplace densities with $\theta = 0$, $\tau = 1$, $\kappa = 0.5, 0.8, 1$

Kotz *et al.* (2001), ch.3, generalize the essential properties of Asymmetric Laplace distribution.

Proposition 1.1 (Moment generating function of AL distribution). If $Y \sim \mathcal{AL}(\theta, \kappa, \tau)$, then the moment generating function of Y is

$$M_{\theta,\kappa,\tau}(t) = \frac{e^{\theta t}}{1 - \frac{1}{2}\tau^2 t^2 - \frac{\tau}{\sqrt{2}}\left(\frac{1}{\kappa} - \kappa\right)t}, \quad -\frac{\sqrt{2}}{\tau\kappa} < t < \frac{\sqrt{2}\kappa}{\tau}. \tag{1–11}$$

Proposition 1.2 (Cumulants of AL distribution). The cumulants of an $\mathcal{AL}(\theta, \kappa, \tau)$ can be stated as

$$\kappa_n(Y) = \begin{cases} \theta + \frac{\tau}{\sqrt{2}}(\kappa^{-1} - \kappa), & \text{if } n = 1; \\ (n-1)!\left(\frac{\tau}{\sqrt{2}}\right)^n(\kappa^{-n} - \kappa^n), & \text{if } n > 1 \text{ is odd}; \\ (n-1)!\left(\frac{\tau}{\sqrt{2}}\right)^n(\kappa^{-n} + \kappa^n) & \text{if n is even.} \end{cases} \tag{1–12}$$

The mean and variance of Y, which coincide with the first and second cumulants, respectively, are

$$\mu = \theta + \frac{\tau}{\sqrt{2}}\left(\frac{1}{\kappa} - \kappa\right), \quad \text{and} \quad \sigma^2 = \frac{\tau^2}{2}(\frac{1}{\kappa^2} + \kappa^2) = (\mu - \theta)^2 + \tau^2. \tag{1–13}$$

Proposition 1.3 (Coefficients of skewness and kurtosis). For a distribution of an random variance Y with a finite third moment and standard deviation greater than zero, the *coefficient of skewness* is a measure of symmetry defined by

$$\gamma_1 = \frac{E(Y - EY)^3}{(E(Y - EY)^2)^{3/2}}. \tag{1–14}$$

For an $\mathcal{AL}(\theta, \kappa, \tau)$ distribution, the coefficient of skewness is

$$\gamma_1 = 2\frac{1/\kappa^3 - \kappa^3}{(1/\kappa^2 + \kappa^2)^{3/2}}. \tag{1–15}$$

The coefficient of skewness is nonzero unless $\kappa = 1$. The absolute value of γ_1 is bounded by two, and as κ increases within the interval $(0, \infty)$, the corresponding value of γ_1 decreases monotonically from 2 to -2.

For an random variable with a finite fourth moment, the *kurtosis* is defined as

$$\gamma_2 = \frac{E(Y - EY)^4}{(\mathrm{Var(Y)})^2} - 3. \tag{1–16}$$

It is a measure of peakedness and of heaviness of the tails (properly adjusted, so that $\gamma_2 = 0$ for a normal distribution) and is independent of the scale. If $\gamma_2 > 0$, the distribution is said to be *Leptokurtic*, with heavy tails and high peakedness; it is *Platykurtic* otherwise.

For an $\mathcal{AL}(\theta, \kappa, \tau)$ distribution, the kurtosis of AL distribution is

$$\gamma_2 = 6 - \frac{12}{(1/\kappa^2 + \kappa^2)^2}. \tag{1–17}$$

The AL distribution is leptokurtic and γ_2 varies from 3 (the least value for the Symmetric Laplace distribution with $\kappa = 1$) to 6 (the greatest value attained for the limiting exponential distribution when $\kappa \to 0$).

Proposition 1.4 (Quantiles of AL distribution). For an $\mathcal{AL}(\theta, \kappa, \tau)$ distribution, the qth quantile, ξ_q, is

$$\xi_q = \begin{cases} \theta + \frac{\tau\kappa}{\sqrt{2}} \log\left\{\frac{1+\kappa^2}{\kappa^2} q\right\}, & \text{for } q \in \left(0, \frac{\kappa^2}{1+\kappa^2}\right], \\ \theta - \frac{\tau}{\kappa\sqrt{2}} \log\{(1+\kappa^2)(1-q)\}, & \text{for } q \in \left(\frac{\kappa^2}{1+\kappa^2}, 1\right). \end{cases} \tag{1–18}$$

Therefore, the VaR and CVaR are then easily derived.

Proposition 1.5 (VaR and CVaR for AL distribution). Let $Y \sim \mathcal{AL}(\theta, \kappa, \tau)$, where θ, κ, τ are unknown parameters. Then for $\frac{\kappa^2}{(1+\kappa^2)} < \alpha < 1$, VaR and CVaR can be obtained as

$$\xi_\alpha(Y) = \theta - \frac{\tau \log[(1+\kappa^2)(1-\alpha)]}{\kappa\sqrt{2}}$$

and

$$\phi_\alpha(Y) = \xi_\alpha(Y) + \frac{\tau}{\kappa\sqrt{2}}. \tag{1–19}$$

CHAPTER 2
APPROXIMATION TO THE DISTRIBUTION OF MLES OF VAR AND CVAR UNDER AL LAW

In this chapter, we give the explicit expressions of the maximum likelihood estimators (MLEs) of VaR and CVaR under Asymmetric Laplace distribution. Large sample asymptotic distributions of MLEs are establishes via Delta methods. Finite sample approximations of MLEs are developed by general saddlepoint approximation. Finally, the accuracy of the approximations is checked via simulations.

We will assume that the location parameter, θ, is known. Both VaR and CVaR are translation invariant and positively homogenous (Pflug, 2000), i.e. $\xi_\alpha(Y) = \theta + \tau\xi_\alpha(X)$ and $\phi_\alpha(Y) = \theta + \tau\phi_\alpha(X)$. Without loss of generality, in this chapter we focus on $X \sim \mathcal{AL}(0, \kappa, \tau)$, provided $\theta = 0$ is known. Consequently, VaR and CVaR can be presented as

$$\begin{aligned} \xi_\alpha(X) &= -\frac{\tau \log[(1+\kappa^2)(1-\alpha)]}{\kappa\sqrt{2}} \\ \phi_\alpha(X) &= \xi_\alpha(X) + \frac{\tau}{\kappa\sqrt{2}}. \end{aligned} \tag{2–1}$$

2.1 Maximum Likelihood Estimations of VaR and CVaR under AL Distribution

Kotz *et al.* (2001), ch.3, give explicit expressions of MLEs of κ and τ provided the value of θ is known. Consider a random sample $X_1, \ldots, X_n$ from $X \sim \mathcal{AL}(0, \kappa, \tau)$. Let $x_{(1)}$ and $x_{(n)}$ be the first and nth order statistics of this random sample. If $x_{(1)} < 0 < x_{(n)}$, the MLEs of κ and τ exist and are available in closed form as

$$\begin{aligned} \tilde{\kappa} &= \left(\frac{\frac{1}{n}\sum_{i=1}^n x_i^-}{\frac{1}{n}\sum_{i=1}^n x_i^+} \right)^{\frac{1}{4}} \\ \tilde{\tau} &= \sqrt{2}\left(\frac{1}{n}\sum_{i=1}^n x_i^+\right)^{\frac{1}{4}} \left(\frac{1}{n}\sum_{i=1}^n x_i^-\right)^{\frac{1}{4}} \left\{ \sqrt{\frac{1}{n}\sum_{i=1}^n x_i^+} + \sqrt{\frac{1}{n}\sum_{i=1}^n x_i^-} \right\}, \end{aligned} \tag{2–2}$$

where $x_i^+ = x_i I_{[x_i \geq 0]}$ and $x_i^- = -x_i I_{[x_i < 0]}$.

Let $\frac{1}{n}\sum_{i=1}^{n} x_i^+ = V_1$ and $\frac{1}{n}\sum_{i=1}^{n} x_i^- = V_2$, the MLEs of κ and τ can be expressed equivalently as

$$\tilde{\kappa} = \left(\frac{V_2}{V_1}\right)^{\frac{1}{4}}, \quad \text{and} \quad \tilde{\tau} = \sqrt{2}\,(V_1V_2)^{\frac{1}{4}}\left(\sqrt{V_1}+\sqrt{V_2}\right). \tag{2–3}$$

Then the MLEs of VaR and CVaR are obtained by equivariance,

$$\tilde{\xi}_\alpha(X) = -\frac{\tilde{\tau}\log[(1+\tilde{\kappa}^2)(1-\alpha)]}{\tilde{\kappa}\sqrt{2}},$$

and

$$\tilde{\phi}_\alpha(X) = \tilde{\xi}_\alpha(X) + \frac{\tilde{\tau}}{\tilde{\kappa}\sqrt{2}}, \tag{2–4}$$

or, equivalently,

$$\begin{aligned}\tilde{\xi}_\alpha(X) &= -\left(V_1+\sqrt{V_1V_2}\right)\left[\log(\sqrt{V_1}+\sqrt{V_2}) - \frac{1}{2}\log V_1 + \log(1-\alpha)\right],\\ \tilde{\phi}_\alpha(X) &= \tilde{\xi}_\alpha(X) + \left(V_1+\sqrt{V_1V_2}\right).\end{aligned} \tag{2–5}$$

2.2 Asymptotic distribution of $\tilde{\xi}_\alpha(X)$ and $\tilde{\phi}_\alpha(X)$

Kotz *et al.* (2001), ch.3, prove the consistency and asymptotic normality of $\tilde{\kappa}$ and $\tilde{\tau}$.

Proposition 2.1 (Consistency and asymptotic normality of $\tilde{\kappa}$ and $\tilde{\tau}$). Let $X_1, \ldots, X_n$ be i.i.d. random sample from distribution $\mathcal{AL}(\theta, \kappa, \tau)$, where the value of θ is known. Then the MLEs of $[\kappa, \tau]$, $[\tilde{\kappa}, \tilde{\tau}]$, given by Eq. 2–1 are

(i) Strongly consistent;

(ii) Asymptotically bivariate normal with the asymptotic covariance matrix

$$\sum(\tilde{\kappa}, \tilde{\tau}) = \begin{bmatrix} \frac{(1+\kappa^2)^2}{8} & \frac{1-\kappa^4}{8\kappa} \\ \frac{1-\kappa^4}{8\kappa} & \frac{\tau^2(1+6\kappa^2+\kappa^4)}{8\kappa^2} \end{bmatrix}, \tag{2–6}$$

(iii) Asymptotically efficient, namely, this asymptotic covariance matrix coincides with the inverse of the Fisher information matrix.

In our case, the fisher information matrix is

$$I(\kappa,\tau) = \begin{bmatrix} \kappa^{-2} + 4(1+\kappa^2)^{-2} & -\frac{1-\kappa^2}{\tau\kappa(1+\kappa^2)} \\ -\frac{1-\kappa^2}{\tau\kappa(1+\kappa^2)} & \tau^{-2} \end{bmatrix}. \tag{2–7}$$

Proposition 2.2 (Multivariate delta method).

Let $g : \mathbf{D}_g \subset \mathbf{R}^k \mapsto \mathbf{R}^m$ be a map defined on a subset of $\mathbf{R}^k$ and differentiable at θ. Let T_n be random vectors taking their values in the domain of g. If $\sqrt{n}(T_n - \theta) \xrightarrow{d} N_k(0,\Sigma)$, then

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} N_m(0, g'_\theta \Sigma (g'_\theta)^T). \tag{2–8}$$

Let $g(T_n) = \begin{bmatrix} \xi_\alpha \\ \phi_\alpha \end{bmatrix}$,

then

$$g'_\theta = \begin{bmatrix} \frac{\partial \xi_\alpha}{\partial \kappa} & \frac{\partial \xi_\alpha}{\partial \tau} \\ \frac{\partial \phi_\alpha}{\partial \kappa} & \frac{\partial \phi_\alpha}{\partial \tau} \end{bmatrix} = \begin{bmatrix} \tau\frac{(1+\kappa^2)\omega_{\alpha,\kappa} - 2\kappa^2}{\sqrt{2}\kappa^2(1+\kappa^2)} & -\frac{\omega_{\alpha,\kappa}}{\kappa\sqrt{2}} \\ \tau\frac{(1+\kappa^2)(\omega_{\alpha,\kappa}-1) - 2\kappa^2}{\sqrt{2}\kappa^2(1+\kappa^2)} & \frac{(1-\omega_{\alpha,\kappa})}{\kappa\sqrt{2}} \end{bmatrix}, \tag{2–9}$$

where

$$\omega_{\alpha,\kappa} \equiv \log[(1+\kappa^2)(1-\alpha)]. \tag{2–10}$$

Applying the multivariate delta method and the asymptotic normalities of $[\tilde{\kappa}, \tilde{\tau}]$ in Proposition 2.1, we get the asymptotic distribution of MLEs of VaR and CVaR

Theorem 2.1 (Asymptotic joint distribution of MLEs of VaR and CVaR) . Let $Y \sim AL(0,\kappa,\tau)$. Define $\omega_{\alpha,\kappa}$ the same as in Eq. 2–10 and the MLEs of VaR and CVaR, $[\tilde{\xi}_\alpha, \tilde{\phi}_\alpha]$, as in Eq. 2–1, respectively, we have

$$\sqrt{n}\left(\begin{bmatrix} \tilde{\xi}_\alpha \\ \tilde{\phi}_\alpha \end{bmatrix} - \begin{bmatrix} \xi_\alpha \\ \phi_\alpha \end{bmatrix}\right) \xrightarrow{d} N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tilde{\sigma}^2(\xi_\alpha) & \tilde{\sigma}(\xi_\alpha,\phi_\alpha) \\ \tilde{\sigma}(\xi_\alpha,\phi_\alpha) & \tilde{\sigma}^2(\phi_\alpha) \end{bmatrix}\right),$$

where,

$$\begin{aligned}
\tilde{\sigma}^2(\xi_\alpha) &= \frac{\tau^2}{4\kappa^2}\left[\kappa^2(\omega_{\alpha,\kappa}-1)^2+2\omega_{\alpha,\kappa}^2\right] \\
\tilde{\sigma}^2(\phi_\alpha) &= \frac{\tau^2}{4\kappa^2}\left[(\kappa^2+2)\omega_{\alpha,\kappa}^2-4(\kappa^2+1)\omega_{\alpha,\kappa}+4\kappa^2+2\right] \\
\tilde{\sigma}(\xi_\alpha,\phi_\alpha) &= \frac{\tau^2(\omega_{\alpha,\kappa}-1)}{4\kappa^2}\left[(\kappa^2+2)\omega_{\alpha,\kappa}-2\kappa^2\right]. \qquad (2\text{–}11)
\end{aligned}$$

2.3 Approximation of Finite Sample Distribution

2.3.1 General Saddlepoint Approximation to L Statistics

Easton and Ronchetti (1986) derive the saddlepoint approximation for the density of a general statistic U_n.

Suppose that $x_1,\ldots,x_n$ are n iid real valued random variables with density f. $U_n(x_1,\ldots,x_n)$ is a real valued statistic with pdf f_n and cdf F_n. Let $M_n(t)=\int e^{tx}f_n(x)dx$ denotes the moment generating function and $K_n(t)=\log M_n(t)$ the cumulant generating function of U_n. Further suppose that the moment generating function of $M_n(t)$ exists for real t in some nonvanishing interval that contains the origin.

Fourier inversion gives

$$\begin{aligned}
f_n(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} M_n(it)e^{-itx}dt \\
&= \frac{n}{2\pi i}\int_{-i\infty}^{i\infty} M_n(nT)e^{-nTx}dT \\
&= \frac{n}{2\pi i}\int_{\lambda-i\infty}^{\lambda+i\infty} e^{n(R_n(T)-Tx)}dT, \qquad (2\text{–}12)
\end{aligned}$$

where λ is any real number in the interval where the moment generating function exists.

Let

$$R_n(T)=K_n(nT)/n, \qquad (2\text{–}13)$$

applying Edgeworth approximation, $R_n(T)$ can be approximated in terms of the first four cumulants of U_n.

$$\tilde{R}_n(T) = \mu_n T + \frac{n\sigma_n^2 T^2}{2} + \frac{\kappa_{3n}\sigma_n^3 n^2 T^3}{6} + \frac{\kappa_{4n}\sigma_n^4 n^3 T^4}{24}, \tag{2–14}$$

where μ_n , σ_n^2 , κ_{3n}, κ_{4n} are the mean, the variance, and the third and fourth cumulant of U_n.

We assume that the Edgeworth expansion up to and including the term of order n^{-1} for f_n exists. Expansions of the form

$$\begin{aligned} \mu_n &= \mu + a_1/n + o(n^{-1}) \\ \sigma_n &= \sigma/n^{1/2} + b_1/n^{3/2} + o(n^{-3/2}), \end{aligned} \tag{2–15}$$

will suffice to keep the same order in the approximation.

Applying the saddlepoint technique to the integral in Eq. 2–12 gives the saddlepoint approximation of f_n with uniform error of order $O(n^{-1})$,

$$f_n(x) \approx n^{\frac{1}{2}} \left\{ 2\pi \tilde{R}_n''(\hat{t}) \right\}^{-\frac{1}{2}} \exp\left\{ n\tilde{R}_n(\hat{t}) - n\hat{t}x \right\}, \tag{2–16}$$

where $\hat{t}$ is the *saddlepoint* and it is the solution to the *saddlepoint equation*, $\tilde{R}_n'(\hat{t}) = x$; $\tilde{R}_n'(\cdot)$, $\tilde{R}_n''(\cdot)$ denote the first and second derivatives of $\tilde{R}_n(t)$.

In addition, if we apply the same technique to Lugannani and Rice (1980) formula, we get the saddlepoint approximation to the cdf of U_n,

$$F_n(x) \approx \Phi(\hat{r}) + \phi(\hat{r}) \left\{ \frac{1}{\hat{r}} - \frac{1}{\hat{q}} \right\}, \tag{2–17}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions with

$$\begin{aligned} \hat{r} &= sgn\left(\hat{t}\right) \left[2n \left\{ \hat{t}x - \tilde{R}_n(\hat{t}) \right\} \right]^{\frac{1}{2}} \\ \hat{q} &= \hat{t} \left\{ n\tilde{R}_n''\left(\hat{t}\right) \right\}^{\frac{1}{2}}. \end{aligned}$$

In Eq. 2–17, at $x = E(U_n)$, the alternate expression used is

$$F_n(x) \approx \frac{1}{2} + (72\pi n)^{-\frac{1}{2}} \tilde{R}_n'''(0) \tilde{R}_n''(0)^{-\frac{3}{2}},$$

where $\tilde{R}_n'''(\cdot)$ is the third derivative of the $\tilde{R}_n(t)$.

2.3.2 Approximation to the First Four Cumulants of MLEs of VaR and CVaR

The first four cumulants of MLEs of VaR and CVaR can not be derived straightforwardly. To apply Easton and Ronchetti method, we consider to use Taylor expansion to approximate the cumulants.

Distributions and moments of x_i^+ and x_i^-.

Consider the auxiliary random vector

$$Z^{(i)} = \left[Z_1^i, Z_2^i\right]', \quad i = 1, 2, \ldots, n, \tag{2–18}$$

where $Z_1^{(i)} = X_i^+$ and $Z_2^{(i)} = X_i^-$. Therefore, $V_1 = \bar{Z}_1^n$ and $V_2 = \bar{Z}_2^n$.

According to Kotz *et al.* (2001), ch.3, $Z^{(i)}$ are independent and identically distributed as,

$$Z^{(i)} \stackrel{d}{=} \begin{pmatrix} \frac{\tau}{\sqrt{2}\kappa}\delta_{1,i}W_{1,i} \\ \frac{\tau\kappa}{\sqrt{2}}\delta_{2,i}W_{2,i} \end{pmatrix},$$

where $\{W_{1,i}\}$ and $\{W_{2,i}\}$ are the standard exponential variables. $\{\delta_{1,i}\}$ and $\{\delta_{2,i}\}$ are iid Bernoulli random variables with success probabilities $1/(1+\kappa^2)$ and $\kappa^2/(1+\kappa^2)$, respectively. W_1, W_2, and (δ_1, δ_2) are mutually independent. By definition, the first four moments of δ_1 are

$$E\delta_1 = E\delta_1^2 = E\delta_1^3 = E\delta_1^4 = \frac{1}{(1+\kappa^2)}. \tag{2–19}$$

The first four moments of δ_2 are

$$E\delta_2 = E\delta_2^2 = E\delta_2^3 = E\delta_2^4 = \frac{\kappa^2}{(1+\kappa^2)}. \tag{2–20}$$

The moment generation function of standard exponential variable is $\frac{1}{(1-t)}$. Therefore, the first four moments of W are $EW = 1$, $EW^2 = 2$, $EW^3 = 6$, $EW^4 = 24$.

Since δ_1 and δ_2 are independent of W_1 and W_2, respectively, the first four moments of Z_1 and Z_2 can easily be obtained,

$$\begin{aligned}
EZ_1 &= \frac{\tau}{\sqrt{2}\kappa(1+\kappa^2)}, \qquad EZ_1^2 = \frac{\tau^2}{\kappa^2(1+\kappa^2)} \\
EZ_1^3 &= \frac{3\tau^3}{\sqrt{2}\kappa^3(1+\kappa^2)}, \qquad EZ_1^4 = \frac{6\tau^4}{\kappa^4(1+\kappa^2)} \\
EZ_2 &= \frac{\tau\kappa^3}{\sqrt{2}(1+\kappa^2)}, \qquad EZ_2^2 = \frac{\tau^2\kappa^4}{(1+\kappa^2)} \\
EZ_2^3 &= \frac{3\tau^3\kappa^5}{\sqrt{2}(1+\kappa^2)}, \qquad EZ_2^4 = \frac{6\tau^4\kappa^6}{(1+\kappa^2)}.
\end{aligned} \tag{2–21}$$

The variances of Z_1 and Z_2 can be represented as

$$\begin{aligned}
\mathrm{Var}Z_1 &= \frac{\tau^2\kappa^2}{2\left(1+\kappa^2\right)^2}\left[\left(\frac{1}{\kappa^2}+1\right)^2 - 1\right] \\
\mathrm{Var}Z_2 &= \frac{\tau^2\kappa^2}{2(1+\kappa^2)^2}\left[\left(\kappa^2+1\right)^2 - 1\right].
\end{aligned} \tag{2–22}$$

Note that $EZ_1^{(i)}Z_2^{(i)} = 0$ since at least one of Z_1^i and Z_2^i equal to zero. The covariance of Z_1, Z_2 can be obtained as

$$\mathrm{Cov}(Z_1, Z_2) = -\frac{\tau^2\kappa^2}{2(1+\kappa^2)^2}. \tag{2–23}$$

Other third and fourth central moments and covariances can also be deduced from the first moments of Z_1 and Z_2 by applying Eq. 2–24.

$$\begin{aligned}
E(Z_1 - EZ_1)^3 &= EZ_1^3 - 3EZ_1^2EZ_1 + 2(EZ_1)^3 \\
E(Z_1 - EZ_1)^2(Z_2 - EZ_2) &= -EZ_1^2EZ_2 + 2(EZ_1)^2EZ_2 \\
E(Z_1 - EZ_1)(Z_2 - EZ_2)^2 &= -EZ_1EZ_2^2 + 2EZ_1(EZ_2)^2 \\
E(Z_2 - EZ_2)^3 &= EZ_2^3 - 3EZ_2^2EZ_2 + 2(EZ_2)^3 \\
E(Z_1 - EZ_1)^4 &= EZ_1^4 - 4EZ_1^3EZ_1 + 6EZ_1^2(EZ_1)^2 - 3(EZ_1)^4
\end{aligned}$$

$$
\begin{aligned}
E(Z_1 - EZ_1)^3(Z_2 - EZ_2) &= -EZ_1^3 EZ_2 + 3EZ_1^2 EZ_1 EZ_2 - 3(EZ_1)^3 EZ_2 \\
E(Z_1 - EZ_1)^2(Z_2 - EZ_2)^2 &= (EZ_1)^2 EZ_2^2 + EZ_1^2 (EZ_2)^2 - 3(EZ_1)^2 (EZ_2)^2 \\
E(Z_1 - EZ_1)(Z_2 - EZ_2)^3 &= -EZ_1 EZ_2^3 + 3EZ_1 EZ_2 EZ_2^2 - 3EZ_1 (EZ_2)^3 \\
E(Z_2 - EZ_2)^4 &= EZ_2^4 - 4EZ_2^3 EZ_2 + 6EZ_2^2 (EZ_2)^2 - 3(EZ_2)^4.
\end{aligned}
\tag{2–24}
$$

Taylor expansion for $\tilde{\xi}_\alpha(X)$ and $\tilde{\phi}_\alpha(X)$.

