Univariate and Bivariate ACD Models for High-Frequency Data Based on Birnbaum-Saunders and Related Distributions

UNIVARIATE AND BIVARIATE ACD MODELS FOR HIGH-FREQUENCY DATA BASED ON BIRNBAUM-SAUNDERS AND RELATED DISTRIBUTIONS

BY

TAO TAN

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

McMaster University

Doctor of Philosophy (2018) McMaster University
(Mathematics) Hamilton, Ontario, Canada

TITLE:	Univariate and Bivariate ACD Models for High-Frequency Data Based on Birnbaum-Saunders and Related Distributions
AUTHOR:	Tao Tan
SUPERVISOR:	Professor N. Balakrishnan
NUMBER OF PAGES:	xv, 132

To my husband Dr. Hanjiang Dong

To my parents Zuwei Tan & Lunzhen Li

To my sister Rong Tan

Abstract

This thesis proposes a new class of bivariate autoregressive conditional median duration models for matched high-frequency data and develops some inferential methods for an existing univariate model as well as the bivariate models introduced here to facilitate model fitting and forecasting.

During the last two decades, the autoregressive conditional mean duration (ACD) model has been playing a dominant role in analyzing irregularly spaced high-frequency financial data. Univariate ACD models have been extensively discussed in the literature. However, some major challenges remain. The existing ACD models do not provide a good distributional fit to financial durations, which are right-skewed and often exhibit unimodal hazard rates. Birnbaum-Saunders (BS) distribution is capable of modeling a wide variety of positively skewed data. Median is not only a robust measure of central tendency, but also a natural scale parameter of the BS distribution. A class of conditional median duration models, the BS-ACD and the scale-mixture BS ACD models based on the BS, BS power-exponential and Student-t BS (BSt) distributions, have been suggested in the literature to improve the quality of the model fit.

The BSt-ACD model is more flexible than the BS-ACD model in terms of kurtosis and skewness. In Chapter 2, we develop the maximum likelihood estimation method

for the BSt-ACD model. The estimation is performed by utilizing a hybrid of optimization algorithms. The performance of the estimates is then examined through an extensive Monte Carlo simulation study. We also carry out model discrimination using both likelihood-based method and information-based criterion. Applications to real trade durations and comparison with existing alternatives are then made.

The bivariate version of the ACD model has not received attention due to non-synchronicity. Although some bivariate generalizations of the ACD model have been introduced, they do not possess enough flexibility in modeling durations since they are conditional mean-based and do not account for non-monotonic hazard rates.

Recently, the bivariate BS (BVBS) distribution has been developed with many desirable properties and characteristics. It allows for unimodal shapes of marginal hazard functions. In Chapter 3, upon using this bivariate BS distribution, we propose the BVBS-ACD model as a natural bivariate extension of the BS-ACD model. It enables us to jointly analyze matched duration series, and also capture the dependence between the two series. The maximum likelihood estimation of the model parameters and associated inferential methods have been developed. A Monte Carlo simulation study is then carried out to examine the performance of the proposed inferential methods. The goodness-of-fit and predictive performance of the model are also discussed. A real bivariate duration data analysis is provided to illustrate the developed methodology.

The bivariate Student-t BS (BVBSt) distribution has been introduced in the literature as a robust extension of the BVBS distribution. It provides greater flexibility in terms of the kurtosis and skewness through the inclusion of an additional shape

parameter. In Chapter 4, we propose the BVBSt-ACD model as a natural extension of the BSt-ACD model to the bivariate case. We then discuss the maximum likelihood estimation of the model parameters. A simulation study is carried out to investigate the performance of these estimators. Model discrimination is then done by using information-based criterion. Methods for evaluating the goodness-of-fit and predictive ability of the model are also discussed. A simulated data example is used to illustrate the proposed model as compared to the BVBS-ACD model.

Finally, in Chapter 5, some concluding comments are made and also some problems for future research are mentioned.

Key words: High-frequency financial data; Autoregressive conditional duration model; Conditional quantile duration; Student-t Birnbaum-Saunders distribution; Bivariate Birnbaum-Saunders distribution; Bivariate Student-t Birnbaum-Saunders distribution; Maximum likelihood estimation; Nelder-Mead algorithm; BFGS method; Monte Carlo simulation; Density forecast; Goodness-of-fit; Model discrimination; Information-based criterion.

Acknowledgements

I would like to express my sincere appreciation to my supervisor, Prof. N. Balakrishnan, for accepting me as his student, for giving me many opportunities for success, for his guidance, support, encouragement and great patience throughout my Ph.D. journey, and also for careful reading of my thesis. Without him, this thesis would not have been completed and I would not be in the position I am today.

I am very grateful to my supervisory committee members, Prof. Shui Feng and Prof. Aaron Childs for their valuable advice which has definitely improved the content of this thesis, along with my external examiner, Prof. Nalini Ravishanker and defense chair Prof. Ben Bolker. I thank Prof. John Maheu for valuable discussions.

I would also like to thank all my friends and fellow students who inspired and helped me in various ways. I am thankful to all the faculty members and staff for all their help during my graduate studies in the Department of Mathematics and Statistics.

Finally, special thanks go to my husband, my parents and my sister for their endless love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 High-frequency financial data

1.1.1 Features of high-frequency financial data

Due to the rapid development of computing power, storage capacity and the adoption of electronic trading systems, a massive amount of financial data has become widely accessible on a transaction-by-transaction basis, so-called high-frequency financial data. High-frequency data are observations taken at a fine time scale (Tsay (2010)). Here, time is often measured in seconds. Typical examples include transaction times, transaction prices, bid and ask quotes and transaction volumes. High-frequency financial data play an important role in market microstructure theory, indicating the information with respect to fundamental asset prices and the behavior of market participants.

The distinctive feature of transaction data is the irregular spacing over time since the arrival times of market events (such as trades and quotes) are random. Durations are the waiting times between market events such as trades, price and volume changes, which are quite informative and become crucial in understanding market activity and the underlying microstructure of the financial market.

Trade duration is the time elapsed between two consecutive trades or transactions of equity. This type of data have some special features. In addition to irregular spacing mentioned above, trade durations display a cluster effect. Short durations tend to be followed by short durations, and likewise for long durations, which leads to positive autocorrelations. Trade durations often exhibit diurnal patterns due to the intraday periodicity of daily trading activities. Trading activities are heavier in the beginning and closing of the trading day, and usually thinner in the middle of the day. Thus, we usually observe short durations during opening and closing hours and long durations during lunch hour. Another important feature of trade durations is that they often show unimodal hazard rates; see Section 1.1.3 for details on assessing the shape of the hazard.

The high-frequency financial data used in this thesis is from the New York Stock Exchange (NYSE) Trade and Quote (TAQ) database (note: the data sets were generously provided by Dr. Chad R. Bhatti, V.P., Citizens Bank). Transactions recorded before 9:30 am and after 4:00 pm were omitted, and simultaneous transactions were treated as a single transaction.

1.1.2 Diurnal adjustment of durations

Trade durations exhibit strong diurnal patterns or time-of-day effects, which indicates high trading activities during opening and closing hours and less activity around noon. In the analysis of duration data and ACD framework, it is common to remove time-of-day effects first and then fit a ACD model to diurnally adjusted data.

The adjusted durations can be obtained by

$$X_i = \tilde{X}_i/\hat{\phi} = \tilde{X}_i/\exp(\hat{s}),$$

where $\tilde{X}_i$ is the raw duration and ϕ is the time-of-day effect. To estimate s, we use quadratic splines and indicator variables on the thirty-minute intervals from 9:30 am to 4:00 pm; see Engle and Russell (1998), Engle (2000), Tsay (2010), Bhatti (2010), Hautsch (2012) and Leiva *et al.* (2014).

1.1.3 Total Time on Test plot

The Total Time on Test (TTT) plot can be used to identify the shape of the hazard function (see Aarset (1987) and Azevedo *et al.* (2012)). Its theoretical counterpart is the scaled TTT transform. If a life distribution $F(y)$ is absolutely continuous and its survival function is $S(y) = 1 - F(y)$, the scaled TTT transform is $W(u) = H^{-1}(u)/H^{-1}(1)$ for $0 < u < 1$, where $H^{-1}(u) = \int_0^{F^{-1}(u)} S(y)dy$. If $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$ denote the order statistics with common pdf $f(y)$ and cdf $F(y)$, the empirical scaled TTT transform is given by $W_n(r/n) = H_n^{-1}(r/n)/H_n^{-1}(1) = \left[\sum_{i=1}^{r} y_{i:n} + (n-r)y_{r:n}\right] / \sum_{i=1}^{n} y_{i:n}$, for $r = 1, \ldots, n$. The TTT plot is obtained by plotting $(r/n, W_n(r/n))$, $r = 1, \ldots, n$. For example, a linear TTT curve, which is randomly around the main diagonal of the unit square (the 45 °-line), implies a constant hazard function, a convex TTT curve indicates a decreasing hazard, and a concave TTT curve demonstrates an increasing hazard. If a TTT curve is concave (convex) and then convex (concave), the corresponding hazard is unimodal (bathtub).

1.2 Some ACD models

1.2.1 Standard ACD and Log-ACD

Engle and Russell (1998) introduced the following ACD(p, q) model, which is one of the primary tools used in modeling durations:

$$X_i = \psi_i \epsilon_i, \qquad \psi_i = \omega + \sum_{j=1}^{p} \alpha_j X_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}, \tag{1.1}$$

where X_i denotes the duration, $\psi_i = E(X_i|\mathcal{F}_{i-1})$ is the conditional expectation given $\mathcal{F}_{i-1}$, the information set available at the ith duration, ϵ_i is a series of i.i.d. positive random variables such that $E(\epsilon_i) = 1$, $\omega > 0$, $\alpha_j \geq 0$ and $\beta_j \geq 0$ are the unknown parameters, and $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$. To relax the positivity constraints on the model parameters, Bauwens and Giot (2000) suggested the Logarithmic ACD model of the form

$$X_i = \psi_i \epsilon_i, \qquad \ln\psi_i = \omega + \alpha\epsilon_{i-1} + \beta\ln\psi_{i-1}, \quad |\beta| < 1. \tag{1.2}$$

In these models, durations are assumed to follow either exponential or Weibull distributions. Therefore, the hazard functions are either constant or monotonic.

1.2.2 GG-ACD

It is necessary to briefly introduce the generalized gamma (GG) distribution before presenting details of the GG-ACD model. The PDF of the GG distribution with shape parameters v and η and scale parameter ω is given by

$$f_X(x) = \frac{\eta}{\omega\Gamma(v)} \left(\frac{x}{\omega}\right)^{v\eta-1} \exp\left[-\left(\frac{x}{\omega}\right)^{\eta}\right], x > 0, v, \omega, \eta > 0, \tag{1.3}$$

where $\Gamma(v) = \int_0^\infty u^{v-1}\exp(-u)du$ is the complete gamma function. The exponential ($v = \eta = 1$), gamma ($\eta = 1$), lognormal ($v \to \infty$) and Weibull ($v = 1$) distributions are all special cases of the GG distribution, denoted by $GG(v, \eta, \omega)$. The CDF of the GG distribution can be expressed as $F_X(x) = \frac{\Gamma((x/\omega)^\eta, v)}{\Gamma(v)}$, where $\Gamma(y, v) = \int_0^y u^{v-1}\exp(-u)du$ is the lower incomplete gamma function.

If $X \sim GG(v, \eta, \omega)$, then $E(X^k) = \frac{\omega^k \Gamma(v+k/\eta)}{\Gamma(v)}$. In particular, $E(X) = \frac{\omega\Gamma(v+1/\eta)}{\Gamma(v)}$ and $V(X) = \omega^2 \left(\frac{\Gamma(v+2/\eta)}{\Gamma(v)} - \left(\frac{\Gamma(v+1/\eta)}{\Gamma(v)}\right)^2\right)$.

The GG distribution can produce all common types of hazard function (see Bhatti (2010)):

Table 1.1: Hazard function shapes for the GG distribution

	$v\eta - 1 < 0$	$v\eta - 1 > 0$	$v\eta - 1 = 0$
$\eta = 1$	Decreasing	Increasing	Constant
$\eta > 1$	Bathub	Increasing	Increasing
$\eta < 1$	Decreasing	Unimodal	Decreasing

To allow for non-monotonic hazards, Lunde (1999) proposed the GG-ACD model. It is one of the most flexible conditional mean duration models, which can be written as

$$X_i = \psi_i \epsilon_i, \qquad \ln \psi_i = \alpha + \beta \ln \psi_{i-1} + \gamma \left[\frac{X_{i-1}}{\psi_{i-1}}\right], \quad |\beta| < 1, \tag{1.4}$$

where $X_i \sim \text{GG}(v, \eta, \psi_i)$, $\epsilon_i \overset{\text{iid}}{\sim} \text{GG}(v, \eta, 1)$ and $\psi_i = E(X_i \mid \mathcal{F}_{i-1}) = \frac{\omega_i \Gamma(v+1/\eta)}{\Gamma(v)}$. The pdf of GG-ACD model is given by

$$f(x_i) = \frac{\eta}{\varphi(v,\eta)\psi_i\Gamma(v)}\left[\frac{x_i}{\varphi(v,\eta)\psi_i}\right]^{v\eta-1}\exp\left(-\left[\frac{x_i}{\varphi(v,\eta)\psi_i}\right]^{\eta}\right), i = 1, \ldots n, \quad (1.5)$$

where $\varphi(v, \eta) = \Gamma(v)/\Gamma(v + 1/\eta)$. The associated survival function is given by

$$S_{\text{GG}}(x_i; v, \eta, \psi_i) = 1 - \Gamma({x_i}^{\eta}[\psi_i\varphi(v,\eta)]^{-\eta}, v)/\Gamma(v), \;\; x_i, v, \eta, \psi_i > 0. \quad (1.6)$$

1.2.3 BS-ACD

Bhatti (2010) introduced the autoregressive conditional median duration model based on the BS distribution. The primary motivation for the BS-ACD model is to provide an alternative to existing ACD models and enhance the quality of the model fit. The BS distribution fits duration data well in terms of the shapes of its density and hazard functions. The BS-ACD model is specified in terms of a time-varying conditional median duration $\sigma_i = {F_{X_i}}^{-1}(0.5|\mathcal{F}_{i-1})$, where $F_{X_i}^{-1}$ is the inverse CDF or quantile function of the model, instead of the conditional mean duration in the usual ACD model, viz.,

$$\ln \sigma_i = \alpha + \beta \ln \sigma_{i-1} + \gamma\left[\frac{X_{i-1}}{\sigma_{i-1}}\right], \quad |\beta| < 1. \quad (1.7)$$

Median is not only a robust measure of central tendency, but also a natural parameter of the BS distribution. The associated PDF with the BS-ACD model is given by

$$f_{X_i}(x_i; \kappa, \sigma_i) = \frac{1}{2\sqrt{2\pi}\kappa\sigma_i}\left[\left(\frac{\sigma_i}{x_i}\right)^{\frac{1}{2}} + \left(\frac{\sigma_i}{x_i}\right)^{\frac{3}{2}}\right]\exp\left[-\frac{1}{2\kappa^2}\left(\frac{x_i}{\sigma_i} + \frac{\sigma_i}{x_i} - 2\right)\right], \quad x_i, \kappa, \sigma_i > 0. \quad (1.8)$$

The corresponding survival function is given by

$$S_{\mathrm{BS}}(x_i, \kappa, \sigma_i) = 1 - \Phi\left[\frac{1}{\kappa}\left(\sqrt{\frac{x_i}{\sigma_i}} - \sqrt{\frac{\sigma_i}{x_i}}\right)\right], \quad i = 1, 2, \ldots, n. \tag{1.9}$$

1.3 A brief literature review

The analysis of high-frequency financial data has received increasing attention recently in the literature. Recent research interest in duration models was largely spurred by the original work of Engle and Russell (1998), which introduced the autoregressive conditional duration (ACD) model to analyze the irregularly spaced trade durations of single assets. See Hautsch (2012), Pacurar (2008) and Saranjeet and Ramanathan (2018) for surveys on this topic.

Univariate ACD models have been well studied in the literature. An important extension is the logarithmic version of the ACD model, introduced by Bauwens and Giot (2000), which relaxes the positivity constraints on the model parameters. There exists different distributional assumptions on the conditional durations in the literature such as the exponential and Weibull distributions in Engle and Russell (1998), and the generalized gamma (GG) distribution in Lunde (1999). The first two distributions have either a flat or monotonic hazard rates. However, the last one is quite flexible and can produce all common types of hazard functions. Thus, the GG-ACD model is regarded as the status quo in the literature of ACD models. The parametric formulation of the ACD models determines the dynamics of the conditional mean duration.

The Birnbaum-Saunders (BS) distribution, introduced by Birnbaum and Saunders (1969), has been used quite successfully to model a wide variety of positively skewed data; see also Johnson *et al.* (1995) for a concise review of the classical BS distribution and Leiva (2016) for an overview on recent theoretical developments and applications. All developments on BS and related distributions have recently been thoroughly reviewed in the discussion paper by Balakrishnan and Kundu (2018). Due to its good properties, Jin and Kawczak (2003) claimed that BS distribution deserves more attention in the study of high-frequency financial data, and these authors then utilized the BS kernel to estimate the probability density function of trade durations. As illustrated in Kundu *et al.* (2008), the hazard rate of BS distribution has a unimodal shape. Considering that the hazard function for duration data is usually unimodal, Bhatti (2010) suggested the BS-ACD model possessing a realistic distribution and an ACD specification in terms of conditional median instead of conditional mean in the orginal ACD model. Median is not only a robust measure of central tendency but also a natural parameter in the BS distribution. Thus, the model fit can be enhanced by modeling the conditional median duration instead of the conditional mean duration as done in the original ACD model. A recent extension of this conditional median-type ACD model is due to Leiva *et al.* (2014), which considered the BS, BS power-exponential and Student-t BS distributions, so-called scale-mixture BS distribution family (see Balakrishnan *et al.* (2009) and Díaz-García and Leiva (2005) for details on SBS distributions) and developed corresponding expectation-maximization (EM) algorithms for SBS-ACD models. Moreover, Saulo and Leão (2017) discussed the conditional median-type ACD model based on the log-symmetric distributions. Saulo *et al.* (2017a) compared the mean-based and median-based BS-ACD models regarding model-fitting, forecasting and influence analysis. In general, their study confirmed that the conditional median-type ACD models based on BS and associated distributions are superior to the existing ACD model (the Generalized Gamma ACD model, see Lunde (1999)) in terms of model-fitting and forecasting.

1.4 Motivation and objectives

Despite the vast literature on duration series of single assets, a bivariate version of the ACD model has not received enough attention because of nonsynchronous trading/asynchronous observations. Only a few papers, with regard to the ACD model, have focused on jointly analyzing the duration series of trades and quotes of a single asset. Engle and Lunde (2003) proposed a censored bivariate ACD model using a semiparametric estimation approach. Mosconi and Olivetti (2005) introduced bivariate Exponential and Weibull ACD models (see Balakrishnan and Lai (2009) for detailed discussions on these bivariate distributions) in the same line.

However, both models share some common drawbacks. First, they only take into account the duration of one stock, and do not include information given by another stock. New available information to the stock market may affect not only a specific stock, but also the related stocks in the same industry sector or even the whole stock market (Simonsen (2007)). As a result, new information may lead to dependence between trade durations in different stocks. Therefore, modeling the dependence between two or multiple duration series remains a challenge. Second, durations often show unimodal hazard rates. The above bivariate ACD models fail to capture this main feature of duration data. A simulation study in Grammig and Maurer (2000) showed that the misspecification of the hazard rate can severely deteriorate the predictive ability of the ACD model. Third, the median is a better measure of central tendency than the mean in a skewed distribution. Hence, the model fit can be improved by replacing the conditional mean in the original ACD model with the conditional median. Furthermore, there is no work till now on the topic of joint modeling of the the duration series for matched data.

Recently, Kundu *et al.* (2010) derived the bivariate Birnbaum-Saunders (BVBS) distribution through a transformation of the bivariate normal distribution (see Kundu *et al.* (2013) and Vilca *et al.* (2014a,b) for generalizations of the BVBS distribution). Its marginals are univariate BS distributions with unimodal hazard functions and its conditional distributions can be expressed in terms of normal distribution. Moreover, it has a correlation parameter which indirectly represents the dependence between the two BS random variables. The bivariate Student-t BS (BVBSt) distribution was suggested by Vilca *et al.* (2014a) as a robust extension of the BVBS distribution, allowing more flexibility in kurtosis and skewness achieved with the inclusion of an additional shape parameter.

Our major objective is to jointly model the trade durations of two assets for matched data, pairs of durations with the same starting time (see Simonsen (2007)). In this thesis, we construct a bivariate ACD model based on BVBS and Student-t BVBS distributions, which facilitate the joint analysis and also to detect the dependence between two matched duration series.

1.5 Distributions for high-frequency trade duration

1.5.1 BS distribution

The CDF of a two-parameter BS random variable X is given by

$$F_X(x;\kappa,\sigma) = \Phi\left[\frac{1}{\kappa}\left(\sqrt{\frac{x}{\sigma}} - \sqrt{\frac{\sigma}{x}}\right)\right], \quad x > 0, \tag{1.10}$$

where $\Phi(\cdot)$ is the standard normal CDF, $\kappa > 0$ is the shape parameter and $\sigma > 0$ is the scale parameter which is also the median. The corresponding PDF is

$$f_X(x;\kappa,\sigma) = \frac{1}{2\sqrt{2\pi}\kappa\sigma}\left[\left(\frac{\sigma}{x}\right)^{\frac{1}{2}} + \left(\frac{\sigma}{x}\right)^{\frac{3}{2}}\right]\exp\left[-\frac{1}{2\kappa^2}\left(\frac{x}{\sigma} + \frac{\sigma}{x} - 2\right)\right], \quad x > 0. \tag{1.11}$$

This distribution is simply denoted by $BS(\kappa,\sigma)$.

The BS distribution possesses the scaling property, namely, for $c > 0$, $cX \sim BS(\kappa, c\sigma)$, and in particular, $\frac{X}{\sigma} \sim BS(\kappa, 1)$, and moreover it has the reciprocal property, namely, $\frac{1}{X} \sim BS(\kappa, \frac{1}{\sigma})$. There is a monotone transformation between the BS and normal distributions given by

$$Z = \frac{1}{\kappa}\left(\sqrt{\frac{X}{\sigma}} - \sqrt{\frac{\sigma}{X}}\right), \tag{1.12}$$

or equivalently

$$X = \sigma\left[\frac{1}{2}\kappa Z + \sqrt{\left(\frac{1}{2}\kappa Z\right)^2 + 1}\right]^2, \tag{1.13}$$

where $Z \sim N(0,1)$. Using the transformation in (1.13), the mean, variance and the coefficients of of skewness (CS) and kurtosis (CK) can be shown to be

$$E(X) = \sigma\left(1 + \frac{1}{2}\kappa^2\right), \qquad V(X) = (\kappa\sigma)^2\left(1 + \frac{5}{4}\kappa^2\right), \tag{1.14}$$

$$CS = \frac{4\kappa(11\kappa^2 + 6)}{(5\kappa^2 + 4)^{\frac{3}{2}}}, \qquad CK = 3 + \frac{6\kappa^2(93\kappa^2 + 40)}{(5\kappa^2 + 4)^2}. \tag{1.15}$$

Kundu *et al.* (2008) have shown that the hazard function of the BS distribution

$$h_X(x;\kappa,\sigma) = \frac{f_X(x;\kappa,\sigma)}{1 - F_X(x;\kappa,\sigma)} \tag{1.16}$$

has an unimodal shape.

1.5.2 Generalized BS Distribution

A random variable X is said to have a generalized Birnbaum-Saunders distribution if it allows the stochastic representation

$$X = \sigma\left[\frac{1}{2}\kappa Z + \sqrt{\left(\frac{1}{2}\kappa Z\right)^2 + 1}\right]^2 \sim GBS(\kappa, \sigma; g) \tag{1.17}$$

and

$$Z = \frac{1}{\kappa}\left(\sqrt{\frac{X}{\sigma}} - \sqrt{\frac{\sigma}{X}}\right) \sim S(g), \tag{1.18}$$

where $\kappa > 0$ is the shape parameter, $\sigma > 0$ is the scale parameter which is also the median and Z follows a standard symmetrical distribution, denoted by $Z \sim S(g)$, with g being the kernel of the p.d.f. of Z (see Díaz-García and Leiva (2005) and Sanhueza *et al.* (2008)).

The GBS distribution has many good properties. For example, if $X \sim GBS(\kappa, \sigma; g)$, then

(1) $cX \sim GBS(\kappa, c\sigma; g)$ with $c > 0$, and $X^{-1} \sim GBS(\kappa, \sigma^{-1}; g)$;

(2) $U = Z^2 = \frac{1}{\kappa^2}\left(\frac{X}{\sigma} + \frac{\sigma}{X} - 2\right) \sim G_{\chi^2}(g)$, a generalized chi-square distribution with one degree of freedom. The pdf of U is given by $f_U(u) = cg(u)u^{-1/2}$, with $u > 0$, where $cg(\cdot)$ is the pdf of $Z \sim S(g)$;

(3) The quantile function of X is given by $x(q) = F_X^{-1}(p) = \frac{\sigma}{4}(\kappa z_q + [\kappa^2 z_q^2 + 4]^{1/2})^2$, where z_q is the qth quantile of $Z \sim S(g)$. If $q = 0.5$, then $x(0.5) = \sigma$ and so σ is the median;

(4) If $E(Z^k)$ exists, then

$$E(X^k) = \sigma^k \sum_{j=0}^{k} \binom{2k}{2j} \sum_{i=0}^{j} \binom{j}{i} E(U^{k+i-j}) \left(\frac{\kappa}{2}\right)^{2(k+i-j)} \tag{1.19}$$

(see Sanhueza *et al.* (2008) and Díaz-García and Leiva (2007)).

Thus, the mean, variance and the coefficients of skewness (CS) and kurtosis (CK) of X are given by

$$E(X) = \frac{\sigma}{2}(2 + u_1\kappa^2), \tag{1.20}$$

$$Var(X) = \frac{\sigma^2\kappa^2}{4}[4u_1 + (2u_2 - u_1^2)\kappa^2], \tag{1.21}$$

$$CS = \frac{4\kappa[(3u_2 - 3u_1^2) + \frac{1}{2}(2u_3 - 3u_1u_2 + u_1^3)\kappa^2]}{[4u_1 + (2u_2 - u_1^2)\kappa^2]^{\frac{3}{2}}}, \tag{1.22}$$

$$CK = \frac{16u_2 + (32u_3 - 48u_1u_2 + 24u_1^3)\kappa^2 + (8u_4 - 16u_1u_3 + 12u_1^2u_2 - 3u_1^4)\kappa^4}{[4u_1 + (2u_2 - u_1^2)\kappa^2]^2}, \tag{1.23}$$

where $u_k = E(U^k)$ and $U \sim G_{\chi^2}(g)$ (see Sanhueza *et al.* (2008)).

Table 1.2: Moment of the Normal and Student-t distribution

Distribution	$u_1(g)$	$u_2(g)$	$u_3(g)$	$u_4(g)$
Normal	1	3	15	105
Student-t	$\frac{\nu}{(\nu-2)}$	$\frac{3\nu^2}{(\nu-2)(\nu-4)}$	$\frac{15\nu^3}{(\nu-2)(\nu-4)(\nu-6)}$	$\frac{105\nu^4}{(\nu-2)(\nu-4)(\nu-6)(\nu-8)}$
	$\nu > 2$	$\nu > 4$	$\nu > 6$	$\nu > 8$

(see Saulo *et al.* (2017b)).

1.5.3 BSt Distribution

In (1.17), if $g(\cdot)$ is normal, then $X \sim BS(\kappa, \sigma)$. On the other hand, if $g(\cdot)$ is t kernel, then $X \sim \mathrm{BS}t(\kappa, \sigma, \nu)$. When ν goes to infinity, the BSt distribution tends to be the BS distribution.

If $X \sim \mathrm{BS}t(\kappa, \sigma, \nu)$, the corresponding PDF of X is given by

$$f_X(x; \kappa, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)\kappa\sigma}\left[\left(\frac{\sigma}{x}\right)^{\frac{1}{2}} + \left(\frac{\sigma}{x}\right)^{\frac{3}{2}}\right]\left[1 + \frac{\left(\frac{x}{\sigma} + \frac{\sigma}{x} - 2\right)}{\nu\kappa^2}\right]^{-\frac{\nu+1}{2}}, \quad (1.24)$$

where $\kappa > 0$ and $\nu > 0$ are shape parameters, ν is also known as the degree of freedom, and $\sigma > 0$ is the scale parameter.