Let $f_1(V_1, V_2) = \tilde{\xi}_\alpha(X)$, $\theta_1 = EZ_1 = EV_1$ and $\theta_2 = EZ_2 = EV_2$. By applying the first five terms of Taylor expansion for the two variables, V_1 and V_2, we can approximate the mean of $\tilde{\xi}_\alpha(X)$ with error term of n^{-3},

$$
\begin{aligned}
& \tilde{\xi}_\alpha(X) = f_1(V_1, V_2) \\
\approx\ & f_1(\theta_1, \theta_2) + \left[(V_1 - \theta_1) \frac{\partial\ f_1}{\partial\ V_1}\bigg|_{V_1=\theta_1} + (V_2 - \theta_2) \frac{\partial\ f_1}{\partial\ V_2}\bigg|_{V_2=\theta_2} \right] \\
+\ & \frac{1}{2}\left[(V_1 - \theta_1)^2 \frac{\partial^2\ f_1}{\partial\ V_1^2}\bigg|_{V_1=\theta_1} + 2(V_1 - \theta_1)(V_2 - \theta_2) \frac{\partial^2\ f_1}{\partial\ V_1 \partial\ V_2}\bigg|_{V_1=\theta_1, V_2=\theta_2} \right. \\
+\ & \left. (V_2 - \theta_2)^2 \frac{\partial^2\ f_1}{\partial\ V_2^2}\bigg|_{V_2=\theta_2} \right] \\
+\ & \frac{1}{3!}\left[(V_1 - \theta_1)^3 \frac{\partial^3\ f_1}{\partial\ V_1^3}\bigg|_{V_1=\theta_1} + 3(V_1 - \theta_1)^2 (V_2 - \theta_2) \frac{\partial^3\ f_1}{\partial\ V_1^2 \partial\ V_2}\bigg|_{V_1=\theta_1, V_2=\theta_2} \right. \\
+\ & \left. 3(V_1 - \theta_1)(V_2 - \theta_2)^2 \frac{\partial^3\ f_1}{\partial\ V_1 \partial\ V_2^2}\bigg|_{V_1=\theta_1, V_2=\theta_2} + (V_2 - \theta_2)^3 \frac{\partial^3\ f_1}{\partial\ V_2^3}\bigg|_{V_2=\theta_2} \right] \\
+\ & \frac{1}{4!}\left[(V_1 - \theta_1)^4 \frac{\partial^4\ f_1}{\partial\ V_1^4}\bigg|_{V_1=\theta_1} + 4(V_1 - \theta_1)^3 (V_2 - \theta_2) \frac{\partial^4\ f_1}{\partial\ V_1^3 \partial\ V_2}\bigg|_{V_1=\theta_1, V_2=\theta_2} \right. \\
+\ & 6(V_1 - \theta_1)^2 (V_2 - \theta_2)^2 \frac{\partial^4\ f_1}{\partial\ V_1^2 \partial\ V_2^2}\bigg|_{V_1=\theta_1, V_2=\theta_2} \\
+\ & 4(V_1 - \theta_1)(V_2 - \theta_2)^3 \frac{\partial^4\ f_1}{\partial\ V_1 \partial\ V_2^3}\bigg|_{V_1=\theta_1, V_2=\theta_2} \\
+\ & \left. (V_2 - \theta_2)^4 \frac{\partial^4\ f_1}{\partial\ V_2^4}\bigg|_{V_2=\theta_2} \right].
\end{aligned}
\tag{2–25}
$$

Take expectations on both sides, the mean of $\tilde{\xi}_\alpha(X)$ can be approximated as

$$
\begin{aligned}
& E\left[\tilde{\xi}_\alpha(X)\right] = Ef_1(V_1,V_2) \\
\approx\ & f_1(\theta_1,\theta_2) \\
+\ & \frac{1}{2}\left[\mathrm{Var}V_1\frac{\partial^2\ f_1}{\partial\ V_1^2}\bigg|_{V_1=\theta_1} + 2\mathrm{Cov}(V_1,V_2)\frac{\partial^2\ f_1}{\partial\ V_1\partial\ V_2}\bigg|_{V_1=\theta_1,V_2=\theta_2} + \mathrm{Var}V_2\frac{\partial^2\ f_1}{\partial\ V_2^2}\bigg|_{V_2=\theta_2}\right] \\
+\ & \frac{1}{6}\left[E(V_1-\theta_1)^3\frac{\partial^3\ f_1}{\partial\ V_1^3}\bigg|_{V_1=\theta_1} + 3E(V_1-\theta_1)^2(V_2-\theta_2)\frac{\partial^3\ f_1}{\partial\ V_1^2\partial\ V_2}\bigg|_{V_1=\theta_1,V_2=\theta_2}\right. \\
+\ & \left. 3E(V_1-\theta_1)(V_2-\theta_2)^2\frac{\partial^3\ f_1}{\partial\ V_1\partial\ V_2^2}\bigg|_{V_1=\theta_1,V_2=\theta_2} + E(V_2-\theta_2)^3\frac{\partial^3\ f_1}{\partial\ V_2^3}\bigg|_{V_2=\theta_2}\right] \\
+\ & \frac{1}{24}\left[E(V_1-\theta_1)^4\frac{\partial^4\ f_1}{\partial\ V_1^4}\bigg|_{V_1=\theta_1} + 4E(V_1-\theta_1)^3(V_2-\theta_2)\frac{\partial^4\ f_1}{\partial\ V_1^3\partial\ V_2}\bigg|_{V_1=\theta_1,V_2=\theta_2}\right. \\
+\ & 6E(V_1-\theta_1)^2(V_2-\theta_2)^2\frac{\partial^4\ f_1}{\partial\ V_1^2\partial\ V_2^2}\bigg|_{V_1=\theta_1,V_2=\theta_2} \\
+\ & 4E(V_1-\theta_1)(V_2-\theta_2)^3\frac{\partial^4\ f_1}{\partial\ V_1\partial\ V_2^3}\bigg|_{V_1=\theta_1,V_2=\theta_2} \\
+\ & \left. E(V_2-\theta_2)^4\frac{\partial^4\ f_1}{\partial\ V_2^4}\bigg|_{V_2=\theta_2}\right] \\
.
\end{aligned}
\tag{2–26}
$$

Since $V_1 = \bar{Z}_1^{(n)}$ and $V_2 = \bar{Z}_2^{(n)}$, this approximation can be further expressed as

$$
\begin{aligned}
& E\left[\tilde{\xi}_\alpha(X)\right] \\
\approx\ & f_1(\theta_1,\theta_2) \\
+\ & \frac{1}{2n}\left[\mathrm{VaR}Z_1\frac{\partial^2\ f_1}{\partial\ V_1^2}\bigg|_{V_1=\theta_1} + 2\mathrm{Cov}(Z_1,Z_2)\frac{\partial^2\ f_1}{\partial\ V_1\partial\ V_2}\bigg|_{V_1=\theta_1,V_2=\theta_2} + \mathrm{VaR}Z_2\frac{\partial^2\ f_1}{\partial\ V_2^2}\bigg|_{V_2=\theta_2}\right] \\
+\ & \frac{1}{6n^2}\left[E(Z_1-\theta_1)^3\frac{\partial^3\ f_1}{\partial\ V_1^3}\bigg|_{V_1=\theta_1} + 3E(Z_1-\theta_1)^2(Z_2-\theta_2)\frac{\partial^3\ f_1}{\partial\ V_1^2\partial\ V_2}\bigg|_{V_1=\theta_1,V_2=\theta_2}\right. \\
+\ & \left. 3E(Z_1-\theta_1)(Z_2-\theta_2)^2\frac{\partial^3\ f_1}{\partial\ V_1\partial\ V_2^2}\bigg|_{V_1=\theta_1,V_2=\theta_2} + E(Z_2-\theta_2)^3\frac{\partial^3\ f_1}{\partial\ V_2^3}\bigg|_{V_2=\theta_2}\right] \\
+\ & \frac{1}{24n^3}\left[E(Z_1-\theta_1)^4\frac{\partial^4\ f_1}{\partial\ V_1^4}\bigg|_{V_1=\theta_1} + 4E(Z_1-\theta_1)^3(Z_2-\theta_2)\frac{\partial^4\ f_1}{\partial\ V_1^3\partial\ V_2}\bigg|_{V_1=\theta_1,V_2=\theta_2}\right. \\
+\ & 6E(Z_1-\theta_1)^2(Z_2-\theta_2)^2\frac{\partial^4\ f_1}{\partial\ V_1^2\partial\ V_2^2}\bigg|_{V_1=\theta_1,V_2=\theta_2} \\
+\ & \left. 4E(Z_1-\theta_1)(Z_2-\theta_2)^3\frac{\partial^4\ f_1}{\partial\ V_1\partial\ V_2^3}\bigg|_{V_1=\theta_1,V_2=\theta_2} + E(Z_2-\theta_2)^4\frac{\partial^4\ f_1}{\partial\ V_2^4}\bigg|_{V_2=\theta_2}\right].
\end{aligned}
\tag{2–27}
$$

Let k_1, k_2, k_3, k_4 be the first four cumulants, μ_1, μ_2, μ_3, μ_4 be the first four moments of $\tilde{\xi}_\alpha(X)$, respectively. Set $f_2(V_1, V_2) = \left(\tilde{\xi}_\alpha(X)\right)^2$, $f_3(V_1, V_2) = \left(\tilde{\xi}_\alpha(X)\right)^3$, $f_4(V_1, V_2) = \left(\tilde{\xi}_\alpha(X)\right)^4$. The same technique can be applied to approximate the second, third and forth moments as

$$\begin{aligned}
\mu_2 &= E\left(\tilde{\xi}_\alpha(X)\right)^2 = Ef_2(V_1, V_2) \\
\mu_3 &= E\left(\tilde{\xi}_\alpha(X)\right)^3 = Ef_3(V_1, V_2) \\
\mu_4 &= E\left(\tilde{\xi}_\alpha(X)\right)^4 = Ef_4(V_1, V_2).
\end{aligned} \tag{2–28}$$

To approximate the first four cumulants of $\tilde{\xi}_\alpha(X)$, we apply the relationship of moments and cumulants,

$$\begin{aligned}
k_1 &= \mu_1 \\
k_2 &= \mu_2 - (\mu_1)^2 \\
k_3 &= 2(\mu_1)^3 - 3\mu_1\mu_2 + \mu_3 \\
k_4 &= -6(\mu_1)^4 + 12(\mu_1)^2\mu_2 - 3(\mu_2)^2 - 4\mu_1\mu_3 + \mu_4.
\end{aligned} \tag{2–29}$$

As the result, the approximation to the distribution of $\tilde{\xi}_\alpha(X)$ and $\tilde{\phi}_\alpha(X)$ can be obtained.

2.3.3 Assessing the Accuracy of the Saddlepoint Approximations

In this section, we compare the saddlepoint approximation to the distribution of MLEs of VaR and CVaR, with empirical values obtained via simulation.

Let $\hat{F}_{sim}(r)$ and $\hat{F}_{sad}(r)$ denote the estimates of the true cdfs of maximum likelihood estimator, obtained via simulations and saddlepoint approximations, respectively. The *Percent relative error* (PRE) of the cdfs is a commonly used technique to measure the accuracy of the saddlepoint estimation. We define the PRE at the quantile r as

$$\mathrm{PRE}(r) = \begin{cases} \frac{\hat{F}_{sad}(r) - \hat{F}_{sim}(r)}{\hat{F}_{sim}(r)} 100, & \hat{F}_{sim}(r) \leq 0.5, \\ \frac{\left(1 - \hat{F}_{sad}(r)\right) - \left(1 - \hat{F}_{sim}(r)\right)}{1 - \hat{F}_{sim}(r)} 100, & \hat{F}_{sim}(r) > 0.5. \end{cases} \tag{2–30}$$

Thus, large absolute values of PRE denote larger discrepancies between the saddlepoint approximation and the true distribution, while a PRE value of 0 indicates perfect agreement.

To assess the accuracy of this saddlepoint estimation, we approximate cdfs of MLEs of VaR and CVaR with parameters $n = 50, 100, \theta = 0, \kappa = 0.8, 1, \tau = 1, \alpha = 0.9, 0.99$. We calculate empirical cdfs through 10^6 simulations. PREs are computed at 10 points of equal distance between the $10\% \sim 90\%$ quantiles.

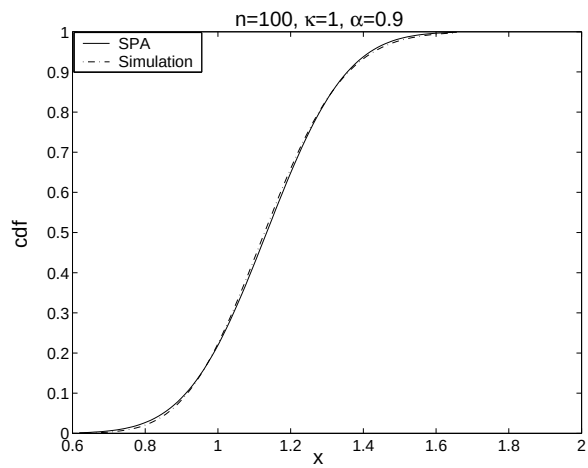

Figure 2-1. Estimated CDFs of $\tilde{\xi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, with n=100, $\theta = 0, \tau = 1, \kappa = 1, \alpha = 0.9$

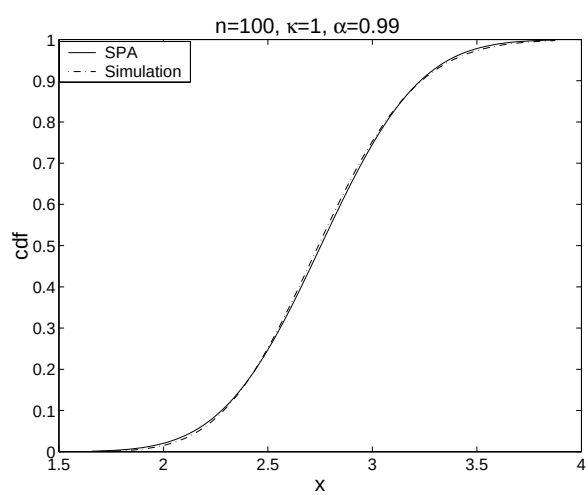

Figure 2-2. Estimated cdfs of $\tilde{\xi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, with n=100, $\theta = 0, \tau = 1, \kappa = 1, \alpha = 0.99$

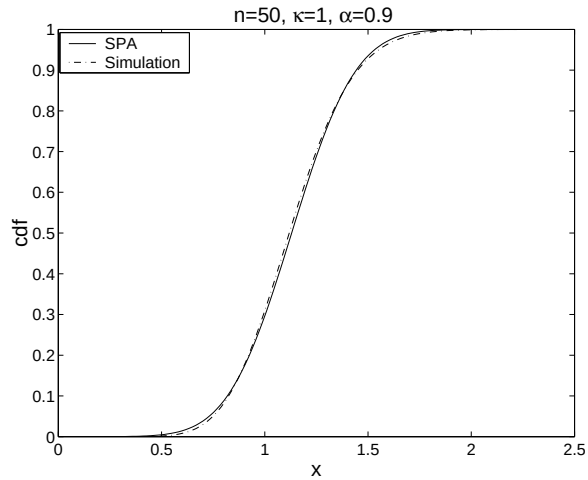

Figure 2-3. Estimated cdfs of the $\tilde{\xi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, with n=50, $\theta = 0, \tau = 1, \kappa = 1, \alpha = 0.9$

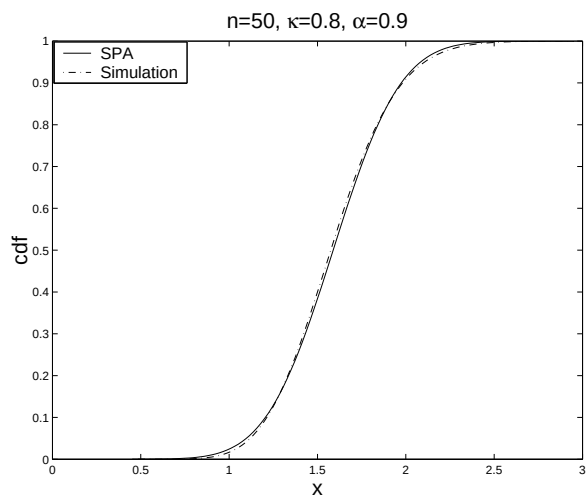

Figure 2-4. Estimated cdfs of the $\tilde{\xi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, with n=50, $\theta = 0, \tau = 1, \kappa = 0.8, \alpha = 0.9$

Calculations based on a few equispaced points between the 10th and 90th percentiles, reveal that PREs for the MLEs of VaR and CVaR are between 1% and 4.5%, with $n = 50, 100, \kappa = 0.8, 1, \alpha = 0.9, 0.95, 0.99$. This indicates that our approach works well to approximate the distributions of MLEs of VaR and CVaR from the Asymmetric Laplace distribution.

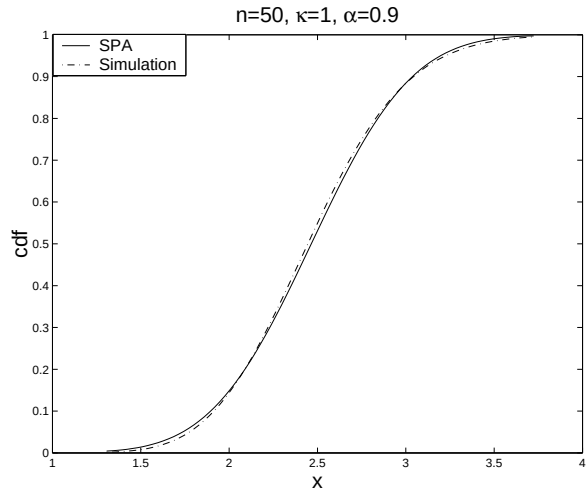

Figure 2-5. Estimated cdfs of $\tilde{\phi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, with n=100, $\theta = 0, \tau = 1, \kappa = 1, \alpha = 0.9$

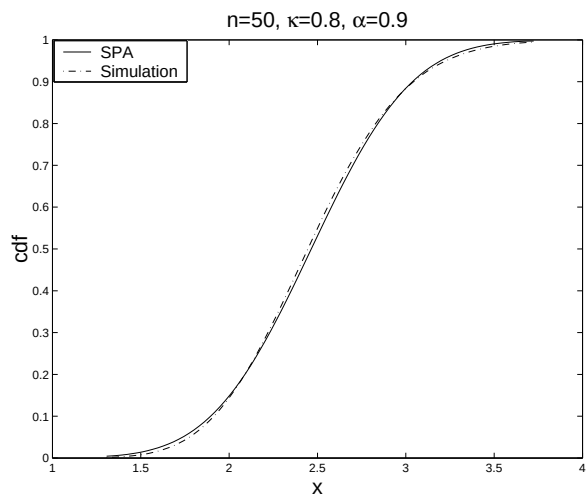

Figure 2-6. Estimated cdfs of $\tilde{\phi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, with n=100, $\theta = 0, \tau = 1, \kappa = 0.8, \alpha = 0.9$

We have chosen $\alpha = 0.9, 0.95, 0.99$ because they are commonly used probability levels in financial risk measures. We choose $\kappa = 0.8$ because it is close to the fitted values obtained from a real data set, while $\kappa = 1$ indicates the symmetric case.

Also, note from Fig. 2-7 and Fig. 2-8, the PREs are less than zero for large value of r, which means that we tend to overestimate $\tilde{\xi}_\alpha(X)$ and $\tilde{\phi}_\alpha(X)$ over the right tails. The approximations are better for larger sample sizes, and also better for symmetric cases.

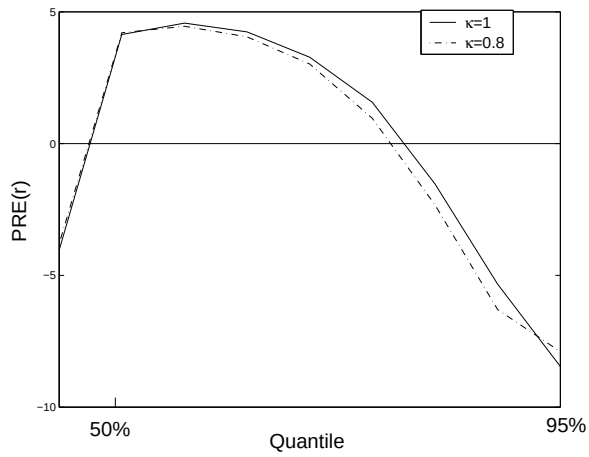

Figure 2-7. Percent relative errors (PREs) of the saddlepoint approximation to the distribution of $\tilde{\xi}_\alpha(X)$ with $n = 100, \theta = 0, \kappa = 1, 0.8, \tau = 1, \alpha = 0.9$, computed at the same quantile values.

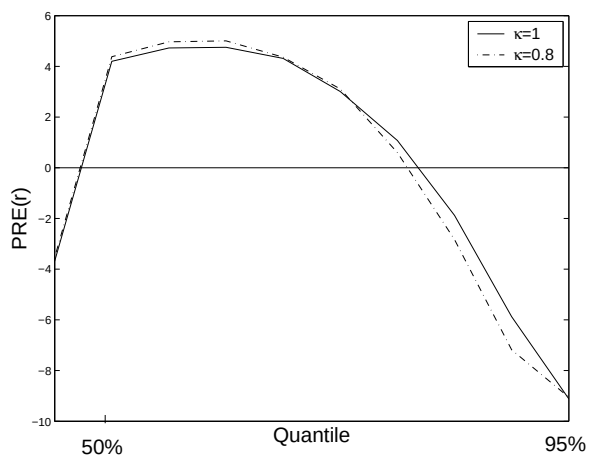

Figure 2-8. Percent relative errors (PREs) for the saddlepoint approximation to the distribution of $\tilde{\phi}_\alpha(X)$ with $n = 100, \theta = 0, \kappa = 1, 0.8, \tau = 1, \alpha = 0.9$, computed at the same quantile values.

CHAPTER 3
APPROXIMATION TO THE DISTRIBUTION OF NONPARAMETRIC ESTIMATORS OF VAR AND CVAR UNDER AL LAW

In this chapter, we study the nonparametric estimators of VaR and CVaR. We approximate the distributions of nonparametric estimators of VaR and CVaR using saddlepoint approximation. We derive the moment generating functions of NPEs and then approximate the distributions of NPEs using saddlepoint approximation. The moment generating functions of VaR and CVaR are mixtures of hypergeometric functions which makes the calculation more computational intensive.

We analyze the performance of the MLEs and NPEs by comparing the saddlepoint approximated distributions. Daily log returns of USD/EUR exchange rate are studied assuming IID AL distribution.

3.1 Nonparametric Estimators of VaR and CVaR

The Nonparametric approach makes no assumption about the distribution the of underlying financial data. Consider a random sample $Y_1, \ldots, Y_n$, let $Y_{(1)} \leq \cdots \leq Y_{(n)}$ denote the corresponding order statistics from this random sample, the NPE of VaR is the αth empirical quantile,

$$\hat{\xi}_\alpha(Y) = Y_{(k_\alpha)}, \tag{3–1}$$

where the $k_\alpha = [n\alpha]$ denotes either of the two integers closest to $n\alpha$. The NPE of CVaR is the corresponding empirical tail mean,

$$\hat{\phi}_\alpha(Y) = \frac{1}{n - k_\alpha + 1} \sum_{r=k_\alpha}^{n} Y_{(r)}. \tag{3–2}$$

Note that $\hat{\xi}_\alpha(Y)$ and $\hat{\phi}_\alpha(Y)$ being linear combinations of order statistics, are known as *L-statistics* (David and Nagaraja, 2003).

3.2 Asymptotic Distribution of $\hat{\xi}_\alpha(X)$ and $\hat{\phi}_\alpha(X)$

Since the NPE of VaR is simply an order statistic of a random sample, the asymptotic distribution of $\hat{\xi}_\alpha(X)$ comes from the standard result of the asymptotic theory of order statistics, for example, David and Nagaraja (2003), ch.10. The asymptotics of $\hat{\phi}_\alpha(X)$ are more complex. This result was first derived by Stigler (1973) in the context of the trimmed mean.

Consistency and the joint asymptotic distribution of $(\hat{\xi}_\alpha, \hat{\phi}_\alpha)$ under iid sampling from a continuous cdf F with pdf f, has recently been established by Giurcanu and Trindade (2005) using the theory of estimating equations. Define σ_α^2 be the variance of the distribution obtained by truncating the distribution of Y at ξ_α, i.e.,

$$\sigma_\alpha^2 = \frac{1}{1-\alpha}\int_{\xi_\alpha}^{\infty}(y-\phi_\alpha)^2 f(y)dy. \tag{3–3}$$

Theorem 3.1 (Asymptotic distribution of NPEs of VaR and CVaR). Under random sampling from Y with cdf F, pdf f, and finite variance, we have the following central limit theorem for the NPEs of VaR and CVaR,

$$\sqrt{n}\left(\begin{bmatrix}\hat{\xi}_\alpha\\ \hat{\phi}_\alpha\end{bmatrix} - \begin{bmatrix}\xi_\alpha\\ \phi_\alpha\end{bmatrix}\right) \xrightarrow{d} N_2\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)} & \frac{\alpha(\phi_\alpha-\xi_\alpha)}{f(\xi_\alpha)}\\ \frac{\alpha(\phi_\alpha-\xi_\alpha)}{f(\xi_\alpha)} & \frac{\sigma_\alpha^2+\alpha(\phi_\alpha-\xi_\alpha)^2}{1-\alpha}\end{bmatrix}\right). \tag{3–4}$$

3.3 Approximation of Finite Sample Distribution

3.3.1 Moment Generating Function of Nonparametric Estimators of VaR and CVaR

In this section, we approximate the distribution of NPEs of VaR and CVaR from iid random sample of AL distribution. We apply saddlepoint approximation starting from the moment generating functions (mgfs) of $\hat{\xi}_\alpha(X)$ and $\hat{\phi}_\alpha(X)$.

Trindade and Zhu (2007) derive the mgfs of $\hat{\xi}_\alpha(X)$ and $\hat{\phi}_\alpha(X)$. First, we define the Gauss hypergeometric function $F_{2,1}(a,b;c;z)$ as

$$F_{2,1}(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 u^{b-1}(1-u)^{c-b-1}(1-uz)^{-a}du, \tag{3–5}$$

where a,b,c,z are real constants and $\Gamma(\cdot)$ is the gamma function. The hypergeometric function converges for $|z| < 1$ provided $c > a + b - 1$.

Lemma 3.1. Let $Y_1, \ldots, Y_n$ be a random sample from $\mathcal{AL}(\theta, \kappa, \tau)$, with corresponding standardizations $X_i = (Y_i - \theta)/\tau \sim \mathcal{AL}(0, \kappa, 1)$, $i = 1, \ldots, n$, if

$$\mathrm{T}_n(\mathrm{Y}) = \sum_{i=1}^n c_i Y_{(i)}, \quad \text{and} \quad \mathrm{T}_n(\mathrm{X}) = \sum_{i=1}^n c_i X_{(i)}$$

are any L-statistics, then we have the following relations between them, their mgfs, and their pdfs:

(i) $\mathrm{T}_n(\mathrm{Y}) = \theta + \tau \mathrm{T}_n(\mathrm{X})$,

(ii)$M_{\mathrm{T}_n(\mathrm{Y})}(t) = e^{t\theta} M_{\mathrm{T}_n(\mathrm{X})}(t\tau)$,

(iii)$f_{\mathrm{T}_n(\mathrm{Y})}(y) = \frac{1}{\tau} f_{\mathrm{T}_n(\mathrm{X})}(\frac{y-\theta}{\tau})$.

Proof:.Straightforward results for location-scale families of distributions.

Lemma 3.2. For any real constants a, b, and c,

$$B(c; a+1, b+1) = \int_0^c u^a(1-u)^b du = \frac{c^{a+1}}{a+1} F_{2,1}(-b, a+1; a+2; c), \tag{3–6}$$

and converges for all $|c| < 1$, provided $b > -2$.

Proof:. This is just the definition of the Incomplete beta function (Abramowitz and Stegun, 1972). The connection with the hypergeometric function is easily derived from Eq. 3–5.

Lemma 3.3 (pdfs of order statistics from iid random sample). Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population

with cdf $\mathrm{F}_X(x)$ and pdf $\mathrm{f}_X(x)$. Then the pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1-F_X(x)]^{n-j}. \tag{3–7}$$

Theorem 3.2 (MGF of an order statistic under AL law). Let $Y_{(r)}$ be the rth order statistic under iid sampling from $\mathcal{AL}(\theta, \kappa, \tau)$, then the mgf of $Y_{(r)}$ is the mixture of hypergeometric functions,

$$M_{Y_{(r)}}(t) = \frac{n! e^{t\theta}}{a_1! a_2!} \sum_{i=1}^{2} \frac{z_i^{n-a_i}}{b_i(t)} F_{2,1}\left(-a_i, b_i(t); 1+b_i(t); z_i\right), \tag{3–8}$$

where

$$\begin{aligned} a_1 &= n-r, \quad b_1(t) = n - a_1 + t\kappa\tau/\sqrt{2}, \qquad z_1 = \frac{\kappa^2}{1+\kappa^2}, \\ a_2 &= r-1, \quad b_2(t) = n - a_2 - t\tau/(\kappa\sqrt{2}), \quad z_2 = \frac{1}{1+\kappa^2}, \end{aligned}$$

and is defined for all $-\infty < t < \infty$.

Proof:. Since $Y_{(r)}$ is an L-statistic, we can simply consider the standard case $X_1, \ldots, X_n \sim$ iid $\mathcal{AL}(0, \kappa, 1)$. Applying Lemma 3.1, the mgf in the general case is

$$M_{Y_{(r)}}(t) = e^{t\theta} M_{X_{(r)}}(t\tau). \tag{3–9}$$

The pdf and cdf of the standard case can be represented as

$$f_X(x) = \frac{\kappa\sqrt{2}}{1+\kappa^2}\left[\exp\left(\frac{\sqrt{2}x}{\kappa} I(x<0)\right) + \exp\left(-\sqrt{2}\kappa x I(x>0)\right)\right], \tag{3–10}$$

and

$$F_X(x) = \left[\frac{\kappa^2}{1+\kappa^2}\exp\left(\frac{\sqrt{2}x}{\kappa}\right)\right] I(x<0) + \left[1 - \frac{1}{1+\kappa^2}\exp\left(-\sqrt{2}\kappa x\right)\right] I(x>0). \tag{3–11}$$

Applying Lemma 3.3, the pdf of $X_{(r)}$ can be expressed as

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x)[F_X(x)]^{r-1}[1-F_X(x)]^{n-r}. \tag{3–12}$$

The mgf of $X_{(r)}$ can be split into two parts,

$$\begin{aligned}
M_{X_{(r)}}(t) &= \int_{-\infty}^{+\infty} e^{tx} f_{X_{(r)}}(x)dx \\
&= \int_{-\infty}^{0} e^{tx} f_{X_{(r)}}(x)dx + \int_{0}^{+\infty} e^{tx} f_{X_{(r)}}(x)dx \\
&= \frac{n!}{(r-1)!(n-r)!}\Bigg\{ \int_{-\infty}^{0} e^{tx} f_X(x) F_X(x)^{(r-1)} \left[1-F_X(x)\right]^{n-r} dx \\
&\quad + \int_{0}^{+\infty} e^{tx} f_X(x) F_X(x)^{(r-1)} \left[1-F_X(x)\right]^{n-r} dx \Bigg\} \\
&\equiv c(n,r)\{J_1(t)+J_2(t)\}. \qquad (3\text{–}13)
\end{aligned}$$

For $J_1(t)$, apply transformation $u = F_X(x)I(x<0) = \frac{\kappa^2}{1+\kappa^2}e^{\frac{\sqrt{2}x}{\kappa}}I(x<0)$, then $0<u<\frac{\kappa^2}{1+\kappa^2}$. Also notice, $x = F_X^{-1}(u) = \log(\frac{1+\kappa^2}{\kappa^2}u)^{\frac{\kappa}{\sqrt{2}}}$, $du = dF_X(x) = f_X(x)dx$. Plugging in all these details, we get

$$\begin{aligned}
J_1(t) &= \int_0^{\frac{\kappa^2}{1+\kappa^2}} e^{t\log\left(\frac{1+\kappa^2}{\kappa^2}u\right)^{\frac{\kappa}{\sqrt{2}}}} u^{r-1}(1-u)^{n-r}du \\
&= \int_0^{\frac{\kappa^2}{1+\kappa^2}} \left(\frac{1+\kappa^2}{\kappa^2}u\right)^{\frac{\kappa t}{\sqrt{2}}} u^{r-1}(1-u)^{n-r}du \\
&= \left(\frac{1+\kappa^2}{\kappa^2}\right)^{\frac{\kappa t}{\sqrt{2}}} \int_0^{\frac{\kappa^2}{1+\kappa^2}} u^{\frac{\kappa t}{\sqrt{2}}+r-1}(1-u)^{n-r}du. \qquad (3\text{–}14)
\end{aligned}$$

Set $a = \frac{\kappa t}{\sqrt{2}}+r-1$, $b = n-r$ and $c = \frac{\kappa^2}{1+\kappa^2}$, apply Lemma 3.2,

$$J_1(t) = \frac{\left(\frac{\kappa^2}{1+\kappa^2}\right)^r}{\frac{\kappa t}{\sqrt{2}}+r} F_{2,1}\left(r-n, \frac{\kappa t}{\sqrt{2}}+r; \frac{\kappa t}{\sqrt{2}}+r+1; \frac{\kappa^2}{1+\kappa^2}\right). \qquad (3\text{–}15)$$

For $J_2(t)$, Let $v = 1-F_X I(x>0) = \frac{1}{1+\kappa^2}\exp(-\sqrt{2}\kappa x)I(x>0)$, then $0<r<\frac{1}{1+\kappa^2}$. $x = \log\left[v(1+\kappa^2)\right]^{-\frac{1}{\sqrt{2}\kappa}}$, $dv = d[1-F_X(x))] = -f_X(x)dx$. Therefore,

$$\begin{aligned}
J_2(t) &= -\int_{\frac{1}{1+\kappa^2}}^{0} e^{t\log\left[v(1+\kappa^2)\right]^{-\frac{1}{\sqrt{2}\kappa}}}(1-v)^{r-1}v^{n-r}dv \\
&= \int_0^{\frac{1}{1+\kappa^2}} \left[v(1+\kappa^2)\right]^{-\frac{t}{\sqrt{2}\kappa}}(1-v)^{r-1}v^{n-r}dv \\
&= \int_0^{\frac{1}{1+\kappa^2}} (1+\kappa^2)^{-\frac{t}{\sqrt{2}\kappa}} v^{n-r-\frac{t}{\sqrt{2}\kappa}}(1-v)^{r-1}dv. \qquad (3\text{–}16)
\end{aligned}$$

Set $a = n - r - \frac{t}{\sqrt{2}\kappa}$, $b = r - 1$ and $c = \frac{1}{1+\kappa^2}$, we get

$$J_2(t) = \frac{(\frac{1}{1+\kappa^2})^{(n-r+1)}}{n - r - \frac{t}{\sqrt{2}\kappa} + 1} F_{2,1}\left(1 - r, n - r - \frac{t}{\sqrt{2}\kappa} + 1, n - r - \frac{t}{\sqrt{2}\kappa} + 2, \frac{1}{1+\kappa^2}\right). \tag{3–17}$$

Applying Lemma 3.1, gives the desired result.

Corollary 3.1 (MGF of NPE of VaR under AL law).

If $\hat{\xi}_\alpha = Y_{(k_\alpha)}$ denotes the NPE of $\mathrm{VaR}_\alpha(Y)$ based on a random sample of size n from $\mathcal{AL}(\theta, \kappa, \tau)$, its mgf is given by Theorem 3.2 with $r = [n\alpha]$.

Theorem 3.3 (MGF of sum of upper order statistics under AL law).

Let $Y_{(n-k+1)} \leq \ldots \leq Y_{(n)}$ be the highest k, $1 \leq k \leq n$, order statistics under iid sampling from $\mathcal{AL}(\theta, \kappa, \tau)$. Let $S_{n,k}(Y) = \sum_{i=n-k+1}^{n} Y_{(i)}$ be their sum. Then the mgf of $S_{n,k}(Y)$ is the mixture of hypergeometric functions,

$$M_{S_{n,k}(Y)}(t) = \frac{n! e^{t\theta}}{a_1! a_2!} \sum_{i=1}^{2} d_i(t) F_{2,1}(-a_i, b_i(t); 1 + b_i(t); z_i(t)), \tag{3–18}$$

where

$$\begin{aligned}
a_1 &= k, \qquad b_1(t) = \frac{\sqrt{2}(n-k)}{\sqrt{2} + t\kappa\tau}, \qquad z_1(t) = \frac{\kappa\sqrt{2} - t\tau}{\kappa\sqrt{2} + \sqrt{2}/\kappa}, \\
a_2 &= n - k - 1, \quad b_2(t) = 1 + k - \frac{tk\tau}{\kappa\sqrt{2}}, \quad z_2(t) = \frac{1}{1+\kappa^2}, \\
d_1(t) &= \left(\frac{\kappa\sqrt{2} z_2(t)}{z_1(t)(t\tau + \sqrt{2}/\kappa)}\right)^{a_1} \frac{(z_2(t)\kappa^2)^{n-a_1}}{n - a_1}, \\
d_2(t) &= \left(\frac{\kappa\sqrt{2} z_2(t)}{\kappa\sqrt{2} - t\tau}\right)^{a_1} \frac{z_2(t)}{b_2(t)},
\end{aligned}$$

and is defined for $-\sqrt{2}/(\kappa\tau) < t < \kappa\sqrt{2}/\tau$.

Proof:. In light of Lemma 3.1, it suffice to consider the standard case, $X_1, \ldots, X_n \sim$ iid $\mathcal{AL}(0, \kappa, 1)$, hence the mgf in the general case is

$$M_{S_{n,k}(Y)}(t) = e^{t\theta} M_{S_{n,k}(X)}(t\tau). \tag{3–19}$$

Since $S_{n,k}(X) \equiv S_{n,k} = \sum_{i=n-k+1}^{n} X_{(i)}$, David and Nagaraja (2003), Section 6.5, show that conditional on $X_{(n-k)}$, $S_{n,k}$ can be written as a sum of k iid random variables. If F and f denote the cdf and pdf of the distribution of X_1, this means that

$$(S_{n,k}|X_{(n-k)} = x) \stackrel{d}{=} \sum_{i=1}^{k} X_i^*, \tag{3–20}$$

where $X_1^*, \ldots, X_k^*$ are iid with pdf $f_x^*(y) = f(y)I(y > x)/(1 - F(x))$. Therefore, the pdf of $S_{n,k}$ is given by

$$f_{S_{n,k}}(y) = \int_{-\infty}^{y} f_{S_{n,k}|X_{(n-k)}=x}(y|x) f_{X_{(n-k)}}(x)dx, \tag{3–21}$$

with corresponding mgf

$$\begin{aligned} M_{S_{n,k}}(t) &= \int_{\mathbf{R}} e^{ty} f_{S_{n,k}}(y)dy \\ &= \int_{\mathbf{R}} e^{ty} \int_{-\infty}^{y} f_{S_{n,k}|X_{(n-k)}=x}(y|x) f_{X_{(n-k)}}(x)dxdy \\ &= \int_{-\infty}^{\infty} \int_{x}^{\infty} e^{ty} f_{S_{n,k}|X_{(n-k)}=x}(y|x) f_{X_{(n-k)}}(x)dydx \\ &= \int_{-\infty}^{\infty} f_{X_{(n-k)}}(x) \int_{x}^{\infty} e^{ty} f_{S_{n,k}|X_{(n-k)}=x}(y|x)dydx \\ &= \int_{\mathbf{R}} [M_x^*(t)]^k f_{X_{(n-k)}}(x)dx, \end{aligned} \tag{3–22}$$

where $M_x^*(t)$ denotes the mgf of X_1^*, which depends on x. This technique, which only works if the integration can be performed analytically, has been used by Alam and Wallenius (1979) for obtaining the distribution of $S_{n,k}$ in a random sample from a gamma distribution. Since $x < y$, there can be three cases: (i) $0 < x < y$, (ii) $x < y < 0$, (iii) $x < 0 < y$. In general,

$$f_x^*(y) = \begin{cases} \kappa\sqrt{2}\exp(-\kappa\sqrt{2}(y-x)) & \text{if } 0 < x < y; \\ \frac{\kappa\sqrt{2}e^{\frac{\sqrt{2}y}{\kappa}}}{1+\kappa^2\left(1-e^{\frac{\sqrt{2}x}{\kappa}}\right)} & \text{if } x < y < 0; \\ \frac{\kappa\sqrt{2}e^{-\sqrt{2}y\kappa}}{1+\kappa^2\left(1-e^{\frac{\sqrt{2}x}{\kappa}}\right)} & \text{if } x < 0 < y. \end{cases} \tag{3–23}$$

The mgf of X_1^* can be considered in two cases $x > 0$ and $x < 0$.

In the case $x > 0$, i.e., $0 < x < y$,

$$\begin{aligned} M_x^*(t) &= \int_x^\infty e^{ty} f_x^*(y)dy \\ &= \int_x^\infty e^{ty}\kappa\sqrt{2}\exp(-\kappa\sqrt{2}(y-x))dy \\ &= \kappa\sqrt{2}e^{\kappa x\sqrt{2}}\int_x^\infty e^{(t-\kappa\sqrt{2})y}dy \\ &= \frac{\kappa\sqrt{2}e^{\kappa x\sqrt{2}}}{t-\kappa\sqrt{2}}e^{(t-\kappa\sqrt{2})y}\Big|_x^\infty. \end{aligned}$$

If $t - \kappa\sqrt{2} < 0$, $e^{(t-\kappa\sqrt{2})y}\big|_{y=\infty}$ goes to 0,

$$\begin{aligned} M_x^*(t) &= \frac{\kappa\sqrt{2}e^{\kappa x\sqrt{2}}}{t-\kappa\sqrt{2}}(0 - e^{(t-\kappa\sqrt{2})x}) \\ &= \frac{\kappa\sqrt{2}e^{tx}}{\kappa\sqrt{2}-t}. \end{aligned} \tag{3–24}$$

In case $x < 0$, $M_x^*(t)$ can be split into two parts,

$$\begin{aligned} M_x^*(t) &= \int_x^0 e^{ty} f_x^*(y)dy + \int_0^\infty e^{ty} f_x^*(y)dy \\ &= \frac{\kappa\sqrt{2}}{1+\kappa^2\left(1-e^{\frac{\sqrt{2}x}{\kappa}}\right)}\left[\int_x^0 e^{y\left(t+\frac{\sqrt{2}}{\kappa}\right)}dy + \int_0^\infty e^{y\left(t-\kappa\sqrt{2}\right)}dy\right] \\ &= \frac{\kappa\sqrt{2}}{1+\kappa^2\left(1-e^{\frac{\sqrt{2}x}{\kappa}}\right)}\left[\frac{e^{y(t+\frac{\sqrt{2}}{\kappa})}\big|_x^0}{t+\frac{\sqrt{2}}{\kappa}} + \frac{e^{y(t-\kappa\sqrt{2})}\big|_0^\infty}{t-\kappa\sqrt{2}}\right] \\ &= \frac{\kappa\sqrt{2}}{1+\kappa^2\left(1-e^{\frac{\sqrt{2}x}{\kappa}}\right)}\left[\frac{1-e^{x(t+\frac{\sqrt{2}}{\kappa})}}{t+\frac{\sqrt{2}}{\kappa}} + \frac{0-1}{t-\kappa\sqrt{2}}\right] \\ &= \frac{\kappa\sqrt{2}}{\left[1+\kappa^2\left(1-e^{\frac{\sqrt{2}}{\kappa}x}\right)\right]\left(t+\frac{\sqrt{2}}{\kappa}\right)}\left[\frac{\kappa\sqrt{2}+\frac{\sqrt{2}}{\kappa}}{\kappa\sqrt{2}-t} - \exp\left(x\left(t+\frac{\sqrt{2}}{\kappa}\right)\right)\right]. \end{aligned} \tag{3–25}$$

Substituting this into Eq. 3–22, and by a series of obvious u-substitutions, the integrands can be reduced to mixtures of hypergeometric functions, and we obtain eventually the

statement of the theorem, which proves the result. The distribution of $\hat{\phi}_\alpha(Y)$ follows immediately be setting $k = n - [n\alpha] + 1$ and $t = \frac{t}{n-[n\alpha]+1}$.

Corollary 3.2 (MGF of NPE of CVaR under AL law).

If $\hat{\phi}_\alpha = \frac{1}{n-[n\alpha]+1} \sum_{i=[n\alpha]}^{n} Y_{(i)}$ denotes the NPE of $\mathrm{CVaR}_\alpha(Y)$ based on a random sample of size n from $\mathcal{AL}(\theta, \kappa, \tau)$, its mgf is given by Theorem 3.3 with $k = n - [n\alpha] + 1$ and $t = t/(n - [n\alpha] + 1)$

3.3.2 Saddlepoint Approximation and Lugannani-Rice Formula

The saddlepoint approximation was first introduced by Daniels (1954). Let $X_1, \ldots, X_n$ be an iid random sample from a distribution with density $f(x)$ and let $S_n(X_1, \ldots, X_n)$ be a real valued statistic with density f_n. Let $M_n(t)$ and $K_n(t)$ be the moment generating function and cumulant generating function of S_n, respectively. For continuous random variables, the saddlepoint approximation to f_n at x is

$$f_n(x) \approx \left\{2\pi K_n''(\hat{t})\right\}^{-\frac{1}{2}} \exp\left\{K_n(\hat{t}) - \hat{t}x\right\}, \tag{3–26}$$

where $\hat{t}$ is the *saddlepoint* and it is the solution to the *saddlepoint equation*, $K_n'(\hat{t}) = x$; and $K_n'(\cdot)$, $K_n''(\cdot)$ denote the first and second derivatives of the cumulant generating function $K_n(t)$. The relative error of the approximation is of order $O(n^{-1})$.

The saddlepoint approximation to the cumulative distribution function of S_n, due to Lugannani and Rice (1980), is given by

$$F_n(x) \approx \Phi(\hat{r}) + \phi(\hat{r}) \left\{\frac{1}{\hat{r}} - \frac{1}{\hat{q}}\right\}, \tag{3–27}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions with

$$\begin{aligned} \hat{r} &= \operatorname{sgn}\left(\hat{t}\right) \left[2\left\{\hat{t}x - K_n\left(\hat{t}\right)\right\}\right]^{\frac{1}{2}} \\ \hat{q} &= \hat{t}\left\{K_n''\left(\hat{t}\right)\right\}^{\frac{1}{2}}. \end{aligned}$$

In Eq. 3–27, at the mean of the distribution, i.e., at $x = E(S_n)$, $q = 0$, so that the alternate expression used is

$$F_n(x) = \frac{1}{2} + (72\pi)^{-\frac{1}{2}} K_n'''(0) K_n''(0)^{-\frac{3}{2}}, \tag{3–28}$$

where $K_n'''(\cdot)$ is the third derivative of the cumulant generating function $K_n(t)$.

Since the mgfs of NPEs of VaR and CVaR are available in closed form, the density and distribution functions of $\hat{\xi}_\alpha(x)$ and $\hat{\phi}_\alpha(x)$ can be approximated by applying the saddlepoint approximation. There is a problem in that both of the mgfs are mixtures of hypergeometric functions, which are computationally burdensome to evaluate explicitly due to slow convergence of the power series expansions defining the hypergeometric function. A computationally more efficient alternative is to employ instead the Laplace approximations of the hypergeometric function developed by Butler and Wood (2002).

3.3.3 Laplace Approximation of Hypergeometric Function

Butler and Wood(2002) provide a Laplace approximation for the hypergeometric function,

$$\hat{F}_{2,1}(a,b;c;x) = c^{c-1/2} r_{2,1}^{-1/2} \left(\frac{\hat{y}}{a}\right)^a \left(\frac{1-\hat{y}}{c-a}\right)^{c-a} (1-x\hat{y})^{-b}, \tag{3–29}$$

where

$$r_{2,1} = \frac{\hat{y}^2}{a} + \frac{(1-\hat{y})^2}{c-a} - \frac{bx^2}{(1-x\hat{y})^2} \frac{\hat{y}^2}{a} \frac{(1-\hat{y})^2}{c-a} \quad .$$

Let $g(y) = -\{alog(y) + (c-a)log(1-y) - blog(1-xy)\}$, $\hat{y}$ is the solution of $g'(y) = 0$,

$$\hat{y} = \frac{2a}{\sqrt{e^2 - 4ax(c-b)} - e} \quad ,$$

where $e = x(b-a) - c$.

Therefore, by applying this approximation we can estimate the distributions of $\hat{\xi}_\alpha(X)$ and $\hat{\phi}_\alpha(X)$ under AL law. To get a more accurate approximation, we normalized the cumulate generating function by subtracting its approximated value at $x = 0$.

Assessing the Accuracy of the Saddlepoint Approximations

In this section, we apply the same technique to estimate the accuracy of the saddlepoint approximation to the distribution of $\hat{\xi}_\alpha(X)$ and $\hat{\phi}_\alpha(X)$. We use the same parameters for simulations. And the percent relative error are defined the same as in Chapter 2.

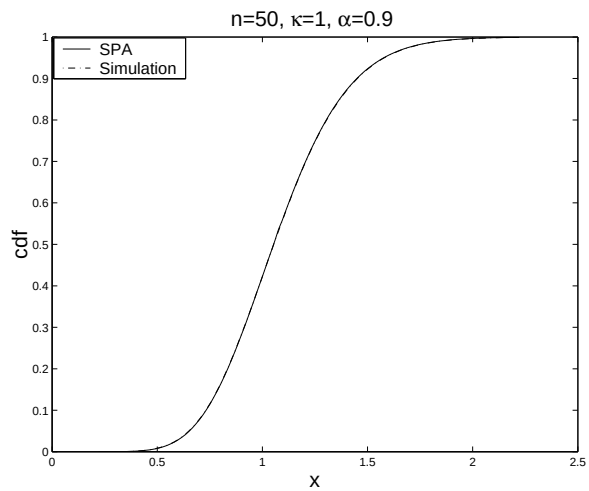

Figure 3-1. Estimated cdfs of $\hat{\xi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, respectively, under AL law with $n = 50, \theta = 0, \tau = 1, \kappa = 1, \alpha = 0.9$

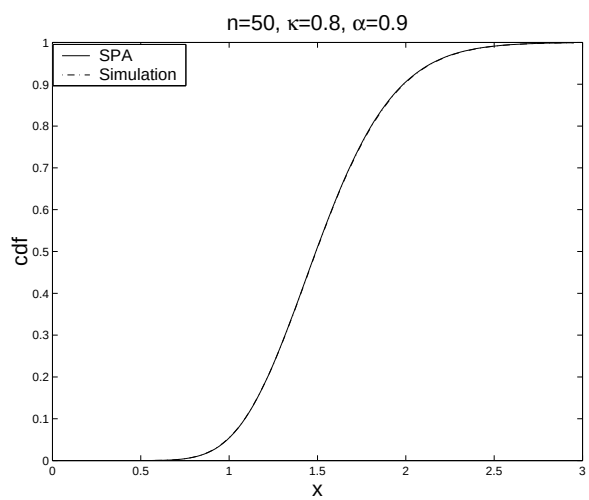

Figure 3-2. Estimated cdfs of $\hat{\xi}_\alpha(X)$, obtained via simulations and saddlepoint approximations, respectively, under AL law with $n = 50, \theta = 0, \tau = 1, \kappa = 0.8, \alpha = 0.9$

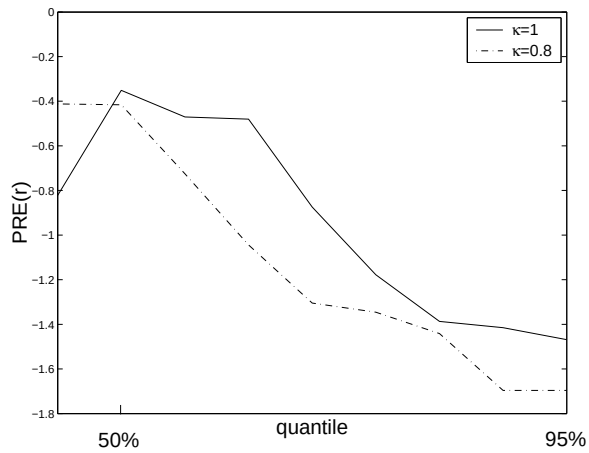

Figure 3-3. Percent relative errors (PREs)for the saddlepoint approximation to the distribution $\hat{\xi}_\alpha(X)$ under AL law with $n = 100, \theta = 0, \kappa = 1, 0.8, \tau = 1, \alpha = 0.9$, computed at the same quantile values.

In general, we have good estimation for NPEs of VaR and CVaR under AL distribution. The PREs for the cdfs of $\hat{\xi}_\alpha(X)$ are less than 2% and PREs for cdfs of $\hat{\phi}_\alpha(X)$ are between 2% to 14%. Note also, in Fig. 3-3, the PREs are less than zero, which indicates a tendency to overestimate $\hat{\xi}_\alpha(X)$ and $\hat{\phi}_\alpha(X)$.

3.4 Comparison of the Distributions of Parametric and Nonparametric Estimators

In this section we compare the MLEs and NPEs of VaR and CVaR under iid sampling from an Asymmetric Laplace law. We consider both the large sample case and finite sample case.

3.4.1 Large Sample Case

In light of the asymptotic normality results of Theorem 2.1 and Theorem 3.1, we compare the estimators of VaR and CVaR through the asymptotic relative efficiencies (AREs) of the MLE with respect to the NPE. If $X \sim \mathcal{AL}(0, \kappa, 1)$, routine calculations give for $\kappa^2/(1+\kappa^2) < \alpha < 1$, $f(\xi_\alpha) = \sqrt{2}\kappa(1-\alpha)$, and $\sigma_\alpha^2 + \alpha(\xi_\alpha - \phi_\alpha)^2 = (1+\alpha)/(2\kappa^2)$. In the general case of $Y = \tau X$ distributed as $\mathcal{AL}(\kappa, \tau)$, these results when substituted into Theorem 3.1 give $\sum(\hat{\xi}_\alpha) = (\alpha\tau^2)/[2\kappa^2(1-\alpha)] = \sum(\hat{\xi}_\alpha, \hat{\phi}_\alpha)$, and $\sum(\tilde{\phi}_\alpha) =$

$[(1+\alpha)\tau^2]/[2\kappa^2(1-\alpha)]$. Using Theorem 2.1, the ARE for the NPEs of VaR is,

$$\mathrm{ARE}(\tilde{\xi}_\alpha, \hat{\xi}_\alpha) = \frac{\sum(\tilde{\xi}_\alpha)}{\sum(\hat{\xi}_\alpha)} = \frac{1-\alpha}{2\alpha}\left[2\omega_{\alpha,\kappa}^2 + \kappa^2(1-\omega_{\alpha,\kappa}^2)\right], \tag{3–30}$$

while the ARE for the NPEs of CVaR is,

$$\mathrm{ARE}(\tilde{\phi}_\alpha, \hat{\phi}_\alpha) = \frac{\sum(\tilde{\phi}_\alpha)}{\sum(\hat{\phi}_\alpha)} = \frac{1-\alpha}{2(1+\alpha)}\left[2+4\kappa^2-4(1+\kappa^2)\omega_{\alpha,\kappa}+(2+\kappa^2)\omega_{\alpha,\kappa}^2\right], \tag{3–31}$$

where $\omega_{\alpha,\kappa}$ is defined the same as in Eq. 2–10.

Fig. 3-4 displays both AREs as a function of $\alpha \geq 0.5$ and $0 < \kappa \leq 1$.

3.4.2 Finite Sample Case

In this section, we compare performance of MLEs and NPEs of VaR and CVaR under an AL distribution in the finite sample case. Fig. 3-5 plot the saddlepoint approximated cdfs of MLEs and NPEs and compare the distributions with the true values of VaR and CVaR.

Note in general, the MLEs are more symmetric and unbiased, while NPEs are right skewed and biased. Also the NPEs are more skewed and biased as the AL distribution becomes more asymmetric.

3.5 Analysis of Exchange Rate Data

In this section, we analyze the exchange rate of the USD to the EUR. The data consists of the daily average "ask price" of 1 USD in EUR from Jan 31, 2005 to Jan 31, 2006 with 366 data points. The source of the data is from oanda.com.

We are interested in analyzing the natural logarithm of the price ratio for two consecutive days, and the data were transformed accordingly to give 365 daily log returns. Summary statistics are as follows: minimum -1.846E-2, median 0E-7, mean 2.049E-4, maximum 1.325E-2.

Fig. 3-6 gives the histogram of the data, box plots by day of the week, and the sample ACFs of the squares and absolute values of the data.

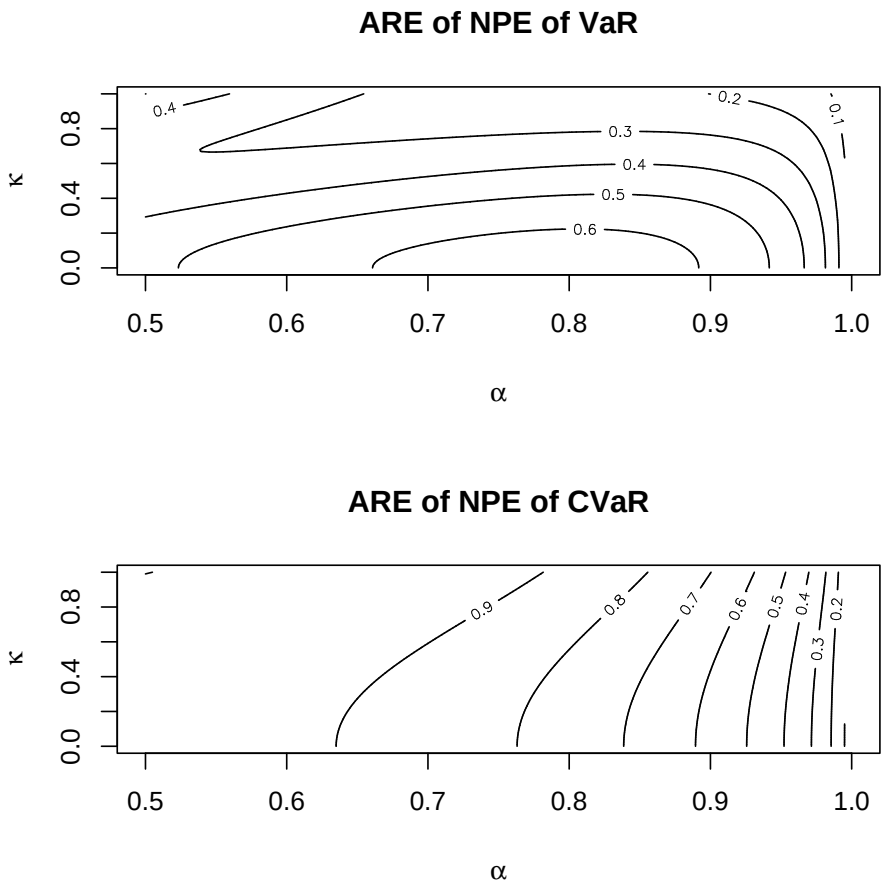

Figure 3-4. Asymptotic relative efficiencies of the maximum likelihood estimators with respect to the nonparametric estimators of VaR and CVaR, under iid sampling from the standard Asymmetric Laplace distribution. The AREs are plotted as a function of the left tail probability level $\alpha \geq 0.5$ and the skewness parameter $0 < \kappa \leq 1$.

The histogram of the log returns suggests high peakedness and heavy tails. The normal Q-Q plot of the data indicates a violation of normality. The Pearson Chi-square normality test reports a p-value of 2.2e-16. And the Kolmogorov-Smirov normality reports a p-value of 5.032e-14. This is strong evidence that the data do not follow a normal distribution.

We consider fitting the daily log returns data using an Asymmetric Laplace distribution. The dotted line superimposed on the histogram shows a fitted AL(0.9679, 4.436E-3) density with parameters estimated via maximum likelihood. Consequently, the

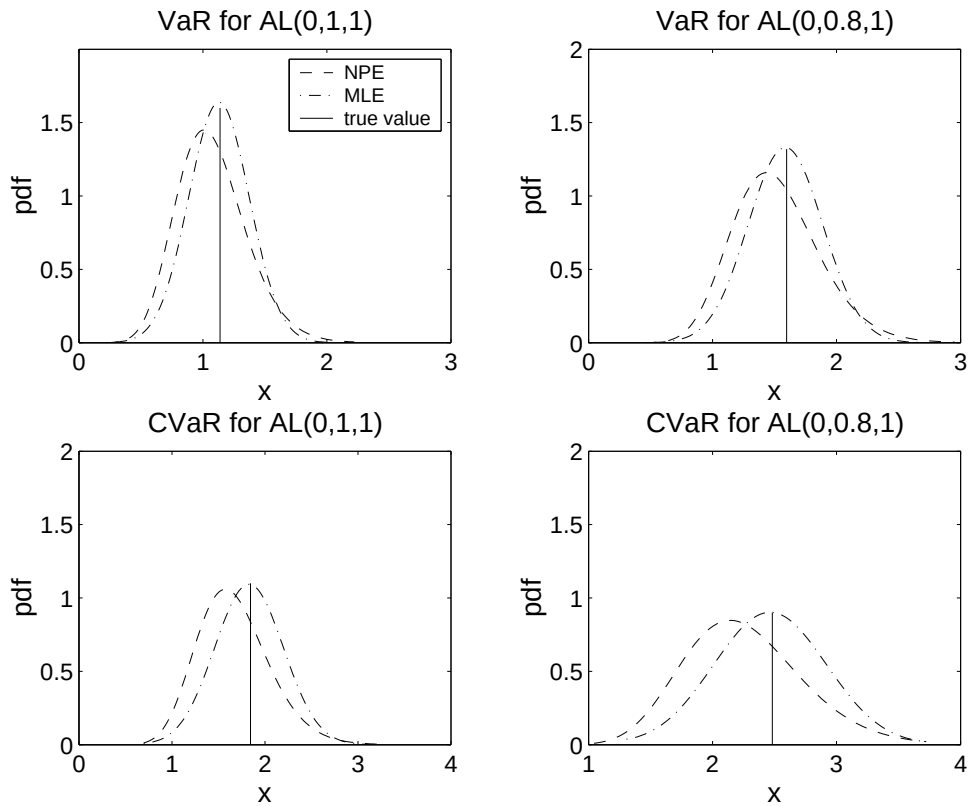

Figure 3-5. The resulting saddlepoint approximated density functions of MLEs and NPEs of VaR and CVaR at n=50, $\theta = 0, \tau = 1, \kappa = 1, 0.8, \alpha = 0.9$. The true values of VaR and CVaR are indicated by the solid vertical line.