The corresponding CDF can be written as

$$F(x; \kappa, \sigma, \nu) = \frac{1}{2}\left\{1 + I_{[a(x)^2]/[a(x)^2+\nu]}\left(\frac{1}{2}, \frac{\nu}{2}\right)\right\}, \quad (1.25)$$

where $a(x) = (\sqrt{x/\sigma} - \sqrt{\sigma/x})/\kappa$ and $I_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt / \int_0^1 t^{a-1}(1-t)^{b-1}dt$ is the incomplete beta function ratio (see Azevedo *et al.* (2012)).

Similarly, there is a monotone transformation between the BSt and Student-t distributions. Using the transformation in (1.17), the mean, variance and the coefficients of skewness and kurtosis can be shown to be

$$E(X) = \frac{\sigma}{2}\left(2 + \frac{\nu}{\nu-2}\kappa^2\right),\ \nu > 2, \tag{1.26}$$

$$Var(X) = \frac{\sigma^2\kappa^2\nu}{4(\nu-2)}\left[4 + \frac{\nu(5\nu-8)\kappa^2}{(\nu-2)(\nu-4)}\right],\ \nu > 4, \tag{1.27}$$

$$CS = \frac{\frac{4\kappa(\nu-1)v^2}{(\nu-4)(\nu-2)^2}\left[6 + \frac{\nu(11\nu-18)\kappa^2}{(\nu-6)(\nu-2)}\right]}{\left[\frac{4\nu}{(\nu-2)} + \frac{\nu^2(5\nu-8)\kappa^2}{(\nu-4)(\nu-2)^2}\right]^{3/2}},\ \nu > 6, \tag{1.28}$$

$$CK = \frac{\frac{3\nu^2}{(\nu-2)(\nu-4)}\left[16 + \frac{8\nu(15\nu^2-42\nu+32)\kappa^2}{(\nu-6)(\nu-2)^2} + \frac{\nu^2(211\nu^3-894\nu^2+1288\nu-640)\kappa^4}{(\nu-8)(\nu-6)(\nu-2)^3}\right]}{\left[\frac{4\nu}{(\nu-2)} + \frac{\nu^2(5\nu-8)\kappa^2}{(\nu-4)(\nu-2)^2}\right]^2},\nu > 8. \tag{1.29}$$

The BSt distribution also possesses unimodal hazard; see Azevedo *et al.* (2012) for a thorough study of hazard functions of this family of distributions.

1.5.4 Bivariate BS distribution

Kundu *et al.* (2010) introduced the bivariate Birnbaum-Saunders (BVBS) distribution. A bivariate random vector (X_1, X_2) is said to follow a BVBS distribution with parameters κ_1, σ_1, κ_2, σ_2, ρ, if its joint CDF can be written as

$$F_{X_1,X_2}(x_1,x_2) = \Phi_2\left[\frac{1}{\kappa_1}\left(\sqrt{\frac{x_1}{\sigma_1}} - \sqrt{\frac{\sigma_1}{x_1}}\right), \frac{1}{\kappa_2}\left(\sqrt{\frac{x_2}{\sigma_2}} - \sqrt{\frac{\sigma_2}{x_2}}\right); \rho\right], \quad x_1, x_2 > 0, \tag{1.30}$$

where $\kappa_1,\ \sigma_1,\ \kappa_2,\ \sigma_2 > 0$, $-1 < \rho < 1$, and $\Phi_2(z_1, z_2; \rho)$ is the joint CDF of a standard bivariate normal vector (Z_1, Z_2) with correlation coefficient ρ. We denote this distribution simply by $BVBS(\kappa_1, \sigma_1, \kappa_2, \sigma_2, \rho)$. The corresponding joint PDF of

(X_1, X_2) is given by

$$f_{X_1,X_2}(x_1,x_2) = \phi_2\left[\frac{1}{\kappa_1}\left(\sqrt{\frac{x_1}{\sigma_1}}-\sqrt{\frac{\sigma_1}{x_1}}\right), \frac{1}{\kappa_2}\left(\sqrt{\frac{x_2}{\sigma_2}}-\sqrt{\frac{\sigma_2}{x_2}}\right); \rho\right] \times \frac{1}{2\kappa_1\sigma_1}\left[\left(\frac{\sigma_1}{x_1}\right)^{\frac{1}{2}}+\left(\frac{\sigma_1}{x_1}\right)^{\frac{3}{2}}\right]\frac{1}{2\kappa_2\sigma_2}\left[\left(\frac{\sigma_2}{x_2}\right)^{\frac{1}{2}}+\left(\frac{\sigma_2}{x_2}\right)^{\frac{3}{2}}\right], \tag{1.31}$$

where $\phi_2(z_1, z_2; \rho)$ is the joint PDF of $(Z_1,\, Z_2)$ defined by

$$\phi_2(z_1,z_2;\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(z_1^2+z_2^2-2\rho z_1 z_2)\right\}. \tag{1.32}$$

If $(X_1, X_2) \sim BVBS(\kappa_1, \sigma_1, \kappa_2, \sigma_2, \rho)$, then $X_j \sim BS(\kappa_j, \sigma_j)$, $j = 1, 2$. The conditional PDF of X_1, given $X_2 = x_2$, is given by

$$f_{X_1|X_2}(x_1|x_2) = \frac{1}{2\kappa_1\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}}\left[\left(\frac{\sigma_1}{x_1}\right)^{\frac{1}{2}}+\left(\frac{\sigma_1}{x_1}\right)^{\frac{3}{2}}\right] \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{1}{\kappa_1}\left(\sqrt{\frac{x_1}{\sigma_1}}-\sqrt{\frac{\sigma_1}{x_1}}\right)-\frac{\rho}{\kappa_2}\left(\sqrt{\frac{x_2}{\sigma_2}}-\sqrt{\frac{\sigma_2}{x_2}}\right)\right]^2\right\}. \tag{1.33}$$

The corresponding CDF of X_1, given $X_2 = x_2$, is

$$F_{X_1|X_2}(x_1|x_2) = \Phi\left\{\frac{\frac{1}{\kappa_1}\left(\sqrt{\frac{x_1}{\sigma_1}}-\sqrt{\frac{\sigma_1}{x_1}}\right)-\frac{\rho}{\kappa_2}\left(\sqrt{\frac{x_2}{\sigma_2}}-\sqrt{\frac{\sigma_2}{x_2}}\right)}{\sqrt{1-\rho^2}}\right\}. \tag{1.34}$$

The relationship between the BVBS and bivariate normal distributions can be represented by $Z_j = \frac{1}{\kappa_j}\left(\sqrt{\frac{X_j}{\sigma_j}}-\sqrt{\frac{\sigma_j}{X_j}}\right)$, or equivalently $X_j = \sigma_j\left[\frac{1}{2}\kappa_j Z_j + \sqrt{\left(\frac{1}{2}\kappa_j Z_j\right)^2+1}\right]^2$, for $j = 1, 2$, where $(Z_1, Z_2) \sim N(\mathbf{0}, \mathbf{\Sigma})$, with $\mathbf{\Sigma} = \left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)$ Thus, random numbers from the BVBS distribution can be easily generated using bivariate normal distribution (see Kundu *et al.* (2010)). Also, $\rho_{X_1,X_2} \approx \rho$ (see Balakrishnan and Zhu (2015)).

1.5.5 Bivariate Student-t BS Distribution

A bivariate random vector (X_1, X_2) is said to follow a bivariate Student-t BS distribution with parameters κ_1, σ_1, κ_2, σ_2, ρ, ν, if its joint CDF can be written as

$$F_{X_1,X_2}(x_1,x_2) = \boldsymbol{T}_2\left[\frac{1}{\kappa_1}\left(\sqrt{\frac{x_1}{\sigma_1}}-\sqrt{\frac{\sigma_1}{x_1}}\right), \frac{1}{\kappa_2}\left(\sqrt{\frac{x_2}{\sigma_2}}-\sqrt{\frac{\sigma_2}{x_2}}\right); \rho, \nu\right], \quad x_1, x_2 > 0, \tag{1.35}$$

where κ_1, σ_1, κ_2, $\sigma_2 > 0$, $-1 < \rho < 1$, $\nu > 2$ and $T_2(z_1, z_2; \rho, \nu)$ is the joint CDF of a bivariate t vector (Z_1, Z_2) with correlation coefficient ρ. We denote this distribution simply by BVBS$t(\kappa_1, \sigma_1, \kappa_2, \sigma_2, \rho, \nu)$.

The corresponding joint PDF of (X_1, X_2) is given by

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) = \boldsymbol{t}_2&\left[\frac{1}{\kappa_1}\left(\sqrt{\frac{x_1}{\sigma_1}}-\sqrt{\frac{\sigma_1}{x_1}}\right), \frac{1}{\kappa_2}\left(\sqrt{\frac{x_2}{\sigma_2}}-\sqrt{\frac{\sigma_2}{x_2}}\right); \rho, \nu\right] \\ &\times \frac{1}{2\kappa_1\sigma_1}\left[\left(\frac{\sigma_1}{x_1}\right)^{\frac{1}{2}}+\left(\frac{\sigma_1}{x_1}\right)^{\frac{3}{2}}\right]\frac{1}{2\kappa_2\sigma_2}\left[\left(\frac{\sigma_2}{x_2}\right)^{\frac{1}{2}}+\left(\frac{\sigma_2}{x_2}\right)^{\frac{3}{2}}\right], \end{aligned} \tag{1.36}$$

where $\boldsymbol{t}_2(z_1, z_2; \rho, \nu)$ is the joint PDF of $(Z_1,\, Z_2)$ defined by

$$\boldsymbol{t}_2(z_1, z_2; \rho, \nu) = \frac{1}{2\pi\sqrt{1-\rho^2}}\left[1+\frac{z_1^2+z_2^2-2\rho z_1 z_2}{\nu(1-\rho^2)}\right]^{-\frac{\nu+2}{2}} \tag{1.37}$$

If $(X_1, X_2) \sim$ BVBS$t(\kappa_1, \sigma_1, \kappa_2, \sigma_2, \rho, \nu)$, then $X_j \sim$ BS$t(\kappa_j, \sigma_j, \nu)$, $j = 1, 2$. The conditional PDF of X_1, given $X_2 = x_2$, is given by

$$f_{X_1|X_2}(x_1|x_2) = f_{q(x_2)}\left([a_{x_1}(\kappa_{1\rho,\sigma_1}) - \mu_1(x_2)]^2\right) A_{x_1}(\kappa_{1\rho,\sigma_1}), \tag{1.38}$$

where $f_{q(x_2)}(u) = \frac{\Gamma(\frac{\nu+2}{2})}{\sqrt{\pi}\Gamma(\frac{\nu+1}{2})}\left[\nu + q(x_2)\right]^{\frac{\nu+1}{2}}\left[\nu + q(x_2) + u\right]^{-\frac{\nu+2}{2}}$, $u \geq 0$, $q(x_2) = a_{x_2}^2(\kappa_2, \sigma_2)$, $a_{x_j}(\kappa_j, \sigma_j) = \frac{1}{\kappa_j}\left(\sqrt{\frac{x_j}{\sigma_j}} - \sqrt{\frac{\sigma_j}{x_j}}\right)$, $j = 1, 2$, $\kappa_{1\rho} = \sqrt{1-\rho^2}\kappa_1$, $\mu_1(x_2) = \rho a_{x_2}(\kappa_2, \sigma_2)$, and $A_{x_j}(\kappa_j, \sigma_j) = \frac{\partial a_{x_j}(\kappa_j,\sigma_j)}{\partial x_j} = \frac{1}{\kappa_j\sigma_j}\left[\left(\frac{\sigma_j}{x_j}\right)^{\frac{1}{2}} + \left(\frac{\sigma_j}{x_j}\right)^{\frac{3}{2}}\right]$.

The corresponding joint CDF of (X_1, X_2) can be written as

$$P(X_1 \leq x_1, X_2 \leq x_2) = G_2(a_{\boldsymbol{t}}(\boldsymbol{\kappa}, \boldsymbol{\sigma}); \rho, \nu) = G_2(a_{x_1}(\kappa_1, \sigma_1), a_{x_2}(\kappa_2, \sigma_2); \rho, \nu), \tag{1.39}$$

where

$$\begin{aligned} G_2(a_{\boldsymbol{x}}(\boldsymbol{\kappa}, \boldsymbol{\sigma}); \rho, \nu) = & \, G_2\left(\sqrt{\frac{\nu-1}{\nu+1}} a_{\boldsymbol{x}}(\boldsymbol{\kappa}, \boldsymbol{\sigma}); \rho, \nu - 1\right) \\ & + C_\nu \left\{ \frac{a_{x_1}(\kappa_1, \sigma_1)}{1 + \nu + a_{x_1}^2(\kappa_1, \sigma_1)} G\left(\frac{\sqrt{\nu}(a_{x_2}(\kappa_2, \sigma_2) - \rho a_{x_1}(\kappa_1, \sigma_1))}{\sqrt{1-\rho^2}\sqrt{1 + \nu + a_{x_1}^2(\kappa_1, \sigma_1)}}; \nu\right) \right. \\ & \left. + \frac{a_{x_2}(\kappa_2, \sigma_2)}{1 + \nu + a_{x_1}^2(\kappa_2, \sigma_2)} G\left(\frac{\sqrt{\nu}(a_{x_1}(\kappa_1, \sigma_1) - \rho a_{x_2}(\kappa_2, \sigma_2))}{\sqrt{1-\rho^2}\sqrt{1 + \nu + a_{x_2}^2(\kappa_2, \sigma_2)}}; \nu\right) \right\} \end{aligned} \tag{1.40}$$

where $C_\nu = \frac{\Gamma(\frac{\nu}{2})(\nu+1)^{(\nu-1)/2}}{2\sqrt{\pi}\Gamma\left(\frac{\nu+1}{2}\right)}$ and $G(.;\nu)$ denotes the cdf of the usual univariate Student-t distribution with ν degrees of freedom.

1.6 Likelihood inference

1.6.1 Maximum likelihood estimation

Likelihood inference is a procedure of making a statement about a data generating process (DGP). The DGP can be represented by a statistical model we select. The goal of likelihood inference is to make a statement about the unknown parameters $\boldsymbol{\theta}$ that governs the model. The likelihood principle states that all the revelent information about the parameters is contained in the observed data, represented by the likelihood function. Therefore, the likelihood function is a function of $\boldsymbol{\theta}$ given the observed data and associated with the probability of the data given the parameters. The maximum likelihood estimator (MLE) is then obtained by finding the value of $\boldsymbol{\theta}$

in the parameter space Θ which is "most likely" to have generated the observed data, i.e.,

$$\hat{\boldsymbol{\theta}}_{MLE} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\, L(\boldsymbol{\theta}|\boldsymbol{x}) \tag{1.41}$$

Setting the first derivative (gradient) of the log-likelihood function to zero

$$\frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{1.42}$$

the MLEs become the solution of the first-order conditions.

In many situations, however, there is no analytical solution or closed-form expressions. Hence, numerical optimization algorithms need to be used. Numerical methods for nonlinear optimization problems are iterative in nature. These algorithms begin by assuming starting values for the unknown parameters and then proceed iteratively. In our estimation method, we implemented a hybrid of numerical optimization algorithms, namely, NM followed by BFGS.

1.6.2 The NM method

Nelder-Mead (NM) simplex algorithm, proposed by Nelder and Mead (1965), is one of the most widely used direct search methods for nonlinear optimization in a multidimensional space

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}}\, f(\boldsymbol{x}) \tag{1.43}$$

which targets finding the minimum of a nonlinear objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, or alternatively the maximum of $-f$. Here, $\boldsymbol{x}^*$ is the solution. This algorithm is derivative-free and only requires the values of f. It preforms simplex search using heuristic ideas.

In geometry, a simplex is defined as the convex hull of a set of $(n+1)$ affinely independent points, $\{\boldsymbol{x_j}\}_{j=1}^{n+1}(\boldsymbol{x_j} \in \mathbb{R}^n)$, called vertices of the simplex. For example, a two-vertex simplex with one dimension is a line segment and a three-vertex simplex is a triangle with two dimensions.

The Nelder-Mead algorithm starts with a nondegenerate simplex and then modifies the simplex at each iteration using the operations of reflection, contraction, expansion and shrinking respectively. The Nelder-Mead method iteratively generates a sequence of transformations of the working simplex to approximate an optimal point of f. At each iteration, the vertices of the simplex are sorted according to the objective function values

$$f(\boldsymbol{x_1}) \leq f(\boldsymbol{x_2}) \leq \cdots \leq f(\boldsymbol{x_{n+1}}), \tag{1.44}$$

where $\boldsymbol{x_1}$ is is called the best vertex and $\boldsymbol{x_{n+1}}$ the worst.

The algorithm attempts to replace the worst vertex $\boldsymbol{x_{n+1}}$ with a new point of the form

$$\boldsymbol{x}(\delta) = (1+\delta)\overline{\boldsymbol{x}} - \delta \boldsymbol{x_{n+1}}, \tag{1.45}$$

where $\overline{\boldsymbol{x}}$ is the centroid of the convex hull of $\{\boldsymbol{x_j}\}_{j=1}^{n}$ given by

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i}.$$

Here, $\delta \in \mathbb{R}$ is a coefficient associated with a particular step, which can be the reflection, contraction and expansion, or shrinking step of the Nelder-Mead iteration. The algorithm terminates if either $f(\boldsymbol{x_{n+1}}) - f(\boldsymbol{x_1})$ is sufficiently small, or the number of function evaluations has exceeded an user-specified limit.

1.6.3 The BFGS method

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is one of the most popular quasi-Newton methods. BFGS not only possesses computational efficiency and good asymptotic convergence, but also very effective self-correcting properties (see Wright and Nocedal (1999)).

Suppose the objective function $f : \mathbb{R}^n \to \mathbb{R}$ in (1.43) is convex and twice continuously differentiable. The well-known Newton method is derivative-based. It uses the quadratic Taylor expansion of f at the current point $\boldsymbol{x_k}$:

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x_k}) + (\boldsymbol{x} - \boldsymbol{x_k})^T \bigtriangledown f(\boldsymbol{x_k}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x_k})^T H(\boldsymbol{x_k})(\boldsymbol{x} - \boldsymbol{x_k}), \tag{1.46}$$

where $H(\boldsymbol{x_k}) = \bigtriangledown^2 f(\boldsymbol{x_k})$ is the Hessian of $f(\boldsymbol{x})$ at $\boldsymbol{x_k}$. The First-Order Necessary Condition requires

$$\bigtriangledown f(\boldsymbol{x_k}) + H(\boldsymbol{x_k})(\boldsymbol{x} - \boldsymbol{x_k}) = 0 \tag{1.47}$$

which yields

$$\boldsymbol{x_{k+1}} = \boldsymbol{x_k} - H(\boldsymbol{x_k})^{-1} \bigtriangledown f(\boldsymbol{x_k}) \tag{1.48}$$

Newton method computes the Hessian directly and converges quickly. However, the method assumes $H(\boldsymbol{x_k})$ is nonsingular and positive-definite at each iteration and there is no guarantee that $f(\boldsymbol{x_{k+1}}) \leq f(\boldsymbol{x_k})$. For large n, the evaluation of H can be computationally expensive.

Various modifications of the Newtons method, called quasi-Newton methods, mitigate some of these problems. Instead of the exact computation of H, BFGS algorithm approximates H at each iteration by the update formula

$$H_{k+1} = H_k - \frac{(H_k s_k)(H_k s_k)^T}{s_k^T H_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{1.49}$$

where $s_k = \boldsymbol{x_{k+1}} - \boldsymbol{x_k}$ and $y_k = \bigtriangledown f(\boldsymbol{x_{k+1}}) - \bigtriangledown f(\boldsymbol{x_k})$. H_{k+1} satisfies the secant equation $H_{k+1} s_k = y_k$ and converges to $H(\boldsymbol{x}^*)$.

The BFGS quasi-Newton method with the analytic gradient will be faster, more stable and lead to more accurate estimates than a numerical gradient (see Bard (1974), Bolker (2008) and Mayorov (2011)).

1.6.4 Standard errors of MLEs

Under certain regularity conditions, the ML estimator $\hat{\boldsymbol{\theta}}$ is $\sqrt{n}$-consistent and the asymptotic distribution is

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, n\mathcal{I}(\boldsymbol{\theta_0})^{-1}\right), \tag{1.50}$$

where $\mathcal{I}(\boldsymbol{\theta_0}) = -E\left[\left.\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right] = -E\left[\boldsymbol{H(\theta_0)}\right]$ and $\boldsymbol{H} = \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta'}}$ is the Hessian matrix. In order to estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$, we consider the well-known consistent estimator $-\boldsymbol{H(\hat{\theta})}^{-1}$, which simply evaluates the second derivative matrix of the log-likelihood function at the MLEs. The square roots of its diagonal elements provide the standard errors of the estimates of the parameters.

1.7 Simulation study

Monte Carlo simulations play an important role in research. Lots of statistical models and methods are introduced or validated based on evidence from Monte Carlo simulations. In this thesis, we conduct simulation studies to investigate the performance of inferential methods developed here. For the BSt-ACD model, given a true parameter vector, various settings based on different sample sizes are considered. For the bivariate BS-ACD and BSt-ACD models, we fix vectors of true parameters. A number of different scenarios concerning sample sizes and levels of correlation are then considered. We first generate 1000 simulated datasets from the true models. For each setting and each estimate, the performance of the MLEs of the models under consideration are then evaluated in terms of the mean, coefficients of skewness and kurtosis, bias and root mean squared error (RMSE) of the MLEs over 1000 replicates.

1.8 Model discrimination

In this thesis, we use model discrimination to evaluate the performance of the BSt-ACD and BVBSt-ACD models, respectively. We consider a set of candidate models, which are the special cases of the concerned general model. Based on the general model, we are then interested in assessing how often the true model gets selected in the set of candidate models. such a model discrimination can be done either by Likelihood-based test or by Information-based criterion.

1.8.1 Likelihood-based method

We study the performance of the likelihood-ratio test for testing the null hypothesis that the data generating process (DGP) can be described by one of the simple models versus an alternative hypothesis that the DGP can be described by a member of the general model family other than the one specified under the null hypothesis. The likelihood-ratio test statistic can be expressed as $\Lambda = 2(\hat{l} - \hat{l_0})$, where $\hat{l_0}$ and $\hat{l}$ are the constrained and unconstrained maximized log-likelihood values. The large sample null distribution of Λ is simple. According to Wilks' theorem, under some regularity conditions, for testing $H_0 : \nu = \nu_0$ versus $H_1 : \nu = \nu_1$, $\Lambda \overset{a}{\sim} \chi^2_1$, where χ^2_1 represents a chi-square distribution with one degree of freedom. However, such results are only valid in the interior of the parameter space. In the case of $H_0 : \nu \to \infty$, where the parameter is at the boundary, the asymptotic null distribution of Λ can be approximated as $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$, i.e., a mixture of a point mass at zero and a χ^2_1 distribution with equal weights and $P[\Lambda \leq \lambda] = \frac{1}{2} + \frac{1}{2}P[\chi^2_1 \leq \lambda]$ (See Self and Liang (1987) and Balakrishnan and Pal (2013)).

1.8.2 Information-based criterion

Two popular model selection criteria, AIC and BIC, are used here. They can be expressed as follows: AIC$= -2l + 2k$ and BIC$= -2l + k\ln(n)$, where l stands for the maximized log-likelihood value, p is the number of model parameters to be estimated, and n is the sample size. The model with the minimum AIC or BIC is then chosen to be the working model.

1.9 Goodness-of-fit test

We can evaluate goodness-of-fit of the models under consideration by using in-sample density forecasts and the genaralized Cox-Snell (GCS) residuals.

Density forecasts have come to play an important role in both macroeconomics and finance (see Tay and Wallis (2000)). Diebold *et al.* (1998) introduced the density forecast technique, which could be employed to evaluate ACD models (see Bauwens *et al.* (2004), Hautsch (2012) and Leiva *et al.* (2014)). Let $\{x_i\}_{i=1}^n$ denote a series generated from the series of true predictive densities $\{f_i(x_i|\mathcal{F}_{i-1})\}_{i=1}^n$. Let $\{p_i(x_i|\mathcal{F}_{i-1}))\}_{i=1}^n$ be a series of one-step-ahead density forecasts. The null hypothesis is that the density forecasts should coincide with the true predictive densities, that is, $\{f_i(x_i|\mathcal{F}_{i-1})\}_{i=1}^n = \{p_i(x_i|\mathcal{F}_{i-1})\}_{i=1}^n$. Next, a series of probability integral transforms (PIT) of $\{x_i\}_{i=1}^n$ is obtained by $z_i = \int_{-\infty}^{x_i} p_i(u)du$. Under the null hypothesis, the distribution of $\{z_i\}_{i=1}^n$ should be i.i.d. $U(0,1)$.

Diebold *et al.* (1999) extended the above idea to the multivariate case. Here, we are especially interested in the bivariate case. Let $\{(x_{1i}, x_{2i})\}_{i=1}^n$ represent a bivariate series and $\{p_i(x_{1i}, x_{2i}|\mathcal{F}_{i-1})\}_{i=1}^n$ be a series of joint density forecasts. The forecasts can be decomposed into $p_i(x_{1i}, x_{2i}) = p_i(x_{1i})p_i(x_{2i}|x_{1i})$ and $p_i(x_{1i}, x_{2i}) = p_i(x_{2i})p_i(x_{1i}|x_{2i})$, $i = 1, \ldots, n$. By applying the PIT to $p_i(x_{1i})$, $p_i(x_{2i})$, $p_i(x_{1i}|x_{2i})$ and $p_i(x_{2i}|x_{1i})$, respectively, the corresponding series we obtain, z_1, z_2, $z_{1|2}$ and $z_{2|1}$, should be i.i.d $U(0,1)$ individually under the null hypothesis. According to Diebold *et al.* (1999), for good bivariate forecasts, not only the z series but also the pooled series $\{z_1, z_{2|1}\}$ and $\{z_2, z_{1|2}\}$ should each be i.i.d. $U(0,1)$.

We can use the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests to assess the uniformity of the z series and the Ljung-Box test to examine the independence of the z series (see Mitchell and Wallis (2011), Rossi and Sekhposyan (2014) and Leiva *et al.* (2014)). The Ljung-Box test here is defined as

$$H_0: \ \{z_i\} \text{ is independently distributed.}$$
$$H_1: \ \{z_i\} \text{ is not independently distributed.}$$

The test statistic is

$$Q_m = n(n+2)\sum_{h=1}^{m}\frac{\hat{\rho}_h^2}{n-h}$$

where $\hat{\rho}_h = \sum_{i=h+1}^{n}(z_i - \bar{z})(z_{i-h} - \bar{z})/\sum_{i=1}^{n}(z_i - \bar{z})^2$ and $\bar{z} = \sum_{i=1}^{n} z_i/n$. If $\{z_i\}$ is an i.i.d. sequence satisfying $E(z_i^2) < \infty$, then

$$(\sqrt{n}\hat{\rho}_1, \ldots, \sqrt{n}\hat{\rho}_m) \xrightarrow{d} N_m(\mathbf{0}, \boldsymbol{I})$$

as $n \to \infty$ (see Anderson (1971), Runde (1997), Fan and Yao (2003), Tsay (2010) and Fan and Yao (2017)), where $\boldsymbol{I}$ is the $m \times m$ identity matrix. Obviously, under H_0, $\sum_{h=1}^{m} n\hat{\rho}_h^2$ is then approximately χ_m^2 distribution. When n is large, $(n+2)/(n-h) \approx 1$. Therefore, under H_0, the Ljung-Box test statistics

$$Q_m = \sum_{h=1}^{m}\frac{n+2}{n-h}n\hat{\rho}_h^2 \approx \sum_{h=1}^{m} n\hat{\rho}_h^2 \xrightarrow{d} \chi_m^2$$

as $n \to \infty$.

Bhatti (2010) suggested to use the GCS residuals to assess the goodness-of-fit of the BS-ACD model. Suppose X is a random variable with survival function

$S(x)$. Regardless of the form of $S(x)$, the distribution of the random variable $Y = -\ln(S(X))$ is exp(1) (see Collett (2003)). The GCS residual is then defined as $r_i^{GCS} = -\ln(\hat{S}(x_i|\mathcal{F}_{i-1}))$, for $i = 1, \ldots, n$, where $\hat{S}(x_i|\mathcal{F}_{i-1})$ is the fitted conditional survival function of the ith trade duration (see Engle and Lunde (2003), Bhatti (2010) and Leiva *et al.* (2014)). Under the assumption that the model is correctly specified, irrespective of the ACD specification, the GCS residuals should be i.i.d. standard exponential distribution.