MLE of VaR, $\tilde{\xi}_\alpha(X)$, equals 7.566E-3 and the MLE of CVaR $\tilde{\phi}_\alpha(X)$, equals 1.08E-2, with $\alpha = 0.95$.

It has long been noted that there is a "trading day effect" in currency exchange rate data (for example, McFarland et al., 1982). The boxplots in Fig. 3-6 decompose the 365 returns by day-of-the-week. The day-of-the-week corresponding to a particular return denotes the log of the ratio of the price on that day to that of the previous day. By inspection of the box plots we can see that there is a higher volatility during the week than on the weekends.

The other two plots are the sample autocorrelations for the squares and absolute values of the log returns. Although the 365 returns appear serially uncorrelated, the sample ACF suggests the presence of a dependence occurring precisely at lags that are multiples of 7. This is evidence of the presence of day-of-the-week effects.

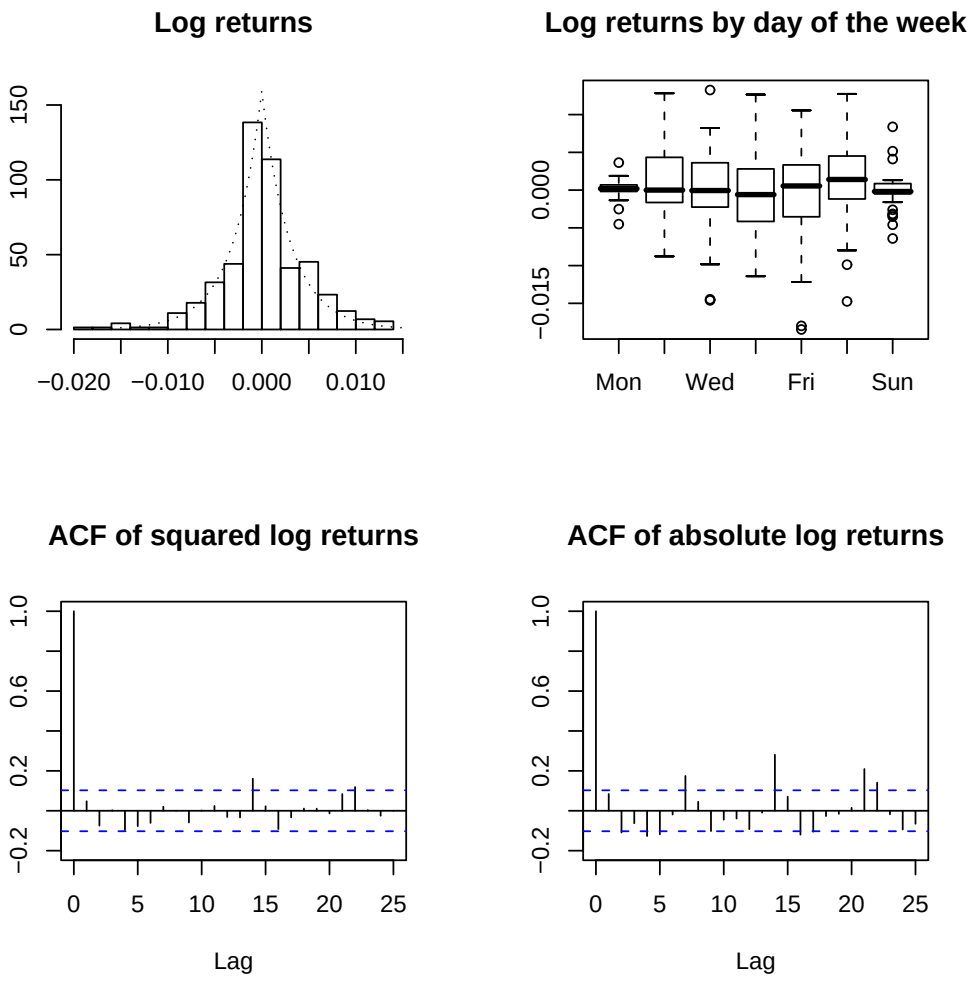

Figure 3-6. Daily log returns of USD/EUR exchange rates from Feb 1, 2005 to Jan 31, 2006. The top plots show the data both for the entire period and by day-of-the-week. The bottom plots display the sample autocorrelation function of the squares and absolute values of the data over the entire period.

We believe the reason for the low volatility during weekends is the lack of institutional investment during weekends. Most of the foreign exchange trades during weekends are over-the-counter (OTC) service by retail banks.

Next we removed the weekends from the data to see whether the daily log returns of weekdays follow a normal distribution. As a consequence, Monday's log return is the log ratio of Monday and the previous Friday's price. Fig. 3-8 is the histogram of daily log returns of the exchange rate date without the weekends. The data are less peaked but still have heavy tails. The Pearson Chi-square normality test reports a p-value of 2.2e-16. And

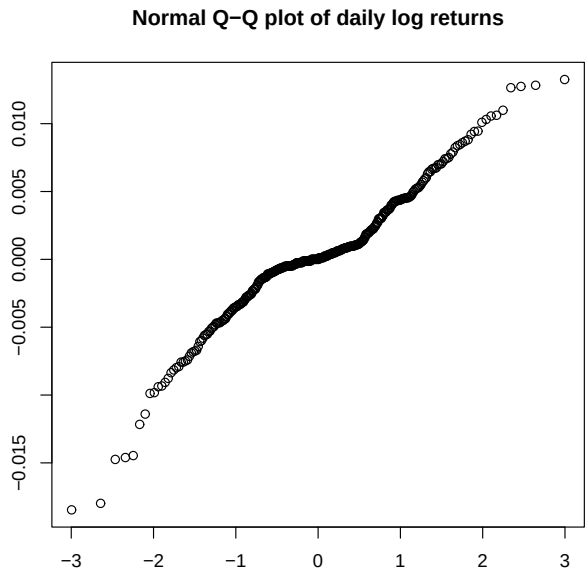

Figure 3-7. Normal Q-Q plot of daily log returns of exchange rate.

the Kolmogorov-Smirov normality reports a p-value of 1.011e-13. Therefore, the data are not normally distributed even when we remove the weekends.

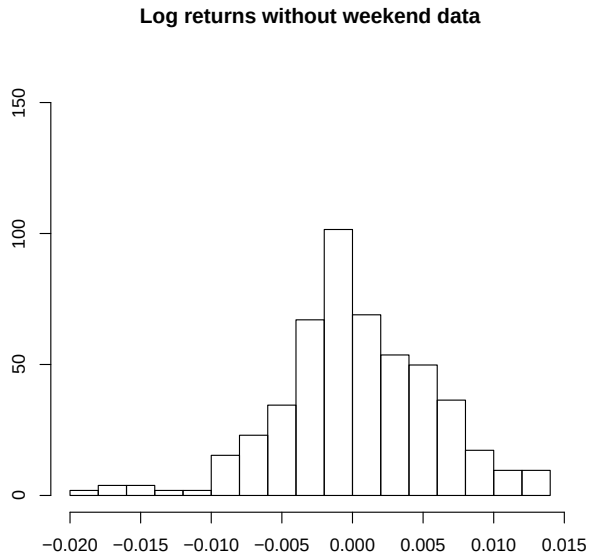

Figure 3-8. Histogram of the daily log returns of the exchange rate without weekends

On observing the "trading-day-effect", we therefore fit the $\mathcal{AL}(\kappa, \tau)$ distributions via maximum likelihood to each of the day-of-the-week returns.

The log returns on Wednesday, Friday and Sunday are approximately symmetric with fitted $\tilde{\kappa}$ equal to 0.9962,1.0218 and 1.032, respectively. While log returns on Monday,

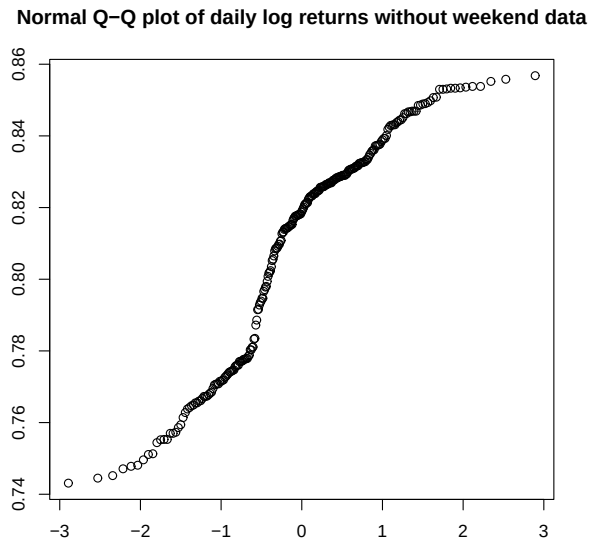

Figure 3-9. Normal Q-Q plot of daily log returns of exchange rate without weekends.

Table 3-1: Maximum likelihood estimates and standard errors (s.e.) for the skewness and scale parameters as an $\mathcal{AL}(\kappa, \tau)$ fit to the log returns of the USD/EUR exchange rates by week day, as well as the MLEs of VaR and CVaR at $\alpha = 0.95$.

Day	$\tilde{\kappa}$	(s.e.)	$\tilde{\tau}$	(s.e.)	$\tilde{\xi}_{0.95}$	(s.e.)	$\tilde{\phi}_{0.95}$	(s.e.)
Mon	0.897	(0.089)	9.987E-4	(1.389E-4)	1.892E-3	(0.0025)	2.680E-3	(0.003)
Tue	0.878	(0.086)	5.031E-3	(6.940E-4)	9.820E-3	(0.0131)	1.387E-2	(0.015)
Wed	0.996	(0.098)	5.196E-3	(7.206E-4)	8.507E-3	(0.012)	1.220E-2	(0.014)
Thu	1.117	(0.110)	5.523E-3	(7.683E-4)	7.645E-3	(0.0117)	1.114E-2	(0.014)
Fri	1.022	(0.100)	6.330E-3	(8.780E-4)	9.991E-3	(0.014)	1.144E-2	(0.017)
Sat	0.836	(0.083)	5.685E-3	(7.948E-4)	1.186E-2	(0.002)	1.667E-2	(0.003)
Sun	1.032	(0.101)	1.794E-3	(2.489E-4)	2.791E-3	(0.004)	4.021E-3	(0.005)

Tuesday and Saturday are right skewed with fitted $\tilde{\kappa}$ equal to 0.8973, 0.8782 and 0.8358, respectively. The $\tilde{\kappa}$ on Thursday is 1.1169, indicating left skewness.

The MLEs of VaR are much larger on Tuesday to Saturday than on Monday and Sunday. Consequently, the same trend is seen for the MLEs of CVaR.

Using the asymptotic covariance matrices in Eq. 2–6 and Eq. 3–4, it is interesting to compare confidence regions for the parametric and nonparametric estimators of the bivariate parameter $(\xi_\alpha, \phi_\alpha)$. Both regions take the form of ellipses; for example, with a confidence level of 95%, the region for the MLEs is given by

$$\left\{ (\xi_\alpha, \phi_\alpha) \in \mathrm{R}^2 : \left[\xi_\alpha - \tilde{\xi}_\alpha, \phi_\alpha - \tilde{\phi}_\alpha \right] \begin{bmatrix} \sum(\tilde{\xi}_\alpha) & \sum(\tilde{\xi}_\alpha, \tilde{\phi}_\alpha) \\ \sum(\tilde{\xi}_\alpha, \tilde{\phi}_\alpha) & \sum(\tilde{\phi}_\alpha) \end{bmatrix}^{-1} \begin{bmatrix} \xi_\alpha - \tilde{\xi}_\alpha \\ \phi_\alpha - \tilde{\phi}_\alpha \end{bmatrix} \leq \frac{\chi^2_{2,0.95}}{n} \right\}. \tag{3–32}$$

While the construction of this region is straightforward in the parametric case by simply "plugging-in" the MLEs themselves wherever they appear in the covariance matrix, that for the nonparametric case is complicated by the need to estimate the inverse of the pdf, a quantity sometimes called the *sparsity function*. Estimation of the sparsity is notoriously difficult and tends to be shied away from in favor of other approaches whenever it occurs (as it frequently does) in nonparametric inference. However, there has been recent renewed interest in this subject since the sparsity features prominently in the asymptotics of quantile regression (Koenker, 2005). Using the method suggested in Koenker (2005, Section 4.10.1) with the Hall and Sheather bandwidth, and plugging in the NPEs whenever they appear in the covariance matrix, nonparametric confidence region construction for $(\xi_\alpha, \phi_\alpha)$ is therefore a feasible proposition.

We employed the above approach in producing Fig. 3-10, which shows the resulting confidence ellipses for the MLEs and NPEs of the $\alpha = 0.9$ right tail (VaR, CVaR) for the week day distribution of the USD/EUR exchange rate log returns data. The high correlation between the MLEs of VaR and CVaR is reflected in the very narrow semiminor axes of each respective ellipse. The downward bias in the NPEs is also immediately apparent, a fact that concurs with the bias noted in the saddlepoint pdf of Fig. 3-5.

The implication of these findings for the practitioner is that it may be preferable to commit to an appropriate parametric model, such as the AL law, when attempting to draw inferences from data of this nature. From a small simulation study which we omit for the sake of brevity, we have also noted that the parametric confidence bands have vastly superior coverage probabilities.

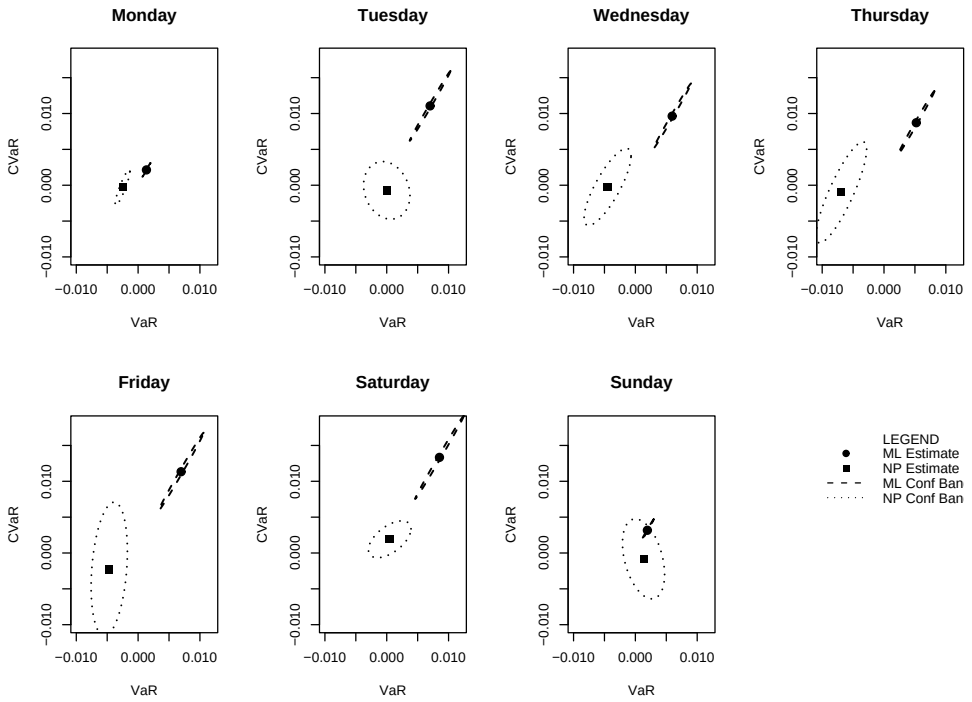

Figure 3-10. Maximum likelihood (circle) and nonparametric (square) bivariate estimators of (VaR, CVaR) for the $\alpha = 0.9$ tail of the week day distributions of the USD/EUR exchange rate log return data. The dashed and dotted lines delineate the boundary of 95% confidence ellipses for the maximum likelihood and nonparametric estimators, respectively.

CHAPTER 4
TIME SERIES ARMA AND GARCH MODELS UNDER AL NOISE

In this chapter, we develop the time series models under AL noise. Traditional time series ARMA models assume Gaussian noise, which implies that the marginal distribution will be Gaussian as well. We have noticed before that financial data are usually heavy tailed which is not consistent with a Gaussian distribution. It is therefore reasonable to extend the AL distribution to time series models.

Early works have applied the Symmetric Laplace distribution to ARMA models. These efforts have focused on two directions: assume marginal Symmetric Laplace distribution, for example, NLAR(1) and NLAR(2) models (Dewald and Lewis, 1985), NAREX(1) model (Novković, 1998), or assume Symmetric Laplace noise, for example, Damsleth and El-Shaarawi (1989). Damsleth and El-Shaarawi (1989) have shown that these two requirements cannot be simultaneously achieved within the class of linear time series models.

Dewald and Lewis (1985) discuss the NLAR(1) and NLAR(2) model assuming a standard Symmetric Laplace marginal distribution. The Symmetric Laplace distribution is a special case of AL distribution with $\kappa = 1$. A Symmetric Laplace distribution is called a standard Symmetric Laplace distribution when $\tau = 1$ and $\theta = 0$.

Random variable X is said to be distributed as Symmetric Laplace distribution (double exponential distribution) with location parameter $-\infty < \theta < \infty$ and scale parameter $\lambda > 0$, if its pdf is of the form

$$f_X(x) = \frac{1}{2\lambda} e^{-\frac{|x-\theta|}{\lambda}}, \qquad -\infty < x < \infty. \tag{4–1}$$

A Symmetric Laplace distribution are called standard Laplace distribution if $\lambda = 1$.

NLAR(1) model starts by assuming $\{X_n\}$ to be a stationary process with standard Laplace marginal distribution, $0 < |\phi| < 1$ and $0 < \alpha < 1$,

$$X_t = \begin{cases} \phi X_{t-1} + Z_t, & \text{w.p.} \quad \alpha, \\ Z_t & \text{w.p.} \quad 1-\alpha. \end{cases} \tag{4–2}$$

Then the noise term can be derived as

$$Z_t = \begin{cases} L_t, & \text{w.p.} \quad 1-p, \\ \sqrt{1-\alpha}|\phi|L_t & \text{w.p.} \quad p, \end{cases} \tag{4–3}$$

where L_t are i.i.d. standard Laplace random variables. And $p = \frac{\alpha\phi^2}{1-(1-\alpha)\phi^2}$. In addition, let $\lambda = (1-\alpha)^{-1/2}|\phi|^{-1}$, the density function of Z_t is

$$f_{Z_t}(x) = \frac{1}{2}(1-p)e^{-|x|} + \frac{1}{2}\lambda p e^{\lambda|x|}, \tag{4–4}$$

which is a convex mixture of Laplace densities.

Similarly, the NLAR(2) model assumes standard Laplace distribution and applies this to the AR(2) model,

$$X_t = \phi_1 K_t' X_{t-1} + \phi_2 K_t'' X_{t-2} + Z_t, \tag{4–5}$$

where $0 < |\phi_i| < 1$, for $i = 1, 2$. $\{K_t', K_t''\}$ is a sequence of i.i.d. discrete bivariate random variables with distribution,

$$\{K_t', K_t''\} = \begin{cases} (1,0), & \text{w.p.} \ \ \alpha_1, \\ (0,1), & \text{w.p.} \ \ \alpha_2, \\ (0,0), & \text{w.p.} \ \ 1-\alpha_1-\alpha_2, \end{cases} \tag{4–6}$$

for $t = 0, \pm 1, \pm 2, \cdots$; $0 < \alpha_i < 1$ for $i = 1, 2$ and $\alpha_1 + \alpha_2 < 1$.

Therefore the noise term can be expressed as

$$Z_t = \begin{cases} L_t, & \text{w.p.} \ \ 1-p_2-p_3, \\ |b_2|L_t, & \text{w.p.} \ \ p_2, \\ |b_3|L_t & \text{w.p.} \ \ p_3, \end{cases} \tag{4–7}$$

where $\{L_t\}$ are i.i.d. standard Laplace random variable; p_2, p_3, b_2, b_3 are functions of $\alpha_1, \alpha_2, \phi_1, \phi_2$.

The NAREX(1) model discussed by Novković (1998) assumes the marginal distribution to be a Symmetric Laplace distribution $L(\lambda)$, as defined in Eq. 4–1, with scale parameter λ,

$$X_t = \begin{cases} \phi_1 X_{t-1} + Z_t, & \text{w.p. } p_0, \\ \phi_2 X_{t-1} + Z_t, & \text{w.p. } p_1, \\ \phi_3 X_{t-1} & \text{w.p. } p_2, \end{cases} \tag{4–8}$$

where $0 \leq p_0, p_1, p_2 \leq 1$, $p_0 + p_1 + p_2 = 1$, $0 < \phi_1, \phi_2, \phi_3 < 1$. After working with the characteristic function of the X_t and Z_t, Z_t can be expressed as a mixture of symmetric Laplace distribution,

$$Z_t = \begin{cases} 0, & \text{w.p. } A_0, \\ L(\lambda), & \text{w.p. } A_1, \\ L(\phi_3 \lambda), & \text{w.p. } A_2, \\ L(\lambda\sqrt{\frac{p_0\phi_2^2 + p_1\phi_1^2}{p_0+p_1}}), & \text{w.p. } \frac{A_3}{p_0+p_1}, \end{cases} \tag{4–9}$$

where A_0, A_1 and A_2 are functions of $\phi_1, \phi_2, \phi_3, p_0, p_1, p_2$.

Damsleth and El-Shaarawi (1989) deduce the marginal distribution of observations generated by an ARMA model assuming Symmetric Laplace noise.

Let $\{Z_t\}$ be a series of i.i.d. Symmetric Laplace distributed random variables with scale parameter λ. Let the observed stationary time series $\{X_t\}$ be generated by the ARMA scheme,

$$\Phi(B)X_t = \Theta(B)Z_t, \tag{4–10}$$

the marginal pdf X_t is

$$f_{X_t}(x) = \frac{1}{2\lambda} \sum_{j=0}^{\infty} \alpha_j |\psi_j|^{-1} \exp\left(-\frac{1}{\lambda}\left|\frac{x}{\psi_j}\right|\right), \tag{4–11}$$

Where the α_j are given by $\alpha_j = \Pi_{\substack{i=0 \\ i\neq j}}^{n}(1 - |\frac{\psi_i}{\psi_j}|)^2$.

On observing the asymmetry property of financial data, it is reasonable to apply the Asymmetric Laplace distribution to time series modeling.

4.1 ARMA (p, q) Model

Definition 4.1: (ARMA(p, q) process).

$\{X_t\}$ is an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \lambda_1 Z_{t-1} + \cdots + \lambda_q Z_{t-q}, \tag{4–12}$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and the polynomial $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \lambda_1 z + \cdots + \lambda_q z^q)$ have no common factors.

The process $\{X_t\}$ is said to be an ARMA(p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process.

Eq. 4–12 can also be expressed as

$$\Phi(B)X_t = \Lambda(B)Z_t, \tag{4–13}$$

where $\Phi(\cdot)$ and $\Lambda(\cdot)$ are the pth and qth-degree polynomials

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\Lambda(z) = 1 + \lambda_1 z - \cdots + \lambda_p z^p$$

and B is the backward shift operator ($B^j X_t = X_{t-j}, B^j Z_t = Z_{t-j}, j = 0, \pm, \ldots$).

The time series $\{X_t\}$ is said to be an *autoregressive process of order* p (or AR(p)) if $\Lambda(z) \equiv 1$, and a *moving-average process of order* q (or MA(q)) if $\Phi(z) \equiv 1$.

$\{X_t\}$ is assumed to be stationary, which means that the autoregressive polynomial $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for all complex z with $|z| = 1$.

If $\Phi(z) \neq 0$ for all z on the unit circle, then there exists $\delta > 0$ such that

$$\frac{1}{\phi(z)} = \sum_{j=-\infty}^{\infty} \chi_j z^j \quad \text{for} \quad 1 - \delta < |z| < 1 + \delta, \tag{4–14}$$

and $\sum_{j=-\infty}^{\infty} |\chi_j| < \infty$. Therefore, we can define

$$\frac{1}{\Phi(B)} = \sum_{j=-\infty}^{\infty} \chi_j B^j. \tag{4–15}$$

Applying the operator $\chi(B) = \frac{1}{\Phi(B)}$ to both sides of Eq. 4–13, we obtain

$$X_t = \chi(B)\Phi(B)X_t = \chi(B)\Lambda(B)Z_t = \Psi(B)Z_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \tag{4–16}$$

where $\psi(z) = \chi(z)\lambda(z) = \sum_{j=-\infty}^{\infty} \psi_j Z^j$.

Definition 4.2: (Causality).

An ARMA(p, q) process $\{X_t\}$ is causal, or a causal function of $\{Z_t\}$, if there exist constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \text{for all t}, \tag{4–17}$$

causality is equivalent to the condition

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad \text{for all} \quad |z| \leq 1. \tag{4–18}$$

Definition 4.3: (Invertibility).

An ARMA(p, q) process $\{X_t\}$ is invertible if there exist constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad \text{for all t}, \tag{4–19}$$

invertibility is equivalent to the condition

$$\lambda(z) = 1 + \lambda_1 z + \cdots + \lambda_p z^p \neq 0 \quad \text{for all} \quad |z| \leq 1. \tag{4–20}$$

4.2 ARMA(p,q) Model under AL Noise

Traditional analysis of ARMA models assumes that the white noise term $\{Z_t\}$ is a series of i.i.d. normal distributions, which we already know is not appropriate for fitting financial data. To apply an ARMA model to financial time series data, we introduce the ARMA model under AL noise.

Definition 4.4: (ARMA(p,q) process under AL noise).

Let $\{Z_t\}$ be a series of i.i.d. Asymmetric Laplace distributed random variables $\mathcal{AL}(\theta,\kappa,\tau)$, with PDF given by

$$f_{Z_t}(z) \;=\; \frac{\sqrt{2}\kappa}{\tau(1+\kappa^2)}\left\{\exp\left(-\frac{\sqrt{2}\kappa}{\tau}|z-\theta|I_{[z\geq\theta]}\right)+\exp\left(-\frac{\sqrt{2}}{\tau\kappa}|z-\theta|I_{[z<\theta]}\right)\right\},$$

$\{X_t\}$ is an ARMA(p,q) process under AL noise if $\{X_t\}$ is stationary and if for every t,

$$X_t-\phi_1X_{t-1}-\cdots-\phi_pX_{t-p}=Z_t+\lambda_1Z_{t-1}+\cdots+\lambda_qZ_{t-q}, \tag{4–21}$$

the polynomial $(1-\phi_1z-\cdots-\phi_pz^p)$ and $(1+\lambda_1z+\cdots+\lambda_qz^q)$ have no common factors.

To ensure $\{Z_t\}$ has zero mean, we require that

$$\theta\equiv\frac{\tau}{\sqrt{2}}\left(\frac{1}{\kappa}-\kappa\right). \tag{4–22}$$

Motivation of our model. Our defined model has the flexibility to describe a stationary time series process with asymmetric turbulence about the mean. Asymmetry is an important property of financial data, for example, the cost to buy a call option is limited but the return can be substantial. An improvement in credit quality brings limited returns to investors, but in case of defaults or downgrades, the loss would be substantial. Again, our model has the flexibility to describe data with heavy tails.

We simulated an ARMA(1,1) model under AL noise with $\kappa>1$, which means that the AL distribution is left skewed. Consequently, the ARMA process is skewed to the downside of the mean. There are some deep drops approximately at time t=140, 180, 300.

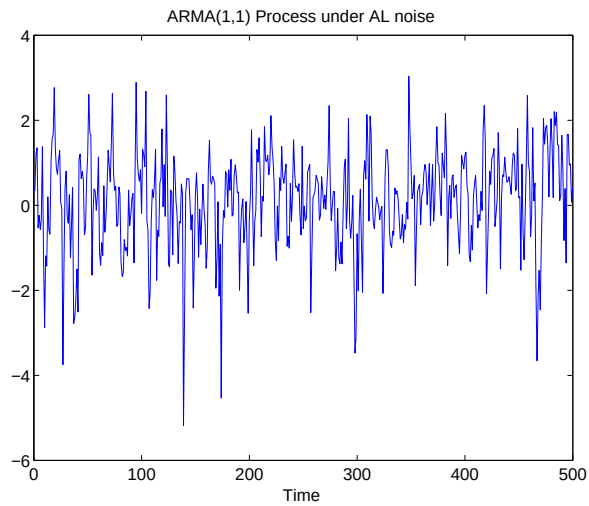

Figure 4-1. Simulated ARMA(1,1) process under AL noise with $\phi_1 = 0.4$, $\lambda_1 = 0.1$, $\kappa = 1.25$, $\tau = 1$.

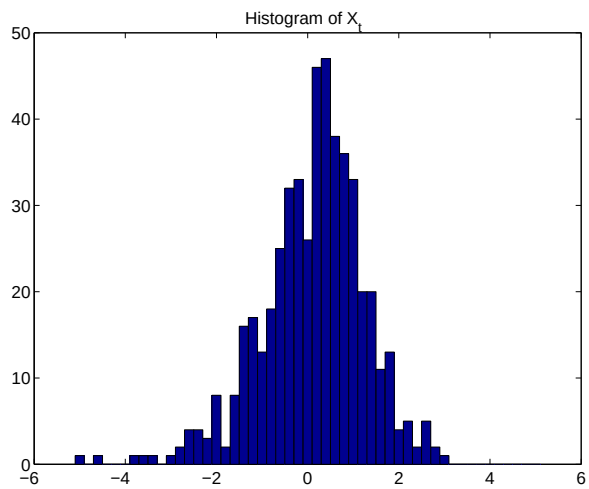

Figure 4-2. Histogram of the Simulated ARMA(1,1) process under AL noise with $\phi_1 = 0.4$, $\lambda_1 = 0.1$, $\kappa = 1.25$, $\tau = 1$.

These drops are more frequent and deeper compared with the upper points. This model is good to describe markets with unexpected and sudden losses, for example, a market crash. Similarly, if we set the skewness parameter $\kappa < 0$, the process will be skewed to the upper side.

The histogram of the $\{X_t\}$ indicates the distribution has heavy tails.

4.2.1 Marginal Distribution of ARMA Model under AL Noise

We derive the pdf of the marginal distribution of an ARMA(p, q) model with Asymmetric Laplace noise. Similar method have been used by Damsleth and EL-Shaarawi (1989) in deriving the marginal distribution of an ARMA model with double-exponential noise. This method is also a special case of Box (1954), where he derives the distribution of any linear combination of independent χ^2 variables with even degrees of freedom.

Characteristic function of AL noise. The characteristic function of $Z_t \sim \mathcal{AL}(\theta, \kappa, \tau)$, is

$$\varphi_{Z_t}(t) = \frac{e^{i\theta t}}{1 + \frac{1}{2}\tau^2 t^2 - i\frac{\tau}{\sqrt{2}}(\frac{1}{\kappa} - \kappa)t}, \tag{4–23}$$

or equivalently,

$$\varphi_{Z_t}(t) = \frac{e^{i\theta t}}{1 + \frac{1}{2}\tau^2 t^2 - i\mu t}, \tag{4–24}$$

where $\mu = \frac{\tau}{\sqrt{2}}(\frac{1}{\kappa} - \kappa)$.

We assume $\{X_t\}$ to be causal, then

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}. \tag{4–25}$$

Let

$$U_n = \sum_{j=0}^{n} \psi_j Z_{t-j} \tag{4–26}$$

and assume $\psi_i \neq \psi_j$ for $i \neq j$, since $Z_1, Z_2, \ldots, Z_n$ are i.i.d., the characteristic function of U_n is

$$\varphi_{U_n}(t) = \prod_{j=0}^{n} \frac{e^{i\theta\psi_j t}}{1 + \frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t}, \tag{4–27}$$

which can be resolved into partial fractions

$$\varphi_{U_n}(t) = \prod_{j=0}^{n} \frac{e^{i\theta\psi_j t}}{1 + \frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t} = \sum_{j=0}^{n} a_j \frac{e^{i\theta\psi_j t}}{1 + \frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t}, \tag{4–28}$$

where the a_j are constants not containing t. We will give an explicit expression for a_j in the next section.

Solve for constant a_j.

The value of the constants a_j can be obtained as follows:

Let $y = 1 + \frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t$, solving this equation, we get

$$t = \frac{i(\mu \pm \text{sign}(\psi_j)\sqrt{\mu^2 + 2\tau^2(1-y)})}{\tau^2\psi_j} =: f_j(y). \tag{4–29}$$

Mathematically, both plus and minus signs in Eq. 4–29 are correct. But we will see in the next step, that they will not necessarily make our derived function a pdf. We identify the correct signs by plotting the pdfs. Actually, the plus sign is correct for the part of pdf below the mean, and the minus sign is correct for the part of pdf above the mean.

Setting $y = 0$, we get

$$f_j(0) = \frac{i(\mu \pm \text{sign}(\psi_j)\sqrt{\mu^2 + 2\tau^2})}{\tau^2\psi_j}. \tag{4–30}$$

Eq. 4–28 can be expressed as

$$\begin{aligned} &\frac{e^{i\theta\psi_j t}}{1 + \frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t} \prod_{\substack{i=0 \\ i\neq j}}^{n} \frac{e^{i\theta\psi_i t}}{1 + \frac{1}{2}\tau^2\psi_i^2 t^2 - i\mu\psi_i t} \\ = \; & a_j \frac{e^{i\theta\psi_j t}}{1 + \frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t} + \sum_{\substack{i=0 \\ i\neq j}}^{n} a_i \frac{e^{i\theta\psi_i t}}{1 + \frac{1}{2}\tau^2\psi_i^2 t^2 - i\mu\psi_i t}, \end{aligned} \tag{4–31}$$

substituting y into Eq. 4–31,

$$\begin{aligned} &\frac{e^{i\theta\psi_j f_j(y)}}{y} \prod_{\substack{i=0 \\ i\neq j}}^{n} \frac{e^{i\theta\psi_i f_j(y)}}{1 + \frac{1}{2}\tau^2\psi_i^2 f_j^2(y) - i\mu\psi_i f_j(y)} \\ = \; & a_j \frac{e^{i\theta\psi_j f_j(y)}}{y} + \sum_{\substack{i=0 \\ i\neq j}}^{n} a_i \frac{e^{i\theta\psi_i f_j(y)}}{1 + \frac{1}{2}\tau^2\psi_i^2 f_j^2(y) - i\mu\psi_i f_j(y)}, \end{aligned} \tag{4–32}$$

multiplying by y on both sides,

$$
\begin{aligned}
& e^{i\theta\psi_j f_j(y)} \prod_{\substack{i=0\\ i\neq j}}^{n} \frac{e^{i\theta\psi_i f_j(y)}}{1+\frac{1}{2}\tau^2\psi_i^2 f_j^2(y) - i\mu\psi_i f_j(y)} \\
= \quad & a_j e^{i\theta\psi_j f_j(y)} + y \sum_{\substack{i=0\\ i\neq j}}^{n} a_i \frac{e^{i\theta\psi_i f_j(y)}}{1+\frac{1}{2}\tau^2\psi_i^2 f_j^2(y) - i\mu\psi_i f_j(y)}, \qquad (4\text{–}33)
\end{aligned}
$$

setting $y = 0$, we get

$$
a_j = \prod_{\substack{i=0\\ i\neq j}}^{n} \frac{e^{i\theta\psi_i f_j(0)}}{1+\frac{1}{2}\tau^2\psi_i^2 f_j^2(0) - i\mu\psi_i f_j(0)}. \qquad (4\text{–}34)
$$

Marginal pdf of X_t . Apply the Fourier inversion formula,

$$
\begin{aligned}
f_{U_n}(u) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_{U_n}(t) e^{-itu} dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \prod_{j=0}^{n} \frac{e^{i\theta\psi_j t}}{1+\frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t} e^{-itu} dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \sum_{j=0}^{n} a_j \frac{e^{i\theta\psi_j t}}{1+\frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t} e^{-itu} dt \\
&= \sum_{j=0}^{n} a_j \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\theta\psi_j t}}{1+\frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t} e^{-itu} dt \right], \qquad (4\text{–}35)
\end{aligned}
$$

where we recognize $\frac{e^{i\theta\psi_j t}}{1+\frac{1}{2}\tau^2\psi_j^2 t^2 - i\mu\psi_j t}$ as the characteristic funtion of $Z_j' \sim \mathcal{AL}(\psi_j\theta, \psi_j\mu, |\psi_j|\tau)$.

Let $\theta_j' = \psi_j\theta$, $\mu_j' = \psi_j\mu$ and $\tau_j' = |\psi_j|\tau$, since $\kappa = \frac{\sqrt{2}\tau}{\mu+\sqrt{2\tau^2+\mu^2}}$, the skewness parameter of Z_j', κ_j' can be obtained as

$$
\begin{aligned}
\kappa_j' &= \frac{\sqrt{2}\tau_j'}{\mu_j' + \sqrt{2\tau_j'^2 + \mu_j'^2}} \\
&= \frac{\sqrt{2}|\psi_j|\tau}{\psi_j\mu + \sqrt{2\psi_j^2\tau^2 + \psi_j^2\mu^2}} \\
&= \frac{\sqrt{2}|\psi_j|\tau}{\psi_j\mu + |\psi_j|\sqrt{2\tau^2 + \mu^2}} \\
&= \frac{\sqrt{2}\tau}{\mathrm{sign}(\psi_j)\mu + \sqrt{2\tau^2 + \mu^2}}. \qquad (4\text{–}36)
\end{aligned}
$$

Note that $\mu = \frac{\tau}{\sqrt{2}}(\frac{1}{\kappa} - \kappa)$, therefore

$$\begin{aligned}
\kappa'_j &= \frac{\sqrt{2}\tau}{\text{sign}(\psi_j)\frac{\tau}{\sqrt{2}}(\frac{1}{\kappa} - \kappa) + \sqrt{2\tau^2 + \frac{\tau^2}{2}(\frac{1}{\kappa} - \kappa)^2}} \\
&= \frac{\sqrt{2}}{\text{sign}(\psi_j)\frac{1}{\sqrt{2}}(\frac{1}{\kappa} - \kappa) + \frac{1+\kappa^2}{\sqrt{2}\kappa}} \\
&= \frac{2}{\text{sign}(\psi_j)(\frac{1}{\kappa} - \kappa) + (\frac{1}{\kappa} + \kappa)}.
\end{aligned} \tag{4–37}$$

If $\psi_j \geq 0$, $\kappa'_j = \kappa$; If $\psi_j < 0$, $\kappa'_j = \frac{1}{\kappa}$.

Consequently, the pdf of Z'_j is

$$f_{Z'_j}(z) = \frac{\sqrt{2}\kappa'_j}{|\psi_j|\tau(1+\kappa'^2_j)}\left\{\exp\left(-\frac{\sqrt{2}\kappa'_j}{|\psi_j|\tau}|z - \psi_j\theta|I_{[z\geq\psi_j\theta]}\right) + \exp\left(-\frac{\sqrt{2}}{|\psi_j|\tau\kappa'_j}|z - \psi_j\theta|I_{[z<\psi_j\theta]}\right)\right\}. \tag{4–38}$$

Back to Eq. 4–35,

$$\begin{aligned}
f_{U_n}(u) &= \sum_{j=0}^{n} a_j \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\theta\psi_j t}}{1 + \frac{1}{2}\tau^2\psi^2 t^2 - i\mu\psi_j t} e^{-itu} dt\right] \\
&= \sum_{j=0}^{n} a_j \left[\frac{1}{2\pi}\varphi_{Z'_j}(t) e^{-itu} dt\right] \\
&= \sum_{j=0}^{n} a_j \frac{\sqrt{2}\kappa'_j}{|\psi_j|\tau(1+\kappa'^2_j)}\left\{\exp\left(-\frac{\sqrt{2}\kappa'_j}{|\psi_j|\tau}|u - \psi_j\theta|I_{[u\geq\psi_j\theta]}\right)\right. \\
&+ \left.\exp\left(-\frac{\sqrt{2}}{|\psi_j|\tau\kappa'_j}|u - \psi_j\theta|I_{[u<\psi_j\theta]}\right)\right\},
\end{aligned} \tag{4–39}$$

where

$$a_j = \prod_{\substack{i=0 \\ i\neq j}}^{n} \frac{e^{i\theta\psi_i f_j(0)}}{1 + \frac{1}{2}\tau^2\psi_i^2 f_j^2(0) - i\mu\psi_i f_j(0)}. \tag{4–40}$$

Now,

$$f(x) = \lim_{n\to\infty} f_n(x). \tag{4–41}$$

Then we obtain the following expression for the marginal pdf of X_t,

$$\begin{aligned} f_{X_t}(x) &= \sum_{j=0}^{\infty} a_j \frac{\sqrt{2}\kappa'_j}{|\psi_j|\tau(1+\kappa_j'^2)} \left\{ \exp\left(-\frac{\sqrt{2}\kappa'_j}{|\psi_j|\tau} |x - \psi_j\theta| I_{[x \geq \psi_j\theta]} \right) \right. \\ &+ \left. \exp\left(-\frac{\sqrt{2}}{|\psi_j|\tau\kappa'_j} |x - \psi_j\theta| I_{[x < \psi_j\theta]} \right) \right\}, \end{aligned} \tag{4–42}$$

where

$$a_j = \prod_{\substack{i=0 \\ i \neq j}}^{\infty} \frac{e^{i\theta\psi_i f_j(0)}}{1 + \frac{1}{2}\tau^2\psi_i^2 f_j^2(0) - i\mu\psi_i f_j(0)}. \tag{4–43}$$

To check the validation of our derived formula, we plot the marginal pdf of AR(1) and ARMA(2,2) models under AL noise. The reason for selecting an AR(1) is because it is a typical model in fitting financial data. We compare our theoretical curve with simulated data. It turns out that our derived curve matches the simulated data perfectly.

Below are some remarks concerning the plotting of the PDF and data simulation.

Remark 1: Our derived marginal pdf is an infinite summation. But in practice, this series converges very fast. The convergence is slower for higher order ARMA models. But generally, it converges within $j = 25$. So we actually use a finite summation.

Remark 2: In Eq. 4–16, $\Psi(B) = \Phi^{-1}(B)\Theta(B)$ can be obtained by expanding $\Phi^{-1}(B)$ using Taylor series expansion, and then calculate the product of $\Phi^{-1}(B)$ and $\Theta(B)$. Some computer programs with symbolic mathematical functions, like Matlab, provide functions for Taylor series expansion and polynomial convolution.

Remark 3: When we generate a simulated realization from an ARMA(p, q) model with a sample size of n, we use zeros for the unknown values $X_1, \ldots, X_p$ in the ARMA(p, q) recursions, run the recursions well beyond n to $n + 500$, and then take the last n values as our sample. By doing this, we diminish the initialization error of the recursions due to the use of zeros.

Fig. 4-3 and Fig. 4-5 are the plots of the derived PDFs of AR(1) and ARMA(2,2). We selected AL noise with $\tau = 1$, $\kappa = 0.8$. To ensure a zero mean noise, θ is predetermined

by $\theta = -\mu = -\frac{\tau}{\sqrt{2}}(\frac{1}{\kappa} - \kappa)$. In Fig. 4-4 and Fig. 4-6, we compare the derived PDFs with the simulated histograms.

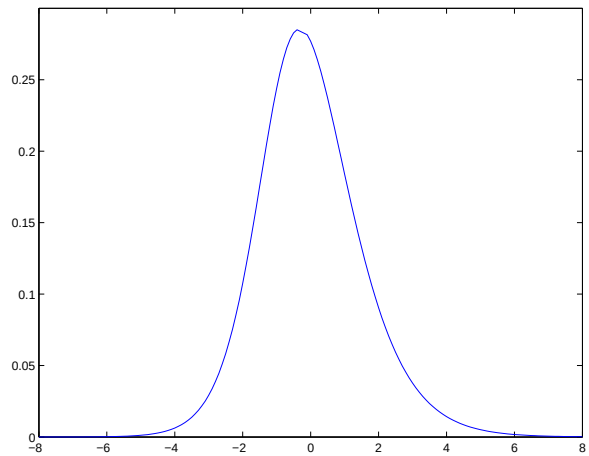

Figure 4-3. Derived marginal pdf of AR(1) model under Asymmetric Laplace noise with $\tau = 1$, $\kappa = 0.8$, $\theta = -0.318$ and $\phi = 0.75$.

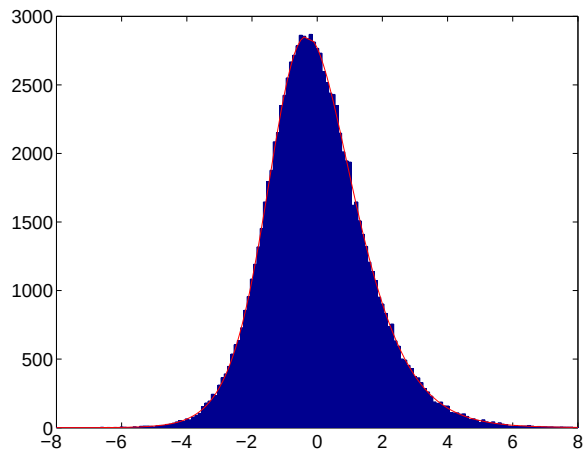

Figure 4-4. Comparison of derived marginal pdf (red line) and simulated histogram of AR(1) model. The theoretic pdf has been inflated by number of replications to match the histogram. $\tau = 1$, $\kappa = 0.8$, $\theta = -0.318$ and $\phi = 0.75$.

4.2.2 Fit AR(p) Model Using Conditional Maximum Likelihood Estimation

Proposition 4.1 (Joint distribution of $(x_1, \ldots, x_n)$).

Let $f(x_1, \ldots, x_n)$ denote the pdf of the joint distribution of the first n observations, $X_1, \ldots, X_n$, under an AR(p) model. Then the likelihood can be written as

$$f(x_1, \ldots, x_n) = f(x_1, \ldots, x_p) f(x_{p+1}|x_t, t \leq p) \cdots f(x_n|x_t, t \leq n-1). \tag{4–44}$$

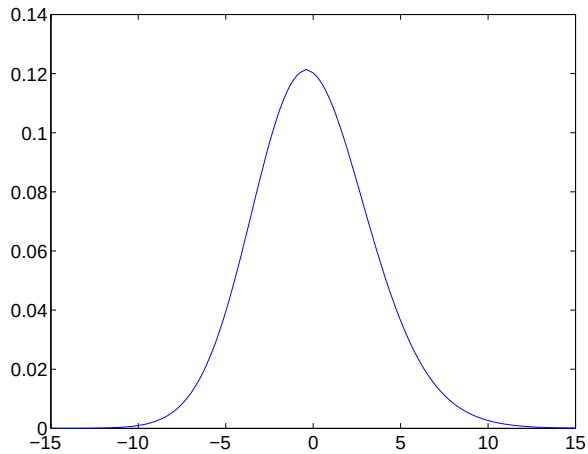

Figure 4-5. Derived marginal pdf of ARMA(2,2) model under Asymmetric Laplace noise.$\tau = 1$, $\kappa = 0.8$, $\theta = -0.318$, $\phi_1 = 0.7$, $\phi_2 = 0.2$, $\lambda_1 = 0.4$, $\lambda_2 = 0.2$, which is causal and invertible.

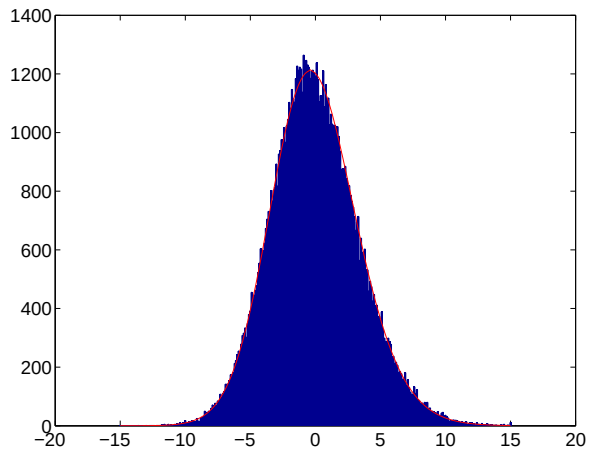

Figure 4-6. Comparison of theoretic marginal pdf (red line) and simulated histogram of ARMA(2,2) model. The theoretic pdf has been inflated by number of replications to match the histogram. $\tau = 1$, $\kappa = 0.8$, $\theta = -0.318$, $\phi_1 = 0.7$, $\phi_2 = 0.2$, $\lambda_1 = 0.4$, $\lambda_2 = 0.2$, which is causal and invertible

This result is straightforward from the definition of the conditional probability.

Since X_t follows an AR(p) process, $X_t|X_s, s < t$, has the same distribution as Z_t, but with a mean equal to $\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p}$. That is, the conditional distribution of X_t given X_s for $s < t$, is the same as the distribution as Z_t but with mean $\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p}$. As a consequence, we can compute the densities of all the pdf on the right hand side of Eq. 4–44, exept that of the first term, $f(x_1, \ldots, x_p)$.

We can not derive explicitly the expression for the first term, $f(x_1, \ldots, x_p)$. When the value of n is large relative to p, we ignore the first term, and base our likelihood estimation on the remaining pieces. This is known as *conditional maximum likelihood estimation*. For example, this is what is done in fitting GARCH models in finance, since it is difficult to compute the distribution of $f(x_1, \ldots, x_p)$.

Fit simulated data. In this section, we simulate data to check the accuracy of our conditional maximum likelihood estimation method. We simulate data based on AR(1), AR(2) and AR(3) models with AL noise. Sample size is $n = 100$. We use the same method to simulate time series data as we did in the last section. The parameters ϕ_i are chosen to ensure a causal AR model.

We then apply a numerical optimization method to maximize our conditional likelihood. The AL parameters are chosen as $\kappa_0 = 1$, and $\tau_0 = 2$. We use the Yule-Walker (YW) estimator as our initial value. This is a standard method of moments estimator.

In the Table. 4-1, Table. 4-2, and Table. 4-3, we list the fitted model parameters of the first five simulations. $\hat{\phi}_{yw}$ indicates the Yule-Walker estimator of each simulation, which is also our initial value of ϕ. The calculation of Means and MSEs is based on 100 simulations.

Table 4-1: Fitted model parameters of AR(1) model under AL noise

parameters	κ	τ	θ	ϕ_1	$\hat{\phi}_{1_{yw}}$
True	0.8	1	-0.3182	0.7	
MLEs_simu1	0.7791	0.7713	-0.2752	0.7398	0.7106
MLEs_simu2	0.9652	0.9990	-0.0501	0.6125	0.6030
MLEs_simu3	0.8981	1.1812	-0.1799	0.6305	0.5256
MLEs_simu4	0.7962	0.9572	-0.3112	0.7087	0.7042
MLEs_simu5	0.7515	1.1285	-0.4622	0.7066	0.6763
mean(MLEs)	0.8002	0.9733	-0.31182	0.7045	
MSE(MLEs)	0.0068	0.0107	0.0184	0.0048	

In general, our method works well in fitting AR(p) models. MSEs for ϕ are less than 1%, with a better fit for models of lower orders. MSEs for AL parameters τ and θ are slightly higher, with an average around 0.01.

Table 4-2: Fitted model parameters of AR(2) model under AL noise

parameters	κ	τ	θ	ϕ_1	ϕ_2	$\hat{\phi}_{1_{yw}}$	$\hat{\phi}_{2_{yw}}$
True	0.8	1	-0.3182	0.7	-0.1		
MLEs_simu1	0.6711	0.9298	-0.5384	0.8011	-0.1761	0.7115	-0.1884
MLEs_simu2	0.6742	0.7489	-0.4283	0.7489	-0.1657	0.8734	-0.3434
MLEs_simu3	0.9057	1.0423	-0.1462	0.8545	-0.1996	0.8189	-0.2161
MLEs_simu4	0.8597	1.0650	-0.2286	0.7099	-0.0831	0.7797	-0.1810
MLEs_simu5	0.6922	0.8903	-0.4737	0.8246	-0.1054	0.8165	-0.2273
mean(MLEs)	0.7890	0.9747	-0.3314	0.6969	-0.1052		
MSE(MLEs)	0.0078	0.0143	0.0190	0.0092	0.0064		

Table 4-3: Fitted model parameters of AR(3) model under AL noise

parameters	κ	τ	θ	ϕ_1	ϕ_2	ϕ_3	$\hat{\phi}_{1_{yw}}$	$\hat{\phi}_{2_{yw}}$	$\hat{\phi}_{3_{yw}}$
True	0.8	1	-0.3182	0.7	-0.2	0.1			
MLEs_simu1	0.6485	0.8390	-0.5300	0.7523	-0.2370	0.2356	0.6332	-0.1432	0.1601
MLEs_simu2	0.6725	0.9841	-0.5667	0.7332	-0.2579	0.1040	0.7251	-0.3287	0.1245
MLEs_simu3	0.8131	0.9301	-0.2741	0.5539	-0.1674	0.1183	0.6046	-0.2339	0.1907
MLEs_simu4	0.7972	1.2948	-0.4185	0.7532	-0.2345	0.0735	0.6538	-0.1676	0.0978
MLEs_simu5	0.6533	0.5996	-0.3720	0.7608	-0.3232	0.1572	0.6343	-0.1573	0.1010
mean(MLEs)	0.8064	0.9685	-0.3021	0.7097	-0.2089	0.1003			
MSE(MLEs)	0.0085	0.0116	0.0219	0.0080	0.0094	0.0053			

4.2.3 Fitting an ARMA(p, q) Model Using Conditional Maximum Likelihood Estimation

Joint distribution of $(X_1, \ldots, X_n)$ under an ARMA(p, q) . The joint distribution $(X_1, \ldots, X_n)$ under an ARMA(p, q) can be calculated in a similar way to Eq. 4–44, except that $\{X_t\}$ is a function of $\{X_s\}$ and $\{Z_s\}$, $s < t$. Therefore the values of $\{Z_s\}$ have to first be calculated recursively.

Let $\{X_t\}$ be an ARMA(p, q) process under AL noise,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \lambda_1 Z_{t-1} + \cdots + \lambda_q Z_{t-q}, \tag{4–45}$$

where $Z_t \sim \mathcal{AL}(\theta, \kappa, \tau)$.

Equivalently, Eq. 4–45 can be expressed as

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \lambda_1 Z_{t-1} + \cdots + \lambda_q Z_{t-q} + Z_t. \tag{4–46}$$

The joint distribution of $(X_1, X_2, \ldots, X_n)$ can be calculated as in Section 4.2.2. using Eq. 4–44.

Now $X_t|X_s, s < t$ has the same AL distribution as Z_t, but with a mean equal to $\phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \lambda_1 Z_{t-1} + \cdots + \lambda_q Z_{t-q} = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \lambda_j Z_{t-j}$.

The value of Z_t can be iteratively calculated from Eq. 4–45 with

$$Z_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \lambda_1 Z_{t-1} - \cdots - \lambda_q Z_{t-q}. \tag{4–47}$$

Empirically, we set $Z_1 = Z_2 = \ldots = Z_p = 0$, using Eq. 4–47, $Z_t, t > p$ can be calculated as

$$\begin{aligned}
Z_{p+1} &= X_{p+1} - \phi_1 X_p - \ldots - \phi_p X_1 \\
Z_{p+2} &= X_{p+2} - \phi_1 X_{p+1} - \ldots - \phi_p X_2 - \lambda_1 Z_{p+1} \\
Z_{p+3} &= X_{p+3} - \phi_1 X_{p+2} - \ldots - \phi_p X_3 - \lambda_1 Z_{p+2} - \lambda_2 Z_{p+1} \\
&\vdots \\
Z_n &= X_n - \phi_1 X_{n-1} - \ldots - \phi_p X_{n-p} - \lambda_1 Z_{n-1} - \lambda_2 Z_{n-2} - \ldots - \lambda_q Z_{n-q+1}
\end{aligned} \tag{4–48}$$

Fitting simulated data. As in the last section, we now simulate data to check the accuracy of our conditional maximum likelihood estimation method. We simulate data from ARMA(1,1) models with AL noise and Sample size of $n = 100$. We use the same method to simulate time series data as in the last section. The parameters ϕ_i and λ_j are chosen to ensure model causality and invertibility.

We apply a numerical optimization method to maximize our conditional likelihood. The AL parameters are chosen as $\kappa_0 = 1$ and $\tau_0 = 2$. The initial values for ϕ and λ come from the ARIMA function in R with Gaussian noise.

In Table. 4-4, we list the fitted model parameters of the first five simulation. The calculation of Means and MSEs is based on 100 simulations.

Table 4-4: Fitted value of ARMA(1,1) model under AL noise

parameters	κ	τ	θ	ϕ_1	λ_1
True	0.8	1	-0.3182	0.7	0.5
MLEs_simu1	0.7398	0.9210	-0.3985	0.7598	0.5045
MLEs_simu2	0.8150	0.9835	-0.2865	0.7111	0.4810
MLEs_simu3	0.8282	1.0287	-0.2759	0.7119	0.4770
MLEs_simu4	0.8357	0.9979	-0.2546	0.6419	0.6056
MLEs_simu5	0.7415	1.0194	-0.4376	0.7237	0.4352
mean(MLEs)	0.8074	0.9853	-0.3035	0.6888	0.5004
MSE(MLEs)	0.0079	0.0112	0.0217	0.0056	0.0088

4.3 ARMA Models Driven by GARCH Noise

When linear models are not appropriate, nonlinear time series models such as GARCH, bilinear models, autoregressive models with random coefficients, and threshold models, are possible alternatives. GARCH models were developed on observing that the volatility of some time series processes are correlated. This is a common situation in financial time series data. As a result, GARCH models are widely used in finance.