1.10 Out-of-sample forecast evaluation

It is important to evaluate the model's predictive performance on the data not used in the model estimation. So, we split the full data set into two subsets. We take the first part of the data set as the training set, or "in-sample" data, for the purpose of estimation and use the rest of the data as the test set or "out-of-sample" data, for prediction. The predictive performance of the model can then be assessed by using density forecasts detailed in Diebold *et al.* (1998) and Diebold *et al.* (1999), detailed above in Section 1.9.

1.11 Scope of the thesis

This thesis considers autoregressive conditional median duration models (ACD) in both univariate and bivariate settings. We first investigate the univariate BSt-ACD model and then propose natural bivariate extensions of the BS-ACD and BSt-ACD models for the analysis of matched durations.

The rest of this thesis proceeds as follows. In Chapter 2, we show how to implement the maximum likelihood estimation method to estimate the BSt-ACD model. Monte Carlo simulation study is conducted to evaluate the proposed method. The performance of the model is further examined by model discrimination. A real data example and comparison with two alternative models are finally provided.

In Chapters 3 and 4, we focus on the bivariate versions of ACD models. In Chapter 3, by making use of the BVBS distribution, we introduce the BVBS-ACD model to analyze two matched duration series jointly and measure the strength of dependence between them. The maximum likelihood estimation of model parameters and associated inferential methods are then discussed. The proposed model and its estimation method are then examined via a simulation study. The goodness-of-fit and predictive performance of the model are evaluated through in-sample and out-of-sample density forecasts. The methodology is illustrated finally by using a real high-frequency data.

In Chapter 4, we propose the BVBSt-ACD model and then discuss the maximum likelihood estimation of the model parameters based on a hybrid of optimization algorithms. We investigate the performance of the inferential methods developed here through a Monte Carlo simulation study. The goodness-of-fit and predictive performance of the model are also discussed. Model discrimination using information-based method is conducted. We give model fitting and forecasting results using the modeling procedure on a simulated bivariate dataset. The results show the superiority of the proposed approach in comparison with the BVBS-ACD model.

In Chapter 5, we give a summary of the thesis and suggest some possible directions for future research.

Chapter 2

BSt-ACD model

2.1 Introduction

The generalized Birnbaum-Saunders (GBS) distribution was introduced by Díaz-García and Leiva (2005), which allows great flexibility in terms of the kurtosis and skewness and can therefore be used as an alternative to the classical BS distribution in robustness studies (see Barros *et al.* (2009), Balakrishnan *et al.* (2007), Leiva *et al.* (2008) and Barros *et al.* (2009)).

Leiva *et al.* (2014) proposed the BSt-ACD model and developed associated EM algorithm for the estimation of model parameters. To estimate the BS-ACD model, Bhatti (2010) developed a maximum likelihood (ML) method based on a mixture of optimization algorithms, namely NM followed by BFGS. In this Chapter, as an extension of the ML estimation procedure developed by Bhatti (2010), we suggest a maximum likelihood method with an extra shape parameter to estimate the BSt-ACD model.

The rest of this Chapter is organized as follows. The BSt-ACD model and the corresponding log-likelihood function is presented in Section 2.2. We derive the first and second derivatives of $\ln L(\boldsymbol{\theta})$ in Sections 2.3 and 2.4. In Section 2.5, we discuss the maximum likelihood based estimation method and associated inference for the model parameters. A simulation study is carried out to examine the performance of these estimates in Section 2.6. Model discrimination is carried out using the likelihood-based method and information-based criteria in Section 2.7. In Section 2.8, we illustrate the proposed methodology using high-frequency data on two stocks from the New York Stock Exchange. We present the applications of the BSt-ACD model to trade durations and then a comparison with existing alternatives, the BS-ACD and GG-ACD models.

2.2 BSt-ACD Model and the log-likelihood function

The BSt-ACD model allows more flexibility in kurtosis and skewness by including an extra parameter, the degrees of freedom of the Student-t distribution. This may potentially facilitate robust parameter estimation.

By working with the density of X_i directly, the BSt-ACD model can be expressed as

$$X_i = \sigma_i \epsilon_i, \qquad \ln \sigma_i = \alpha + \beta \ln \sigma_{i-1} + \gamma \left[\frac{X_{i-1}}{\sigma_{i-1}} \right], \quad |\beta| < 1, \tag{2.1}$$

where $X_i \sim \text{BS}t(\kappa, \sigma_i, \nu)$ and $\epsilon_i \overset{\text{iid}}{\sim} \text{BS}t(\kappa, 1, \nu)$.

The pdf associated with the BSt-ACD model is given by

$$\begin{aligned} f_{X_i}(x_i;\boldsymbol{\theta}) =& \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)\kappa\sigma_i}\left[\left(\frac{\sigma_i}{x_i}\right)^{\frac{1}{2}}+\left(\frac{\sigma_i}{x_i}\right)^{\frac{3}{2}}\right]\left[1+\frac{\left(\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2\right)}{\nu\kappa^2}\right]^{-\frac{\nu+1}{2}} \\ =& \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)\kappa x_i}\left[\left(\frac{x_i}{\sigma_i}\right)^{\frac{1}{2}}+\left(\frac{\sigma_i}{x_i}\right)^{\frac{1}{2}}\right]\left[1+\frac{\left(\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2\right)}{\nu\kappa^2}\right]^{-\frac{\nu+1}{2}}, \end{aligned} \tag{2.2}$$

here, $\boldsymbol{\theta}=(\alpha,\beta,\gamma,\kappa,\nu)'$ is the model parameter vector. $x_i>0,\nu>0,i=1,\ldots,n$.

The corresponding survival function can be expressed as

$$S_{\mathrm{BS}t}(x_i;\kappa,\sigma_i,\nu)=1-\frac{1}{2}\left\{1+I_{[a(x_i)^2]/[a(x_i)^2+\nu]}\left(\frac{1}{2},\frac{\nu}{2}\right)\right\}, \tag{2.3}$$

where $a(x_i)=(\sqrt{x_i/\sigma_i}-\sqrt{\sigma_i/x_i})/\kappa$ and $I_x(a,b)=\int_0^x t^{a-1}(1-t)^{b-1}dt/\int_0^1 t^{a-1}(1-t)^{b-1}dt$ is the incomplete beta ratio (see Leiva *et al.* (2014)).

For $i=1,2,\ldots,n$, the individual log-likelihood function can be expressed as

$$\begin{aligned} \ln l_i(\boldsymbol{\theta}) =& -\ln 2\sqrt{\pi}-\ln(\kappa)-\ln(x_i)+\ln\Gamma(\frac{\nu+1}{2})-\frac{1}{2}\ln(\nu)-\ln\Gamma(\frac{\nu}{2}) \\ &+\ln\left[\left(\frac{x_i}{\sigma_i}\right)^{\frac{1}{2}}+\left(\frac{\sigma_i}{x_i}\right)^{\frac{1}{2}}\right]-\frac{\nu+1}{2}\ln\left[1+\frac{\left(\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2\right)}{\nu\kappa^2}\right] \end{aligned}$$

The log-likelihood function is then given by

$$\begin{aligned} \ln L(\boldsymbol{\theta}) =& \sum_{i=1}^{n}\left\{-\ln 2\sqrt{\pi}-\ln\kappa-\ln x_i+\ln\Gamma\left(\frac{\nu+1}{2}\right)-\ln\Gamma\left(\frac{\nu}{2}\right)-\frac{1}{2}\ln(\nu)\right. \\ &\left.+\ln\left[\left(\frac{x_i}{\sigma_i}\right)^{\frac{1}{2}}+\left(\frac{\sigma_i}{x_i}\right)^{\frac{1}{2}}\right]-\frac{\nu+1}{2}\ln\left(1+\frac{\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2}{\nu\kappa^2}\right)\right\}. \end{aligned} \tag{2.4}$$

2.3 The first derivatives of ln$L(\boldsymbol{\theta})$

The first derivatives of ln$L(\boldsymbol{\theta})$ with respect to α, β and γ are given by

$$\begin{aligned}
\frac{\partial \ln L}{\partial \alpha} &= \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \alpha} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \alpha}, \\
\frac{\partial \ln L}{\partial \beta} &= \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \beta} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \beta}, \\
\frac{\partial \ln L}{\partial \gamma} &= \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \gamma} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \gamma},
\end{aligned}$$

where

$$\frac{\partial \ln l_i}{\partial \sigma_i} = \frac{x_i - \sigma_i}{2\sigma_i} \left(\frac{(\nu+1)(x_i+\sigma_i)}{\kappa^2 x_i \sigma_i \nu + (x_i - \sigma_i)^2} - \frac{1}{x_i + \sigma_i} \right), \tag{2.5}$$

and by Lemma 2.3 in Mayorov (2011),

$$\frac{\partial \sigma_i}{\partial \alpha} = \sigma_i \left(1 + \frac{1}{\sigma_{i-1}} \frac{\partial \sigma_{i-1}}{\partial \alpha} \left(\beta - \gamma \frac{x_{i-1}}{\sigma_{i-1}} \right) \right), \tag{2.6}$$

$$\frac{\partial \sigma_i}{\partial \beta} = \sigma_i \left(\ln \sigma_{i-1} + \frac{1}{\sigma_{i-1}} \frac{\partial \sigma_{i-1}}{\partial \beta} \left(\beta - \gamma \frac{x_{i-1}}{\sigma_{i-1}} \right) \right), \tag{2.7}$$

$$\frac{\partial \sigma_i}{\partial \gamma} = \sigma_i \left(\frac{x_{i-1}}{\sigma_{i-1}} + \frac{1}{\sigma_{i-1}} \frac{\partial \sigma_{i-1}}{\partial \gamma} \left(\beta - \gamma \frac{x_{i-1}}{\sigma_{i-1}} \right) \right); \tag{2.8}$$

also, from by Lemma 2.4 in Mayorov (2011),

$$\frac{\partial\sigma_i}{\partial\alpha} =\sigma_i\left(1+\sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta-\gamma\frac{x_l}{\sigma_l}\right)\right), i\geq 1, \tag{2.9}$$

$$\frac{\partial\sigma_i}{\partial\beta} =\sigma_i\left(\ln\sigma_{i-1}+\sum_{k=0}^{i-2}\ln\sigma_k\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_l}{\sigma_l}\right)\right), i\geq 2, \tag{2.10}$$

$$\frac{\partial\sigma_i}{\partial\gamma} =\sigma_i\left(\frac{x_{i-1}}{\sigma_{i-1}}+\sum_{k=0}^{i-2}\frac{x_k}{\sigma_k}\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_l}{\sigma_l}\right)\right), i\geq 2, \tag{2.11}$$

with $\nabla\sigma_0=(0,0,0)'$ and $\nabla\sigma_1=(\sigma_1,\sigma_1\ln\sigma_0,\sigma_1 x_0/\sigma_0)'$.

The first derivatives of $\ln L(\boldsymbol{\theta})$ with respect to κ and ν are given by

$$\frac{\partial\ln L}{\partial\kappa} =\sum_{i=1}^{n}\frac{\partial\ln l_i}{\partial\kappa},$$
$$\frac{\partial\ln L}{\partial\nu} =\sum_{i=1}^{n}\frac{\partial\ln l_i}{\partial\nu},$$

where

$$\frac{\partial\ln l_i}{\partial\kappa} =\frac{1}{\kappa}\left(\frac{(\nu+1)(x_i-\sigma_i)^2}{\kappa^2 x_i\sigma_i\nu+(x_i-\sigma_i)^2}-1\right), \tag{2.12}$$

$$\frac{\partial\ln l_i}{\partial\nu} =\frac{1}{2}\left\{\Psi\left(\frac{\nu+1}{2}\right)-\Psi\left(\frac{\nu}{2}\right)\right.$$
$$\left.+\frac{\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2-\kappa^2}{\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2+\nu\kappa^2}-\ln\left(1+\frac{\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2}{\nu\kappa^2}\right)\right\}. \tag{2.13}$$

with $\Psi(x)=\frac{d}{dx}\ln\left(\Gamma(x)\right)=\frac{\Gamma'(x)}{\Gamma(x)}$ denoting the digamma function.

2.4 The second derivatives of $\ln L(\boldsymbol{\theta})$

The second derivatives of $\ln L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \kappa, \nu)'$ are as follows:

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \alpha}}{\partial \alpha} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_i^2} \left(\frac{\partial \sigma_i}{\partial \alpha} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial^2 \sigma_i}{\partial \alpha^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \beta}}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_i^2} \left(\frac{\partial \sigma_i}{\partial \beta} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial^2 \sigma_i}{\partial \beta^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \gamma^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \gamma}}{\partial \gamma} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_i^2} \left(\frac{\partial \sigma_i}{\partial \gamma} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial^2 \sigma_i}{\partial \gamma^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha \, \partial \beta} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha \, \partial \beta} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \alpha}}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_i^2} \frac{\partial \sigma_i}{\partial \beta} \frac{\partial \sigma_i}{\partial \alpha} + \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial^2 \sigma_i}{\partial \alpha \, \partial \beta} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha \, \partial \gamma} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha \, \partial \gamma} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \alpha}}{\partial \gamma} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_i^2} \frac{\partial \sigma_i}{\partial \gamma} \frac{\partial \sigma_i}{\partial \alpha} + \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial^2 \sigma_i}{\partial \alpha \, \partial \gamma} \right],$$

$$\frac{\partial^2 \ln L}{\partial \beta \, \partial \gamma} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta \, \partial \gamma} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \beta}}{\partial \gamma} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_i^2} \frac{\partial \sigma_i}{\partial \gamma} \frac{\partial \sigma_i}{\partial \beta} + \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial^2 \sigma_i}{\partial \beta \, \partial \gamma} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha \, \partial \kappa} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha \, \partial \kappa} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \alpha}}{\partial \kappa} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \kappa} \frac{\partial \sigma_i}{\partial \alpha},$$

$$\frac{\partial^2 \ln L}{\partial \beta \, \partial \kappa} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta \, \partial \kappa} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \beta}}{\partial \kappa} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \kappa} \frac{\partial \sigma_i}{\partial \beta},$$

$$\frac{\partial^2 \ln L}{\partial \gamma \, \partial \kappa} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma \, \partial \kappa} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \gamma}}{\partial \kappa} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \kappa} \frac{\partial \sigma_i}{\partial \gamma},$$

$$\frac{\partial^2 \ln L}{\partial \alpha \, \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha \, \partial \nu} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \alpha}}{\partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \nu} \frac{\partial \sigma_i}{\partial \alpha},$$

$$\frac{\partial^2 \ln L}{\partial \beta \, \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta \, \partial \nu} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \beta}}{\partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \nu} \frac{\partial \sigma_i}{\partial \beta},$$

$$\frac{\partial^2 \ln L}{\partial \gamma \, \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma \, \partial \nu} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \gamma}}{\partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \nu} \frac{\partial \sigma_i}{\partial \gamma},$$

$$\frac{\partial^2 \ln L}{\partial \kappa^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \kappa^2},$$
$$\frac{\partial^2 \ln L}{\partial \kappa \, \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \kappa \, \partial \nu},$$
$$\frac{\partial^2 \ln L}{\partial \nu^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \nu^2},$$

where

$$\frac{\partial^2 \ln l_i}{\partial \sigma_i^2} = \frac{1}{2\sigma_i} \left\{ \frac{(\nu+1)\,(x_i-\sigma_i)\,(x_i+\sigma_i)\,(\kappa^2 \nu \, x_i - 2\,(x_i-\sigma_i))}{\left((x_i-\sigma_i)^2 + \kappa^2 \sigma_i \, \nu \, x_i\right)^2} + \frac{(\nu+1)\,({x_i}^2+{\sigma_i}^2)}{\sigma_{\mathrm{i}} \left((x_i-\sigma_i)^2 + \kappa^2 \sigma_i \, \nu \, x_i\right)} - \frac{{x_i}^2 + 2\,\sigma_i \, x_i - {\sigma_i}^2}{\sigma_i \, (x_i+\sigma_i)^2} \right\}, \tag{2.14}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \kappa} = \frac{\kappa \, \nu \,(\nu+1)\, x_i \,({\sigma_i}^2 - {x_i}^2)}{\left((x_i-\sigma_i)^2 + \kappa^2 \sigma_i \, \nu \, x_i\right)^2} \tag{2.15}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_i \, \partial \nu} = \frac{\left({\sigma_i}^2 - {x_i}^2\right) \left(\kappa^2 \, \sigma_i \, \nu \, x_i - \nu \left(x_i - \sigma_i\right)^2\right)}{2 \, \sigma_i \, \nu \left(\left(x_i - \sigma_i\right)^2 + \kappa^2 \, \sigma_i \, \nu \, x_i\right)^2}, \tag{2.16}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa^2} = \frac{1}{\kappa^2} \left\{ 1 - \frac{(\nu + 1) \, \left(x_i - \sigma_i\right)^2 \, \left(\left(x_i - \sigma_i\right)^2 + 3 \, \kappa^2 \, \sigma_i \, \nu \, x_i\right)}{\left(\left(x_i - \sigma_i\right)^2 + \kappa^2 \, \sigma_i \, \nu \, x_i\right)^2} \right\}, \tag{2.17}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa \, \partial \nu} = \frac{\left(x_i - \sigma_i\right)^2 \, \left(1 + \nu - \left(x_i - \sigma_i\right)^2 - \kappa^2 \, \sigma_i \, \nu \, x_i\right)}{\kappa \, \nu \left(\left(x_i - \sigma_i\right)^2 + \kappa^2 \, \sigma_i \, \nu \, x_i\right)^2}, \tag{2.18}$$

$$\frac{\partial^2 \ln l_i}{\partial \nu^2} = \frac{1}{2} \left\{ \frac{\left(x_i - \sigma_i\right)^2 \, \left((\nu - 1) \, \left(x_i - \sigma_i\right)^2 - 2 \, \kappa^2 \, \sigma_i \, \nu \, x_i\right)}{\nu^2 \left(\left(x_i - \sigma_i\right)^2 + \kappa^2 \, \sigma_i \, \nu \, x_i\right)^2} + \frac{\Psi_1\left(\frac{\nu+1}{2}\right)}{2} - \frac{\Psi_1\left(\frac{\nu}{2}\right)}{2} + \frac{1}{\nu^2} \right\}, \tag{2.19}$$

$\Psi_1(x)$ is the second derivative of $\ln\left(\Gamma(x)\right)$ function.

$$\frac{\partial^2\sigma_i}{\partial\alpha^2} = \frac{\partial\sigma_i}{\partial\alpha}\left(1+\sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)\right) + \sigma_i\gamma\prod_{l=k}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k}^{i-1}\frac{x_l\,\frac{\partial\sigma_{jl}}{\partial\alpha}}{\sigma_{jl}^2\left(\beta-\frac{\gamma\, x_{jl}}{\sigma_{jl}}\right)}, \quad i\geq 1, \quad (2.20)$$

$$\frac{\partial^2\sigma_i}{\partial\beta^2} = \frac{\partial\sigma_i}{\partial\gamma}\left\{\ln\sigma_{j,i-1}+\sum_{k=0}^{i-2}\ln\sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)\right\}$$
$$+\sigma_i\left\{\sum_{k=0}^{i-2}\left(\ln\sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k+1}^{i-1}\frac{\frac{\gamma x_{jl}\frac{\partial\sigma_{jl}}{\partial\gamma}}{\sigma_{jl}^2}+1}{\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}}+\frac{\frac{\partial\sigma_{jk}}{\partial\gamma}\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)}{\sigma_{jk}}\right)+\frac{\frac{\partial\sigma_{j,i-1}}{\partial\gamma}}{\sigma_{j,i-1}}\right\}, \quad i\geq 2, \quad (2.21)$$

$$\frac{\partial^2\sigma_i}{\partial\gamma^2} = \frac{\partial\sigma_i}{\partial\gamma}\left\{\frac{x_{j,i-1}}{\sigma_{j,i-1}}+\sum_{k=0}^{i-2}\frac{x_{jk}}{\sigma_{jk}}\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)\right\}$$
$$+\sigma_i\left\{\sum_{k=0}^{i-2}\left(\frac{x_{jk}}{\sigma_{jk}}\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k+1}^{i-1}\frac{\frac{\gamma x_{jl}\frac{\partial\sigma_{jl}}{\partial\gamma}}{\sigma_{jl}^2}+1}{\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}}+\frac{\frac{\partial\sigma_{jk}}{\partial\gamma}\prod_{l=k+1}^{i-1}\left(\beta-\gamma\frac{x_{jl}}{\sigma_{jl}}\right)}{\sigma_{jk}}\right)+\frac{\frac{\partial\sigma_{j,i-1}}{\partial\gamma}}{\sigma_{j,i-1}}\right\}, \quad i\geq 2, \quad (2.22)$$

$$\frac{\partial^2 \sigma_i}{\partial\,\alpha\,\partial\,\beta} = \frac{\partial \sigma_i}{\partial \beta}\left(1 + \sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}\right)\right) + \sigma_i \sum_{l=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k}^{i-1}\frac{\frac{\gamma\,x_{jl}\,\frac{\partial \sigma_{jl}}{\partial \beta}}{\sigma_{jl}^2} + 1}{\beta - \frac{\gamma\,x_{jl}}{\sigma_{jl}}}, \quad i \geq 1, \quad (2.23)$$

$$\frac{\partial^2 \sigma_i}{\partial\,\alpha\,\partial\,\gamma} = \frac{\partial \sigma_i}{\partial \gamma}\left(1 + \sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}\right)\right) + \sigma_i \sum_{l=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k}^{i-1}\frac{\frac{\gamma\,x_{jl}\,\frac{\partial \sigma_{jl}}{\partial \beta}}{\sigma_{jl}^2} - \frac{x_{jl}}{\sigma_{jl}}}{\beta - \frac{\gamma\,x_{jl}}{\sigma_{jl}}}, \quad i \geq 1, \quad (2.24)$$

$$\frac{\partial^2 \sigma_i}{\partial\,\beta\,\partial\,\gamma} = \frac{\partial \sigma_i}{\partial \gamma}\left\{\ln \sigma_{j,i-1} + \sum_{k=0}^{i-2}\ln \sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}\right)\right\}$$
$$+\sigma_i\left\{\sum_{k=0}^{i-2}\left(\ln \sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k+1}^{i-1}\frac{\frac{\gamma x_{jl}\frac{\partial \sigma_{jl}}{\partial \gamma}}{\sigma_{jl}^2} - \frac{x_{jl}}{\sigma_{jl}}}{\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}} + \frac{\frac{\partial \sigma_{jk}}{\partial \gamma}\prod_{l=k+1}^{i-1}\left(\beta - \gamma\frac{x_{jl}}{\sigma_{jl}}\right)}{\sigma_{jk}}\right) + \frac{\frac{\partial \sigma_{j,i-1}}{\partial \gamma}}{\sigma_{j,i-1}}\right\}. \quad i \geq 2. \quad (2.25)$$

2.5 Estimation and inference

We estimate the parameter $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \kappa, \nu)'$ of the BSt-ACD model by maximizaing the log-likelihood function in equation (2.4). We propose the following two-step procedure for this purpose of estimation. In the first step, we estimate the ACD parameters, α, β, and γ, by using the NM algorithm with κ and ν fixed at their initial values. Then in the second step, we estimate over the whole parameter space by employing the BFGS algorithm with analytical gradients. We have derived and implemented the analytical gradients (the first derivatives in Section 2.3) in the second step.

Under certain regularity conditions (smoothness and boundariness), the ML estimator $\hat{\boldsymbol{\theta}}$ is $\sqrt{n}$-consistent and the asymptotic distribution of it is

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N_5\left(\mathbf{0}, n\mathcal{I}(\boldsymbol{\theta}_\mathbf{0})^{-1}\right) \tag{2.26}$$

where $\mathcal{I}(\boldsymbol{\theta}_\mathbf{0}) = -E\left[\left.\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\mathbf{0}}\right] = -E\left[\boldsymbol{H}(\boldsymbol{\theta}_\mathbf{0})\right]$ with $\boldsymbol{H} = \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ being the Hessian matrix. In order to estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$, we consider the well-known consistent estimator $-\boldsymbol{H}(\hat{\boldsymbol{\theta}})^{-1}$, which simply evaluates the second derivative matrix of the log-likelihood function at the MLEs. The square roots of its diagonal elements then provide the standard errors of the estimates.

2.6 Simulation study

In this Section, we present the results of a simulation study carried out for evaluating the performance of the maximum likelihood estimates determined by the procedure described in Section 2.5.

The simulation scenarios considered are as follows: sample sizes $n \in \{3000, 5000, 10000, 20000\}$ and the vector of parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \kappa, \nu)' = (0.1, 0.9, 0.1, 1.0, 12)'$. To estimate the BS$t$-ACD model, we set the starting values for the ACD parameters to be $(\alpha_0, \beta_0, \gamma_0)' = (0.01, 0.80, 0.01)'$. κ_0 is determined as $\kappa_0 = \sqrt{2\left(\frac{\overline{X}}{med(X)} - 1\right)\frac{\nu-2}{\nu}}$. We set $\nu_0 = 3$.

For each sample size and each estimate, we computed the mean, coefficients of skewness and kurtosis, bias and root mean squared error (RMSE) of the MLEs. Tables 2.1-2.4 summarize the distribution of the MLEs over 1000 simulation trials. From the simulated values presented in Tables 2.1-2.4, we observe that the bias, MSE and RMSE are consistently small and tend to 0 as sample size increases from 3000 to 20,000. It reveals that the MLEs are asymptotically unbiased and consistent. Moreover, the standard errors of the estimates become smaller, and also the sample skewness and kurtosis tend to 0 and 3, respectively, with increasing sample size. This reveals that the asymptotic distribution of the MLEs are normal, as suggested by the asymptotic distribution of the MLEs stated above.

Table 2.1: Simulation results for BSt-ACD model when $n = 3000$

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\kappa}$	$\hat{v}$
trueparam	0.1000	0.9000	0.1000	1.1000	12.0000
mean	0.0994	0.8972	0.1042	1.0909	10.9443
skew	0.7005	-0.6133	-0.0666	0.1661	3.4636
kurt	3.9595	3.9989	2.9727	3.2400	22.2730
bias	-0.0006	-0.0028	0.0042	-0.0091	-1.0557
MSE	0.0031	0.0005	0.0002	0.0016	26.4100
RMSE	0.0558	0.0226	0.0126	0.0399	5.1391
SE	0.0558	0.0225	0.0119	0.0389	5.0320

Table 2.2: Simulation results for BSt-ACD model when $n = 5000$

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\kappa}$	$\hat{v}$
trueparam	0.1000	0.9000	0.1000	1.1000	12.0000
mean	0.1010	0.8991	0.1007	1.0946	11.2733
skew	0.2384	-0.2098	-0.0053	0.1486	1.2490
kurt	2.9921	2.9386	2.9319	2.8445	5.2743
bias	0.0010	-0.0009	0.0007	-0.0054	-0.7267
MSE	0.0006	0.0001	0.0000	0.0004	4.0278
RMSE	0.0238	0.0096	0.0051	0.0193	2.0069
SE	0.0238	0.0096	0.0050	0.0185	1.8717

Table 2.3: Simulation results for BSt-ACD model when $n = 10000$

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\kappa}$	$\hat{v}$
trueparam	0.1000	0.9000	0.1000	1.1000	12.0000
mean	0.1000	0.8997	0.1003	1.0953	11.3754
skew	0.1613	-0.1726	0.0819	0.1590	0.9950
kurt	3.1191	3.1408	2.8582	2.9853	4.4802
bias	0.0000	-0.0003	0.0003	-0.0047	-0.6246
MSE	0.0003	0.0000	0.0000	0.0002	2.3332
RMSE	0.0168	0.0068	0.0036	0.0141	1.5275
SE	0.0168	0.0068	0.0036	0.0133	1.3947

Table 2.4: Simulation results for BSt-ACD model when $n = 20000$

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\kappa}$	$\hat{v}$
trueparam	0.1000	0.9000	0.1000	1.1000	12.0000
mean	0.0996	0.9001	0.1001	1.0967	11.5472
skew	0.1929	-0.1420	0.1562	-0.0386	0.2609
kurt	3.1248	3.3047	2.9425	2.9998	2.9496
bias	-0.0004	0.0001	0.0001	-0.0033	-0.4528
MSE	0.0001	0.0000	0.0000	0.0001	1.4330
RMSE	0.0115	0.0046	0.0025	0.0106	1.1971
SE	0.0115	0.0046	0.0025	0.0100	1.1087

2.7 Model discrimination

The BSt-ACD model provides more flexibility in terms of kurtosis and skewness by the inclusion of an additional parameter, namely, the degrees of freedom ν of the Student-t distribution. The BSt_7, BSt_9, BSt_{12} and BS-ACD models are some special cases, for example. Using the BSt-ACD model, it will then be of great interest to examine how often the true model gets selected in the set of candidate models and to choose a simple model that provides an adequate fit to the data. This model evaluation technique is called model discrimination. One may refer to the book by McLachlan and Peel (2000), and the recent papers by Balakrishnan and Peng (2006), Balakrishnan and Pal (2013), Balakrishnan and Pal (2016), and Balakrishnan *et al.* (2017) in this direction in different contexts. Two types of common selection criteria are used in the literature for this purpose, namely, likelihood ratio tests (LRTs) and information criterion.

2.7.1 Likelihood-based method

We study the performance of the likelihood ratio test (LRT) for testing the null hypothesis that the data generating process (DGP) can be described by one of the BSt_7 ($H_0 : \nu = 7$), BSt_9 ($H_0 : \nu = 9$), BSt_{12} ($H_0 : \nu = 12$) and BS ($H_0 : \nu \to \infty$) ACD models versus an alternative hypothesis that the DGP can be described by a member of the BSt-ACD family other than the one specified in the null hypothesis.

First, we assume that the true data generating processes (DGP) is the BSt_7, BSt_9, BSt_{12} and BS-ACD models (true models), respectively. Next we simulate datasets from each true model and fit the BSt_7, BSt_9, BSt_{12}, BS and BSt-ACD models to these datasets, respectively. We use 1000 replications of sample sizes of $n \in$

$\{2000, 3000, 5000\}$ and the same parameter setting $(\alpha, \beta, \gamma, \kappa)' = (0.1, 0.9, 0.1, 1.0)'$ for all true models. For each simulated sample, we obtain the LRT statistics of the fitted BSt_7, BSt_9, BSt_{12} and BS-ACD models versus the fitted general BSt-ACD model. We then evaluate the observed significance levels and powers of the LRT by calculating the proportion of the 1000 simulation trials that reside in the rejection of the null hypothesis at a nominal significance level of 0.05.

The results so obtained are reported in Table 2.5. The observed levels lie on the main diagonal, shown in bold, while other numbers are the observed powers. As the sample size increases, the observed levels decrease and the observed powers increase, which show an improvement in performance with sample size.

It is easy to see that the χ_1^2 distribution provides only a reasonable approximation to the null distribution of the LRT when testing for the BSt_{12}-ACD model as the observed levels are not so close to the nominal level, varying from 8.2% to 9.5%. But, the χ_1^2 distribution provides a good approximation to the null distribution of the LRT when testing for the BSt_7 and BSt_9-ACD models as the observed levels are indeed close to the nominal level. However, when testing for the BS-ACD model, the mixture chi-square distribution doesn't provide a good approximation to the null distribution of the LRT as the observed levels are found to be considerably above the nominal level ranging from 37.5% to 42.4%. The asymptotic mixture chi-square form is not sufficient for this result. However, the observed levels decrease as the sample size increases. For substantially larger sample size, they come down to the nominal level, 5%.

The observed powers vary in different situations and increase as sample size increases. When the true model is BSt_7-ACD, the test has high power to reject BS (the observed power is 100% in all cases) and BSt_{12}-ACD (the observed powers are from 89.7% to 99.8%) and moderate power to reject BSt_9-ACD (the observed powers take values from 48.8% to 73.7%). In all cases, the test has high power to reject BSt_7, BSt_9 and BSt_{12}-ACD models if the true DGP is BS-ACD. When the true model is the BSt_9-ACD, the test has high power to reject BS-ACD (the observed power is 100% in all cases), moderate power to reject BSt_{12}-ACD (the observed powers range from 45.7% to 71.0%) and low power to reject BSt_7-ACD (the observed powers vary from 22.0% to 56.4%). When the true DGP is the BSt_{12}-ACD, the test has high power to reject BS-ACD (the observed powers take values from 99.3% to 100%), good power to reject BSt_7-ACD (the observed powers are from 64.8% to 99.0%) and low power to reject BSt_9-ACD (the observed powers range from 14.3% to 42.4%).

Thus, the difference in model fitting between the BSt_7, BSt_9, BSt_{12} or BS-ACD models is significant enough to be detected by the LRT under the BSt-ACD setup.

2.7.2 Information-based criterion

We consider the same parameter settings as mentioned in Section 2.7.1 and generate data from the true models. For each dataset, we calculate the AIC and BIC values of the fitted the BSt_7, BSt_9, BSt_{12} and BS-ACD models. Then, the performance of the AIC and BIC can be evaluated by the selection rates for each of the fitted models. Since all fitted models have the same number of parameters, the AIC and BIC choose the same model for the same dataset simulated from the true model, and in fact they correspond to the model with the largest maximized log-likelihood value.

These results are presented in Table 2.6. The selection rates for the correct models lie on the main diagonal, shown in bold, while other numbers are the selection rates for the wrong models. The information criteria can distinguish between the BSt_7, BSt_9, BSt_{12} and BS-ACD models with a relatively high selection rates for the true models. As the sample size increases, the selection rates for the correct models increase and the selection rates for the wrong models decrease, which show an improvement in performance with sample size.

The information criteria perform well when the true DGP is BSt_7-ACD and BS-ACD models. The selection rates of the correct models are from 84.7% to 91.5% if the true distribution is BSt_7 and from 71.3% to 92.2% if the true distribution is BS. The information criteria perform moderately when the true DGP is BSt_9-ACD and BSt_{12}-ACD models. The selection rates of the right model are from 49.3% to 72.9% if the true model is BSt_9-ACD and from 55.2% to 77.3% if the true DGP is BSt_{12}-ACD.

As expected, when the true models are BSt_7-ACD and BSt_9-ACD, the selection rates for BSt_9-ACD and BSt_7-ACD are higher than those of BSt_{12}-ACD, which implies that BSt_7-ACD is closer to BSt_9-ACD than BSt_{12}-ACD, which are consistent with those results obtained earlier by the LRT.

In general, the BSt_7-ACD and BS-ACD models are quite different in terms of skewness and kurtosis. For a small sample size, $n = 2000$, the information criteria can distinguish between the BSt_7 and BS-ACD models with a high selection rates for the true models. The BSt_9-ACD and BSt_{12}-ACD models are very close to each other. Thus, the corresponding selection rates of the correct models are low. But for a larger sample size, $n = 5000$, the information criteria can distinguish between them with a relatively high selection rates for the true models.

Table 2.5: Model Discrimination by LR test

	True BSt-ACD Model			
	n=2000			
Fitted ACD Model	BSt_7	BSt_9	BSt_{12}	BS
BSt_7	**0.071**	0.202	0.648	1.000
BSt_9	0.488	**0.064**	0.143	0.995
BSt_{12}	0.897	0.457	**0.095**	0.883
BS	1.000	1.000	0.993	**0.424**
	n=3000			
Fitted ACD Model	BSt_7	BSt_9	BSt_{12}	BS
BSt_7	**0.052**	0.296	0.885	1.000
BSt_9	0.576	**0.053**	0.234	1.000
BSt_{12}	0.967	0.554	**0.082**	0.987
BS	1.000	1.000	1.000	**0.399**
	n=5000			
Fitted ACD Model	BSt_7	BSt_9	BSt_{12}	BS
BSt_7	**0.056**	0.564	0.990	1.000
BSt_9	0.737	**0.054**	0.424	1.000
BSt_{12}	0.998	0.710	**0.083**	1.000
BS	1.000	1.000	1.000	**0.375**

Table 2.6: Model Discrimination by AIC and BIC

	True BSt-ACD Model			
	n=2000			
Fitted ACD Model	BSt_7	BSt_9	BSt_{12}	BS
BSt_7	**0.847**	0.353	0.050	0.000
BSt_9	0.142	**0.493**	0.389	0.001
BSt_{12}	0.010	0.154	**0.552**	0.286
BS	0.001	0.000	0.009	**0.713**
	n=3000			
Fitted ACD Model	BSt_7	BSt_9	BSt_{12}	BS
BSt_7	**0.866**	0.263	0.012	0.000
BSt_9	0.132	**0.611**	0.337	0.000
BSt_{12}	0.002	0.126	**0.650**	0.190
BS	0.000	0.000	0.001	**0.810**
	n=5000			
Fitted ACD Model	BSt_7	BSt_9	BSt_{12}	BS
BSt_7	**0.915**	0.181	0.002	0.000
BSt_9	0.085	**0.729**	0.225	0.000
BSt_{12}	0.000	0.090	**0.773**	0.078
BS	0.000	0.000	0.000	**0.922**

2.8 Application to trade duration data

In this section, we present the applications of the BSt-ACD model to some trade durations and a comparison with existing alternatives, namely, the BS-ACD and GG-ACD models.

2.8.1 Data description

We consider the trade durations of International Business Machines (IBM) and Johnson & Johnson (JNJ) stocks in the period of 28 consecutive trading days from January 2, 2002 to February 11, 2002. There are 91,819 (IBM) and 61,188 (JNJ) observations for each asset. The original data sets were studied in detail by Bhatti (2010) and Leiva *et al.* (2014).

After restricting the adjusted durations between 10:00 am and 4:00 pm, we obtain 89,171 (IBM) and 59,390 (JNJ) observations for each asset. The initial values σ_0 for each day were set to be the median trade duration from 9:50 to 10:00 am for BSt-ACD model, and ψ_0 to be the mean trade duration from 9:50 to 10:00 am for GG-ACD model.

As seen in Table 2.7, the two adjusted duration series are positive and right-skewed with high degree of kurtosis indicating heavy tails.

Table 2.7: Summary statistics for the adjusted durations

Data	Min	Median	Mean	Max	SD	skew	kurt
IBM	0.170	1.034	1.394	27.480	1.279	3.018	21.543
JNJ	0.131	0.979	1.532	27.930	1.616	3.042	20.083

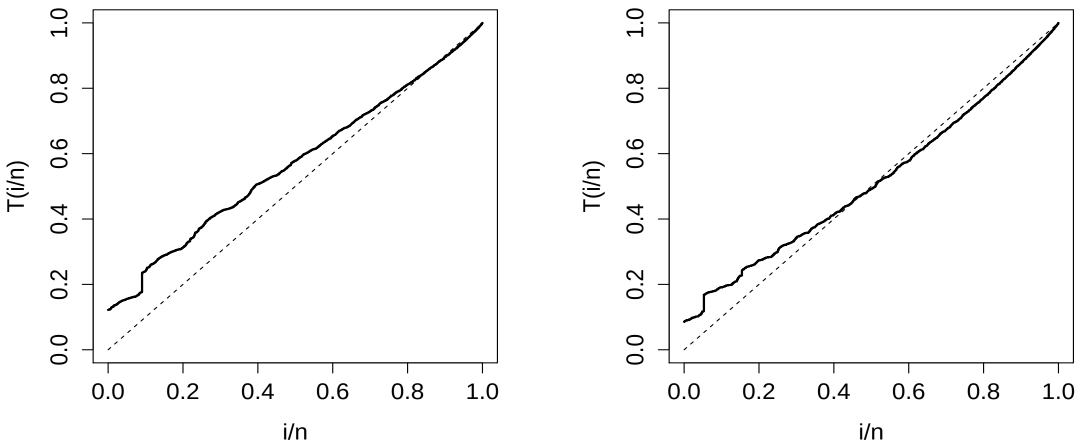

Figure 2.1: TTT plots for IBM (left) and JNJ (right)

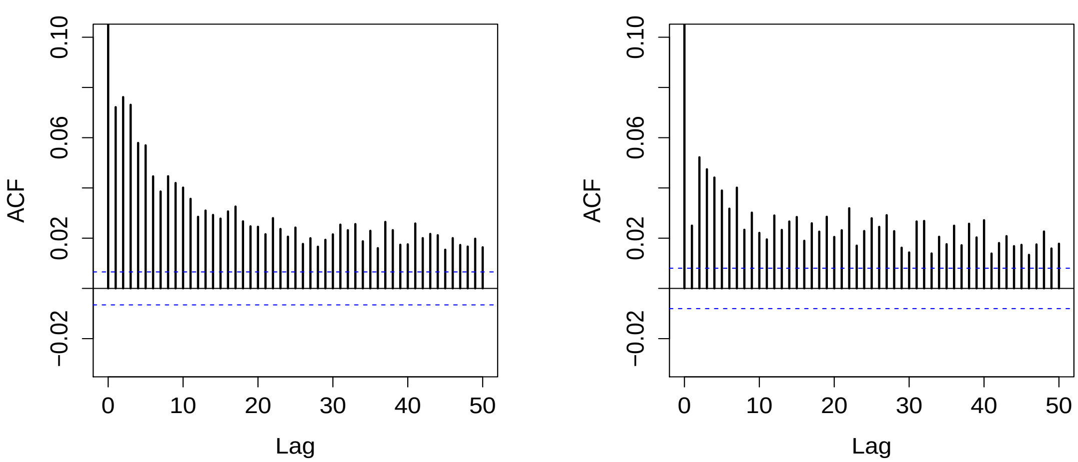

Figure 2.2: ACF plots for IBM (left) and JNJ (right)

In Figure 2.1, the empirical scaled TTT transforms are plotted, from which they are seen to be first concave and then convex, revealing that both marginals may have unimodal shaped hazard rates. Therefore, the BSt distribution could be a good fit to the duration data due to these shapes of its marginal density functions and hazard rates.

The Autocorrelation Function (ACF) plots in Figure 2.2 indicate that there are significant positive autocorrelations up to long lags in both series, which suggests that a ACD specification may be a reasonable choice (see Tsay (2010), Hautsch (2012) and Leiva *et al.* (2014)). Note: the plots are trimmed to be in the interval up to 0.1 even though the value should be 1 at 0. The ACF of the series represents the strength of linear dependence between X_i and X_{i-h} for $0 \leq h < n$. Here, the lag-h autocorrelation of X_i is defined as

$$\rho_h = \frac{Cov(X_i, X_{i-h})}{\sqrt{Var(X_i)Var(X_{i-h})}} = \frac{Cov(X_i, X_{i-h})}{Var(X_i)}. \tag{2.27}$$

Let $\overline{X} = \sum_{i=1}^{n} X_i/n$. Then the lag-$h$ sample autocorrelation of X_i is

$$\hat{\rho}_h = \frac{\sum_{i=h+1}^{n}(X_i - \overline{X})(X_{i-h} - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{2.28}$$

2.8.2 Estimation results

We employ the method described earlier in Section 2.5 to estimate the BSt-ACD model for the IBM an JNJ duration series, respectively. For comparison purpose, we also consider two alternatives, namely, the BS-ACD and GG-ACD models, using the two-step estimation procedures as detailed in Bhatti (2010). The obtained results are shown in Table 2.8. All estimates are statistically significant at 1% level. Overall, the BS-ACD and BSt-ACD models perform much better than the GG-ACD model in terms of AIC and BIC values. The BSt-ACD model outperforms all the considered models and the BS-ACD model provides close values according to AIC values for both series.

Table 2.8: Estimation results for three ACD models

	BSt-ACD model							
Data	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\kappa}$	$\hat{\nu}$	maxlnL	AIC	BIC
IBM	-0.041	0.950	0.029	0.871	81.069	-108306.40	216622.80	216669.80
	(0.0020)	(0.0044)	(0.0014)	(0.0021)	(0.1664)			
JNJ	-0.016	0.979	0.011	1.012	142.443	-79467.52	158945.00	158990.00
	(0.0015)	(0.0032)	(0.0010)	(0.0030)	(0.4033)			
	BS-ACD model							
Data	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\kappa}$		maxlnL	AIC	BIC
IBM	-0.040	0.950	0.029	0.882		-108314.50	216637.00	216674.60
	(0.0020)	(0.0044)	(0.0014)	(0.0021)				
JNJ	-0.016	0.978	0.011	1.019		-79468.83	158945.70	158981.60
	(0.0015)	(0.0032)	(0.0010)	(0.0030)				
	GG-ACD model							
Data	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\upsilon}$	$\hat{\eta}$	maxlnL	AIC	BIC
IBM	-0.029	0.949	0.045	45.377	0.182	-108531.00	217072.00	217119.00
	(0.0014)	(0.0043)	(0.0021)	(0.1520)	(0.0005)			
JNJ	-0.009	0.984	0.016	48.251	0.155	-79939.19	159888.40	159933.30
	(0.0009)	(0.0025)	(0.0015)	(0.4665)	(0.0009)			

The in-sample predictive model for the IBM series based on BSt-ACD model is given by

$$\hat{\sigma}_i = \exp\left(-0.041 + 0.950 \ln \hat{\sigma}_{i-1} + 0.029 \left[\frac{X_{i-1}}{\hat{\sigma}_{i-1}}\right]\right),$$

and for the JNJ series it is given by

$$\hat{\sigma}_i = \exp\left(-0.016 + 0.979 \ln \hat{\sigma}_{i-1} + 0.011 \left[\frac{X_{i-1}}{\hat{\sigma}_{i-1}}\right]\right).$$

2.8.3 Goodness-of-fit

We examine the distribution of GCS residuals for all the considered models through QQ plots. In general, QQ plots in Figure 2.3 indicate that GCS residual of BSs and BSt-ACD models seem to follow exp(1) distribution, considering that the 95th and 99th percentiles of this distribution are around 3.0 and 4.6, respectively. Furthermore, the BSt-ACD model seems to provide a slightly better fit to the data than the BS-ACD model. However, based on the GCS residuals, the GG-ACD model does not fit the JNJ series well although it provides a fairly good fit to the IBM series.

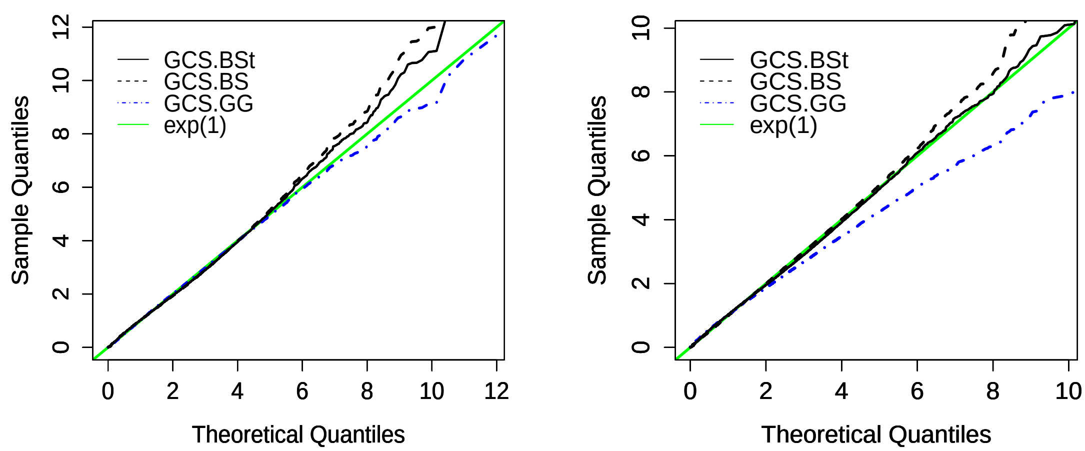

Figure 2.3: GCS plots for IBM (left) and JNJ (right) stocks

2.8.4 Out-of-sample forecast evaluation

We use the first nine tenth of the JNJ series, “in-sample” observations 1 to 53,451, for estimating the model parameters and then employ the resulting predictive model to the rest of the data, “out-of-sample” observations 53,452 to 59,390, to form density forecasts. The corresponding evaluation results are shown in Table 2.9. For the BS-ACD and BSt-ACD models, at 1% significance level, the null hypothesis that the z series is $i.i.d.$ $U(0,1)$ can not be rejected according to the KS, AD and LB tests. The KS and AD tests show that the z series is from the uniform distribution, U(0, 1). The LB tests indicate no serial correlation in the associated z and z^2 series. Overall, the BSt-ACD model yields best forecasts, and the BS-ACD model provides close values, but the GG-ACD model is the worst one with regard to out-of-sample forecast evaluation.

Table 2.9: P-values for out-of-sample tests

	JNJ							
Model	AD	KS	LB(10)	LB(15)	LB(20)	LB^2(10)	LB^2(15)	LB^2(20)
BSt-ACD	0.037	0.056	0.495	0.774	0.720	0.755	0.815	0.743
BS-ACD	0.035	0.044	0.490	0.771	0.717	0.753	0.814	0.741
GG-ACD	< 0.001	< 0.001	0.531	0.788	0.742	0.762	0.786	0.725

Note: KS represents Kolmogorov-Smirnov test, AD represents Anderson-Darling test, LB(l) represents Ljung-Box test for z series over l lags. LB^2(l) represents Ljung-Box test for z^2 series over l lags.

For point-wise forecasts, one can also use measures like mean absolute error (MAE), mean absolute percentage error (MAPE) or mean squared error (MSE) to compare models based on out-of-sample forecast accuracy. We assess the predictive performance of the models by evaluating out-of-sample density forecasts of the models since density forecasts are popular in both macroeconomics and finance.

Chapter 3

Bivariate BS-ACD model

3.1 Introduction

Kundu *et al.* (2010) derived the bivariate Birnbaum-Saunders (BVBS) distribution through a transformation of the bivariate normal distribution (see Kundu *et al.* (2013) and Vilca *et al.* (2014a,b) for generalizations of the BVBS distribution). Its marginals are univariate BS distributions and its conditional distributions can be expressed in terms of normal distribution. Moreover, it has a correlation parameter which indirectly represents the dependence between the two BS random variables.

Our goal now is to construct a bivariate autoregressive conditional duration model based on the bivariate BS distribution, which would then facilitate us to jointly analyze and measure the strength of dependence between two matched duration series, pairs of durations with the same starting time (see Simonsen (2007)).

The rest of this Chapter is organized as follows. In Section 3.2, we present the BVBS-ACD model and the corresponding log-likelihood function. We derive the first and second derivatives of $\ln L(\boldsymbol{\theta})$ in Sections 3.3 and 3.4. In Section 3.5, we discuss the maximum likelihood estimation and associated inference for the model parameters. A Monte Carlo simulation study is carried out in Section 3.6 to examine the properties of the MLEs. In Section 3.7, we provide an application of the BVBS-ACD model to a trade duration data set.

3.2 Bivariate BS-ACD model and the log-likelihood function

By working with the joint density of (X_{1i}, X_{2i}) directly, we propose the following BVBS-ACD model:

$$X_{1i} = \sigma_{1i}\epsilon_{1i}, \qquad \ln\sigma_{1i} = \alpha_1 + \beta_1 \ln\sigma_{1,i-1} + \gamma_1\left[\frac{X_{1,i-1}}{\sigma_{1,i-1}}\right], \quad |\beta_1| < 1, \tag{3.1}$$

$$X_{2i} = \sigma_{2i}\epsilon_{2i}, \qquad \ln\sigma_{2i} = \alpha_2 + \beta_2 \ln\sigma_{2,i-1} + \gamma_2\left[\frac{X_{2,i-1}}{\sigma_{2,i-1}}\right], \quad |\beta_2| < 1, \tag{3.2}$$

where $(X_{1i}, X_{2i}) \sim BVBS(\kappa_1, \sigma_{1i}, \kappa_2, \sigma_{2i}, \rho)$, $\epsilon_{1i} \overset{\text{iid}}{\sim} BS(\kappa_1, 1)$ and $\epsilon_{2i} \overset{\text{iid}}{\sim} BS(\kappa_2, 1)$.

The associated joint PDF of the BVBS-ACD model is given by

$$\begin{aligned}
&f_{X_{1i},X_{2i}}(x_{1i}, x_{2i}; \boldsymbol{\theta})\\
=&\frac{1}{2\kappa_1\sigma_{1i}}\left[\left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{3}{2}}\right]\frac{1}{2\kappa_2\sigma_{2i}}\left[\left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{3}{2}}\right]\\
&\times\phi_2\left[\frac{1}{\kappa_1}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right), \frac{1}{\kappa_2}\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right); \rho\right]\\
=&\frac{1}{2\kappa_1 x_{1i}}\left[\left(\frac{x_{1i}}{\sigma_{1i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}}\right]\frac{1}{2\kappa_2 x_{2i}}\left[\left(\frac{x_{2i}}{\sigma_{2i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}}\right]
\end{aligned} \tag{3.3}$$

$$\times \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{1}{\kappa_1^2}\left(\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2\right)+\frac{1}{\kappa_2^2}\left(\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2\right)-\frac{2\rho}{\kappa_1\kappa_2}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)\right]\right\},$$

here, $\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho)'$ is the model parameter. ϕ_2 denotes the standard bivariate normal density function.

For $i = 1, 2, \ldots, n$, the individual log-likelihood function can be expressed as

$$\begin{aligned}
\ln l_i(\boldsymbol{\theta}) = &- \ln(\kappa_1) - \ln(\kappa_2) - \frac{1}{2}\ln(1-\rho^2) \\
&+ \ln\left[\left(\frac{x_{1i}}{\sigma_{1i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}}\right] - \frac{1}{2(1-\rho^2)\kappa_1^2}\left(\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2\right) \\
&+ \ln\left[\left(\frac{x_{2i}}{\sigma_{2i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}}\right] - \frac{1}{2(1-\rho^2)\kappa_2^2}\left(\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2\right) \\
&+ \frac{\rho}{(1-\rho^2)\kappa_1\kappa_2}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right).
\end{aligned} \tag{3.4}$$

The log-likelihood function, without additive constant, is then given by

$$\begin{aligned}
\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln l_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} \Bigg\{ &- \ln(\kappa_1) - \ln(\kappa_2) - \frac{1}{2}\ln(1-\rho^2) \\
&+ \ln\left[\left(\frac{x_{1i}}{\sigma_{1i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}}\right] - \frac{1}{2(1-\rho^2)\kappa_1^2}\left(\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2\right) \\
&+ \ln\left[\left(\frac{x_{2i}}{\sigma_{2i}}\right)^{\frac{1}{2}} + \left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}}\right] - \frac{1}{2(1-\rho^2)\kappa_2^2}\left(\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2\right) \\
&+ \frac{\rho}{(1-\rho^2)\kappa_1\kappa_2}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)\Bigg\}.
\end{aligned} \tag{3.5}$$

3.3 The first derivatives of $\ln L(\boldsymbol{\theta})$

The first derivatives of $\ln L(\boldsymbol{\theta})$ with respect to α_j, β_j and γ_j are given by

$$\begin{aligned}
\frac{\partial \ln L}{\partial \alpha_j} &= \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}, \\
\frac{\partial \ln L}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}, \\
\frac{\partial \ln L}{\partial \gamma_j} &= \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial \ln l_i}{\partial \sigma_{ji}} =& -\frac{x_{ji} - \sigma_{ji}}{2\sigma_{ji}(x_{ji} + \sigma_{ji})} + \frac{(x_{ji} - \sigma_{ji})(x_{ji} + \sigma_{ji})}{2x_{ji}{\sigma_{ji}}^2(j - \rho^2)\kappa_j^2} \\
& - \frac{\rho}{2\sigma_{ji}(j - \rho^2)\kappa_j\kappa_{j'}} \left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}} - \sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right);
\end{aligned} \tag{3.6}$$

here $j = 1, 2$, $j' = \{1, 2\} - \{j\}$ and

$$\frac{\partial \sigma_{ji}}{\partial \alpha_j} = \sigma_{ji}\left(1 + \sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}}\right)\right), i \geq 1, \tag{3.7}$$

$$\frac{\partial \sigma_{ji}}{\partial \beta_j} = \sigma_{ji}\left(\ln \sigma_{j,i-1} + \sum_{k=0}^{i-2} \ln \sigma_{jk} \prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}}\right)\right), i \geq 2, \tag{3.8}$$

$$\frac{\partial \sigma_{ji}}{\partial \gamma_j} = \sigma_{ji}\left(\frac{x_{j,i-1}}{\sigma_{j,i-1}} + \sum_{k=0}^{i-2} \frac{x_{jk}}{\sigma_{jk}} \prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}}\right)\right), i \geq 2, \tag{3.9}$$

with $\nabla\sigma_{j0} = (0, 0, 0)'$ and $\nabla\sigma_{j1} = (\sigma_{j1}, \sigma_{j1} \ln \sigma_{j0}, \sigma_{j1}x_{j0}/\sigma_{j0})'$.