Engle(1982) introduced the ARCH(p) process where the volatilities are dependent on the past volatilities. Bollerslev (1986) introduced a generalization of the ARCH(p) process, the GARCH process.

4.3.1 ARMA Model Driven by GARCH noise

Definition 4.5: ARMA(p,q) model driven by GARCH(u,v)noise. $\{X_t\}$ is an ARMA(p,q) process driven by GARCH(u,v) noise if

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \lambda_j Z_{t-j}, \tag{4–49}$$

where

$$Z_t = \sqrt{h_t} e_t, \qquad \{e_t\} \sim \text{IID N}(0,1), \tag{4–50}$$

where h_t is the positive function of $Z_s, s < t$, defined by

$$h_t = \alpha_0 + \sum_{i=1}^{u} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{v} \beta_j h_{t-j}, \tag{4–51}$$

with $\alpha_0 > 0$, and $\alpha_j, \beta_j \geq 0$, $j = 1, 2, \ldots$

4.3.2 Conditional Maximum Likelihood Estimation of GARCH model

The model parameters of a GARCH can be estimated using conditional maximum likelihood estimation. Let $f(x_1, \ldots, x_n)$ denote the pdf of the joint distribution of the first n observations, $X_1, \ldots, X_n$, under an AR(p) model. Then the likelihood can be written as

$$f(x_1, \ldots, x_n) = f(x_1, \ldots, x_p) f(x_{p+1} | x_t, t \leq p) \cdots f(x_n | x_t, t \leq n-1). \tag{4–52}$$

When the value of n is large relative to p, we ignore the first term, and base our likelihood estimation on the remaining pieces. This is known as *conditional maximum likelihood estimation.*

Therefore, using the property of a location-scale family, the joint likelihood of an ARMA model driven by GARCH noise can be expressed as,

$$L(\phi_1, \ldots, \phi_p, \lambda_1, \ldots, \lambda_q, \alpha_0, \ldots, \alpha_u, \beta_0, \ldots, \beta_v) = \prod_{t=p+1}^{n} \frac{1}{\sigma_t} \phi\left(\frac{X_t - \mu_t}{\sigma_t}\right), \tag{4–53}$$

where $\phi(\cdot)$ denotes the standard normal density function.

$\{Z_t\}$ can be derived similarly as in Section 4.2.3, with

$$Z_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \lambda_1 Z_{t-1} - \cdots - \lambda_q Z_{t-q}. \tag{4–54}$$

The mean of $\{X_t\}$, μ_t, can be calculated in the same way as for an ARMA(p, q) model,

$$\mu_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \lambda_j Z_{t-j}. \tag{4–55}$$

Standard deviations $\sigma_t = \sqrt{h_t}$, $t \geq 0$, can be computed recursively from Eq. 4–50 and Eq. 4–51 with $Z_t = 0$ and $h_t = \hat{\sigma}^2$ for all $t \leq 0$. $\hat{\sigma}^2$ is the sample variance of $\{Z_1, \ldots, Z_t\}$.

We now have the joint likelihood function of the $\{X_t\}$. MLEs can be found using numerical optimization methods. For the standard Gaussian GARCH, the model can be easily fitted using computational packages, like ITSM, and the GARCH toolbox in Matlab.

4.3.3 ARMA Models Driven by GARCH AL Noise

We now introduce AL noise into the GARCH model. We define an ARMA model driven by AL GARCH noise in exactly the same way as we defined the Gaussian GARCH model of Eq. 4–49, except that $\{e_t\}$ has an Asymmetric Laplace distribution.

Definition 4.6: ARMA (p, q) model driven by GARCH(u, v) AL noise. $\{X_t\}$ is an ARMA(p, q) process driven by GARCH(u, v) AL noise if

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \lambda_j Z_{t-j}, \tag{4–56}$$

where

$$Z_t = \sqrt{h_t} e_t, \qquad \{e_t\} \sim \text{IID}\mathcal{AL}(\theta, \kappa, \tau), \tag{4–57}$$

where h_t is the positive function of $Z_s, s < t$, defined by

$$h_t = \alpha_0 + \sum_{i=1}^{u} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{v} \beta_j h_{t-j}, \tag{4–58}$$

with $\alpha_0 > 0$, and $\alpha_j, \beta_j \geq 0$, $j = 1, 2, \ldots$ and

$$\theta \equiv \frac{\tau}{\sqrt{2}} \left(\frac{1}{\kappa} - \kappa \right). \tag{4–59}$$

The joint likelihood function of the GARCH AL model can then be expressed as

$$L(\phi_1, \ldots, \phi_p, \lambda_1, \ldots, \lambda_q, \alpha_0, \ldots, \alpha_u, \beta_0, \ldots, \beta_v) = \prod_{t=p+1}^{n} \frac{1}{\sigma_t} f_{AL} \left(\frac{X_t - \mu_t}{\sigma_t} \right), \tag{4–60}$$

where $f_{AL}(\cdot)$ is the density function of $\mathcal{AL}(\kappa, \tau, \theta)$.

This model can be fit use conditional likelihood estimation, in the same way as in the last section, except that the conditional likelihood functions of $X_t | X_s, s < t$ are different now. The joint conditional likelihood function can be maximized by using numerical optimization methods.

4.4 Analysis Real Estate Mutual Fund Data

In this section, we fit some real financial data to check the usefulness of our model. We analyze the returns of a real estate mutual fund that is managed by TIAA-CREF. The data range from Jan 1, 2000 to Dec 31, 2006, a total of 1807 daily values. The reason for selecting this data set is that the histogram of the data indicates heavy tails and right skewness.

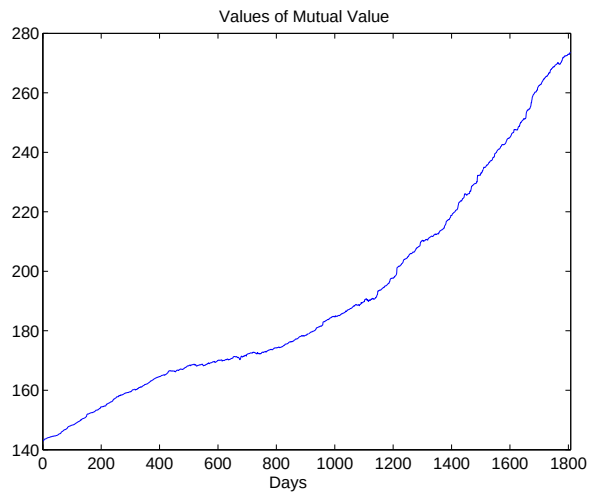

Figure 4-7. Daily values of the mutual fund managed by TIAA-CREF. The data range from Jan 1, 2000 to Dec 31, 2006, a total of 1807 values.

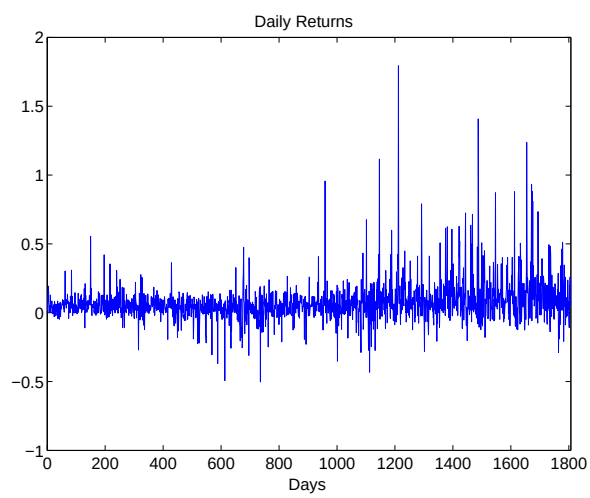

Figure 4-8. Daily returns of the mutual fund managed by TIAA-CREF. The data range from Jan 1, 2000 to Dec 31, 2006, a total of 1807 value.

Obviously, the data are not stationary. Therefore, we differentiated the data at lag 1. The resulting data are the daily returns of the mutual fund.

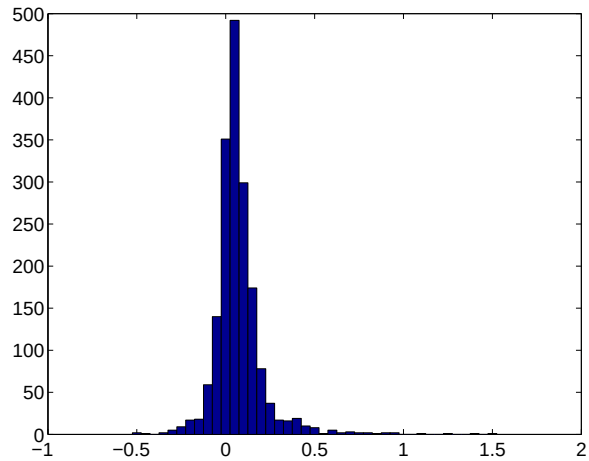

Figure 4-9. Histogram of the daily returns of the real estate mutual fund managed by TIAA-CREF.

The histogram of Fig. 4-9 indicates high peakedness compared with a normal distribution. Also, the data are asymmetric with right skewness. The distribution of the daily returns is close to the shape of an Asymmetric Laplace distribution as we have already discussed. Therefore, we consider this to be a good example to apply our time series models to.

We use four methods to analyze this mutual fund data: an ARMA(p, q) model under Gaussian noise, an ARMA(p, q) model under Asymmetric Laplace noise, an ARMA(p, q) model driven by GARCH Gaussian noise, and an ARMA(p, q) model driven by GARCH AL noise. The first two models are linear models. The other two are nonlinear models which assume that the variances are dependent.

We plot the sample ACF of the residuals to check the validation of the models. If there is no correlation in the sample ACF, we check the sample ACF of the absolute values and squares of residuals. No correlation in the sample ACF of the residuals does not necessarily mean that they are independent. If correlation is detected in the absolute values or squares of residuals, that suggests they are dependent.

Since these models have different numbers of parameters, the log-likelihood by itself is not a good criterion for model selection. We apply the bias-corrected Akaike Information Criterion (AICc) to evaluate our models.

Definition 4.7: (Akaike Information Criterion (AIC)).For a model based on parameters Θ, let $L(\hat{\Theta})$ be the maximized likelihood function, and k the number of free parameters in the model.The Akaike information criterion (AIC) is defined as

$$\text{AIC} = -2\log\left(L(\hat{\Theta})\right) + 2k. \tag{4–61}$$

Definition 4.8: (Bias-corrected Akaike Information Criterion (AICc)).For a model based on parameters Θ, let $L(\hat{\Theta})$ be the maximized likelihood function, and k the number of free parameters in the model. The bias-corrected Akaike information criterion (AICc) is defined as

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}. \tag{4–62}$$

AIC uses the $2k$ term as the penalty for adding more parameters into the model. Usually, the model with minimum AIC value is chosen as the best model to fit to the data.

In time series model selection, we use the AICc criterion. AICc is the empirical correction for small sample sizes. Since AICc converges to AIC as n gets large, AICc should be employed regardless of sample size. In our case, since the sample size 1807 is large enough, AIC is very close to AICc.

Method 1: ARMA(p, q) model under Gaussian noise. This is the traditional time series model as we stated in Section 4.1. We first fit the data up to the order of ARMA(7,7) using the autofit function in ITSM2000. It follows that ARMA(1,3) is the best model with largest log likelihood=995.57, lowest AICc=-1981. Below is the fitted value and standard errors of the ARMA(1,3) from the ARIMA function in the R package.

When looking at the first plot, the sample ACF of the residuals, sample ACFs are out of the boundary at lags 11, 20 and 22. There are also four others close to the boundary.

Table 4-5: Fitted Value of ARMA$(1,3)$ under Gaussian noise

parameters	$\hat{\phi}_1$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
MLEs	0.98	-0.8636	-0.0505	-0.0715
s.e.	1.968E-6	5.545E-4	9.405E-4	5.707E-4

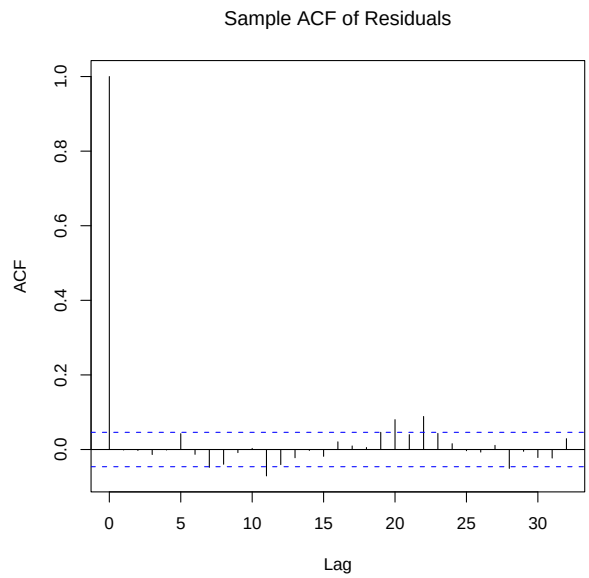

Figure 4-10. Sample ACF of residuals of ARMA(1,3) model under Gaussian noise.

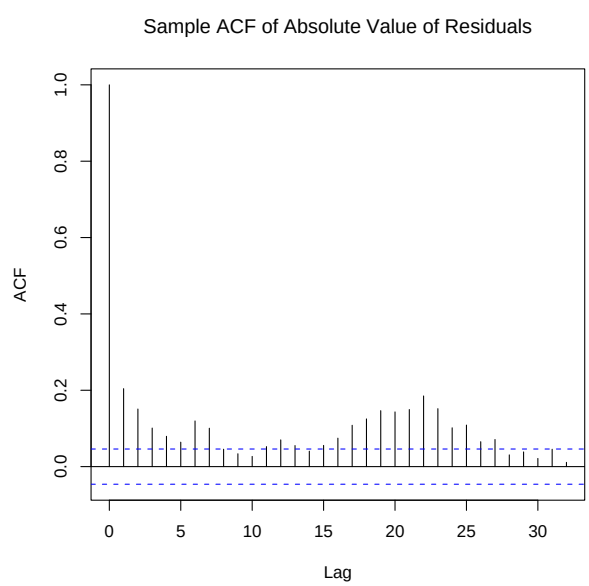

Figure 4-11. Sample ACF of absolute values of residuals of ARMA(1,3) model under Gaussian noise.

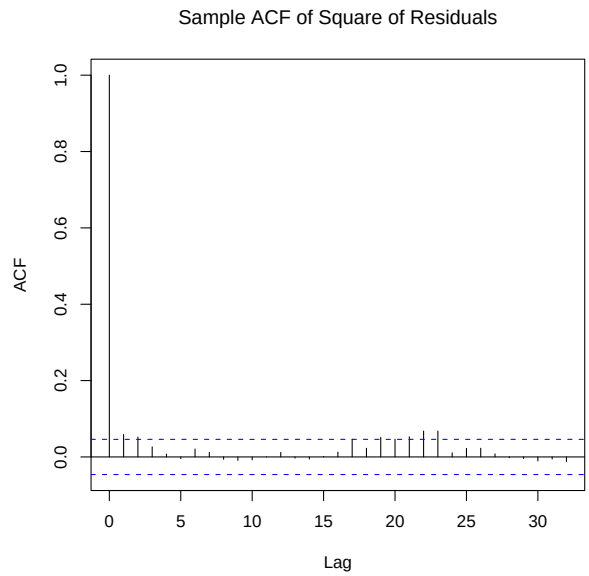

Figure 4-12. Sample ACF of squares of residuals of ARMA(1,3) model under Gaussian noise.

When analyzing the sample ACF of absolute values and squares of residuals, we found an obvious correlation in the residuals. Therefore, it is necessary to consider fitting a GARCH model.

Method 2: ARMA(p, q) model driven by GARCH Gaussian noise.

We fit the ARMA(p, q) model driven by GARCH Gaussian noise, as defined in Section 4.3. The GARCH model is fitted using the matlab GARCH toolbox. We search for the best GARCH model up to the order of ARMA(7,7) and GARCH(2,2). The best model is an ARMA(1,3) driven by GARCH(1,1) noise. The log likelihood of the fitted model is 1135.9, with AICc equal to -2263.8.

Table 4-6: Fitted parameters of ARMA$(1, 3)$ driven by GARCH(1,1) Gaussian noise.

parameters	$\hat{\phi}_1$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\alpha}_1$	$\hat{\beta}_1$
MLEs	0.998	-0.907	0.005	-0.083	0.484	0.353
s.e.	1.47E-3	3.237E-2	3.718E-2	2.589E-2	2.468E-2	2.217E-2

Method 3: ARMA(p, q) model under AL noise.

We now consider fitting an ARMA(p, q) model under AL noise, as discussed in Section 4.2. The model is fitted using the method of conditional maximum likelihood

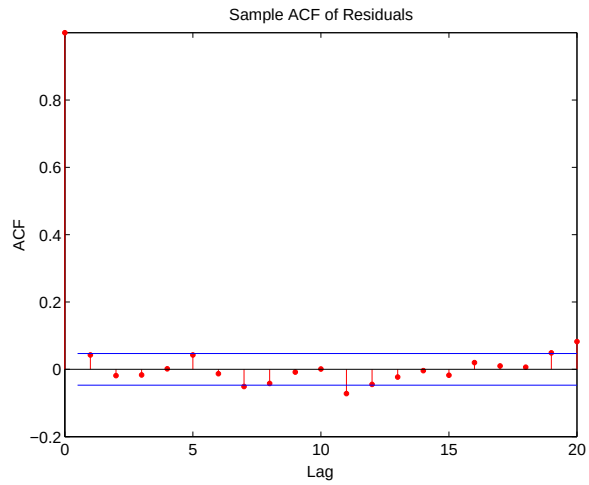

Figure 4-13. Sample ACF of residuals of ARMA(1,3) model driven by GARCH(1,1) Gaussian noise.

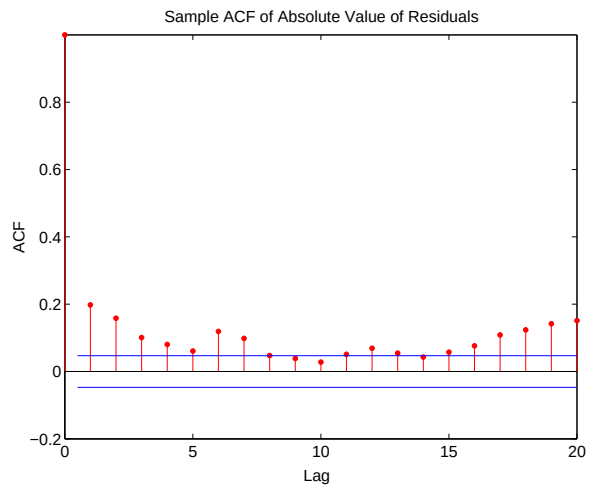

Figure 4-14. Sample ACF of absolute values of residuals of ARMA(1,3) model driven by GARCH(1,1) Gaussian noise.

estimation described in section 4.2. Empirically, we did a numerical optimization in Matlab, using the Simplex method. The initial values of the ARMA parameters are given by fitting a regular ARMA model under Gaussian noise. The initial values of κ and τ are both set to 1.

We search for the best ARMA(p, q) up to the order of ARMA(7,7). The best model is an ARMA(2,6) with log likelihood=1439.8, AICc=-2859.

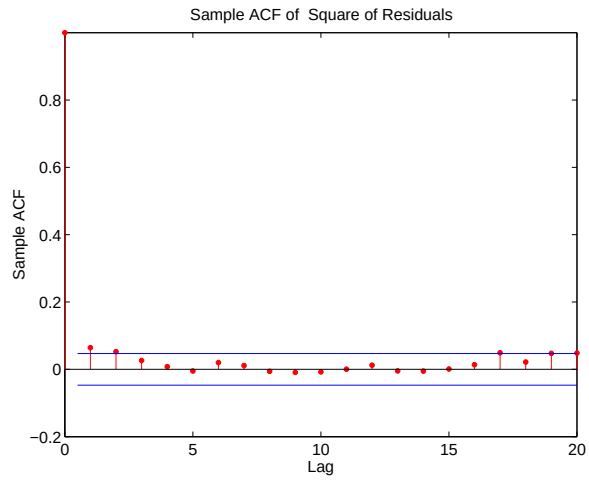

Figure 4-15. Sample ACF of squares of residuals of ARMA(1,3) model driven by GARCH(1,1) Gaussian noise.

Lehmann(1983) indicates that maximum likelihood estimators of ARMA(p, q) processes are approximately normally distributed with variance at least as small as those of other asymptotically normally distributed estimators. Brockwell and Davis (2002) prove that the large-sample distribution of maximum likelihood estimators of ARMA coefficients is the same for $\{Z_t\} \sim \text{IID}(0, \sigma^2)$, regardless of whether or not $\{Z_t\}$ is Gaussian. Therefore, we can use the standard errors of MLEs from Gaussian noise as the standard errors of the MLEs in our model based on AL noise.

The standard deviation of κ and τ are obtained by bootstrapping. For detailed information about bootstrapping on time series models, please refer to Shumway and Stoffer (2000).

Table 4-7: Fitted parameters of ARMA$(2, 6)$ under AL noise

Para	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	$\hat{\lambda}_6$	$\hat{\kappa}$	$\hat{\tau}$
MLEs	0.3526	0.5423	-0.2989	-0.4524	-0.0889	-0.0610	0.0192	0.0142	0.919	0.129
s.e.	1.12E-1	1.11E-1	1.11E-1	8.45E-2	1.23E-3	1.75E-3	9.9E-4	5.72E-4	0.227	0.019

Method 4: ARMA(p, q) model driven by GARCH AL noise. We fit an ARMA(p, q) model driven by GARCH AL noise as defined in Section 4.3. The model is fitted using conditional maximum likelihood similarly as in method 2, except that the conditional variances are dependent. The conditional variances can be found recursively.

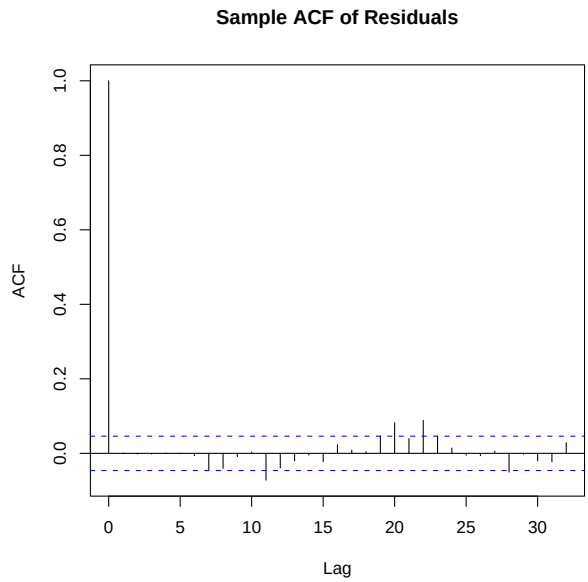

Figure 4-16. Sample ACF of residuals of ARMA(2,6) model under AL noise.

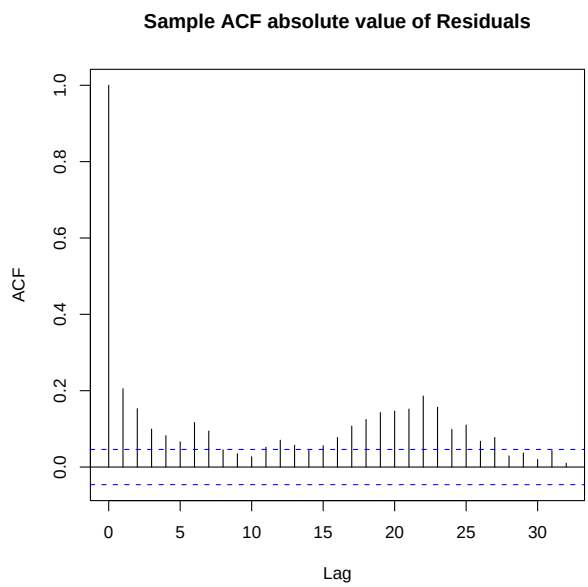

Figure 4-17. Sample ACF of absolute values of residuals of ARMA(2,6) model under AL noise.

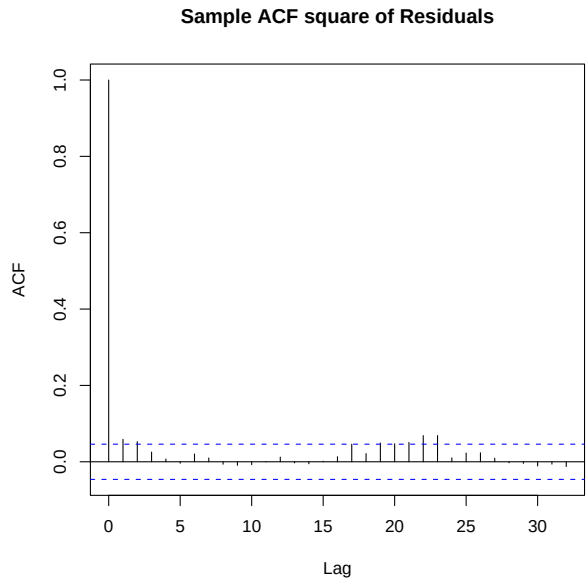

Figure 4-18. Sample ACF of squares of residuals of ARMA(2,6) model under AL noise.

We have searched up to the order of ARMA(7,7). The best model is an ARMA(1,2) driven by GARCH(1,1) noise, with log likelihood and AICc equal to 1458.7 and -2971.3, respectively.

Table 4-8: Fitted parameters of ARMA$(1,3)$ driven by GARCH(1,1) AL noise.

parameters	$\hat{\phi}_1$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\kappa}$	$\hat{\tau}$
MLEs	0.11248	0.977	-0.922	-0.040	0.335	0.342	0.823	0.112
s.e.	1.472E-3	3.240E-2	3.712E-2	2.589E-2	2.468E-2	2.217E-2	0.0165	0.344

Table 4-9: Summary of AICc of the four methods

Model	Noise	AICc
ARMA(1,3)	Gaussian	-1981
ARMA(2,6)	AL	-2859
ARMA(1,3)	GARCH Gaussian	-2263.8
ARMA(1,2)	GARCH AL	-2971.3

According to Table. 4-9, the AICc values of our models based on AL noise are much lower than models based on traditional Gaussian noise. This is an indication that models assuming AL noise provide a better fit for the daily returns of the mutual fund. The assumption of AL noise provides a better description of the data.

Comparing the first two linear models, the ARMA(p,q) under AL noise has a much lower value of AICc. Similarly for the two nonlinear models. An even more attractive

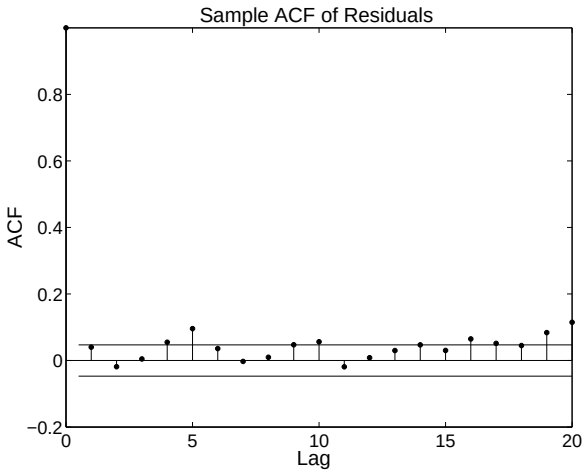

Figure 4-19. Sample ACF of residuals of ARMA(1,3) model driven GARCH(1,1) AL noise.

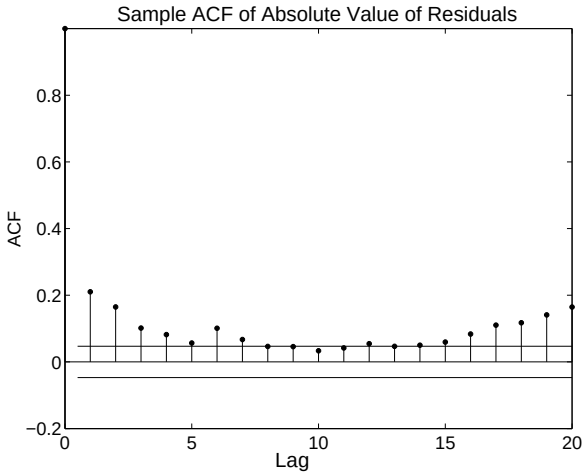

Figure 4-20. Sample ACF of absolute value of residuals of ARMA(1,3) model driven GARCH(1,1) AL noise.

feature of these results is that our linear model has a better fit than the nonlinear GARCH model.