The first derivatives of $\ln L(\boldsymbol{\theta})$ with respect to κ_j, $j = 1, 2$ and ρ are given by

$$\frac{\partial \ln L}{\partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \kappa_j},$$
$$\frac{\partial \ln L}{\partial \rho} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \rho},$$

where

$$\begin{aligned}\frac{\partial \ln l_i}{\partial \kappa_j} =& -\frac{1}{\kappa_j} + \frac{1}{(1-\rho^2)\kappa_j^3}\left(\frac{x_{ji}}{\sigma_{ji}} + \frac{\sigma_{ji}}{x_{ji}} - 2\right) \\ &-\frac{\rho}{(1-\rho^2)\kappa_j^2\kappa_{j'}}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right), \end{aligned} \tag{3.10}$$

$$\begin{aligned}\frac{\partial \ln l_i}{\partial \rho} =& \frac{\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}\left[\frac{1}{\kappa_1^2}\left(\frac{x_{1i}}{\sigma_{1i}} + \frac{\sigma_{1i}}{x_{1i}} - 2\right) + \frac{1}{\kappa_2^2}\left(\frac{x_{2i}}{\sigma_{2i}} + \frac{\sigma_{2i}}{x_{2i}} - 2\right)\right] \\ &+\frac{1+\rho^2}{(1-\rho^2)^2\kappa_1\kappa_2}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right). \end{aligned} \tag{3.11}$$

3.4 The second derivatives of $\ln L(\boldsymbol{\theta})$

The second derivatives of $\ln L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho)'$ are given by

$$\frac{\partial^2 \ln L}{\partial \alpha_j{}^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j{}^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \alpha_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \left(\frac{\partial \sigma_{ji}}{\partial \alpha_j} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \alpha_j^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \beta_j{}^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j{}^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \left(\frac{\partial \sigma_{ji}}{\partial \beta_j} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \beta_j^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \gamma_j{}^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j{}^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \gamma_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \left(\frac{\partial \sigma_{ji}}{\partial \gamma_j} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \gamma_j^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \, \partial \beta_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \, \partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \frac{\partial \sigma_{ji}}{\partial \beta_j} \frac{\partial \sigma_{ji}}{\partial \alpha_j} + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \alpha_j \, \partial \beta_j} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \gamma_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \frac{\partial \sigma_{ji}}{\partial \gamma_j} \frac{\partial \sigma_{ji}}{\partial \alpha_j} + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \alpha_j \, \partial \gamma_j} \right],$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \gamma_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \frac{\partial \sigma_{ji}}{\partial \gamma_j} \frac{\partial \sigma_{ji}}{\partial \beta_j} + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \beta_j \, \partial \gamma_j} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha_1 \, \partial \alpha_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_1 \, \partial \alpha_2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{1i}} \frac{\partial \sigma_{1i}}{\partial \alpha_1}}{\partial \alpha_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \, \partial \sigma_{2i}} \frac{\partial \sigma_{2i}}{\partial \alpha_2} \frac{\partial \sigma_{1i}}{\partial \alpha_1},$$

$$\frac{\partial^2 \ln L}{\partial \beta_1 \, \partial \beta_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_1 \, \partial \beta_2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{1i}} \frac{\partial \sigma_{1i}}{\partial \beta_1}}{\partial \beta_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \, \partial \sigma_{2i}} \frac{\partial \sigma_{2i}}{\partial \beta_2} \frac{\partial \sigma_{1i}}{\partial \beta_1},$$

$$\frac{\partial^2 \ln L}{\partial \gamma_1 \, \partial \gamma_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_1 \, \partial \gamma_2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{1i}} \frac{\partial \sigma_{1i}}{\partial \gamma_1}}{\partial \gamma_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \, \partial \sigma_{2i}} \frac{\partial \sigma_{2i}}{\partial \gamma_2} \frac{\partial \sigma_{1i}}{\partial \gamma_1},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \, \partial \beta_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \, \partial \beta_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \beta_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \, \partial \sigma_{j'i}} \frac{\partial \sigma_{j'i}}{\partial \beta_{j'}} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \, \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \, \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \, \partial \sigma_{j'i}} \frac{\partial \sigma_{j'i}}{\partial \gamma_{j'}} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \, \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \, \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \, \partial \sigma_{j'i}} \frac{\partial \sigma_{j'i}}{\partial \gamma_{j'}} \frac{\partial \sigma_{ji}}{\partial \beta_j},$$

$$
\begin{aligned}
\frac{\partial^2 \ln L}{\partial \alpha_j \partial \kappa_j} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_j} \frac{\partial \sigma_{ji}}{\partial \alpha_j}, \\
\frac{\partial^2 \ln L}{\partial \beta_j \partial \kappa_j} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_j} \frac{\partial \sigma_{ji}}{\partial \beta_j}, \\
\frac{\partial^2 \ln L}{\partial \gamma_j \partial \kappa_j} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_j} \frac{\partial \sigma_{ji}}{\partial \gamma_j}, \\
\frac{\partial^2 \ln L}{\partial \alpha_j \partial \kappa_{j'}} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_{j'}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}, \\
\frac{\partial^2 \ln L}{\partial \beta_j \partial \kappa_{j'}} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_{j'}} \frac{\partial \sigma_{ji}}{\partial \beta_j}, \\
\frac{\partial^2 \ln L}{\partial \gamma_j \partial \kappa_{j'}} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_{j'}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}, \\
\frac{\partial^2 \ln L}{\partial \alpha_j \partial \rho} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} \frac{\partial \sigma_{ji}}{\partial \alpha_j}, \\
\frac{\partial^2 \ln L}{\partial \beta_j \partial \rho} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} \frac{\partial \sigma_{ji}}{\partial \beta_j}, \\
\frac{\partial^2 \ln L}{\partial \gamma_j \partial \rho} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} \frac{\partial \sigma_{ji}}{\partial \gamma_j}, \\
\frac{\partial^2 \ln L}{\partial {\kappa_j}^2} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial {\kappa_j}^2}, \\
\frac{\partial^2 \ln L}{\partial \kappa_1 \partial \kappa_2} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \kappa_1 \partial \kappa_2}, \\
\frac{\partial^2 \ln L}{\partial \kappa_j \partial \rho} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \kappa_j \partial \rho}, \\
\frac{\partial^2 \ln L}{\partial \rho^2} &= \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \rho^2},
\end{aligned}
$$

where $j = 1, 2,\ j' = \{1, 2\} - \{j\}$ and

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} = \frac{\rho\left(3\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}} - \sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{4\,\kappa_1\,\kappa_2\,\left(1-\rho^2\right)\,{\sigma_{ji}}^2} + \frac{{x_{ji}}^2 + 2\,\sigma_{ji}\,x_{ji} - {\sigma_{ji}}^2}{2\,{\sigma_{ji}}^2\,\left(x_{ji}+\sigma_{ji}\right)^2} + \frac{x_{ji}}{{\kappa_j}^2\,\left(\rho^2-1\right)\,{\sigma_{ji}}^3}, \tag{3.12}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \partial \sigma_{2i}} = \frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} + \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} + \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{4\kappa_1\kappa_2(1-\rho^2)\sigma_{1i}\sigma_{2i}}, \tag{3.13}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_j} = \frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}} - \sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{2\kappa_j^2\,\kappa_{j'}\,\left(1-\rho^2\right)\,\sigma_{ji}} - \frac{\left(x_{ji}-\sigma_{ji}\right)\,\left(x_{ji}+\sigma_{ji}\right)}{{\kappa_j}^3\,\left(1-\rho^2\right)\,{\sigma_{ji}}^2\,x_{ji}}, \tag{3.14}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_{j'}} = \frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}} - \sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{2\kappa_j\,{\kappa_{j'}}^2\,\left(1-\rho^2\right)\,\sigma_{ji}}, \tag{3.15}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} = \frac{\rho\,(x_{ji}-\sigma_{ji})\,(x_{ji}+\sigma_{ji})}{\kappa_j{}^2\,(1-\rho^2)^2\,\sigma_{ji}{}^2\,x_{ji}} - \frac{(\rho^2+1)\,\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\,\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{2\,\kappa_1\,\kappa_2\,(\rho-1)^2\,(\rho+1)^2\,\sigma_{ji}}, \tag{3.16}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa_j{}^2} = \frac{2\,\rho\,\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\,\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_j{}^3\,\kappa_{j'}\,(1-\rho^2)} - \frac{3\,\left(\frac{x_{ji}}{\sigma_{ji}}+\frac{\sigma_{ji}}{x_{ji}}-2\right)}{\kappa_j{}^4\,(1-\rho^2)} + \frac{1}{\kappa_j{}^2}, \tag{3.17}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa_1 \partial \kappa_2} = \frac{\rho\,\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\,\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1{}^2\,\kappa_2{}^2\,(1-\rho^2)}, \tag{3.18}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa_j \partial \rho} = \frac{2\,\rho\,\left(\frac{x_{ji}}{\sigma_{ji}}+\frac{\sigma_{ji}}{x_{ji}}-2\right)}{\kappa_j{}^3\,(1-\rho^2)^2} - \frac{(\rho^2+1)\,\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\,\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_j{}^2\,\kappa_{j'}\,(\rho-1)^2\,(\rho+1)^2}, \tag{3.19}$$

$$\frac{\partial^2 \ln l_i}{\partial \rho^2} = \frac{2\,\rho\,(\rho^2+3)\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1\,\kappa_2\,(1-\rho^2)^3} - \frac{(3\,\rho^2+1)\left(\frac{\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2}{{\kappa_2}^2}+\frac{\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2}{{\kappa_1}^2}\right)}{(1-\rho^2)^3} + \frac{\rho^2+1}{(1-\rho^2)^2}, \tag{3.20}$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \alpha_j^2} = \frac{\partial \sigma_{ji}}{\partial \alpha_j}\left(1+\sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right)+\sigma_{ji}\gamma\prod_{l=k}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k}^{i-1}\frac{x_l\,\frac{\partial\sigma_{jl}}{\partial\alpha_j}}{\sigma_{jl}^2\left(\beta_j-\frac{\gamma_j\,x_{jl}}{\sigma_{jl}}\right)}, \quad i \geq 1, \tag{3.21}$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \beta_j^2} = \frac{\partial \sigma_{ji}}{\partial \gamma_j}\left\{\ln\sigma_{j,i-1}+\sum_{k=0}^{i-2}\ln\sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right\}$$
$$+\sigma_{ji}\left\{\sum_{k=0}^{i-2}\left(\ln\sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k+1}^{i-1}\frac{\frac{\gamma_j x_{jl}\frac{\partial\sigma_{jl}}{\partial\gamma_j}}{\sigma_{jl}^2}+1}{\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}}+\frac{\frac{\partial\sigma_{jk}}{\partial\gamma_j}\prod_{l=k+1}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)}{\sigma_{jk}}\right)+\frac{\frac{\partial\sigma_{j,i-1}}{\partial\gamma_j}}{\sigma_{j,i-1}}\right\}, \quad i \geq 2, \tag{3.22}$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \gamma_j^2} = \frac{\partial \sigma_{ji}}{\partial \gamma_j} \left\{ \frac{x_{j,i-1}}{\sigma_{j,i-1}} + \sum_{k=0}^{i-2} \frac{x_{jk}}{\sigma_{jk}} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \right\}$$
$$+\sigma_{ji} \left\{ \sum_{k=0}^{i-2} \left(\frac{x_{jk}}{\sigma_{jk}} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \sum_{l=k+1}^{i-1} \frac{\frac{\gamma_j x_{jl} \frac{\partial \sigma_{jl}}{\partial \gamma_j}}{\sigma_{jl}^2} + 1}{\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}}} + \frac{\frac{\partial \sigma_{jk}}{\partial \gamma_j} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right)}{\sigma_{jk}} \right) + \frac{\frac{\partial \sigma_{j,i-1}}{\partial \gamma_j}}{\sigma_{j,i-1}} \right\},$$
$$i \geq 2, \quad (3.23)$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \alpha_j \, \partial \beta_j} = \frac{\partial \sigma_{ji}}{\partial \beta_j} \left(1 + \sum_{k=1}^{i-1} \prod_{l=k}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \right) + \sigma_{ji} \sum_{l=1}^{i-1} \prod_{l=k}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \sum_{l=k}^{i-1} \frac{\frac{\gamma_j \, x_{jl} \, \frac{\partial \sigma_{jl}}{\partial \beta_j}}{\sigma_{jl}^2} + 1}{\beta_j - \frac{\gamma_j \, x_{jl}}{\sigma_{jl}}},$$
$$i \geq 1, \quad (3.24)$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \alpha_j \, \partial \gamma_j} = \frac{\partial \sigma_{ji}}{\partial \gamma_j} \left(1 + \sum_{k=1}^{i-1} \prod_{l=k}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \right) + \sigma_{ji} \sum_{l=1}^{i-1} \prod_{l=k}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \sum_{l=k}^{i-1} \frac{\frac{\gamma_j \, x_{jl} \, \frac{\partial \sigma_{jl}}{\partial \beta_j}}{\sigma_{jl}^2} - \frac{x_{jl}}{\sigma_{jl}}}{\beta_j - \frac{\gamma_j \, x_{jl}}{\sigma_{jl}}},$$
$$i \geq 1, \quad (3.25)$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \beta_j \, \partial \gamma_j} = \frac{\partial \sigma_{ji}}{\partial \gamma_j} \left\{ \ln \sigma_{j,i-1} + \sum_{k=0}^{i-2} \ln \sigma_{jk} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \right\}$$
$$+\sigma_{ji} \left\{ \sum_{k=0}^{i-2} \left(\ln \sigma_{jk} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \sum_{l=k+1}^{i-1} \frac{\frac{\gamma_j x_{jl} \frac{\partial \sigma_{jl}}{\partial \gamma_j}}{\sigma_{jl}^2} - \frac{x_{jl}}{\sigma_{jl}}}{\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}}} + \frac{\frac{\partial \sigma_{jk}}{\partial \gamma_j} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right)}{\sigma_{jk}} \right) + \frac{\frac{\partial \sigma_{j,i-1}}{\partial \gamma_j}}{\sigma_{j,i-1}} \right\}.$$
$$i \geq 2. \quad (3.26)$$

3.5 Estimation and inference

We estimate the parameter $\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho)'$ of the BVBS-ACD model by maximizing the log-likelihood function (without the additive constant) in equation (3.5). We do the estimation by using a hybrid of optimization algorithms, namely, Nelder-Meade (NM) followed by Broyden-Fletcher-Goldfarb-Shanno (BFGS) (see Bhatti (2010) and Leiva *et al.* (2014) for recent details). First, we apply the NM algorithm to estimate the ACD parameters α_j, β_j, and $\gamma_j, j = 1, 2$ by fixing κ_1, κ_2 and ρ at their initial values. Next, we use BFGS algorithm to estimate over the entire parameter space. We have derived and implemented the analytical gradient (the first derivatives in Section 3.3) in the second step. The BFGS quasi-Newton method with the analytic gradients will be faster, more stable and lead to more accurate estimates than a numerical gradient method (see Bard (1974), Bolker (2008) and Mayorov (2011)).

The standard errors (SEs) of the MLEs of the model parameters can then be calculated as the square root of the diagonal elements of the negative of the inverse Hessian matrix, which is an estimator of the asymptotic covariance matrix.

Leiva *et al.* (2014) discussed the Wald test to check the statistical significance of individual model parameters. Let θ denote a parameter of the BVBS-ACD model. The Wald statistic $W = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}$ can be used to test the hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Then, under H_0, W is asymptotically $N(0, 1)$, under the usual regularity conditions.

3.6 Simulation study

We examine the properties of the MLEs of the parameters of the BVBS-ACD model through a Monte Carlo simulation study. We use 1000 simulated samples of sample sizes n =500, 1000, 3000 and 5000 from the BVBS-ACD model with the vector of true parameters

$$\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho)' = (0.1, 0.9, 0.1, -0.1, 0.95, 0.1, 1.1, 1.2, \rho)'.$$

Here, we choose ρ =0, 0.25, 0.5, 0.95. To estimate the BVBS-ACD model, we set the starting values for the ACD parameters

$$(\alpha_{10}, \beta_{10}, \gamma_{10}, \alpha_{20}, \beta_{20}, \gamma_{20})' = (0.01, 0.70, 0.01, -0.01, 0.8, 0.01)'.$$

$(\kappa_{10}, \kappa_{20}, \rho_0)$ then have well-defined initial values obtained by

$$\kappa_{j0} = \sqrt{2\left(\frac{\overline{X}_j}{med(X_j)} - 1\right)}, \quad j = 1, 2$$

(see (1.14)) and

$$\rho_0 = \text{Corr}\left(\frac{1}{\kappa_{10}}\left(\sqrt{\frac{X_1}{\hat{\sigma}_1}} - \sqrt{\frac{\hat{\sigma}_1}{X_1}}\right), \frac{1}{\kappa_{20}}\left(\sqrt{\frac{X_2}{\hat{\sigma}_2}} - \sqrt{\frac{\hat{\sigma}_2}{X_2}}\right)\right),$$

where $\hat{\sigma}_j$ is estimated by

$$\hat{\sigma}_{ji} = \exp\left(\alpha_{j0} + \beta_{j0} \ln \hat{\sigma}_{j,i-1} + \gamma_{j0}\left[\frac{X_{j,i-1}}{\hat{\sigma}_{j,i-1}}\right]\right),$$

$j = 1, 2$, $i = 1, 2, \ldots, n$. For each setting of the parameter values and each sample size, we calculate the mean, coefficients of skewness and kurtosis, bias and root mean squared error (RMSE) of the MLEs over 1000 replications. These results are presented in Tables 3.1-3.4. For different correlation coefficients, all the estimates possess lowe Bias and RMSE which decrease with increased sample size, tending towards 0. This empirically displays the asymptotic unbiasedness, consistency and efficiency of the MLEs. Moreover, as n increases, the empirical distributions of all the estimators become close to normal distribution, which may be seen from the values of skewness and kurtosis, for example.

3.7 Application to trade duration data

In this section, we present the application of BVBS-ACD model to a high-frequency trade duration data.

3.7.1 Data description

We consider the trade durations of Johnson & Johnson (JNJ) and Procter & Gamble Company (PG) stocks in the period of 30 consecutive trading days from January 2, 2002 to February 13, 2002. There are 65,127 (JNJ) and 57,335 (PG) observations for each asset. The original data sets were studied in detail by Bhatti (2010) and Leiva *et al.* (2014).

After restricting two adjusted duration series for JNJ and PG stocks between 10:00 am and 4:00 pm, we obtain two matched series with 6,222 pairs of durations with the same starting time.

Table 3.1: Simulation results for BVBS-ACD model when $\rho = 0.95$

Parameter	Statistics	n			
		500	1000	3000	5000
$\alpha_1 = 0.1$	Mean	0.0979	0.0988	0.0995	0.0994
	Skew	0.2000	0.3171	0.1825	0.1627
	Kurt	6.4886	3.1753	2.9457	2.8743
	Bias	-0.0021	-0.0012	-0.0005	-0.0006
	RMSE	0.0498	0.0329	0.0189	0.0146
$\beta_1 = 0.9$	Mean	0.8923	0.8961	0.8986	0.8993
	Skew	-0.3145	-0.3093	-0.2455	-0.1321
	Kurt	4.0404	3.0194	3.0778	2.9054
	Bias	-0.0077	-0.0039	-0.0014	-0.0007
	RMSE	0.0206	0.0137	0.0077	0.0058
$\gamma_1 = 0.1$	Mean	0.1112	0.1057	0.1020	0.1012
	Skew	0.0839	0.2243	0.0684	-0.0216
	Kurt	3.0643	3.0290	2.9507	2.9264
	Bias	0.0112	0.0057	0.0020	0.0012
	RMSE	0.0173	0.0107	0.0054	0.0040
$\alpha_2 = -0.1$	Mean	-0.1065	-0.1036	-0.1011	-0.1008
	Skew	-0.0990	0.1592	0.0814	0.0995
	Kurt	6.2791	3.4112	3.1276	3.1182
	Bias	-0.0065	-0.0036	-0.0011	-0.0008
	RMSE	0.0218	0.0139	0.0077	0.0058
$\beta_2 = 0.95$	Mean	0.9456	0.9481	0.9492	0.9496
	Skew	-0.5418	-0.5210	-0.2765	-0.2595
	Kurt	4.5363	3.6419	3.0362	3.0692
	Bias	-0.0044	-0.0019	-0.0008	-0.0004
	RMSE	0.0117	0.0073	0.0040	0.0029
$\gamma_2 = 0.1$	Mean	0.1058	0.1030	0.1009	0.1006
	Skew	0.1677	0.2066	0.1627	-0.0288
	Kurt	3.3693	3.2666	3.1211	3.1027
	Bias	0.0058	0.0030	0.0009	0.0006
	RMSE	0.0127	0.0082	0.0045	0.0033
$\kappa_1 = 1.1$	Mean	1.1270	1.1151	1.1048	1.1032
	Skew	0.4710	0.2482	0.0683	-0.0002
	Kurt	4.0018	3.3397	3.0630	2.8358
	Bias	0.0270	0.0151	0.0048	0.0032
	RMSE	0.0514	0.0318	0.0154	0.0116
$\kappa_2 = 1.2$	Mean	1.2059	1.2037	1.2007	1.2006
	Skew	0.2258	0.2272	0.1380	0.1175
	Kurt	3.7508	3.3847	3.0840	2.8101
	Bias	0.0059	0.0037	0.0007	0.0006
	RMSE	0.0419	0.0283	0.0159	0.0120
$\rho = 0.95$	Mean	0.9460	0.9479	0.9493	0.9496
	Skew	-0.4434	-0.3525	-0.1592	-0.1344
	Kurt	3.3355	3.4547	2.9353	2.8687
	Bias	-0.0040	-0.0021	-0.0007	-0.0004
	RMSE	0.0067	0.0041	0.0020	0.0015

Table 3.2: Simulation results for BVBS-ACD model when $\rho = 0.5$

Parameter	Statistics	n			
		500	1000	3000	5000
$\alpha_1 = 0.1$	Mean	0.1138	0.1077	0.1025	0.1009
	Skew	1.1298	0.6483	0.3646	0.2623
	Kurt	5.4679	3.3928	3.0443	3.0993
	Bias	0.0138	0.0077	0.0025	0.0009
	RMSE	0.0836	0.0552	0.0309	0.0232
$\beta_1 = 0.9$	Mean	0.8880	0.8936	0.8978	0.8989
	Skew	-0.8633	-0.5516	-0.4102	-0.2501
	Kurt	4.4549	3.2770	3.1868	2.9495
	Bias	-0.0120	-0.0064	-0.0022	-0.0011
	RMSE	0.0359	0.0237	0.0131	0.0097
$\gamma_1 = 0.1$	Mean	0.1090	0.1047	0.1017	0.1011
	Skew	0.1115	0.2722	0.0996	-0.0113
	Kurt	3.0447	3.1798	2.9707	2.8346
	Bias	0.0090	0.0047	0.0017	0.0011
	RMSE	0.0199	0.0130	0.0070	0.0053
$\alpha_2 = -0.1$	Mean	-0.0966	-0.0989	-0.0991	-0.0996
	Skew	0.5843	0.3731	0.2382	0.1851
	Kurt	4.1691	3.6212	3.0768	3.2444
	Bias	0.0034	0.0011	0.0009	0.0004
	RMSE	0.0278	0.0189	0.0106	0.0078
$\beta_2 = 0.95$	Mean	0.9436	0.9472	0.9487	0.9493
	Skew	-1.0474	-0.6883	-0.4069	-0.3821
	Kurt	5.6827	3.8742	3.1820	3.3638
	Bias	-0.0064	-0.0028	-0.0013	-0.0007
	RMSE	0.0195	0.0124	0.0068	0.0050
$\gamma_2 = 0.1$	Mean	0.1027	0.1014	0.1004	0.1003
	Skew	0.2234	0.2921	0.2723	0.0024
	Kurt	3.2393	3.4492	3.1149	2.9713
	Bias	0.0027	0.0014	0.0004	0.0003
	RMSE	0.0149	0.0099	0.0056	0.0041
$\kappa_1 = 1.1$	Mean	1.1257	1.1144	1.1049	1.1033
	Skew	0.3089	0.1493	0.0360	-0.0686
	Kurt	3.3437	3.1565	2.9667	3.0989
	Bias	0.0257	0.0144	0.0049	0.0033
	RMSE	0.0503	0.0313	0.0153	0.0115
$\kappa_2 = 1.2$	Mean	1.2042	1.2027	1.2002	1.2003
	Skew	0.0982	0.2395	0.0876	0.1018
	Kurt	3.3152	3.4159	3.0412	2.9583
	Bias	0.0042	0.0027	0.0002	0.0003
	RMSE	0.0411	0.0284	0.0164	0.0121
$\rho = 0.5$	Mean	0.5089	0.5054	0.5014	0.5009
	Skew	0.0150	-0.0573	-0.0271	-0.1053
	Kurt	2.9122	3.0392	2.7894	2.8607
	Bias	0.0089	0.0054	0.0014	0.0009
	RMSE	0.0357	0.0249	0.0133	0.0104

Table 3.3: Simulation results for BVBS-ACD model when $\rho = 0.25$

Parameter	Statistics	n			
		500	1000	3000	5000
$\alpha_1 = 0.1$	Mean	0.1128	0.1074	0.1025	0.1011
	Skew	1.3299	0.7418	0.4268	0.2879
	Kurt	6.4049	3.5942	3.1606	3.1271
	Bias	0.0128	0.0074	0.0025	0.0011
	RMSE	0.0928	0.0606	0.0343	0.0255
$\beta_1 = 0.9$	Mean	0.8877	0.8933	0.8977	0.8988
	Skew	-1.0251	-0.6212	-0.4544	-0.2828
	Kurt	5.1294	3.4056	3.2236	2.9648
	Bias	-0.0123	-0.0067	-0.0023	-0.0012
	RMSE	0.0397	0.0259	0.0145	0.0106
$\gamma_1 = 0.1$	Mean	0.1102	0.1054	0.1019	0.1012
	Skew	0.1325	0.2836	0.1095	0.0027
	Kurt	3.0767	3.1226	2.9154	2.8908
	Bias	0.0102	0.0054	0.0019	0.0012
	RMSE	0.0220	0.0143	0.0077	0.0057
$\alpha_2 = -0.1$	Mean	-0.0966	-0.0990	-0.0991	-0.0997
	Skew	0.6223	0.3911	0.2302	0.1596
	Kurt	4.1643	3.6524	3.0860	3.2282
	Bias	0.0034	0.0010	0.0009	0.0003
	RMSE	0.0310	0.0210	0.0117	0.0085
$\beta_2 = 0.95$	Mean	0.9428	0.9470	0.9486	0.9492
	Skew	-1.2030	-0.6811	-0.4479	-0.3517
	Kurt	6.3884	3.7265	3.2563	3.2519
	Bias	-0.0072	-0.0030	-0.0014	-0.0008
	RMSE	0.0218	0.0136	0.0075	0.0054
$\gamma_2 = 0.1$	Mean	0.1034	0.1017	0.1005	0.1004
	Skew	0.2714	0.2902	0.2840	-0.0007
	Kurt	3.2682	3.3968	3.0376	2.8688
	Bias	0.0034	0.0017	0.0005	0.0004
	RMSE	0.0164	0.0108	0.0060	0.0045
$\kappa_1 = 1.1$	Mean	1.1253	1.1140	1.1049	1.1034
	Skew	0.3513	0.1474	0.0373	-0.0764
	Kurt	3.3813	3.1881	2.9441	3.1389
	Bias	0.0253	0.0140	0.0049	0.0034
	RMSE	0.0505	0.0313	0.0153	0.0114
$\kappa_2 = 1.2$	Mean	1.2040	1.2022	1.2001	1.2002
	Skew	0.2705	0.1860	0.0394	0.0589
	Kurt	4.3483	3.3121	2.9944	2.9714
	Bias	0.0040	0.0022	0.0001	0.0002
	RMSE	0.0416	0.0282	0.0164	0.0121
$\rho = 0.25$	Mean	0.2656	0.2592	0.2526	0.2516
	Skew	0.0799	-0.0045	-0.0007	-0.0876
	Kurt	2.9182	3.0210	2.7867	2.8551
	Bias	0.0156	0.0092	0.0026	0.0016
	RMSE	0.0462	0.0319	0.0167	0.0130

Table 3.4: Simulation results for BVBS-ACD model when $\rho = 0$

Parameter	Statistics	n			
		500	1000	3000	5000
$\alpha_1 = 0.1$	Mean	0.1189	0.1072	0.1015	0.1005
	Skew	1.9843	1.0356	0.5140	0.3007
	Kurt	12.4607	5.4310	3.3849	2.8156
	Bias	0.0189	0.0072	0.0015	0.0005
	RMSE	0.1074	0.0651	0.0345	0.0254
$\beta_1 = 0.9$	Mean	0.8841	0.8930	0.8980	0.8991
	Skew	-1.7870	-0.8688	-0.4545	-0.1977
	Kurt	10.5058	4.9659	3.3610	2.7907
	Bias	-0.0159	-0.0070	-0.0020	-0.0009
	RMSE	0.0465	0.0279	0.0148	0.0108
$\gamma_1 = 0.1$	Mean	0.1123	0.1060	0.1021	0.1011
	Skew	0.4415	0.1910	0.2933	0.1809
	Kurt	3.8432	3.5541	3.3634	3.2788
	Bias	0.0123	0.0060	0.0021	0.0011
	RMSE	0.0244	0.0152	0.0079	0.0060
$\alpha_2 = -0.1$	Mean	-0.0953	-0.0979	-0.0993	-0.0995
	Skew	0.6719	0.3909	0.3022	0.0918
	Kurt	5.2592	3.3247	3.2052	2.8505
	Bias	0.0047	0.0021	0.0007	0.0005
	RMSE	0.0345	0.0218	0.0119	0.0088
$\beta_2 = 0.95$	Mean	0.9411	0.9456	0.9484	0.9492
	Skew	-1.3671	-0.6772	-0.4254	-0.2484
	Kurt	6.8392	3.7528	3.4520	3.0809
	Bias	-0.0089	-0.0044	-0.0016	-0.0008
	RMSE	0.0241	0.0142	0.0077	0.0057
$\gamma_2 = 0.1$	Mean	0.1038	0.1022	0.1009	0.1004
	Skew	0.1192	0.1136	-0.0729	-0.0438
	Kurt	3.3123	2.7931	2.9934	2.8415
	Bias	0.0038	0.0022	0.0009	0.0004
	RMSE	0.0169	0.0108	0.0060	0.0047
$\kappa_1 = 1.1$	Mean	1.1260	1.1123	1.1046	1.1030
	Skew	0.2113	-0.0531	0.0977	0.2128
	Kurt	3.4990	3.2449	2.9873	2.8949
	Bias	0.0260	0.0123	0.0046	0.0030
	RMSE	0.0509	0.0297	0.0154	0.0119
$\kappa_2 = 1.2$	Mean	1.2058	1.2043	1.2017	1.2010
	Skew	0.3055	0.0370	0.0812	0.0353
	Kurt	4.3661	2.8992	2.8667	2.8285
	Bias	0.0058	0.0043	0.0017	0.0010
	RMSE	0.0434	0.0283	0.0156	0.0120
$\rho = 0$	Mean	0.0186	0.0102	0.0048	0.0028
	Skew	0.0696	0.0890	0.1807	-0.0745
	Kurt	3.0213	3.0437	3.0644	3.1268
	Bias	0.0186	0.0102	0.0048	0.0028
	RMSE	0.0493	0.0335	0.0187	0.0144

As seen in Figure 3.1 and Table 3.5, the two matched duration series are positive and right-skewed with high degree of kurtosis indicating heavy tails.

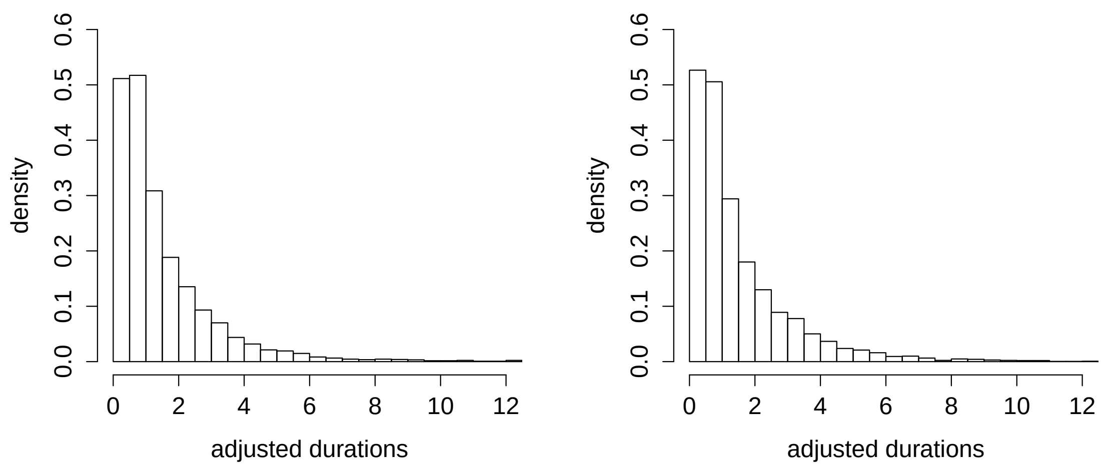

Figure 3.1: Histograms for matched duration series of JNJ (left) and PG (right)

Table 3.5: Summary statistics for matched data

Data	Min	Median	Mean	Max	SD	skew	kurt
JNJ	0.131	0.973	1.499	18.248	1.594	2.886	15.699
PG	0.113	0.963	1.522	14.967	1.624	2.603	12.840

In Figure 3.2, the empirical scaled TTT transforms are presented first concave and then convex, revealing that both marginals may have unimodal hazard rates.

Hence, the BVBS distribution could be a good fit for the duration data due to the shape of its marginal density functions and hazard rates.

The ACF plots in Figure 3.3 indicate that there is a positive autocorrelation in both series, which suggests that a ACD specification is a reasonable choice, as well.

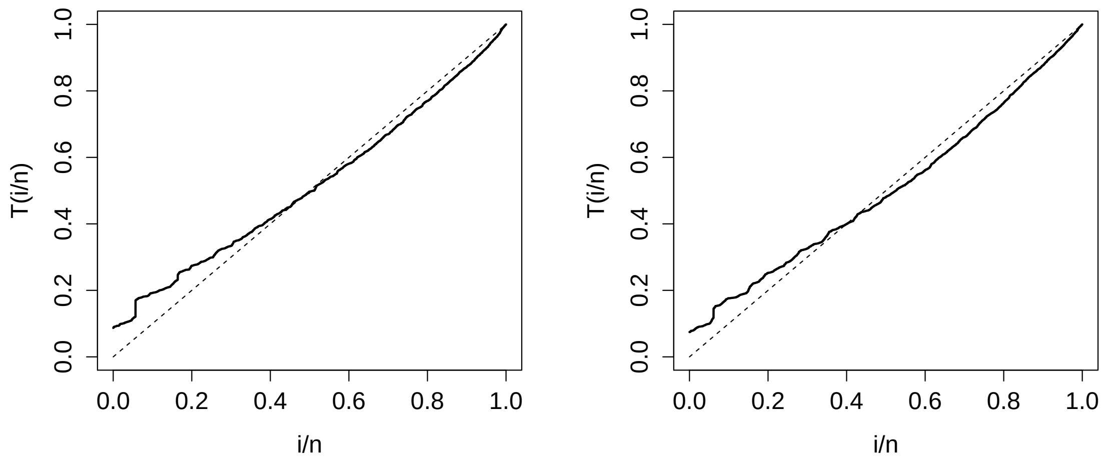

Figure 3.2: TTT plots for matched duration series of JNJ (left) and PG (right)

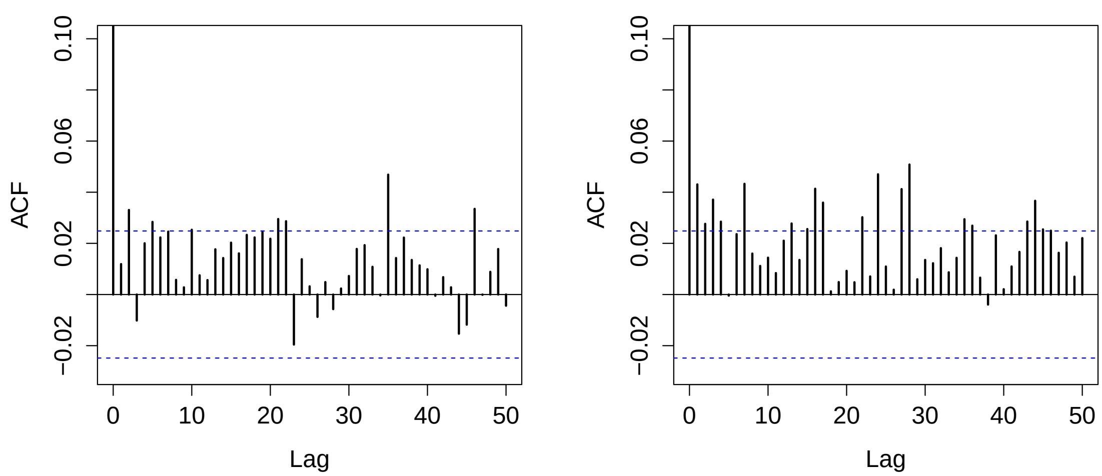

Figure 3.3: ACF plots for matched duration series of JNJ (left) and PG (right)

3.7.2 Estimation results

Table 3.6: Estimation results for the BVBS-ACD model.

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$
MLE	-0.007	0.991	0.004	-0.010	0.991	0.006	1.024	1.077	0.041
SE	0.002	0.005	0.001	0.002	0.004	0.001	0.009	0.010	0.013

We employ the method described earlier in Section 3.5 to estimate the BVBS-ACD model with the matched duration data. The obtained results are shown in Table 3.6. All estimates are statistically significant at 1% level according to the Wald test. The estimated correlation coefficient is 0.041, which implies a weak positive correlation between the two matched duration series. The in-sample predictive model is given by

$$\hat{\sigma}_{1i} = \exp\left(-0.007 + 0.991 \ln \hat{\sigma}_{1,i-1} + 0.004\left[\frac{X_{1,i-1}}{\hat{\sigma}_{1,i-1}}\right]\right),$$

$$\hat{\sigma}_{2i} = \exp\left(-0.010 + 0.991 \ln \hat{\sigma}_{2,i-1} + 0.006\left[\frac{X_{2,i-1}}{\hat{\sigma}_{2,i-1}}\right]\right).$$

3.7.3 Model Comparison

To verify whether it is necessary to include the correlation parameter ρ in the model, we conduct model comparison by using likelihood ratio test and Akaike information criterion (AIC).

We re-estimate the BVBS-ACD model by taking $\rho = 0$. Under $H_0 : \rho = 0$, the likelihood ratio test statistic is defined by $2(\ln L_1 - \ln L_0)$. AIC is given by $-2\ln L + 2k$, where $\ln L$ is the maximized log-likelihood value of the model of interest and k is the corresponding number of estimated parameters. As seen in Table 3.7, the p-value of the likelihood ratio test is 0.001, which does not provide sufficient evidence towards the restricted model. The difference in AIC values between the two models is 8.502, which indicates that the AIC value of the restricted model is substantially larger than that of the unrestricted model. Both these findings suggest that the proposed BVBS-ACD model is a better choice than the restricted model in terms of overall model fit.

Table 3.7: LR test and AICs.

	Maximized log-likelihood	AIC
Restricted model	3098.554	-6181.108
Unrestricted model	3103.805	-6189.610
Difference	5.251	-8.502
p-value of LR test	0.001	—

3.7.4 Goodness-of-fit

To evaluate the goodness-of-fit of the BVBS-ACD model, we investigate the in-sample one-step-ahead density forecasts implied by the predictive model. We examine the uniformity of the z series by Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests along with histogram plots, and also check the independence of the z series by Ljung-Box (LB) test along with ACF plots. Results for these tests are shown in Table 3.8. At 1% significance level, the null hypothesis that the series z_1, z_2, $z_{1|2}$ and $z_{2|1}$ are all i.i.d. $U(0,1)$ can not be rejected. The KS and AD tests, along with the histograms in Figures 3.4 and 3.5 seem to coincide with uniformity of the corresponding z series. The LB tests,along with the ACF plots in Figures 3.6 and 3.7, show the absence of serial correlation in the associated z and z^2 series.

Table 3.8: P-values for in-sample goodness of fit tests for the BVBS-ACD model based on density forecasts

z series	KS	AD	LB(10)	LB(15)	LB(20)	LB^2(10)	LB^2(15)	LB^2(20)
z_1	0.012	0.017	0.119	0.177	0.195	0.435	0.616	0.730
z_2	0.032	0.037	0.682	0.747	0.480	0.421	0.561	0.422
$z_{1\|2}$	0.011	0.019	0.111	0.174	0.185	0.404	0.597	0.705
$z_{2\|1}$	0.025	0.033	0.701	0.782	0.481	0.431	0.590	0.413

Note: LB(l) represents LB test for z series over l lags, where $l = 10, 15$ and 20. LB^2(l) represents LB test for z^2 series over l lags.

We also investigate the estimated GCS residuals obtained by (1.9). We examine the distribution and independence of the GCS residuals by QQ and ACF plots, respectively. The QQ plots in Figure 3.8 show that the GCS residuals seem to follow the exp(1) distribution considering that the 95th and 99th percentiles of this distribution are around 3.0 and 4.6, respectively. The ACF plots in Figure 3.9 demonstrate a lack of serial correlation in the GCS residuals. We can, therefore, conclude that the BVBS-ACD model may provide a good fit for these data.

3.7.5 Out-of-sample forecast evaluation

We use the first six seventh of the matched data, i.e., "in-sample" observations 1 to 5333, to estimate the model parameters, and then employ the resulting predictive model to the rest of the data, i.e., "out-of-sample" observations 5334 to 6222, to form density forecasts. These evaluation results are presented in Table 3.9. At 10% significance level, the null hypothesis that the series z_1, z_2, $z_{1|2}$ and $z_{2|1}$ are all *i.i.d.* $U(0,1)$ can not be rejected. The KS and AD tests both seem to support the uniformity of the corresponding z series. The LB tests indicate no serial correlation is presented

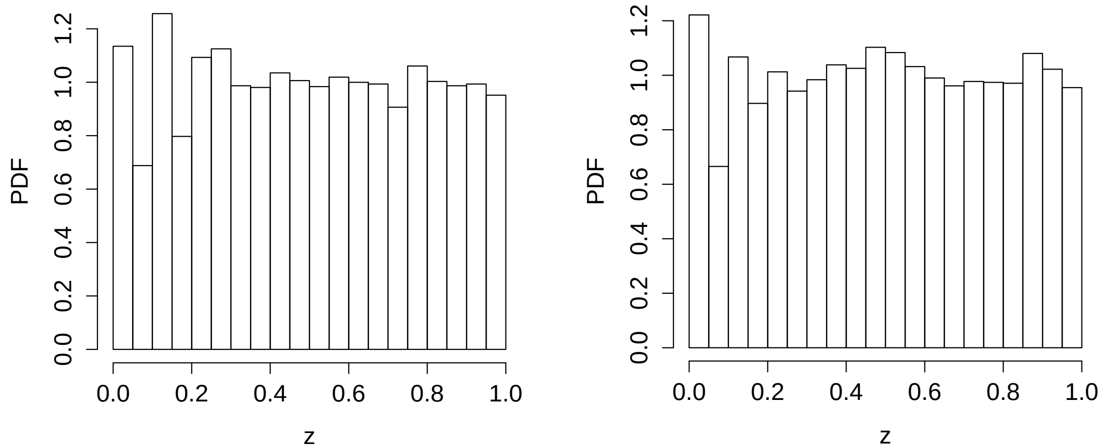

Figure 3.4: Histograms for z_1 (left) and z_2 (right)

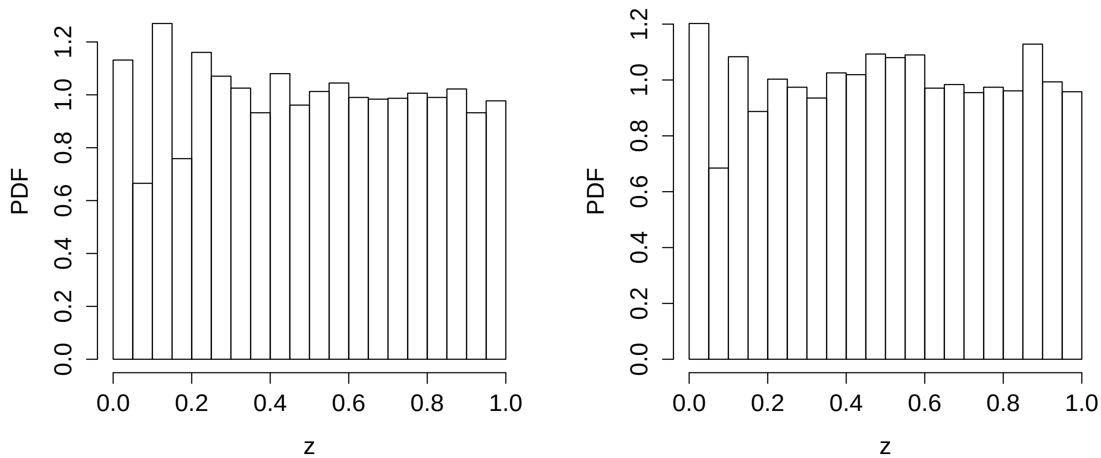

Figure 3.5: Histograms for $z_{1|2}$ (left) and $z_{2|1}$ (right)

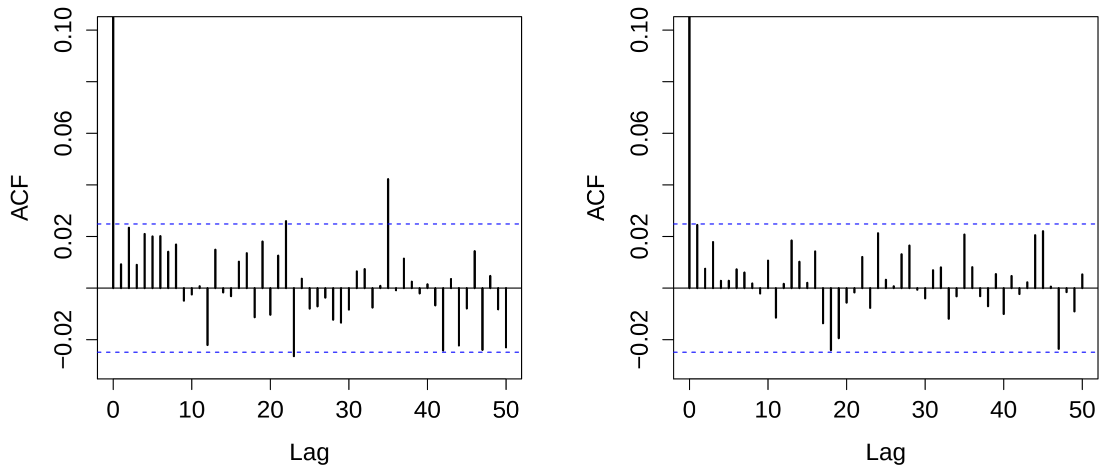

Figure 3.6: ACF plots for z_1 (left) and z_2 (right)

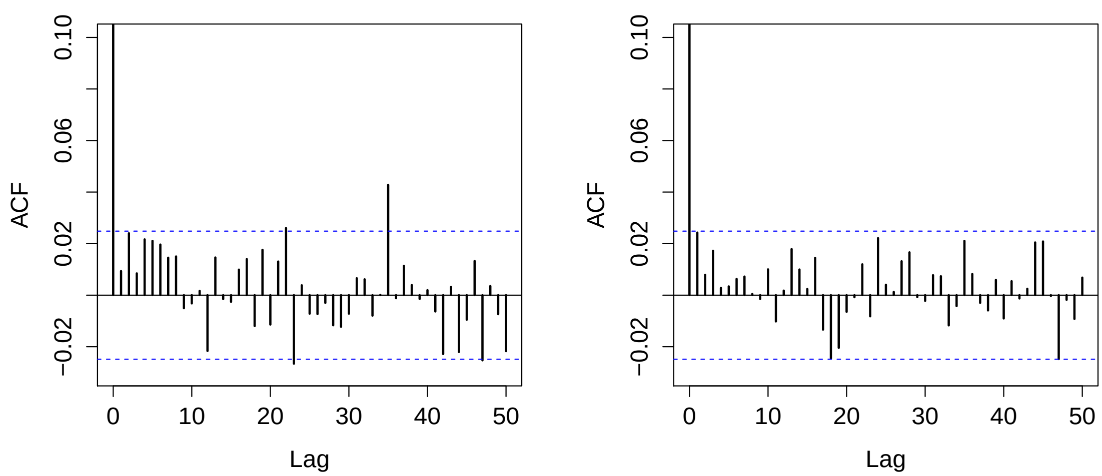

Figure 3.7: ACF plots for $z_{1|2}$ (left) and $z_{2|1}$ (right)

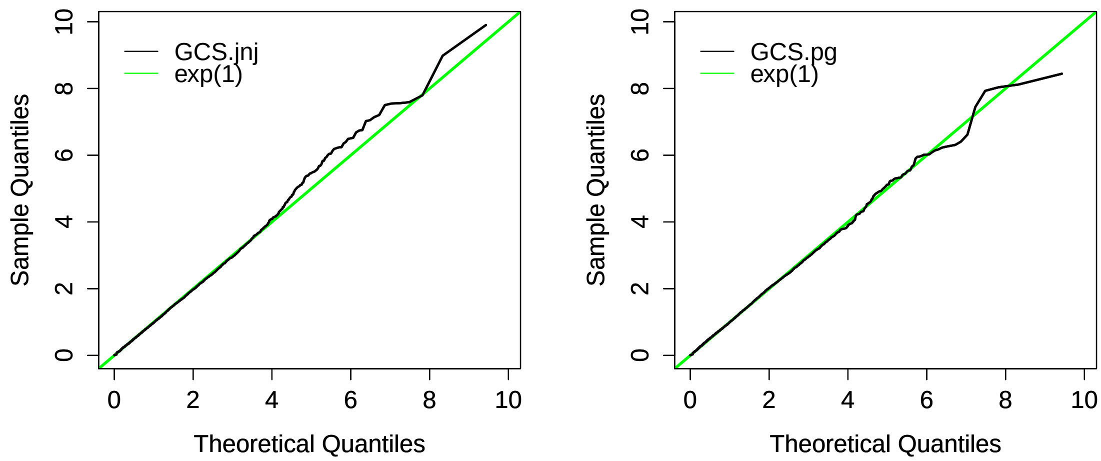

Figure 3.8: QQ plots for GCS residuals of JNJ (left) and PG (right) stocks

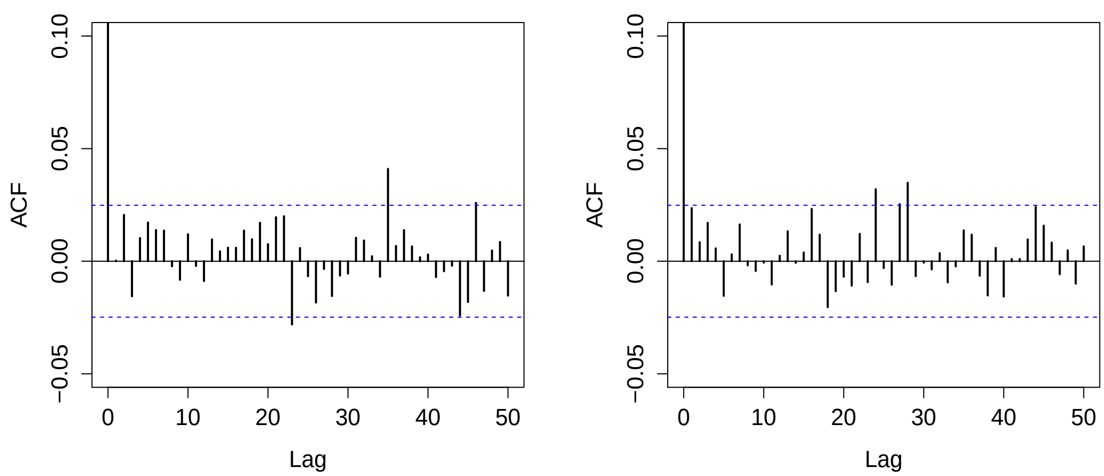

Figure 3.9: ACF plots for GCS residuals of JNJ (left) and PG (right) stocks

in the associated z and z^2 series.

Furthermore, the p-values from the KS, AD, LB(15) and $\text{LB}^2(15)$ tests for $\{z_1, z_{2|1}\}$ are 0.400, 0.230, 0.229 and 0.425. The corresponding values for $\{z_2, z_{1|2}\}$ are 0.452, 0.222, 0.215 and 0.418. Consequently, the null hypothesis that the pooled series $\{z_1, z_{2|1}\}$ and $\{z_2, z_{1|2}\}$ are *i.i.d.* $U(0, 1)$ can not be rejected at level 10%. These results all reveal that the BVBS-ACD model may provide good bivariate forecasts.

Table 3.9: P-values for out-of-sample tests for the BVBS-ACD model based on density forecasts

z series	KS	AD	LB(10)	LB(15)	LB(20)	$\text{LB}^2(10)$	$\text{LB}^2(15)$	$\text{LB}^2(20)$
z_1	0.179	0.106	0.529	0.300	0.258	0.816	0.433	0.561
z_2	0.478	0.278	0.400	0.361	0.303	0.152	0.170	0.160
$z_{1\|2}$	0.185	0.102	0.567	0.316	0.238	0.839	0.453	0.559
$z_{2\|1}$	0.508	0.267	0.401	0.390	0.297	0.139	0.174	0.149

Note: LB(l) represents LB test for z series over l lags, where $l = 10, 15$ and 20. $\text{LB}^2(l)$ represents LB test for z^2 series over l lags.

Chapter 4

Bivariate BSt-ACD model

4.1 Introduction

Kundu *et al.* (2010) extended the BS distribution to the bivariate case and proposed the bivariate Birnbaum-Saunders (BVBS) distribution. The bivariate Student-t BS (BVBSt) distribution was suggested by Vilca *et al.* (2014a) as a robust extension of the BVBS distribution; see Kundu *et al.* (2013) for generalized multivariate Birnbaum-Saunders distributions. Saulo *et al.* (2017b) studied the moment estimation of the parameters of the BVBSt distribution within the generalized bivariate Birnbaum-Saunders family.

Our goal here is to jointly model the trade durations of two assets based on matched data, pairs of durations with the same starting time (see Simonsen (2007)). In the last Chapter, we constructed a bivariate ACD model based on BVBS distribution which allowed us to jointly analyze and measure the strength of the dependence of two matched duration series. In the present Chapter, we propose a bivariate ACD model based on BVBSt distribution which is quite flexible and accounts for higher kurtosis with the inclusion of an additional parameter, namely, the degrees of freedom of the Student-t distribution. This can potentially facilitate robust parameter estimation of the model parameters.

The rest of this Chapter is organized as follows. In Section 4.2, we propose the BVBSt-ACD model and the corresponding log-likelihood function. We derive the first and second derivatives of $\ln L(\boldsymbol{\theta})$ in Sections 4.3 and 4.4. In Section 4.5, we discuss the maximum likelihood estimation and associated inference for the model parameters. A Monte Carlo simulation study is carried out in Section 4.6 to examine the properties of the MLEs. Model discrimination is discussed in Section 4.7. In Section 4.8 and 4.9, we provide an application of the BVBSt-ACD model to bivariate datasets, and then compare the results with those of the BVBS-ACD model.

4.2 Bivariate BSt-ACD model and the log-likelihood function

By working with the joint density of (X_{1i}, X_{2i}) directly, we propose the following BVBSt-ACD model:

$$X_{1i} = \sigma_{1i}\epsilon_{1i}, \qquad \ln\sigma_{1i} = \alpha_1 + \beta_1 \ln\sigma_{1,i-1} + \gamma_1 \left[\frac{X_{1,i-1}}{\sigma_{1,i-1}}\right], \quad |\beta_1| < 1, \tag{4.1}$$

$$X_{2i} = \sigma_{2i}\epsilon_{2i}, \qquad \ln\sigma_{2i} = \alpha_2 + \beta_2 \ln\sigma_{2,i-1} + \gamma_2 \left[\frac{X_{2,i-1}}{\sigma_{2,i-1}}\right], \quad |\beta_2| < 1, \tag{4.2}$$

where $(X_{1i}, X_{2i}) \sim$ BVBS$t(\kappa_1, \sigma_{1i}, \kappa_2, \sigma_{2i}, \rho, \nu)$, $\epsilon_{1i} \overset{\text{iid}}{\sim}$ BS$t(\kappa_1, 1, \nu)$ and $\epsilon_{2i} \overset{\text{iid}}{\sim}$ BS$t(\kappa_2, 1, \nu)$.