It is a well known fact that the volatility of financial data is highly correlated. ARMA models assume constant variance, which leads to correlated residuals. Our ARMA model under AL noise can not remove all the correlation in the residuals. But it has significantly improved the likelihood and AICc. And the GARCH model under AL noise has a significantly better fit than the GARCH model under Gaussian noise.

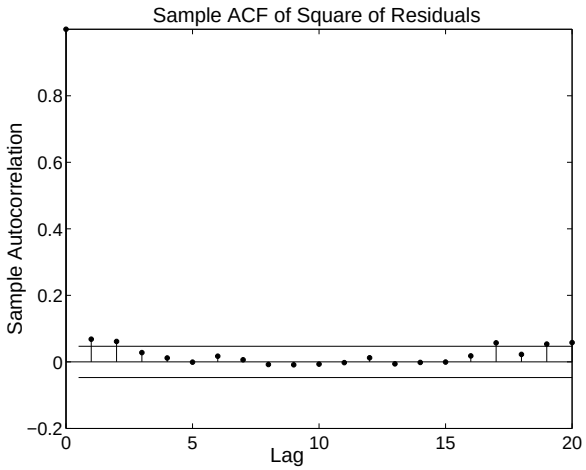

Figure 4-21. Sample ACF of squares of residuals of ARMA(1,3) model driven GARCH(1,1) AL noise.

APPENDIX A
SAR(P) MODEL WITH MULTIVARIATE AL MARGINAL DISTRIBUTION

In Chapter 2, we have analyzed the "trading day effect", that the volatilities of log returns of exchange rates are higher in the middle of the week and lower during the weekends. And sample ACF indicates that the log returns are correlated at the lag of 7 or multiple of 7. SAR(p) models assume the coefficients of the first $p-1$ lags of a AR(p) are zero. Therefore, $\{X_t\}$'s are correlated at the lag p and a multiple of p.

Now we start from assuming that the marginal distribution of a time series model has a multivariate Asymmetric Laplace distribution and approximate the distribution of the generalized estimator of ϕ using saddlepoint approximation. We have tried three methods to approximate the PDFs of the generalized estimator. Since there are relatively large deviations between our approximations and the simulated data, we arrange this part as appendix.

A.1 SAR(p) Model with Multivariate AL Marginal Distribution

A.1.1 SAR(p) Model

SAR(p) model is a special case of AR(p) model, where the coefficients of the first $p-1$ lags are zero. Therefore, $\{X_t\}$ are correlated at lag p and a multiple of p. Paige and Trindade (2003) define the SAR(p) model.

Definition A.1 (SAR(p)model) . The zero-mean process $\{X_t\}$, $t = \ldots, -1, 0, 1, \ldots$ has SAR(p) model if

$$X_t = \phi X_{t-p} + V_t, \qquad \{V_t\} \sim \text{iid } \mathrm{N}(0, \sigma^2), \tag{A–1}$$

with $|\phi| < 1$. It follows that X_t is stationary with autocovariance function (ACVF) at lag $h \geq 0$,

$$\gamma_h := E(X_t X_{t+h}) = \begin{cases} \sigma^2 \phi^k/(1-\phi^2), & \text{if } h = kp, k = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases} \tag{A–2}$$

A realization $X = [X_1, \ldots, X_n]'$ from the model has a multivariate normal distribution, $X \sim N_n(\mathbf{0}, \mathbf{\Gamma}_n)$, where the (i,j)th entry of covariance matrix $\mathbf{\Gamma}_n$ is just $\gamma_{|i-j|}$.

Now we define the SAR(p) with multivariate Asymmetric Laplace marginal distribution. Let $\{Y_t - m\}$ to be the zero-mean process, $t = \ldots, -1, 0, 1, \ldots$, with multivariate Asymmetric Laplace distribution,

$$Y_t = \phi Y_{t-p} + Z_t. \tag{A–3}$$

Therefore, a realization $Y = [Y_1, \ldots, Y_n]'$ from this model has a multivariate Asymmetric Laplace distribution, $Y \sim \mathcal{AL}(m, \mathbf{\Gamma}_n)$. We do not specify the distribution of Z_t here. It does not follow an Asymmetric Laplace distribution, as we have stated earlier in this chapter.

A.1.2 Generalized Estimator of ϕ

Given observations $x_1, \ldots, x_n$ from a stationary time series, let

$$\hat{\gamma}_h = \frac{1}{n}\sum_{t=1}^{n-h} x_t x_{t+h}, \quad \text{and} \quad \hat{\gamma}_0^{(p)} = \frac{1}{n}\sum_{t=p+1}^{n-p} x_t^2, \tag{A–4}$$

denote respectively the sample-based estimates of the ACVF at lag h, and the truncated ACVF at lag 0 obtained by omitting the first and last p observation. For any given nonnegative constants c_1 and c_2, we define the generalized estimator of ϕ,

$$\hat{\phi}_{c_1,c_2} = \frac{\sum_{t=1+p}^{n} x_t x_{t-p}}{c_1 \sum_{t=1}^{p} x_t^2 + \sum_{t=p+1}^{n-p} x_t^2 + c_2 \sum_{t=n-p+1}^{n} x_t^2} =: \frac{S}{cT_1 + T_2 + c_2 T_3}. \tag{A–5}$$

Some of the more common estimators of ϕ and σ^2, can be shown to be the following special cases of Equation(2.11) (Brockwell, Dahlhaus, and Trindade, 2005):

Least Squares,

$$\hat{\phi}_{LS} = \frac{S}{T_1 + T_2} = \hat{\phi}_{1,0}, \tag{A–6}$$

Yule-Walker,

$$\hat{\phi}_{YW} = \frac{\hat{\gamma}_p}{\hat{\gamma}_0} = \frac{S}{T_1 + T_2 + T_3} = \hat{\phi}_{1,1}, \tag{A–7}$$

Burg,

$$\hat{\phi}_{BG} = \frac{2\hat{\gamma}_p}{\hat{\gamma}_0 + \hat{\gamma}_0^{(p)}} = \frac{S}{T_1/2 + T_2 + T_3/2} = \hat{\phi}_{1/2,1/2}. \tag{A–8}$$

We define the $n \times n$ matrix A to be zero everywhere, except when $|i - j| = p$, in which case it is equal to 1/2. Similarly, define $B(c_1, c_2)$ to be the $n \times n$ identity matrix, with the first (last) p diagonal elements multiplied by $c_1(c_2)$. If $X = [X_1, \ldots, X_n]'$ is a realization from model (2.9), we can then express the generalized estimator $\hat{\phi}_{c_1,c_2}$ as a ratio of quadratic forms in normal random variables,

$$\hat{\phi}_{c_1,c_2} = \frac{X'AX}{X'B(c_1, c_2)X} \equiv \frac{P}{Q(c_1, c_2)}. \tag{A–9}$$

A.1.3 Multivariate Asymmetric Laplace Distribution

Kotz *el al.* (2001) define the multivariate Asymmetric Laplace distribution.

Definition A.2: (Multivariate Asymmetric Laplace Distribution). A random vector of $\mathbf{Y}$ in $\mathbf{R}^d$ is said to have a multivariate Asymmetric Laplace distribution (AL) if its characteristic function is given by

$$\Psi(\mathbf{t}) = \frac{1}{1 + \frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t} - i\mathbf{m}'\mathbf{t}}, \tag{A–10}$$

where $\mathbf{m} \in \mathbf{R}^d$ and $\mathbf{\Sigma}$ is a $d \times d$ nonnegative definite symmetric matrix.

We use the notation of $\mathcal{AL}_d(\mathbf{m}, \mathbf{\Sigma})$ to denote the distribution of $\mathbf{Y}$, and write $\mathbf{Y} \sim \mathcal{AL}_d(\mathbf{m}, \mathbf{\Sigma})$. If the matrix $\mathbf{\Sigma}$ is positive-definite, the distribution is truly d-dimensional and has a probability density function. Otherwise, it is degenerate and the probability mass of the distribution is concentrated in a linear proper subspace of the d-dimensional space.

For $\mathbf{m} = 0$ the distribution $\mathcal{AL}_d(\mathbf{0}, \mathbf{\Sigma})$ reduces to the symmetric multivariate Laplace distribution $\mathcal{L}_d(\mathbf{\Sigma})$.

Proposition A.1 (Density function of Multivariate AL Distribution). If $\mathbf{Y} \sim \mathcal{AL}_d(\mathbf{m}, \mathbf{\Sigma})$, the density function of $\mathbf{Y}$ is

$$g(\mathbf{y}) = (2\pi)^{-d/2}|\mathbf{\Sigma}|^{-1/2} \int_0^\infty \exp\left(-\frac{(\mathbf{y} - z\mathbf{m})'\mathbf{\Sigma}^{-1}(\mathbf{y} - z\mathbf{m})}{2z} - z\right) z^{-d/2} dz. \tag{A–11}$$

Proposition A.2 (Presentation of multivariate AL distribution). Let $\mathbf{Y} \sim \mathcal{AL}_d(\mathbf{m}, \mathbf{\Sigma})$ and let $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{\Sigma})$. Let W be an exponentially distributed random variable with mean 1, independent of $\mathbf{X}$. Then,

$$Y \stackrel{d}{=} \mathbf{m}W + W^{1/2}\mathbf{X}. \tag{A–12}$$

Corollary A.1 (Presentation of component Y_i multivariate AL distribution). Let $\mathbf{Y} \sim \mathcal{AL}_d(\mathbf{m}, \mathbf{\Sigma})$, each component $\mathbf{Y}_i$ of $\mathbf{Y}$ admits the representation

$$Y_i \stackrel{d}{=} m_i W + W^{1/2} X_i, \tag{A–13}$$

where X_i is ith component of $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{\Sigma})$. W is an exponentially distributed random variable with mean 1, independent of X_i.

Let σ_{ii} to be the ith diagonal element of $\mathbf{\Sigma}$, and σ_{ii} to be (i, j)th element of $\mathbf{\Sigma}$. Then $X_i \sim \mathcal{N}(0, \sigma_{ii})$ and $E(X_i X_J) = \sigma_{ij}$.

By applying the representation in proposition 1.2, we can derive the mean vector and covariance matrix of multivariate AL distribution.

We have $EY_i = m_i$. since $E(X_i X_j) = \sigma_{ij}$ and $EW^2 = 2$, we have

$$\begin{aligned} E(Y_i Y_j) &= E\left[\left(m_i W + W^{1/2} X_i\right)\left(m_j W + W^{1/2} X_j\right)\right] \\ &= m_i m_j EW^2 + E(W)E(X_i X_j) \\ &= 2m_i m_j + \sigma_{ij}. \end{aligned}$$

Therefore,

$$\mathrm{Cov}(Y_i, Y_j) = E(Y_i Y_j) - EY_i EY_j = 2m_i m_j + \sigma_{ij} - m_i m_j = m_i m_j + \sigma_{ij}.$$

Proposition A.3 (Mean Vector and covariance matrix). Let $\mathbf{Y} \sim \mathcal{AL}_d(\mathbf{m}, \mathbf{\Sigma})$, $E\mathbf{Y}$ is the mean vector of $\mathbf{Y}$, and $\mathbf{Cov}(\mathbf{Y})$ is the covariance matrix of $\mathbf{Y}$, then

$$E(\mathbf{Y}) = \mathbf{m}, \tag{A–14}$$

and

$$\mathbf{Cov}(\mathbf{Y}) = \mathbf{\Sigma} + \mathbf{m}\mathbf{m}'. \tag{A–15}$$

Saddlepoint Approximation to the PDF of $\hat{\phi}_{c1,c2}$

Paige and Trindade(2003) approximate the distribution of $\hat{\phi}_{c_1,c_2}$ by applying saddlepoint approximation to equations. Now we want to apply the similar method to approximate the distribution of $\hat{\phi}_{c_1,c_2}$ under the SAR(p) with multivariate Asymmetric Laplace distribution.

The generalized estimator of ϕ, $\hat{\phi}_{c_1,c_2}$, is a ratio of two quadratic forms. Under normal case, the moment generating function of the quadratic forms can be derived explicitly. But under multivariate Asymmetric Laplace case, since the pdf of distribution is in a complicated form with a function of integration, we can not derive the moment generating function of the quadratic form directly.

We consider to use the generalized saddlepoint approximation method developed by Easton and Ronchetti(1986), as we did in Chapter 2. To apply this method, we need to derive the first four cumulants of the quadratic forms and then approximate the cumulative generating function using Edgeworth expansion. Detail of this generalized saddlepoint approximation method is available in Chapter 2.

We will now derive the expression for EP, EQ, $E(P^2)$, $E(PQ)$, $E(Q^2)$, $E(P^3)$, $E(P^2Q)$, $E(PQ^2)$, $E(Q^3)$, $E(P^4)$, $E(P^3Q)$, $E(P^2Q^2)$, $E(PQ^3)$, $E(Q^4)$, which will be used to develop the cumulants in the next section.

The expression for EP and EQ is quite straightforward. Let column $\mathbf{Y}_n$ has mean μ and variance-covariance matrix $\mathbf{\Sigma}$, and $\mathbf{G}$ is $n \times n$ matrix. then

$$E(\mathbf{Y}'\mathbf{G}\mathbf{Y}) = \mu'\mathbf{G}\mu + \text{trace}(\mathbf{G}\mathbf{\Sigma}). \tag{A–16}$$

McCullaph(1987, Ch3) specifies the expression of generalized moments using tensor method,

$$\begin{aligned}
E(P^2) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})^2 = \sum_{i,j,k,l} a_{i,j}a_{k,l}EY_iY_jY_kY_l \\
E(PQ) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})(\mathbf{Y}'\mathbf{B}\mathbf{Y}) = \sum_{i,j,k,l} a_{i,j}b_{k,l}EY_iY_jY_kY_l \\
E(Q^2) &= E(\mathbf{Y}'\mathbf{B}\mathbf{Y})^2 = \sum_{i,j,k,l} b_{i,j}b_{k,l}EY_iY_jY_kY_l.
\end{aligned} \tag{A–17}$$

Then we have the third moments,

$$\begin{aligned}
E(P^3) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})^3 = \sum_{i,j,k,l,c,d} a_{i,j}a_{k,l}a_{c,d}EY_iY_jY_kY_lY_cY_d \\
E(P^2Q) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})^2(\mathbf{Y}'\mathbf{B}\mathbf{Y}) = \sum_{i,j,k,l,c,d} a_{i,j}a_{k,l}b_{c,d}EY_iY_jY_kY_lY_cY_d \\
E(PQ^2) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})(\mathbf{Y}'\mathbf{B}\mathbf{Y})^2 = \sum_{i,j,k,l,c,d} a_{i,j}b_{k,l}b_{c,d}EY_iY_jY_kY_lY_cY_d \\
E(Q^3) &= E(\mathbf{Y}'\mathbf{B}\mathbf{Y})^3 = \sum_{i,j,k,l,c,d} b_{i,j}b_{k,l}b_{c,d}EY_iY_jY_kY_lY_cY_d.
\end{aligned} \tag{A–18}$$

The fourth moments,

$$\begin{aligned}
E(P^4) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})^4 = \sum_{i,j,k,l,c,d,e,f} a_{i,j}a_{k,l}a_{c,d}a_{e,f}EY_iY_jY_kY_lY_cY_dY_eY_f \\
E(P^3Q) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})^3(\mathbf{Y}'\mathbf{B}\mathbf{Y}) = \sum_{i,j,k,l,c,d,e,f} a_{i,j}a_{k,l}a_{c,d}b_{e,f}EY_iY_jY_kY_lY_cY_dY_eY_f \\
E(P^2Q^2) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})^2(\mathbf{Y}'\mathbf{B}\mathbf{Y})^2 = \sum_{i,j,k,l,c,d,e,f} a_{i,j}a_{k,l}b_{c,d}b_{e,f}EY_iY_jY_kY_lY_cY_dY_eY_f \\
E(PQ^3) &= E(\mathbf{Y}'\mathbf{A}\mathbf{Y})(\mathbf{Y}'\mathbf{B}\mathbf{Y})^3 = \sum_{i,j,k,l,c,d,e,f} a_{i,j}b_{k,l}b_{c,d}b_{e,f}EY_iY_jY_kY_lY_cY_dY_eY_f
\end{aligned}$$

$$E(Q^4) \;=\; E(\mathbf{Y}'\mathbf{B}\mathbf{Y})^4 = \sum_{i,j,k,l,c,d,e,f} b_{i,j}b_{k,l}b_{c,d}b_{e,f}EY_iY_jY_kY_lY_cY_dY_eY_f, \tag{A–19}$$

where $a_{i,j}$ and $b_{i,j}$ are the (i,j) components of matrix A and B. We now apply the representation of multivariate AL distribution in Corollary(1.1), the product of Y_i can be calculate as,

$$\begin{aligned}
& EY_iY_jY_kY_l \\
= \;& E\left[(m_iW + W^{1/2}X_i)(m_jW + W^{1/2}X_j)(m_kW + W^{1/2}X_k)(m_lW + W^{1/2}X_l)\right] \\
= \;& E\left[W^4 m_im_jm_km_l + W^{7/2}(m_im_jm_kX_l)[4] + W^3(m_im_jX_kX_l)[6]\right. \\
& \left. + W^{5/2}(m_iX_jX_kX_l)[4] + W^2X_iX_jX_kX_l\right] \\
= \;& m_im_jm_km_lEW^4 + m_im_jm_kEW^{7/2}EX_l[4] + m_im_jEW^3EX_kX_l[6] \\
& + m_iEW^{5/2}EX_jX_kX_l[4] + EW^2EX_iX_jX_kX_l,
\end{aligned} \tag{A–20}$$

where $X = [X_i, \ldots, X_n]$ has multivariate normal distribution, $X \sim N(0, \mathbf{\Gamma}_n)$. W is an exponential distribution with mean zero with $EW^p = p!$.

The terms followed by numbers in parenthesis indicate the summation of all the permutation in that type. For example,

$$(m_im_jm_kX_l)[4] = m_im_jm_kX_l + m_im_jX_km_l + m_iX_jm_km_l + X_im_jm_km_l. \tag{A–21}$$

We know for normally distributed X with mean zero, the odd moments and cumulants have the values of zero. Therefore, the terms with EX_i, $EX_iX_jX_k$ equal to zero.

To calculate $EX_iX_jX_kX_l$, we introduce the notation by McCullaph(1987, Ch3).

Let $\mathbf{X}$ be the random variable whose components are $X_1, \ldots, X_p$ with moment generating function $M_X(\xi)$ and cumulative generating function $K_X(\xi)$, we have the expansion,

$$\begin{aligned} M_X(\xi) &= 1 + \xi_i\kappa^i + \xi_i\xi_j\kappa^{ij}/2! + \xi_i\xi_j\xi_k\kappa^{ijk}/3! + \xi_i\xi_j\xi_k\xi_l\kappa^{ijkl}/4! + \ldots \\ K_X(\xi) &= 1 + \xi_i\kappa^i + \xi_i\xi_j\kappa^{i,j}/2! + \xi_i\xi_j\xi_k\kappa^{i,j,k}/3! + \xi_i\xi_j\xi_k\xi_l\kappa^{i,j,k,l}/4! + \ldots \end{aligned} \tag{A–22}$$

This expansion implicitly defines all the moments $\kappa^i, \kappa^{ij}, \kappa^{ijk}$ and cumulants $\kappa^i, \kappa^{i,j}, \kappa^{i,j,k}$ of $\mathbf{X}$. We can also express $\kappa^i, \kappa^{ij}, \kappa^{ijk}$ with the expectation forms which we are more familiar with,

$$\kappa^i = E(X^i), \quad \kappa^{ij} = E(X^iX^j), \quad \kappa^{ijk} = E(X^iX^jX^k) \tag{A–23}$$

and

$$\begin{aligned} \kappa^{i,j} &= \kappa^{ij} - \kappa^i\kappa^j \\ \kappa^{i,j,k} &= \kappa^{ijk} - \kappa^i\kappa^{jk}[3] + 2\kappa^i\kappa^j\kappa^k \\ \kappa^{i,j,k,l} &= \kappa^{ijkl} - \kappa^i\kappa^{jkl}[4] - \kappa^{ij}\kappa^{kl}[3] + 2\kappa^i\kappa^j\kappa^{kl}[6] - 6\kappa^i\kappa^j\kappa^k\kappa^l. \end{aligned} \tag{A–24}$$

Applying tensor method, we can express $EX_iX_jX_kX_l$ as

$$\begin{aligned} & EX_iX_jX_kX_l \\ = \; & \kappa^{ijkl} \\ = \; & \kappa^{i,j,k,l} + \kappa^i k^{j,k,l}[4] + \kappa^{i,j}\kappa^{k,l}[3] + \kappa^i\kappa^j\kappa^{k,l}[6] + \kappa^i k^j k^k k^l \\ = \; & \kappa^{i,j}\kappa^{k,l}[3] \\ = \; & \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{kj}. \end{aligned}$$

Therefore,

$$
\begin{aligned}
& EY_iY_jY_kY_l \\
= \;& 24\, m_im_jm_km_l + 6m_im_j\sigma_{kl}[6] + 2(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{lk}\sigma_{il}) \\
= \;& 24\, m_im_jm_km_l + 6\left[m_im_j\sigma_{kl} + m_im_k\sigma_{jl} + m_im_l\sigma_{jk} + m_jm_k\sigma_{il} + m_jm_l\sigma_{ik} + m_km_l\sigma_{ij}\right] \\
& + 2(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{lk}\sigma_{il}).
\end{aligned} \tag{A–25}
$$

In case of $\mathbf{Y}$ has symmetric multivariate Laplace distribution, $m_i = 0$. Eq. A–25 reduces to $EY_iY_jY_kY_l = 2(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{kj})$.

Similar to Eq. A–25, we get

$$
EX_iX_jX_kX_lX_cX_d = k^{i,j}k^{k,l}k^{c,d}[15] = \sigma_{ij}\sigma_{kl}\sigma_{cd}[15] \tag{A–26}
$$

and

$$
EX_iX_jX_kX_lX_cX_dX_eX_f = k^{i,j}k^{k,l}k^{c,d}k^{e,f}[105] = \sigma_{ij}\sigma_{kl}\sigma_{cd}\sigma_{ef}[105]. \tag{A–27}
$$

The higher order of moments of $\mathbf{Y}$ can be calculated as

$$
\begin{aligned}
& E\; Y_iY_jY_kY_lY_cY_d \\
= \;& E\Big[(m_iW + W^{1/2}X_i) + (m_jW + W^{1/2}X_j) + (m_kW + W^{1/2}X_k) \\
& + (m_lW + W^{1/2}X_l) + (m_cW + W^{1/2}X_c) + (m_dW + W^{1/2}X_d)\Big] \\
= \;& E\Big[W^6m_im_jm_km_lm_cm_d + W^{1/2}m_im_jm_km_lm_cX_b[6] + W^5m_im_jm_km_lX_cX_d[15] \\
& + W^{9/2}m_im_jm_kX_lX_cX_d[20] + W^4m_im_jX_kX_lX_cX_d[15] \\
& + W^{7/2}m_iX_jX_kX_lX_cX_d[6] + W^3X_iX_jX_kX_lX_cX_d\Big] \\
= \;& m_im_jm_km_lm_cm_dEW^6 + m_im_jm_km_lEW^5EX_cX_d[15] \\
& + m_im_jEW^4EX_kX_lX_cX_d[15] + EW^3EX_iX_jX_kX_lX_cX_d \\
= \;& 720\, m_im_jm_km_lm_cm_d + 120m_im_jm_km_l\sigma_{cd}[15] + 24m_im_j\sigma_{kl}\sigma_{cd}[45] + 6\sigma_{ij}\sigma_{kl}\sigma_{cd}[15],
\end{aligned} \tag{A–28}
$$

$$
\begin{aligned}
& EY_iY_jY_kY_lY_cY_dY_eY_f \\
= \quad & E\Big[W^8 m_i m_j m_k m_l m_c m_d m_e m_f + W^{15/2} m_i m_j m_k m_l m_c m_d m_e X_f[8] \\
& + W^7 m_i m_j m_k m_l m_c m_d X_e X_f[28] + W^{13/2} m_i m_j m_k m_l m_c X_d X_e X_f[56] \\
& + W^6 m_i m_j m_k m_l X_c X_d X_e X_f[70] + W^{11/2} m_i m_j m_k X_l X_c X_d X_e X_f[56] \\
& + W^5 m_i m_j X_k X_l X_c X_d X_e X_f[28] + W^{9/2} m_i X_j X_k X_l X_c X_d X_e X_f[8] \\
& + W^4 X_i X_j X_k X_l X_c X_d X_e X_f\Big] \\
= \quad & m_i m_j m_k m_l m_c m_d m_e m_f EW^8 + m_i m_j m_k m_l m_c m_d EW^7 EX_e X_f[28] \\
& + m_i m_j m_k m_l EW^6 EX_c X_d X_e X_f[70] + m_i m_j EW^5 EX_k X_j X_c X_d X_e X_f[28] \\
& + EW^4 EX_i X_j X_k X_l X_c X_d X_e X_f \\
= \quad & 40320 m_i m_j m_k m_l m_c m_d m_e m_f + 5040 m_i m_j m_k m_l m_c m_d \sigma_{cd}[28] \\
& + 720 m_i m_j m_k m_l \sigma_{cd}\sigma_{ef}[210] + 120 m_i m_j \sigma_{kl}\sigma_{cd}\sigma_{ef}[420] \\
& + 24 \sigma_{ij}\sigma_{kl}\sigma_{cd}\sigma_{ef}[105].
\end{aligned}
\tag{A–29}
$$

In case of $\mathbf{m} = 0$, these two equations reduce to

$$
\begin{aligned}
& EY_iY_jY_kY_lY_cY_d = 6\sigma_{ij}\sigma_{kl}\sigma_{cd}[15] \\
& EY_iY_jY_kY_lY_cY_dY_eY_f = 24\sigma_{ij}\sigma_{kl}\sigma_{cd}\sigma_{ef}[105].
\end{aligned}
\tag{A–30}
$$

To simplify our calculation, we will concentrate in the case $\mathbf{m} = 0$. The calculation of the first four moments of P and Q is computationally intensive. But it can be done using some computer programs, even it is very time-consuming.

We try three methods to approximate the distribution of $\hat{\phi}_{c_1,c_2}$.

A.1.4 Saddlepoint Approximation to the Estimating Equation

Daniels (1983) extend the saddlepoint method to approximate the distribution of an estimator defined by an estimating equation. We can define $\hat{\phi}_{c_1,c_2}$ as the unique root in r of the estimating equation

$$\eta(r) = P - rQ(c_1, c_2). \tag{A–31}$$

Since $\eta(r)$ is a monotonically decreasing function in r for every realization $\mathbf{Y}$, we have $\hat{\phi}_{c_1,c_2} \geq r \Leftrightarrow \eta(r) \geq 0$, which leads to the device $P(\hat{\phi}_{c_1,c_2} \geq r) = P(\eta(r) \geq 0)$.

Let $K_{\eta(r)}(T)$ be the cumulative generating function of $\eta(r)$. we can not derive $K_{\eta(r)}(T)$ directly, as we have stated before. But we can approximate $K_{\eta(r)}(T)$ using Edgeworth expansion. The cumulative generating function of $\eta(r)$, $K_{\eta(r)}(T)$, can be approximated as,

$$K_{\eta(r)}(T) = \mu_n T + \frac{n\sigma_n^2 T^2}{2} + \frac{\kappa_{3n}\sigma_n^3 n^2 T^3}{6} + \frac{\kappa_{4n}\sigma_n^4 n^3 T^4}{24}, \tag{A–32}$$

where μ_n , σ_n^2 , κ_{3n}, κ_{4n} are the mean, the variance, and the third and fourth cumulant of $\eta(r)$. The first four moments of $\eta(r)$ can be derived from the first four moments of P and Q,

$$\begin{aligned}
E[\eta(r)] &= E(P - rQ) = EP - rEQ \\
E[\eta(r)]^2 &= E(P - rQ)^2 = EP^2 - 2rEPQ + r^2EQ^2 \\
E[\eta(r)]^3 &= E(P - rQ)^3 = EP^3 - 3rEP^2Q + 3r^2EPQ^2 - r^3EQ^3 \\
E[\eta(r)]^4 &= E(P - rQ)^4 = EP^4 - 4rEP^3Q + 6r^2EP^2Q^2 - 4r^3EPQ^3 - r^4EQ^4.
\end{aligned} \tag{A–33}$$

The relationship of cumulants and moments are available in chapter 2.

According to Daniels(1983), the saddlepoint approximation to the pdf of $\hat{\phi}_{c_1,c_2}$ at r is,

$$f_{\hat{\phi}_{c_1,c_2}}(r|\phi) \approx n^{\frac{1}{2}}\{\dot{K}_{\eta(r)}(T_0)/T_0\}\left\{2\pi K''_{\eta(r)}(T_0)\right\}^{-\frac{1}{2}} \exp\left\{nK_{\eta(r)}(T_0)\right\}, \tag{A–34}$$

where $\dot{K}_{\eta(r)}(T)$ indicates the first derivative of $K_{\eta(r)}(T)$ with respect to r, $\dot{K}_{\eta(r)}(T) = \partial K_{\eta(r)}(T)/\partial r$. $K'_{\eta(r)}(T)$ and $K''_{\eta(r)}(T)$ indicate the first and second derivatives of $K_{\eta(r)}(T)$ with respect to T, respectively.