The associated joint PDF of the BVBSt-ACD model is given by

$$\begin{aligned} & f_{X_{1i},X_{2i}}(x_{1i},x_{2i};\boldsymbol{\theta}) \\ =& \frac{1}{2\kappa_1\sigma_{1i}}\left[\left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{3}{2}}\right]\frac{1}{2\kappa_2\sigma_{2i}}\left[\left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{3}{2}}\right] \\ \times& \boldsymbol{t_2}\left[\frac{1}{\kappa_1}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right),\frac{1}{\kappa_2}\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right);\rho,\nu\right] \\ =& \frac{1}{2\kappa_1 x_{1i}}\left[\left(\frac{x_{1i}}{\sigma_{1i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}}\right]\frac{1}{2\kappa_2 x_{2i}}\left[\left(\frac{x_{2i}}{\sigma_{2i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}}\right] \end{aligned} \tag{4.3}$$

$$\times\frac{1}{2\pi\sqrt{1-\rho^2}}\left[1+\frac{\frac{1}{\kappa_1^2}\left(\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2\right)+\frac{1}{\kappa_2^2}\left(\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2\right)-\frac{2\rho}{\kappa_1\kappa_2}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\nu(1-\rho^2)}\right]^{-\frac{\nu+2}{2}},$$

here, $\boldsymbol{\theta}=(\alpha_1,\beta_1,\gamma_1,\alpha_2,\beta_2,\gamma_2,\kappa_1,\kappa_2,\rho,\nu)'$ is the model parameters; in the above, $\boldsymbol{t_2}$ denotes the bivariate t density function.

For $i=1,2,\ldots,n$, the individual log-likelihood function can be expressed as

$$\begin{aligned} \ln l_i(\boldsymbol{\theta}) =& -\ln(\kappa_1)-\ln(\kappa_2)-\frac{1}{2}\ln(1-\rho^2)+\ln\left[\left(\frac{x_{1i}}{\sigma_{1i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}}\right]+\ln\left[\left(\frac{x_{2i}}{\sigma_{2i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}}\right] \\ & -\frac{\nu+2}{2}\ln\left[1+\frac{\frac{1}{\kappa_1^2}\left(\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2\right)+\frac{1}{\kappa_2^2}\left(\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2\right)+\frac{2\rho}{\kappa_1\kappa_2}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\nu(1-\rho^2)}\right]. \end{aligned}$$

The log-likelihood function, without the additive constant, is then given by

$$\begin{aligned} \ln L(\boldsymbol{\theta}) =& \sum_{i=1}^{n}\left\{-\ln(\kappa_1)-\ln(\kappa_2)-\frac{1}{2}\ln(1-\rho^2)+\ln\left[\left(\frac{x_{1i}}{\sigma_{1i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{1i}}{x_{1i}}\right)^{\frac{1}{2}}\right]+\ln\left[\left(\frac{x_{2i}}{\sigma_{2i}}\right)^{\frac{1}{2}}+\left(\frac{\sigma_{2i}}{x_{2i}}\right)^{\frac{1}{2}}\right]\right. \\ & \left.-\frac{\nu+2}{2}\ln\left[1+\frac{\frac{1}{\kappa_1^2}\left(\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2\right)+\frac{1}{\kappa_2^2}\left(\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2\right)+\frac{2\rho}{\kappa_1\kappa_2}\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\nu(1-\rho^2)}\right]\right\}. \end{aligned} \tag{4.4}$$

4.3 The first derivatives of ln$L(\boldsymbol{\theta})$

The first derivatives of ln$L(\boldsymbol{\theta})$ with respect to α_j, β_j and γ_j are given by

$$\frac{\partial \ln L}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$
$$\frac{\partial \ln L}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j},$$
$$\frac{\partial \ln L}{\partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j},$$

where

$$\frac{\partial \ln l_i}{\partial \sigma_{ji}} = -\frac{\frac{\nu+2}{2}\left(\frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{\kappa_1\,\kappa_2\,\sigma_{ji}} + \frac{\frac{1}{x_{ji}}-\frac{x_{ji}}{\sigma_{ji}{}^2}}{\kappa_j{}^2}\right)}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} + \frac{\sigma_{ji} - x_{ji}}{2\,\sigma_{ji}\,(x_{ji}+\sigma_{ji})}; \tag{4.5}$$

here $j = 1, 2$, $j' = \{1, 2\} - \{j\}$ and

$$Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) = -\frac{2\,\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1\,\kappa_2} + \frac{\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2}{\kappa_2{}^2} + \frac{\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2}{\kappa_1{}^2} + \nu\,\left(1-\rho^2\right); \tag{4.6}$$

also,

$$\begin{aligned}
\frac{\partial \sigma_{ji}}{\partial \alpha_j} =& \sigma_{ji}\left(1+\sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right), i \geq 1, && (4.7)\\
\frac{\partial \sigma_{ji}}{\partial \beta_j} =& \sigma_{ji}\left(\ln\sigma_{j,i-1}+\sum_{k=0}^{i-2}\ln\sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right), i \geq 2, && (4.8)\\
\frac{\partial \sigma_{ji}}{\partial \gamma_j} =& \sigma_{ji}\left(\frac{x_{j,i-1}}{\sigma_{j,i-1}}+\sum_{k=0}^{i-2}\frac{x_{jk}}{\sigma_{jk}}\prod_{l=k+1}^{i-1}\left(\beta_j-\gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right), i \geq 2, && (4.9)
\end{aligned}$$

with $\nabla\sigma_{j0} = (0,0,0)'$ and $\nabla\sigma_{j1} = (\sigma_{j1}, \sigma_{j1}\ln\sigma_{j0}, \sigma_{j1}x_{j0}/\sigma_{j0})'$.

The first derivatives of $\ln L(\boldsymbol{\theta})$ with respect to κ_j, $j = 1, 2$, ρ and ν are given by

$$\begin{aligned}
\frac{\partial \ln L}{\partial \kappa_j} &= \sum_{i=1}^{n}\frac{\partial \ln l_i}{\partial \kappa_j},\\
\frac{\partial \ln L}{\partial \rho} &= \sum_{i=1}^{n}\frac{\partial \ln l_i}{\partial \rho},\\
\frac{\partial \ln L}{\partial \nu} &= \sum_{i=1}^{n}\frac{\partial \ln l_i}{\partial \nu},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial \ln l_i}{\partial \kappa_j} =& -\frac{(\nu+2)\left(\frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^2\,\kappa_{j'}}-\frac{\frac{x_{ji}}{\sigma_{ji}}+\frac{\sigma_{ji}}{x_{ji}}-2}{{\kappa_j}^3}\right)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}-\frac{1}{\kappa_j},\\
\frac{\partial \ln l_i}{\partial \rho} =& -\frac{\rho\ (\nu+2)\ (Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)-(1-\rho^2)\nu)}{(1-\rho^2)\ Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}\\
&+\frac{(\nu+2)\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1\,\kappa_2\,Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}+\frac{\rho}{1-\rho^2},\\
\frac{\partial \ln l_i}{\partial \nu} =& \frac{(\nu+2)\ (Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)-(1-\rho^2)\nu)}{2\nu\,Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}\\
&-\frac{1}{2}\log\left(\frac{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}{(1-\rho^2)\nu}\right); && (4.10)
\end{aligned}$$

in the above, $Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)$ is defined as (4.6).

4.4 The second derivatives of $\ln L(\boldsymbol{\theta})$

The second derivatives of $\ln L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho, \nu)'$ are as follows:

$$\frac{\partial^2 \ln L}{\partial \alpha_j{}^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j{}^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \alpha_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \left(\frac{\partial \sigma_{ji}}{\partial \alpha_j} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \alpha_j^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \beta_j{}^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j{}^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \left(\frac{\partial \sigma_{ji}}{\partial \beta_j} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \beta_j^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \gamma_j{}^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j{}^2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \gamma_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \left(\frac{\partial \sigma_{ji}}{\partial \gamma_j} \right)^2 + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \gamma_j^2} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \, \partial \beta_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \, \partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \frac{\partial \sigma_{ji}}{\partial \beta_j} \frac{\partial \sigma_{ji}}{\partial \alpha_j} + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \alpha_j \, \partial \beta_j} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \gamma_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \frac{\partial \sigma_{ji}}{\partial \gamma_j} \frac{\partial \sigma_{ji}}{\partial \alpha_j} + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \alpha_j \, \partial \gamma_j} \right],$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \, \partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \gamma_j} = \sum_{i=1}^{n} \left[\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} \frac{\partial \sigma_{ji}}{\partial \gamma_j} \frac{\partial \sigma_{ji}}{\partial \beta_j} + \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial^2 \sigma_{ji}}{\partial \beta_j \, \partial \gamma_j} \right],$$

$$\frac{\partial^2 \ln L}{\partial \alpha_1 \, \partial \alpha_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_1 \, \partial \alpha_2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{1i}} \frac{\partial \sigma_{1i}}{\partial \alpha_1}}{\partial \alpha_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \, \partial \sigma_{2i}} \frac{\partial \sigma_{2i}}{\partial \alpha_2} \frac{\partial \sigma_{1i}}{\partial \alpha_1},$$

$$\frac{\partial^2 \ln L}{\partial \beta_1 \, \partial \beta_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_1 \, \partial \beta_2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{1i}} \frac{\partial \sigma_{1i}}{\partial \beta_1}}{\partial \beta_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \, \partial \sigma_{2i}} \frac{\partial \sigma_{2i}}{\partial \beta_2} \frac{\partial \sigma_{1i}}{\partial \beta_1},$$

$$\frac{\partial^2 \ln L}{\partial \gamma_1 \, \partial \gamma_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_1 \, \partial \gamma_2} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{1i}} \frac{\partial \sigma_{1i}}{\partial \gamma_1}}{\partial \gamma_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \, \partial \sigma_{2i}} \frac{\partial \sigma_{2i}}{\partial \gamma_2} \frac{\partial \sigma_{1i}}{\partial \gamma_1},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial \beta_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \beta_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \beta_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \sigma_{j'i}} \frac{\partial \sigma_{j'i}}{\partial \beta_{j'}} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \sigma_{j'i}} \frac{\partial \sigma_{j'i}}{\partial \gamma_{j'}} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \gamma_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \sigma_{j'i}} \frac{\partial \sigma_{j'i}}{\partial \gamma_{j'}} \frac{\partial \sigma_{ji}}{\partial \beta_j},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_j} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_j} \frac{\partial \sigma_{ji}}{\partial \beta_j},$$

$$\frac{\partial^2 \ln L}{\partial \gamma_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j \partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \kappa_j} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_j} \frac{\partial \sigma_{ji}}{\partial \gamma_j},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_{j'}} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_{j'}} \frac{\partial \sigma_{ji}}{\partial \beta_j},$$

$$\frac{\partial^2 \ln L}{\partial \gamma_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j \partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \kappa_{j'}} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \kappa_{j'}} \frac{\partial \sigma_{ji}}{\partial \gamma_j},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} \frac{\partial \sigma_{ji}}{\partial \beta_j},$$

$$\frac{\partial^2 \ln L}{\partial \gamma_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} \frac{\partial \sigma_{ji}}{\partial \gamma_j},$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \alpha_j \partial \nu} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \alpha_j}}{\partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \nu} \frac{\partial \sigma_{ji}}{\partial \alpha_j},$$

$$\frac{\partial^2 \ln L}{\partial \beta_j \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \beta_j \partial \nu} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \beta_j}}{\partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \nu} \frac{\partial \sigma_{ji}}{\partial \beta_j},$$

$$\frac{\partial^2 \ln L}{\partial \gamma_j \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \gamma_j \partial \nu} = \sum_{i=1}^{n} \frac{\partial \frac{\partial \ln l_i}{\partial \sigma_{ji}} \frac{\partial \sigma_{ji}}{\partial \gamma_j}}{\partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \nu} \frac{\partial \sigma_{ji}}{\partial \gamma_j},$$

$$\frac{\partial^2 \ln L}{\partial {\kappa_j}^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial {\kappa_j}^2},$$

$$\frac{\partial^2 \ln L}{\partial \kappa_1 \partial \kappa_2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \kappa_1 \partial \kappa_2},$$

$$\frac{\partial^2 \ln L}{\partial \kappa_j \partial \rho} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \kappa_j \partial \rho},$$

$$\frac{\partial^2 \ln L}{\partial \kappa_j \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \kappa_j \partial \nu},$$

$$\frac{\partial^2 \ln L}{\partial \rho \partial \nu} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \rho \partial \nu},$$

$$\frac{\partial^2 \ln L}{\partial \rho^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \rho^2},$$

$$\frac{\partial^2 \ln L}{\partial \nu^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln l_i}{\partial \nu^2},$$

where $Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)$ is defined as (4.6); also,

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{ji}^2} = \frac{(\nu+2)}{2Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \left\{ \frac{\left(\frac{\frac{1}{x_{ji}} - \frac{x_{ji}}{\sigma_{ji}^2}}{\kappa_j^2} + \frac{\rho \left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}} \right) \left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}} - \sqrt{\frac{\sigma_{j'i}}{x_{j'i}}} \right)}{\kappa_1 \kappa_2 \sigma_{ji}} \right)^2}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \right.$$
$$\left. - \frac{\rho \left(3\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}} \right) \left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}} - \sqrt{\frac{\sigma_{j'i}}{x_{j'i}}} \right)}{2\kappa_1 \kappa_2 \sigma_{ji}^2} - \frac{2 x_{ji}}{\kappa_j^2 \sigma_{ji}^3} \right\}$$
$$- \frac{\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} - \sqrt{\frac{\sigma_{ji}}{x_{ji}}} \right)^2}{4\sigma_{ji}^4 \left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}} \right)^2} + \frac{3\sqrt{\frac{x_{ji}}{\sigma_{ji}}} - \sqrt{\frac{\sigma_{ji}}{x_{ji}}}}{4\sigma_{ji}^2 \left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}} \right)}, \tag{4.11}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{1i} \partial \sigma_{2i}} = \frac{(\nu+2)}{2Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \left\{ \frac{\rho \left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} + \sqrt{\frac{\sigma_{1i}}{x_{1i}}} \right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} + \sqrt{\frac{\sigma_{2i}}{x_{2i}}} \right)}{2\kappa_1 \kappa_2 \sigma_{1i} \sigma_{2i}} \right.$$
$$\left. + \frac{\left(\frac{\rho \left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} + \sqrt{\frac{\sigma_{1i}}{x_{1i}}} \right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}} \right)}{2\kappa_1 \kappa_2 \sigma_{1i}} + \frac{\frac{1}{x_{1i}} - \frac{x_{1i}}{\sigma_{1i}^2}}{\kappa_1^2} \right) \left(\frac{\rho \left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}} \right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} + \sqrt{\frac{\sigma_{2i}}{x_{2i}}} \right)}{2\kappa_1 \kappa_2 \sigma_{2i}} + \frac{\frac{1}{x_{2i}} - \frac{x_{2i}}{\sigma_{2i}^2}}{\kappa_2^2} \right)}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \right\}, \tag{4.12}$$

$$\frac{\partial^2 \ln l_i}{\partial\,\sigma_{ji}\,\partial\,\kappa_j} = \frac{(\nu+2)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}\left\{\frac{\rho\,\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\,\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{2\,{\kappa_j}^2\,\kappa_{j'}\,\sigma_{ji}} + \frac{\left(\frac{1}{x_{ji}}-\frac{x_{ji}}{{\sigma_{ji}}^2}\right)}{{\kappa_j}^3}\right.$$
$$\left. + \frac{\left(\frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^2\,\kappa_{j'}} - \frac{\left(\frac{x_{ji}}{\sigma_{ji}}+\frac{\sigma_{ji}}{x_{ji}}-2\right)}{{\kappa_j}^3}\right)\left(\frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{\kappa_1\,\kappa_2\,\sigma_{ji}} + \frac{\frac{1}{x_{ji}}-\frac{x_{ji}}{{\sigma_{ji}}^2}}{{\kappa_j}^2}\right)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}\right\}, \tag{4.13}$$

$$\frac{\partial^2 \ln l_i}{\partial\,\sigma_{ji}\,\partial\,\kappa_{j'}} = \frac{(\nu+2)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}\left\{\frac{\rho\,\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\,\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{2\,\kappa_j\,\kappa_{j'}^2\,\sigma_{ji}}\right.$$
$$\left. + \frac{\left(\frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_j\,\kappa_{j'}^2} - \frac{\left(\frac{x_{j'i}}{\sigma_{j'i}}+\frac{\sigma_{j'i}}{x_{j'i}}-2\right)}{{\kappa_{j'}}^3}\right)\left(\frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{\kappa_1\,\kappa_2\,\sigma_{ji}} + \frac{\frac{1}{x_{ji}}-\frac{x_{ji}}{{\sigma_{ji}}^2}}{{\kappa_j}^2}\right)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)}\right\}, \tag{4.14}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \rho} = \frac{(\nu+2)}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \left\{ \frac{\rho \left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}} + \sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right) \left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}} - \sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{2\,\kappa_1\,\kappa_2\,\sigma_{ji}} \right.$$
$$- \frac{\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right) \left(\frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{\kappa_1\,\kappa_2\,\sigma_{ji}} + \frac{\frac{1}{x_{ji}} - \frac{x_{ji}}{\sigma_{ji}^2}}{\kappa_j^2} \right)}{\kappa_1\,\kappa_2\,Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)}$$
$$+ \frac{\rho\,\left(Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) - (1-\rho^2)\nu\right) \left(\frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{\kappa_1\,\kappa_2\,\sigma_{ji}} + \frac{\frac{1}{x_{ji}} - \frac{x_{ji}}{\sigma_{ji}^2}}{\kappa_j^2} \right)}{(1-\rho^2)Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)}$$
$$\left. - \frac{\rho \left(\frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{\kappa_1\,\kappa_2\,\sigma_{ji}} + \frac{\frac{1}{x_{ji}} - \frac{x_{ji}}{\sigma_{ji}^2}}{\kappa_j^2} \right)}{(1-\rho^2)} \right\}, \tag{4.15}$$

$$\frac{\partial^2 \ln l_i}{\partial \sigma_{ji} \partial \nu} = \frac{\left(\frac{\rho\left(\sqrt{\frac{x_{ji}}{\sigma_{ji}}}+\sqrt{\frac{\sigma_{ji}}{x_{ji}}}\right)\left(\sqrt{\frac{x_{j'i}}{\sigma_{j'i}}}-\sqrt{\frac{\sigma_{j'i}}{x_{j'i}}}\right)}{\kappa_1\,\kappa_2\,\sigma_{ji}} + \frac{\frac{1}{x_{ji}} - \frac{x_{ji}}{\sigma_{ji}^2}}{\kappa_j^2} \right)}{\nu Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \left\{1 - \frac{(\nu+2)\left(Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) - (1-\rho^2)\nu\right)}{2Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)}\right\}, \tag{4.16}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa_j^2} = \frac{(\nu+2)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)} \left\{ \frac{2\left(\frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^2\,\kappa_{j'}} - \frac{\frac{x_{ji}}{\sigma_{ji}}+\frac{\sigma_{ji}}{x_{ji}}-2}{{\kappa_j}^3}\right)^2}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)} \right.$$
$$\left. - \frac{2\,\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^3\,\kappa_{j'}} - \frac{3\left(\frac{x_{ji}}{\sigma_{ji}}+\frac{\sigma_{ji}}{x_{ji}}-2\right)}{{\kappa_j}^4} \right\} + \frac{1}{{\kappa_j}^2}, \tag{4.17}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa_1\,\partial \kappa_2} = \frac{(\nu+2)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)} \left\{ \frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_1}^2\,{\kappa_2}^2} \right.$$
$$\left. + \frac{2\left(\frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1\,{\kappa_2}^2} - \frac{\frac{x_{2i}}{\sigma_{2i}}+\frac{\sigma_{2i}}{x_{2i}}-2}{{\kappa_2}^3}\right)\left(\frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_1}^2\,\kappa_2} - \frac{\frac{x_{1i}}{\sigma_{1i}}+\frac{\sigma_{1i}}{x_{1i}}-2}{{\kappa_1}^3}\right)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)} \right\}, \tag{4.18}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa_j\,\partial \rho} = \frac{(\nu+2)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)} \left\{ -\frac{\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^2\,\kappa_{j'}} \right.$$
$$\left. + \frac{2\left(-\frac{\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1\,\kappa_2} - \rho\,\nu\right)\left(\frac{\rho\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}}-\sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right)\left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}}-\sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^2\,\kappa_{j'}} - \frac{\frac{x_{ji}}{\sigma_{ji}}+\frac{\sigma_{ji}}{x_{ji}}-2}{{\kappa_j}^3}\right)}{Q(x_{1i},x_{2i};\kappa_1,\kappa_2,\rho,\nu)} \right\}, \tag{4.19}$$

$$\frac{\partial^2 \ln l_i}{\partial \kappa_j \, \partial \nu} = \frac{1}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \left\{ -\frac{\rho \left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^2 \, \kappa_{j'}} \right.$$
$$\left. + \frac{\frac{x_{ji}}{\sigma_{ji}} + \frac{\sigma_{ji}}{x_{ji}} - 2}{{\kappa_j}^3} + \frac{(1-\rho^2)\,(v+2)\left(\frac{\rho \left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{{\kappa_j}^2 \, \kappa_{j'}} - \frac{\frac{x_{ji}}{\sigma_{ji}} + \frac{\sigma_{ji}}{x_{ji}} - 2}{{\kappa_j}^3}\right)}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \right\}, \tag{4.20}$$

$$\frac{\partial^2 \ln l_i}{\partial \rho \, \partial \nu} = \frac{1}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)\nu} \left\{ \frac{(\nu+2)\,(Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) - (1-\rho^2)\,\nu)}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} - 2 \right\}$$
$$\times \left\{ \frac{\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1 \, \kappa_2} - \frac{\rho\,(Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) - (1-\rho^2)\,\nu)}{(1-\rho^2)} \right\}, \tag{4.21}$$

$$\frac{\partial^2 \ln l_i}{\partial \rho^2} = \frac{(\nu+2)}{Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \left\{ \frac{2\,\rho \left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1 \, \kappa_2 \, (1-\rho^2)} \right.$$
$$- \frac{\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right) \left(-\frac{2\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1 \, \kappa_2} - 2\,\rho\,\nu\right)}{\kappa_1 \, \kappa_2 \, Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)}$$
$$- \frac{2\rho \left(\frac{\left(\sqrt{\frac{x_{1i}}{\sigma_{1i}}} - \sqrt{\frac{\sigma_{1i}}{x_{1i}}}\right) \left(\sqrt{\frac{x_{2i}}{\sigma_{2i}}} - \sqrt{\frac{\sigma_{2i}}{x_{2i}}}\right)}{\kappa_1 \, \kappa_2} + \rho\,\nu\right)(Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) - (1-\rho^2)\nu)}{(1-\rho^2)}$$
$$\left. - \frac{(1+\rho^2)\,(Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) - (1-\rho^2)\nu)}{{(1-\rho^2)}^2} \right\} + \frac{1}{1-\rho^2} + \frac{2\,\rho^2}{{(1-\rho^2)}^2}, \tag{4.22}$$

$$\frac{\partial^2 \ln l_i}{\partial\, \nu^2} = \frac{(Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu) - (1-\rho^2)\nu)}{\nu Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \left\{ -\frac{(\nu+2)}{2\nu} + 1 - \frac{(\nu+2)\,(1-\rho^2)}{2Q(x_{1i}, x_{2i}; \kappa_1, \kappa_2, \rho, \nu)} \right\}, \tag{4.23}$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \alpha_j^2} = \frac{\partial \sigma_{ji}}{\partial \alpha_j} \left(1 + \sum_{k=1}^{i-1} \prod_{l=k}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \right) + \sigma_{ji} \gamma \prod_{l=k}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \sum_{l=k}^{i-1} \frac{x_l \frac{\partial \sigma_{jl}}{\partial \alpha_j}}{\sigma_{jl}^2 \left(\beta_j - \frac{\gamma_j\, x_{jl}}{\sigma_{jl}} \right)}, \quad i \geq 1, \tag{4.24}$$

$$\frac{\partial^2 \sigma_{ji}}{\partial \beta_j^2} = \frac{\partial \sigma_{ji}}{\partial \gamma_j} \left\{ \ln \sigma_{j,i-1} + \sum_{k=0}^{i-2} \ln \sigma_{jk} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \right\}$$
$$+ \sigma_{ji} \left\{ \sum_{k=0}^{i-2} \left(\ln \sigma_{jk} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right) \sum_{l=k+1}^{i-1} \frac{\frac{\gamma_j x_{jl} \frac{\partial \sigma_{jl}}{\partial \gamma_j}}{\sigma_{jl}^2} + 1}{\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}}} + \frac{\frac{\partial \sigma_{jk}}{\partial \gamma_j} \prod_{l=k+1}^{i-1} \left(\beta_j - \gamma_j \frac{x_{jl}}{\sigma_{jl}} \right)}{\sigma_{jk}} \right) + \frac{\frac{\partial \sigma_{j,i-1}}{\partial \gamma_j}}{\sigma_{j,i-1}} \right\}, \quad i \geq 2, \tag{4.25}$$

$$\frac{\partial^2\sigma_{ji}}{\partial\gamma_j^2} = \frac{\partial\sigma_{ji}}{\partial\gamma_j}\left\{\frac{x_{j,i-1}}{\sigma_{j,i-1}} + \sum_{k=0}^{i-2}\frac{x_{jk}}{\sigma_{jk}}\prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right\}$$
$$+\sigma_{ji}\left\{\sum_{k=0}^{i-2}\left(\frac{x_{jk}}{\sigma_{jk}}\prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k+1}^{i-1}\frac{\frac{\gamma_j x_{jl}\frac{\partial\sigma_{jl}}{\partial\gamma_j}}{\sigma_{jl}^2}+1}{\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}} + \frac{\frac{\partial\sigma_{jk}}{\partial\gamma_j}\prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)}{\sigma_{jk}}\right) + \frac{\frac{\partial\sigma_{j,i-1}}{\partial\gamma_j}}{\sigma_{j,i-1}}\right\},$$
$$i \geq 2, \quad (4.26)$$

$$\frac{\partial^2\sigma_{ji}}{\partial\,\alpha_j\,\partial\,\beta_j} = \frac{\partial\sigma_{ji}}{\partial\beta_j}\left(1 + \sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right) + \sigma_{ji}\sum_{l=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k}^{i-1}\frac{\frac{\gamma_j\,x_{jl}\,\frac{\partial\sigma_{jl}}{\partial\beta_j}}{\sigma_{jl}^2}+1}{\beta_j - \frac{\gamma_j\,x_{jl}}{\sigma_{jl}}},$$
$$i \geq 1, \quad (4.27)$$

$$\frac{\partial^2\sigma_{ji}}{\partial\,\alpha_j\,\partial\,\gamma_j} = \frac{\partial\sigma_{ji}}{\partial\gamma_j}\left(1 + \sum_{k=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right) + \sigma_{ji}\sum_{l=1}^{i-1}\prod_{l=k}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k}^{i-1}\frac{\frac{\gamma_j\,x_{jl}\,\frac{\partial\sigma_{jl}}{\partial\beta_j}}{\sigma_{jl}^2}-\frac{x_{jl}}{\sigma_{jl}}}{\beta_j - \frac{\gamma_j\,x_{jl}}{\sigma_{jl}}},$$
$$i \geq 1, \quad (4.28)$$

$$\frac{\partial^2\sigma_{ji}}{\partial\,\beta_j\,\partial\,\gamma_j} = \frac{\partial\sigma_{ji}}{\partial\gamma_j}\left\{\ln\sigma_{j,i-1} + \sum_{k=0}^{i-2}\ln\sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\right\}$$
$$+\sigma_{ji}\left\{\sum_{k=0}^{i-2}\left(\ln\sigma_{jk}\prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)\sum_{l=k+1}^{i-1}\frac{\frac{\gamma_j x_{jl}\frac{\partial\sigma_{jl}}{\partial\gamma_j}}{\sigma_{jl}^2}-\frac{x_{jl}}{\sigma_{jl}}}{\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}} + \frac{\frac{\partial\sigma_{jk}}{\partial\gamma_j}\prod_{l=k+1}^{i-1}\left(\beta_j - \gamma_j\frac{x_{jl}}{\sigma_{jl}}\right)}{\sigma_{jk}}\right) + \frac{\frac{\partial\sigma_{j,i-1}}{\partial\gamma_j}}{\sigma_{j,i-1}}\right\}.$$
$$i \geq 2. \quad (4.29)$$

4.5 Estimation and inference

We estimate the parameter $\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho, \nu)'$ of the BVBSt-ACD model by maximizing the log-likelihood function (without the additive constant) in equation (4.4). We maximize the above log-likelihood function by a two-step procedure. In the first step, we fix κ_1, κ_2, ρ and ν at their initial values and maximize $\ln L(\boldsymbol{\theta})$ with respect to the ACD parameters α_j, β_j, and $\gamma_j, j = 1, 2$, by utilizing the NM algorithm. Then, in the second step, we estimate over the full parameter space by employing the BFGS algorithm. We have derived and implemented the analytical gradients (the first derivatives in Section 4.3) in the second step. The BFGS quasi-Newton method with the analytic gradient will be faster, more stable and lead to more accurate estimates than a numerical gradient (see Bard (1974), Bolker (2008) and Mayorov (2011)).