T_0 is the saddlepoint solving saddlepoint equation

$$\partial K_{\eta(r)}(T)/\partial T|_{T=T_0} =: K'_{\eta(r)}(T_0) = 0. \tag{A–35}$$

To check the validation of our method, we approximate the SAR(3) model, with $\phi = 0.5$, $\mathbf{m} = \mathbf{0}$. The variance-covariance matrix $\mathbf{\Sigma}$ is defined as Eq. A–2 with $\sigma = 1$. We choose the Burg estimator, i.e.,$c_1 = 0.5$, $c_2 = 0.5$. Sample size $n = 10$ and 50.

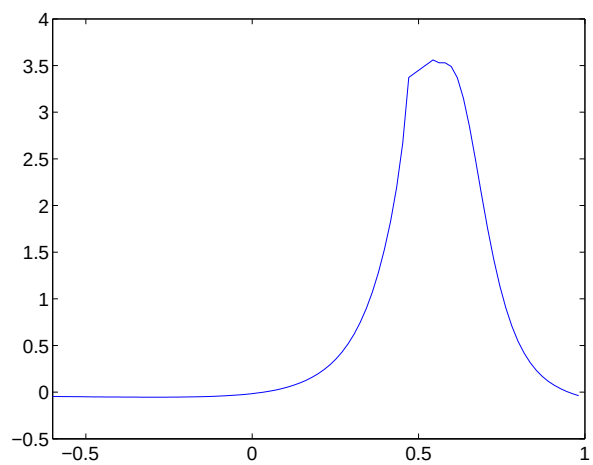

Figure A-1. Approximated pdf of burg estimator in SAR(3) model with Symmetric Laplace marginal distribution using saddlepoint approximation to the equations. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 10$.

Obviously, this is not a good approximation. Our approximated pdf has much higher peak compared with the simulated data. And there are some discontinuous point around the mode in our approximated pdf.

One thing to notice is that the approximated pdf with sample size 50 are much closer to the simulated histogram than the one with sample size 10. This is a good sign, since we are expecting the approximated pdf converge to the real value when sample size getting larger. We did not try the approximation with larger sample size since we are more interested in the small sample approximation.

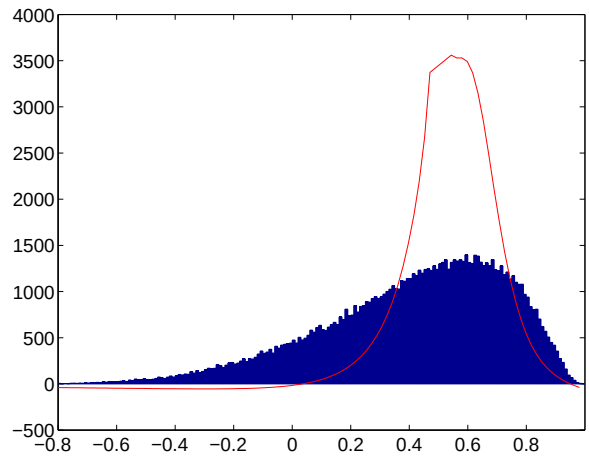

Figure A-2. Approximated pdf and simulated histogram using saddlepoint approximation to the equations. We have inflated the pdf by number of replication to match the histogram. The histogram is based on $10,000$ simulations. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 10$.

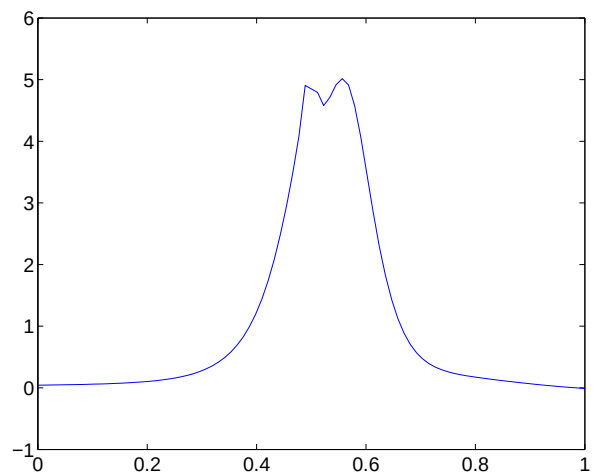

Figure A-3. Approximated pdf of burg estimator in SAR(3) model with Symmetric Laplace marginal distribution using saddlepoint approximation to the equations. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 50$.

Laplace Approximation to the Moments of a Ratio of Quadratic Forms

In this section, we want to approximate the moments of $\eta(r)$ using Laplace approximation and then apply generalized saddlepoint approximation directly.

Lieberman(1994) develop the Laplace approximation to the moments of a ratio of two quadratic forms. Let $\mathbf{X}$ be $n \times 1$ random vector with density function $f(x)$. Let matrices F and G be nonstochastic $n \times n$, F symmetric and G positive definite. Define the statistic

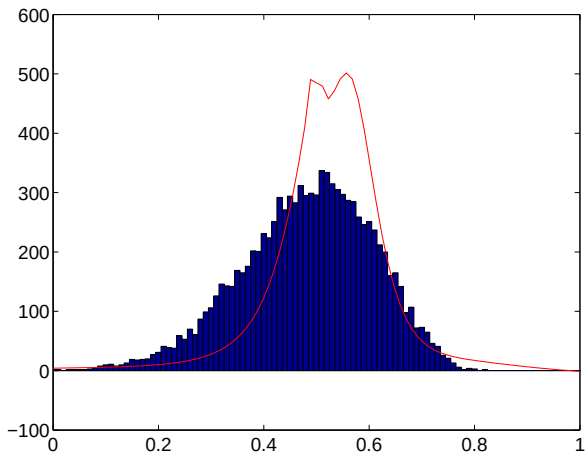

Figure A-4. Comparison of the approximated pdf and simulated histogram using saddlepoint approximation to the equations. We have inflated the pdf by number of replication to match the histogram. The histogram is based on $10,000$ simulations. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 50$.

under consideration is

$$s = \frac{X'FX}{X'GX}. \tag{A–36}$$

The Laplace approximation for the kth moment of r is

$$E_L(s^k) \approx \frac{E[(X'FX)^k]}{[E(X'GX)]^k}. \tag{A–37}$$

This is the Laplace approximation up to the order of $O(n^{-1})$. The approximation to the higher order of expansion is

$$E(s^k) = E_L(s^k) + b_{n1} + b_{n2} + O(n^{-3}), \tag{A–38}$$

where

$$b_{n1} = \frac{k(k+1)}{2}\left\{\frac{E[(X'FX)^k]\kappa_2}{[E(X'GX)]^{k+2}}\right\} - k\left\{\frac{\kappa_{k1}}{[E(X'GX)]^{k+1}}\right\} = O(n^{-1}),$$

$$\begin{aligned} b_{n2} = & \frac{k(k+1)}{2}\left\{\frac{\kappa_{k2}}{[E(X'GX)]^{k+2}}\right\} - \frac{k(k+1)(k+2)}{2}\left\{\frac{3E[(X'FX)^k]\kappa_3 + \kappa_{k1}\kappa_2}{[E(X'GX)]^{k+3}}\right\} \\ & + \frac{k(k+1)(k+2)(k+3)}{8}\left[\frac{E[(X'FX)^k]\kappa_2^2}{[E(X'GX)]^{k+4}}\right] = O(n^{-2}), \end{aligned}$$

where κ_p denotes the pth cumulant of $X'GX$, κ_{km} denotes the generalized cumulant of the product of $(X'FX)^k$ and $(X'GX)^m$.

Specifically,

$$\begin{aligned}\kappa_{k1} =& E[(X'FX)^k(X'GX)] - E[(X'FX)^k]E[X'GX] \\ \kappa_{k2} =& E[(X'FX)^k(X'GX)^2] - 2E[X'GX]E[(X'FX)^k(X'GX)] \\ & - E[(X'GX)^2]E[(X'FX)^k] + 2[E(X'GX)]^2E[(X'GX)^k].\end{aligned}$$

The advantage of the approximation lies in that it does not requite normality of X. It can be applied on any distribution.

We approximated the moments of ratio $r = \frac{Y'AY}{Y'BY}$ up to the order of $O(n^3)$. To make it comparable, we used the same SAR(3) model and same parameters as we did in last section. $\phi = 0.5$, $\mathbf{m} = \mathbf{0}$. The variance-covariance matrix $\mathbf{\Sigma}$ is defined as Eq. A–2 with $\sigma = 1$. We choose the Burg estimator, i.e.,$c_1 = 0.5$, $c_2 = 0.5$. Sample size $n = 10$.

But the approximated moments of r are not close to the simulated value under $100,000$ simulations. To our understanding, this problem lies in that the small sample size. We can see the trend of converge when we increase the size of Y. But since we are considering the small sample size, this is not a good method.

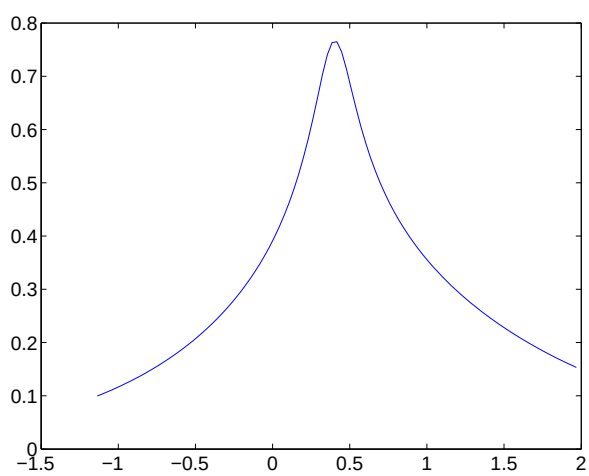

Figure A-5. Saddlepoint approximated pdf of burg estimator in SAR(3) model with Symmetric Laplace marginal distribution. We approximated moments of r using Laplace expansion.$\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 10$.

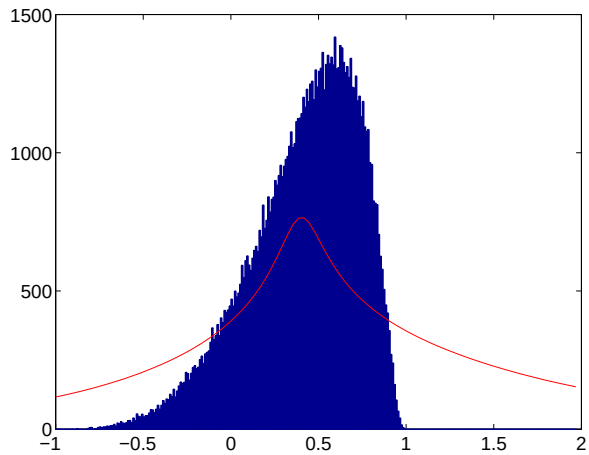

Figure A-6. Comparison of the approximated pdf using Laplace approximation and simulated histogram. The ratio of moments are approximated by Laplace approximation. We inflated the pdf by number of replication to match the histogram. The histogram is based on $10,000$ simulations. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 10$.

A.1.5 Approximate the Moments of r by Taylor Expansion

Then we go back to our old method, approximate the first four moments of $r = \frac{Y'AY}{Y'BY}$ by Taylor series expansion. Then use generalized saddlepoint approximation.

Let $f_1(P,Q) = \frac{P}{Q}$, we expand $f_1(P,Q)$ using Taylor series expansion at the mean of P, EP, and mean of Q, EQ. We will only state the first three terms of the expansion. For higher order of expansion, please refer to the formula in Chapter 2.

$$
\begin{aligned}
& f_1(P,Q) \\
\approx \quad & f_1(EP,EQ) + \left[(P-EP)\frac{\partial\ f_1}{\partial\ P}\bigg|_{P=EP} + (Q-EQ)\frac{\partial\ f_1}{\partial\ Q}\bigg|_{Q=EQ}\right] \\
& +\frac{1}{2}\left[(P-EP)^2\frac{\partial^2\ f_1}{\partial\ P^2}\bigg|_{P=EP} + 2(P-EP)(Q-EQ)\frac{\partial^2\ f_1}{\partial\ P\partial\ Q}\bigg|_{P=EP,Q=EQ}\right. \\
& \left.+(Q-EQ)^2\frac{\partial^2\ f_1}{\partial\ Q^2}\bigg|_{Q=EQ}\right] \\
& +\frac{1}{3!}\left[(P-EP)^3\frac{\partial^3\ f_1}{\partial\ P^3}\bigg|_{P=EP} + 3(P-EP)^2(Q-EQ)\frac{\partial^3\ f_1}{\partial\ P^2\partial\ Q}\bigg|_{P=EP,Q=EQ}\right. \\
& \left.+3(P-EP)(Q-EQ)^2\frac{\partial^3\ f_1}{\partial\ P\partial\ Q^2}\bigg|_{P=EP,Q=EQ} + (Q-EQ)^3\frac{\partial^3\ f_1}{\partial\ Q^3}\bigg|_{Q=EQ}\right],
\end{aligned}
\tag{A–39}
$$

take expectations on both sides,

$$
\begin{aligned}
&Ef_1(P,Q) \approx \ f_1(EP,EQ) \\
&+\frac{1}{2}\left[\mathrm{Var}P\frac{\partial^2\ f_1}{\partial\ P^2}\bigg|_{P=EP} + 2\mathrm{Cov}(P,Q)\frac{\partial^2\ f_1}{\partial\ P\partial\ Q}\bigg|_{P=EP,Q=EQ} + \mathrm{Var}Q\frac{\partial^2\ f_1}{\partial\ Q^2}\bigg|_{Q=EQ}\right] \\
&+\frac{1}{6}\left[E(P-EP)^3\frac{\partial^3\ f_1}{\partial\ P^3}\bigg|_{P=EP} + 3E(P-EP)^2(Q-EQ)\frac{\partial^3\ f_1}{\partial\ P^2\partial\ Q}\bigg|_{P=EP,Q=EQ}\right. \\
&\left.+3E(P-EP)(Q-EQ)^2\frac{\partial^3\ f_1}{\partial\ P\partial\ Q^2}\bigg|_{P=EP,Q=EQ} + E(Q-EQ)^3\frac{\partial^3\ f_1}{\partial\ Q^3}\bigg|_{Q=EQ}\right].
\end{aligned}
\tag{A–40}
$$

We use the similar formula to calculate the moments of P and Q. The only exception is that the cross terms of P and Q do not equal to zero anymore. Therefore, we have more complicated formula,

$$
\begin{aligned}
&E(P-EP)^3 = EP^3 - 3EP^2EP + 2(EP)^3 \\
&E(P-EP)^2(Q-EQ) = EP^2Q - 2EPQEP + 2(EP)^2EQ - EP^2EQ \\
&E(P-EP)(Q-EQ)^2 = EPQ^2 - 2EPQEQ + 2EP(EQ)^2 - EPEQ^2 \\
&E(Q-EQ)^3 = EQ^3 - 3EQ^2EQ + 2(EQ)^3,
\end{aligned}
$$

and

$$
\begin{aligned}
E(P-EP)^4 &= EP^4 - 4EP^3EP + 6EP^2(EP)^2 - 3(EP)^4 \\
E(P-EP)^3(Q-EQ) &= EP^3Q - 3EP^2QEP + 3EPQ(EP)^2 - EP^3EQ \\
&\quad +3EP^2EPEQ - 3(EP)^3EQ \\
E(P-EP)^2(Q-EQ)^2 &= EP^2Q^2 - 2EPQ^2 - 2EP^2EQ + 4EPQEPEQ \\
&\quad +(EP)^2EQ^2 + EP^2(EQ)^2 - 3(EP)^2(EQ)^2 \\
E(P-EP)(Q-EQ)^3 &= EPQ^3 - 3EPQ^2EQ + 3EPQ(EQ)^2 - EPEQ^3 \\
&\quad +3EPEQEQ^2 - 3EP(EQ)^3
\end{aligned}
$$

$$E(Q-EQ)^4 = EQ^4 - 4EQ^3EQ + 6EQ^2(EQ)^2 - 2(EQ)^4. \tag{A–41}$$

We approximated moments of $r = \frac{Y'AY}{Y'BY}$ using the first four terms of Taylor expansion. To make it comparable, we used the same SAR(3) model and same parameters as we did in last two section. $\phi = 0.5$, $\mathbf{m} = \mathbf{0}$. The variance-covariance matrix $\mathbf{\Sigma}$ is defined as Eq. A–2 with $\sigma = 1$. We choose the Burg estimator, i.e.,$c_1 = 0.5$, $c_2 = 0.5$. Sample size $n = 10$.

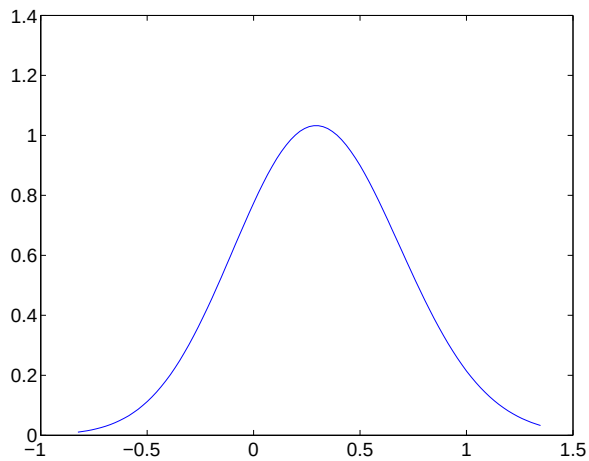

Figure A-7. Saddlepoint approximated pdf of burg estimator of ϕ in SAR(3) model with Symmetric Laplace marginal distribution. We approximate the moments of r using Taylor series expansion. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 10$.

Our approximation is not successful this time. We have used the same method to approximate the pdf of MLEs of VaR and CVaR under Asymmetric Laplace distribution in Chapter 2. And the approximation are very accurate according to our calculated PREs.

This method is not working now. Probably the reason is that the value of higher order moments of quadratic forms are very large. For example, the fourth moments of P and Q are to the order of 10^8. Under this situation, the approximation to the moments of r does not converge.

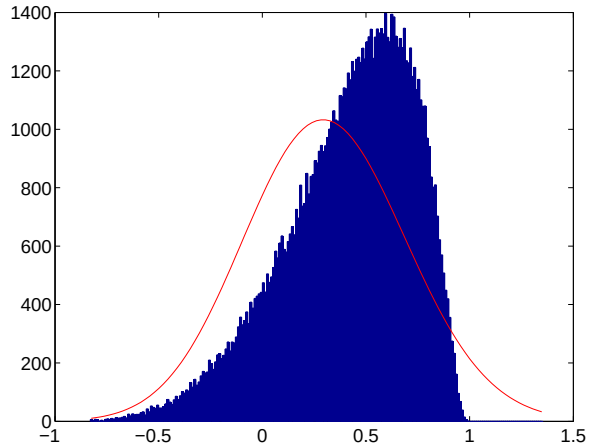

Figure A-8. Comparison of the approximated pdf and simulated histogram. The moments are approximated by Taylor expansion. We inflated the pdf by number of replication to match the histogram. The histogram is based on $10,000$ simulations. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 10$.

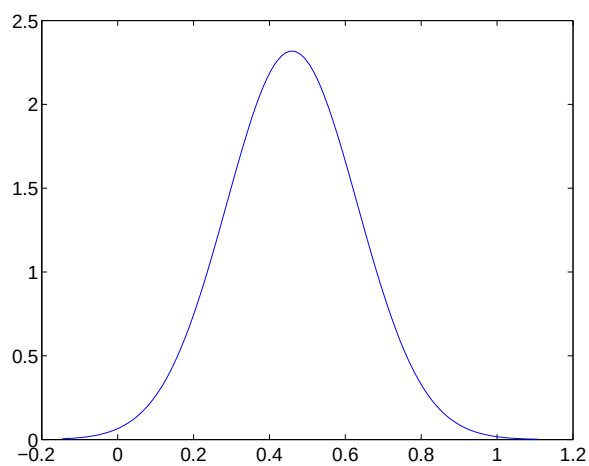

Figure A-9. Saddlepoint approximated pdf of burg estimator of ϕ in SAR(3) model with Symmetric Laplace marginal distribution.We approximated the moments of r using Taylor series expansion. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 50$.

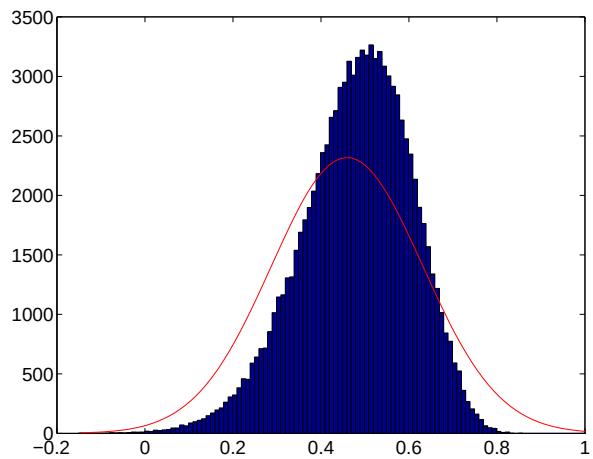

Figure A-10. Comparison of the approximated pdf and simulated histogram. The moments are approximated by Taylor expansion. We have inflated the pdf by number of replication to match the histogram. The histogram is based on $10,000$ simulations. $\phi = 0.5$,$\mathbf{m} = \mathbf{0}$, $\sigma = 1$, $c_1 = 0.5$, $c_2 = 0.5$, $n = 50$.

REFERENCES

[1] Abramowitz, M., and Stegun, I. (1972), *Handbook of Mathematical Functions*, New York: Dover, 9th edition.

[2] Acerbi, C., and Tasche, D. (2002), "Expected Shortfall: a Natural Coherent Alternative to Value at Risk," *Economic Notes*, 31, 379–388.

[3] Alam, K., and Wallenius, K. (1979), "Distribution of a Sum of Order Statistics," *Scandinavian Journal of Statistics*, 6, 123–126.

[4] Artzner,P., Delbaen, F., Eber, J., and Heath, D. (1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 203-228.

[5] Ayebo, A., and Kozubowski, T.J., (2003), "An Asymmetric Generalization of Gaussian and Laplace Laws," *Journal of Probability and Statistical Science*, 1, 187–210.

[6] Bain, L.J., and Engelhardt, M., (1973), "Interval Estimation for the Two-Parameter Double Exponential Distribution," *Technometrics*, 15, 875–887.

[7] Balakrishnan, N., and Basu, A.P., (1995), *The Exponential Distribution: Theory, Methods and Applications*, Gordon and Breach.

[8] T. Bollerslev (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31,307–327.

[9] Box, G.E.P., (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance of the One-Way Classification," *The Annals of Mathematical statistics*, 25, 290–302.

[10] Brockwell, P.J., Dahlhaus, R., and Trindade, A.A., (2005), "Modified Burg Algorithms for Multivariate Subset Autoregression," *Stat Sinica*, 15, 197–213.

[11] Brockwell, P.J. and Devis, R.A. (1991), *Time Series: Theory and Methods*, Springer-Verlag, New York, 2th edition.

[12] BBrockwell, P.J., and Davis, R.A., (2002), *Introduction to Time Series and Forecasting* , New York: Springer-Verlag, 2nd edition.

[13] Burnham, K.P., and Anderson, D.R., (1998), *Model Selection and Inference: a practical information-Theoretic Approach*, New York: Springer-Verlag.

[14] Butler, R.W. and Wood, A.T.A. (2002), "Laplace Approximations for Hypergeometric Functions with Matrix Argument," *Annals of Statistics*, 30, 1155–1177.

[15] David, H., and Nagaraja, H. (2003), *Order Statistics*, Hoboken: Wiley.

[16] Damsleth, E., and El-Shaarawi, A.H., (1989), "ARMA Models with Double-Exponentially Distributed Noise," *Journal of Royalty Statistical Society B*, 51, 61–69.

[17] Daniels, H.E. (1954), "Saddlepoint Approximations in Statistics,," *Annals of Mathematical Statistics*, 25, 631–650.

[18] Daniels, H.E., (1983), "Saddlepoint Approximations for Estimating Equations," *Biometrika*, 70, 89–96.

[19] Dewald,L.S., and Lewis, P.A.W., (1985) "A New Laplace Second-Order Autoregressive Time-Series Model-NLAR(2)," *IEEE Transactions on Information theory*, IT-31, 5, 115–124.

[20] Duffie, D., and Pan, J. (1997), "An Overview of Value-at-Risk," *Journal of Derivatives*, 4, 7–49.

[21] Easton, G.S., and Ronchetti, E. (1986), "General Saddlepoint Approximations with Applications to L Statistics," *Journal of the American Statistical Association*, 81, 420–430.

[22] Engle, R.F.(1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987–1007.

[23] Fernandez, C., Osiewalski, J., and Steel, M.F.J., (1995), "MModeling and Inference with v-Spherical Distribution," *Journal of the American Statistical Association*, 90, 1331–1340.

[24] Fernandez, C. and Steel, M.J.F. (1998), "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359–371.

[25] Giurcanu, M., and Trindade, A.A. (2005), "Establishing Consistency of M-estimators under Concavity with an Application to some Financial Risk Measures," *Technical Report 2005–024*.

[26] Goutis, C. and Casella, G. (1999), "Explaining the Saddlepoint Approximation", *the American Statistician*, 53, 216–224.

[27] Hsu, H., (2005), "Comparing Estimators of VaR and CVaR under the Asymmetric Laplace Distribution," *Master's Thesis*,Dept. of Statistics, University of Florida.

[28] Huzurbazar, S. (1999), "Practical Saddlepoint Approximations," *the American Statistician*, 53, 225–232.

[29] Koenker,R. (2005), *Quantile Regression*, Econometric Society Monograph Series. Cambridge University Press, Cambridge.

[30] Kotz,S., Kozubowski, T., and Podgórski, K.(2001), *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Boston: Birkhäuser.

[31] Komunjer,I., (2006), "Asymmetric Power Distribution: Theory and Application to Risk Measurement," *Journal of Applied Econometrics*, Acceped.

[32] Kozubowski, T., and Podgórski, K. (1999), "A Class of Asymmetric Distributions," *Actuarial Research Clearing House*, 1, 113–134.

[33] Kuester, K., Mittnik, S., and Paolella, M.S., (2006), "Value-at-Risk Prediction: A Comparison of Alternative Strategies," *Journal of Financial Econometrics*, 4, 53–89.

[34] Lehmann, E.L. (1983), *Theory of Point Estimation*, John Wiley, New York.

[35] Lieberman, O., (1994), "A Laplace Approximation to the Moments of a Ratio of Quadratic Forms," *Biometrika*, 81, 681–690.

[36] Lugannani, R., and Rice, S. (1980), "Saddlepoint Approximation for the Distribution of Sum of Independent Random Variables," *Advances in Applied Probability*, 12, 475–490.

[37] Markowitz, H. (1952), "Portfolio Selection," *Journal of Finance*, 7.

[38] McCullagh, P., (1987), *Tensor Methods in Statistics*, London: Chapman and Hall.

[39] McFarland, J.W., Pettit, R.R., Sung, S.K., 1982, "The Distribution of Foreign Exchange Price Changes: Trading Day Effects and Risk Measure," *Journal of Finance*, 37, 693–715.

[40] Morgan,J.P. (1997), *RiskMetrics Technical Documents*, 4th edition.

[41] Novković, M., (1998), "On Laplace Autoregressive Time Series Models," working paper.

[42] Paige, R., and Trindade, A.A., (2003), "Practical Small Sample Inference for Single Lag Subset Autoregressive Models," *Technical Report 2003-15*, Dept.of statistics, University of Florida.

[43] Patton, A.J., (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2, 130–168.

[44] Pflug, G. (2000), *Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk", in Probabilistic Constrained Optimization: Methodology and Applications*, S.Uryasev (Ed.), Dordrecht: Kluwer Academic Publishers.

[45] Rochafellar, R., and Uryasev, S.(2002), "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking and Finance*, 26. 1443–1471.

[46] Rochafellar, R.T., and Uryasev, S. (2000), "Optimization of Conditional Value-at-Risk," *Journal of Risk*, 2, 21–41.

[47] Ruppert, D. (2004), *Statistics and Finance: An Introduction*, Spring Verlag. Germany.

[48] Shumway.R., and Stoffer, D., (2000) *Time Series analysis and its applications*, Springer.

[49] Stigler, S. (1973), "The Asymptotic Distribution of the Trimmed Mean," *The Annals of Statistics*, 1(3), 472–477.

[50] Trindade, A.A., and Zhu,Y., (2007), "Approximating the Distributions of Estimators of Financial Risk under an asymmetric Laplace Law," *Computational Statistics & Data Analysis*, Vol 51, 3433-3447.

BIOGRAPHICAL SKETCH

Yun Zhu was born in Chengdu, China. She received her Ph.D. degree in statistics from University of Florida. She received her bachelor's in finance with concentration in international finance from the Southwestern University of Finance and Economics in China. She is interested in applying statistical methodologies in financial modeling.