Under the regularity conditions, the ML estimator $\hat{\boldsymbol{\theta}}$ is consistent and $\hat{\boldsymbol{\theta}} \overset{a}{\sim} N_{10}\left(\boldsymbol{\theta}_0, \mathcal{I}(\boldsymbol{\theta_0})^{-1}\right)$, where $\mathcal{I}(\boldsymbol{\theta_0}) = -E\left[\left.\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right] = -E\left[\boldsymbol{H}(\boldsymbol{\theta_0})\right]$; $\boldsymbol{H} = \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ is the Hessian matrix. In order to estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$, we consider the consistent estimator $-\boldsymbol{H}(\hat{\boldsymbol{\theta}})^{-1}$. The standard errors can then be approximated by the square roots of the diagonal elements of this matrix.

4.6 Simulation study

In order to assess the performance of the maximum likelihood estimators of the parameters of the BVBSt-ACD model, we carry out a simulation study for different sample sizes and correlation parameter ρ, based on the method detailed in Section 4.5. For sample sizes n=1000, 3000, 5000 and 10000, we simulate 1000 samples from the BVBSt-ACD model with the vector of true parameters

$$\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho, \nu)' = (0.1, 0.9, 0.1, -0.1, 0.95, 0.1, 1.0, 1.05, \rho, 12)'.$$

Here, the value of ρ are chosen as 0.0, 0.2, 0.5 and 0.9. We choose the starting value for ACD parameters to be

$$(\alpha_{10}, \beta_{10}, \gamma_{10}, \alpha_{20}, \beta_{20}, \gamma_{20})' = (0.01, 0.80, 0.01, -0.01, 0.80, 0.01)'.$$

The initial values of shape parameters $\kappa_{j0}, j = 1, 2$, and the correlation parameter ρ_0 are given by

$$\kappa_{j0} = \sqrt{2\left(\frac{\overline{X}_j}{med(X_j)} - 1\right)\frac{\nu-2}{\nu}}, \;\; j = 1, 2,$$

and

$$\rho_0 = \text{Corr}\left(\frac{1}{\kappa_{10}}\left(\sqrt{\frac{X_1}{\hat{\sigma}_1}} - \sqrt{\frac{\hat{\sigma}_1}{X_1}}\right), \frac{1}{\kappa_{20}}\left(\sqrt{\frac{X_2}{\hat{\sigma}_2}} - \sqrt{\frac{\hat{\sigma}_2}{X_2}}\right)\right),$$

where $\hat{\sigma}_j$ is obtained as

$$\hat{\sigma}_{ji} = \exp\left(\alpha_{j0} + \beta_{j0}\ln\hat{\sigma}_{j,i-1} + \gamma_{j0}\left[\frac{X_{j,i-1}}{\hat{\sigma}_{j,i-1}}\right]\right),$$

$j = 1, 2,\; i = 1, 2, \ldots, n$. We set $\nu_0 = 3$.

The performance of the MLEs are examined in terms of the mean, coefficients of skewness and kurtosis, bias and root mean squared error (RMSE) of the MLEs we computed over 1000 replications for each sample size and for each level of correlation.

These results are presented in Tables 4.1-4.16. From these tables, we see that all estimators perform well across different levels of correlation. The bias, MSE and

RMSE decrease toward 0 quickly as sample size increases, which demonstrates the asymptotic unbiasedness and consistency properties of the MLEs. An increase in the sample size shrinks the standard errors of the mean. The skewness and kurtosis obtained from the estimators are close to their limiting values of 0 and 3 for a reasonable sample size. Clearly, the asymptotic normal approximation for the estimators improves with increasing sample size.

4.7 Model discrimination

As mentioned earlier, the BVBSt-ACD model allows more flexibility in terms of the kurtosis through an extra parameter, namely, the degrees of freedom of the Student-t distribution. For the purpose of model discrimination, we consider a collection of its special cases, BVBSt_7, BVBSt_9, BVBSt_{12} and BVBS-ACD models. From the general BVBSt-ACD model, we are interested in investigating how often the true model gets selected in the set of candidate models and select a parsimonious model that fits the data adequately. This model evaluation technique is called model discrimination. One may refer to the book by McLachlan and Peel (2000), and the recent papers by Balakrishnan and Peng (2006), Balakrishnan and Pal (2013), Balakrishnan and Pal (2016), and Balakrishnan *et al.* (2017).

Two well-known model selection criteria, AIC and BIC, are utilized here. They are given by AIC $= -2l + 2k$ and BIC $= -2l + k\log(n)$, where l stands for the maximized log-likelihood value, p is the number of model parameters to be estimated, and n is the sample size. The model with the lowest AIC or BIC will be selected. Given that the true data generating process (DGP) are the BVBSt_7, BVBSt_9, BVBSt_{12} and BV-

Table 4.1: Simulation results for BVBSt-ACD model when $n = 1000$ and $\rho = 0.0$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1047	0.8952	0.1044	-0.0998	0.9468	0.1021	1.0020	1.0437	0.0093	11.4213
Skew	1.1843	-1.0812	0.1918	0.3639	-0.6107	0.0669	0.0534	-0.0130	-0.1307	1.6467
Kurt	7.4287	7.3898	3.3430	3.3304	3.6136	3.3713	3.1052	2.8840	3.1350	6.8536
Bias	0.0047	-0.0048	0.0044	0.0002	-0.0032	0.0021	0.0020	-0.0063	0.0093	-0.5787
MSE	0.0034	0.0006	0.0002	0.0004	0.0002	0.0001	0.0009	0.0010	0.0012	8.8907
RMSE	0.0586	0.0249	0.0132	0.0196	0.0130	0.0116	0.0307	0.0320	0.0345	2.9817
SE	0.0584	0.0244	0.0124	0.0196	0.0126	0.0114	0.0306	0.0314	0.0332	2.9265

Table 4.2: Simulation results for BVBSt-ACD model when $n = 1000$ and $\rho = 0.2$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1014	0.8964	0.1044	-0.0974	0.9468	0.1007	1.0029	1.0452	0.2052	11.6583
Skew	0.6768	-0.5418	0.1008	0.7212	-1.0215	0.0819	0.0092	0.0513	-0.0403	1.6153
Kurt	4.1275	3.8341	3.1454	4.2014	5.3795	2.9682	2.8602	3.2408	2.8969	8.1023
Bias	0.0014	-0.0036	0.0044	0.0026	-0.0032	0.0007	0.0029	-0.0048	0.0052	-0.3417
MSE	0.0027	0.0005	0.0002	0.0004	0.0002	0.0001	0.0009	0.0010	0.0011	8.8651
RMSE	0.0521	0.0221	0.0131	0.0198	0.0138	0.0106	0.0297	0.0314	0.0332	2.9774
SE	0.0521	0.0218	0.0124	0.0196	0.0134	0.0106	0.0295	0.0310	0.0328	2.9592

Table 4.3: Simulation results for BVBSt-ACD model when $n = 1000$ and $\rho = 0.5$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1052	0.8956	0.1034	-0.0991	0.9474	0.1012	1.0034	1.0446	0.5038	11.5242
Skew	0.6547	-0.5637	0.1934	0.3029	-0.5333	0.0845	0.0651	0.0091	-0.2359	1.9076
Kurt	3.8604	3.8578	3.1611	3.2348	3.5331	3.3936	3.1519	2.9583	3.2381	8.7965
Bias	0.0052	-0.0044	0.0034	0.0009	-0.0026	0.0012	0.0034	-0.0054	0.0038	-0.4758
MSE	0.0026	0.0005	0.0001	0.0003	0.0001	0.0001	0.0009	0.0010	0.0006	9.3740
RMSE	0.0514	0.0219	0.0116	0.0171	0.0113	0.0100	0.0304	0.0320	0.0253	3.0617
SE	0.0511	0.0214	0.0111	0.0171	0.0110	0.0100	0.0302	0.0315	0.0250	3.0260

Table 4.4: Simulation results for BVBSt-ACD model when $n = 1000$ and $\rho = 0.9$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1004	0.8974	0.1032	-0.1010	0.9484	0.1014	0.9941	1.0372	0.8989	10.4064
Skew	0.0467	0.1223	-0.0223	0.0741	-0.0413	0.0024	0.0872	0.3339	-0.3515	1.8263
Kurt	4.3680	4.9159	3.7957	3.3658	4.7117	3.8622	3.2550	5.2337	3.3913	8.9198
Bias	0.0004	-0.0026	0.0032	-0.0010	-0.0016	0.0014	-0.0059	-0.0128	-0.0011	-1.5936
MSE	0.0013	0.0002	0.0001	0.0002	0.0001	0.0001	0.0009	0.0012	0.0000	8.9736
RMSE	0.0361	0.0150	0.0093	0.0131	0.0079	0.0079	0.0306	0.0349	0.0066	2.9956
SE	0.0361	0.0148	0.0087	0.0130	0.0078	0.0078	0.0300	0.0324	0.0065	2.5378

Table 4.5: Simulation results for BVBSt-ACD model when $n = 3000$ and $\rho = 0.0$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1002	0.8988	0.1015	-0.0990	0.9487	0.1004	0.9998	1.0474	0.0030	11.6888
Skew	0.4678	-0.3376	0.1518	0.2680	-0.4674	0.2388	0.0792	0.1310	0.1439	0.8277
Kurt	3.2923	3.0950	2.9312	3.1852	3.3588	2.9981	2.8968	3.0442	2.8641	3.9152
Bias	0.0002	-0.0012	0.0015	0.0010	-0.0013	0.0004	-0.0002	-0.0026	0.0030	-0.3112
MSE	0.0010	0.0002	0.0001	0.0001	0.0001	0.0000	0.0003	0.0004	0.0004	2.4541
RMSE	0.0316	0.0136	0.0073	0.0111	0.0075	0.0061	0.0171	0.0188	0.0192	1.5666
SE	0.0316	0.0136	0.0071	0.0111	0.0074	0.0061	0.0171	0.0186	0.0190	1.5361

Table 4.6: Simulation results for BVBSt-ACD model when $n = 3000$ and $\rho = 0.2$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1015	0.8985	0.1014	-0.0989	0.9488	0.1002	0.9999	1.0469	0.2026	11.7139
Skew	0.3924	-0.3209	-0.0016	0.2728	-0.4342	-0.0300	-0.0064	-0.0297	0.1409	0.7954
Kurt	3.0132	3.0105	3.4154	3.4049	3.2703	3.1339	3.1564	2.9972	3.0141	3.8202
Bias	0.0015	-0.0015	0.0014	0.0011	-0.0012	0.0002	-0.0001	-0.0031	0.0026	-0.2861
MSE	0.0009	0.0002	0.0001	0.0001	0.0000	0.0000	0.0003	0.0003	0.0004	2.5165
RMSE	0.0306	0.0129	0.0076	0.0105	0.0068	0.0059	0.0177	0.0185	0.0195	1.5864
SE	0.0305	0.0128	0.0074	0.0104	0.0067	0.0059	0.0177	0.0183	0.0193	1.5611

Table 4.7: Simulation results for BVBSt-ACD model when $n = 3000$ and $\rho = 0.5$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1017	0.8985	0.1012	-0.0989	0.9489	0.1001	1.0001	1.0471	0.5014	11.7283
Skew	0.3472	-0.3114	0.0211	0.2399	-0.3930	-0.0215	-0.0215	-0.0344	0.1039	0.8159
Kurt	2.9604	3.0222	3.3860	3.3000	3.3349	3.0297	3.1442	3.0421	3.0008	3.9104
Bias	0.0017	-0.0015	0.0012	0.0011	-0.0011	0.0001	0.0001	-0.0029	0.0014	-0.2717
MSE	0.0008	0.0001	0.0000	0.0001	0.0000	0.0000	0.0003	0.0003	0.0002	2.5618
RMSE	0.0277	0.0117	0.0068	0.0095	0.0062	0.0054	0.0176	0.0186	0.0151	1.6006
SE	0.0276	0.0116	0.0067	0.0094	0.0061	0.0054	0.0176	0.0184	0.0151	1.5781

Table 4.8: Simulation results for BVBSt-ACD model when $n = 3000$ and $\rho = 0.9$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1006	0.8990	0.1011	-0.0999	0.9494	0.1003	0.9964	1.0439	0.8996	11.2408
Skew	0.2912	-0.2795	0.0928	0.1622	-0.3255	0.0311	-0.0189	0.0220	0.0609	0.7550
Kurt	2.9438	2.9762	3.2166	3.0037	3.3401	2.9659	3.1098	3.1162	2.9959	3.6387
Bias	0.0006	-0.0010	0.0011	0.0001	-0.0006	0.0003	-0.0036	-0.0061	-0.0004	-0.7592
MSE	0.0004	0.0001	0.0000	0.0001	0.0000	0.0000	0.0003	0.0004	0.0000	2.7348
RMSE	0.0201	0.0082	0.0052	0.0074	0.0043	0.0044	0.0179	0.0194	0.0039	1.6537
SE	0.0201	0.0082	0.0051	0.0074	0.0043	0.0044	0.0175	0.0184	0.0038	1.4699

Table 4.9: Simulation results for BVBSt-ACD model when $n = 5000$ and $\rho = 0.0$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1004	0.8994	0.1007	-0.0995	0.9492	0.1002	0.9995	1.0477	0.0019	11.7867
Skew	0.3106	-0.3392	0.0134	0.1315	-0.2489	0.1334	0.0488	-0.0035	0.0406	0.5259
Kurt	3.1250	3.1619	3.0099	2.9958	3.1815	3.0125	2.9455	2.9299	3.3382	3.0521
Bias	0.0004	-0.0006	0.0007	0.0005	-0.0008	0.0002	-0.0005	-0.0023	0.0019	-0.2133
MSE	0.0006	0.0001	0.0000	0.0001	0.0000	0.0000	0.0002	0.0002	0.0002	1.3949
RMSE	0.0242	0.0101	0.0055	0.0082	0.0054	0.0046	0.0137	0.0136	0.0156	1.1811
SE	0.0242	0.0101	0.0055	0.0082	0.0054	0.0046	0.0137	0.0134	0.0155	1.1622

Table 4.10: Simulation results for BVBSt-ACD model when $n = 5000$ and $\rho = 0.2$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1005	0.8994	0.1006	-0.0995	0.9492	0.1002	0.9998	1.0479	0.2015	11.8201
Skew	0.2878	-0.3073	0.0163	0.1145	-0.2483	0.1261	0.0415	-0.0242	0.0179	0.5267
Kurt	3.1018	3.1005	2.9518	2.9385	3.2203	3.0140	2.9628	2.9244	3.3238	3.0966
Bias	0.0005	-0.0006	0.0006	0.0005	-0.0008	0.0002	-0.0002	-0.0021	0.0015	-0.1799
MSE	0.0006	0.0001	0.0000	0.0001	0.0000	0.0000	0.0002	0.0002	0.0002	1.3950
RMSE	0.0235	0.0098	0.0054	0.0080	0.0053	0.0045	0.0137	0.0136	0.0150	1.1811
SE	0.0235	0.0097	0.0053	0.0080	0.0053	0.0045	0.0137	0.0135	0.0149	1.1679

Table 4.11: Simulation results for BVBSt-ACD model when $n = 5000$ and $\rho = 0.5$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1006	0.8994	0.1005	-0.0996	0.9494	0.1001	0.9994	1.0475	0.5007	11.7411
Skew	0.2462	-0.2660	0.0434	0.1069	-0.2151	0.0791	0.0179	-0.0691	-0.0148	0.5491
Kurt	3.0779	3.0373	2.8908	2.8677	3.2043	3.0117	3.0171	2.9668	3.3108	3.2238
Bias	0.0006	-0.0006	0.0005	0.0004	-0.0006	0.0001	-0.0006	-0.0025	0.0007	-0.2589
MSE	0.0004	0.0001	0.0000	0.0001	0.0000	0.0000	0.0002	0.0002	0.0001	1.3561
RMSE	0.0209	0.0087	0.0048	0.0072	0.0048	0.0040	0.0135	0.0137	0.0117	1.1645
SE	0.0209	0.0087	0.0047	0.0071	0.0047	0.0040	0.0135	0.0135	0.0116	1.1359

Table 4.12: Simulation results for BVBSt-ACD model when $n = 5000$ and $\rho = 0.9$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.0999	0.8996	0.1005	-0.1001	0.9497	0.1002	0.9970	1.0454	0.8997	11.4350
Skew	0.1434	-0.1861	0.1117	0.0959	-0.1760	0.0386	-0.0659	-0.1003	-0.0624	0.6420
Kurt	3.0144	2.9651	3.0163	2.6881	3.1966	2.9702	3.0726	3.1128	3.2961	3.5677
Bias	-0.0001	-0.0004	0.0005	-0.0001	-0.0003	0.0002	-0.0030	-0.0046	-0.0003	-0.5650
MSE	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0002	0.0000	1.6033
RMSE	0.0149	0.0060	0.0036	0.0054	0.0033	0.0032	0.0136	0.0144	0.0030	1.2662
SE	0.0149	0.0060	0.0036	0.0054	0.0033	0.0032	0.0133	0.0136	0.0030	1.1337

Table 4.13: Simulation results for BVBSt-ACD model when $n = 10000$ and $\rho = 0.0$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1004	0.8996	0.1003	-0.1000	0.9498	0.1002	0.9993	1.0486	0.0008	11.8615
Skew	0.1088	-0.1241	-0.0315	0.1266	-0.1636	0.0570	-0.0760	-0.0029	0.1040	0.3358
Kurt	2.7984	2.8015	3.1167	2.9711	3.0264	2.9084	3.0068	2.7009	3.0058	3.2362
Bias	0.0004	-0.0004	0.0003	0.0000	-0.0002	0.0002	-0.0007	-0.0014	0.0008	-0.1385
MSE	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.6708
RMSE	0.0170	0.0072	0.0038	0.0055	0.0036	0.0033	0.0096	0.0099	0.0107	0.8190
SE	0.0170	0.0072	0.0037	0.0055	0.0036	0.0033	0.0095	0.0098	0.0107	0.8076

Table 4.14: Simulation results for BVBSt-ACD model when $n = 10000$ and $\rho = 0.2$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1006	0.8995	0.1003	-0.1000	0.9498	0.1002	0.9994	1.0487	0.2006	11.8732
Skew	0.1258	-0.1388	-0.0411	0.1137	-0.1852	0.0540	-0.0768	0.0053	0.0893	0.3892
Kurt	2.8940	2.8965	3.0904	2.9903	3.1591	2.9323	3.0054	2.7508	3.0131	3.3732
Bias	0.0006	-0.0005	0.0003	0.0000	-0.0002	0.0002	-0.0006	-0.0013	0.0006	-0.1268
MSE	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.6929
RMSE	0.0168	0.0071	0.0037	0.0054	0.0035	0.0032	0.0096	0.0099	0.0103	0.8324
SE	0.0168	0.0071	0.0037	0.0054	0.0035	0.0032	0.0096	0.0098	0.0103	0.8231

Table 4.15: Simulation results for BVBSt-ACD model when $n = 10000$ and $\rho = 0.5$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1007	0.8995	0.1002	-0.1000	0.9499	0.1001	0.9996	1.0488	0.5002	11.9009
Skew	0.1474	-0.1495	-0.0162	0.0890	-0.1908	0.0513	-0.0747	0.0340	0.0718	0.3706
Kurt	3.0963	3.0914	3.0280	3.0979	3.3308	2.9971	3.0099	2.8639	3.0212	3.3272
Bias	0.0007	-0.0005	0.0002	0.0000	-0.0001	0.0001	-0.0004	-0.0012	0.0002	-0.0991
MSE	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.6768
RMSE	0.0152	0.0064	0.0034	0.0048	0.0032	0.0029	0.0095	0.0099	0.0080	0.8227
SE	0.0152	0.0064	0.0034	0.0049	0.0032	0.0029	0.0095	0.0098	0.0080	0.8171

Table 4.16: Simulation results for BVBSt-ACD model when $n = 10000$ and $\rho = 0.9$

	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$
Mean	0.1001	0.8998	0.1003	-0.1001	0.9499	0.1002	0.9982	1.0475	0.8998	11.7047
Skew	0.0573	-0.0204	0.0619	-0.0404	-0.0449	0.0759	-0.0337	0.0265	0.0353	0.3554
Kurt	3.2270	3.1912	2.8567	3.1645	3.1138	2.9806	3.0160	3.0525	3.0439	3.2162
Bias	0.0001	-0.0002	0.0003	-0.0001	-0.0001	0.0002	-0.0018	-0.0025	-0.0002	-0.2953
MSE	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0000	0.7371
RMSE	0.0109	0.0045	0.0026	0.0038	0.0023	0.0023	0.0096	0.0102	0.0020	0.8586
SE	0.0109	0.0045	0.0026	0.0038	0.0023	0.0023	0.0094	0.0099	0.0020	0.8066

BS-ACD models (true models), respectively, we fit the BVBSt_7, BVBSt_9, BVBSt_{12} and BVBS-ACD models to the simulated datasets generated from each true model, respectively.

Table 4.17: Model Discrimination by AIC and BIC

	True BVBSt-ACD Model			
	n=5000			
Fitted ACD Model	BVBSt_7	BVBSt_9	BVBSt_{12}	BVBS
BVBSt_7	**0.963**	0.084	0.000	0.000
BVBSt_9	0.037	**0.878**	0.120	0.001
BVBSt_{12}	0.000	0.038	**0.880**	0.032
BVBS	0.000	0.000	0.000	**0.967**

For sample sizes $n = 5000$, we simulated 1000 samples from the BVBSt-ACD model with the vector of true parameters

$$\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho, \nu)' = (0.1, 0.9, 0.1, -0.1, 0.95, 0.1, 1.0, 1.05, 0.5, \nu)'.$$

Here, we chose the degree of freedom ν to be $7, 9, 12$ and $+\infty$. We then assess the performance of the AIC and BIC by the selection rates for each of the fitted model.

As seen in Table 4.17, AIC/BIC can distinguish between the BVBSt_7, BVBSt_9, BVBSt_{12} and BVBS-ACD models with a relatively high selection rates for the true models. AIC/BIC perform best when the true models are BVBSt_7-ACD and BVBS-ACD. The selection rate of the correct model is 96.3% if the true DGP is BVBSt_7-ACD and 96.7% if the true DGP is BVBS-ACD. AIC/BIC perform well when the true models are BVBSt_9-ACD and BVBSt_{12}-ACD. The selection rate of the right model is 87.8% if the true DGP is BVBSt_9-ACD and 88.0% if the true DGP is BVBSt_{12}-ACD.

As expected, when the true models are BVBSt_7-ACD and BVBSt_9-ACD, the selection rates for BVBSt_7-ACD are higher than those of BVBSt_{12}-ACD, which implies that BVBSt_7-ACD is closer to BVBSt_9-ACD than BVBSt_{12}-ACD.

4.8 Illustrative example 1: Example revisited

Does the BVBS-ACD model provide the best fit for the real data in Chapter 3 or can we improve the fit using the BVBSt-ACD model? In this section, we fit the BVBSt-ACD model with different degree of freedom to the matched data. we choose ν =7, 9, 12 and $+\infty$. The corresponding maximized log-likelihood and AIC values, for different choices of ν, are presented in Table 4.18. As ν increases, the maximized log-likelihood value monotonically increases and the AIC value monotonically decreases. Thus, the BVBS-ACD model gives the best fit.

Table 4.18: The maximized log-likelihood value and AIC versus degrees of freedom ν.

ν	Maximized log-likelihood	AIC
7	2962.930	-5907.859
9	3009.919	-6001.838
12	3045.486	-6072.972
$+\infty$	**3103.805**	**-6189.610**

4.9 Illustrative example 2: Simulated data

In this section, we provide an example based on a simulated dataset to illustrate the BVBSt-ACD model developed here. We compare the results of the BVBSt-ACD model with those of the BVBS-ACD model.

4.9.1 Data description

For sample size $n = 5000$, we generated a bivariate dataset from a BVBSt_9-ACD model with the vector of true parameters

$$\begin{aligned}\boldsymbol{\theta} &= (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \kappa_1, \kappa_2, \rho, \nu)' \\ &= (-0.007, 0.991, 0.004, -0.010, 0.991, 0.008, 1.024, 1.077, 0.2, 9)'.\end{aligned} \tag{4.30}$$

As seen in Figure 4.1 and Table 4.19, the two simulated series are positive and right-skewed with higher degree of kurtosis indicating heavy tails.

Table 4.19: Summary statistics for the simulated bivariate dataset

Data	Min	Median	Mean	Max	SD	skew	kurt
Series 1	0.025	1.019	1.708	28.140	2.202	4.212	31.134
Series 2	0.018	1.505	2.692	57.880	3.590	4.206	33.626

In Figure 4.2, the empirical scaled TTT transforms are first concave and then convex, revealing that both marginals may possess unimodal shaped hazard rates.

So, the BVBSt distribution could be a good fit to the duration data due to the shape of its marginal density functions and hazard rates.

The ACF plots in Figure 4.3 indicate that there is a positive autocorrelation in both series, which suggests that a ACD specification may be a reasonable choice.

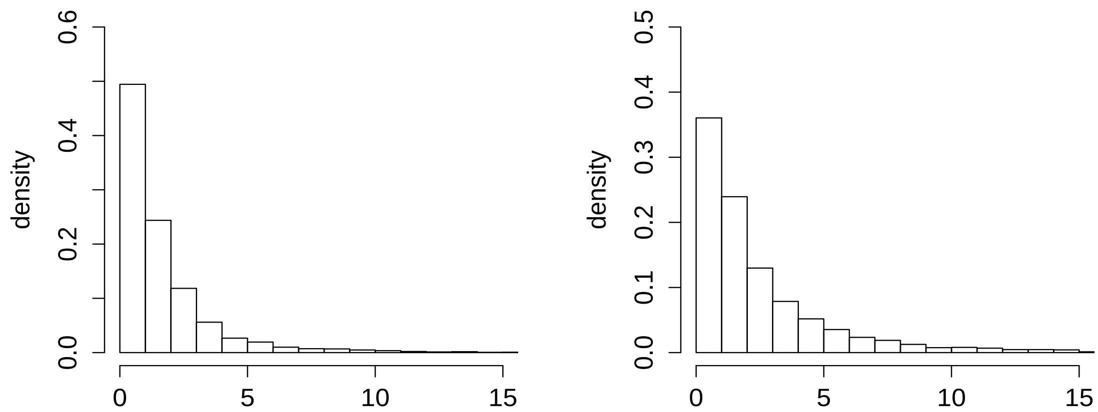

Figure 4.1: Histograms for the simulated bivariate dataset of Series 1 (left) and Series 2 (right)

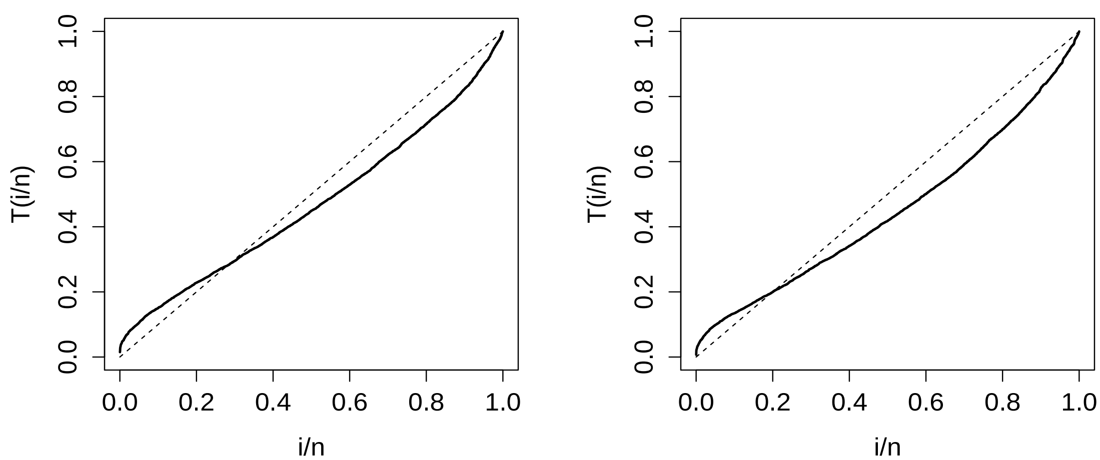

Figure 4.2: TTT plots for the simulated bivariate dataset of Series 1 (left) and Series 2 (right)

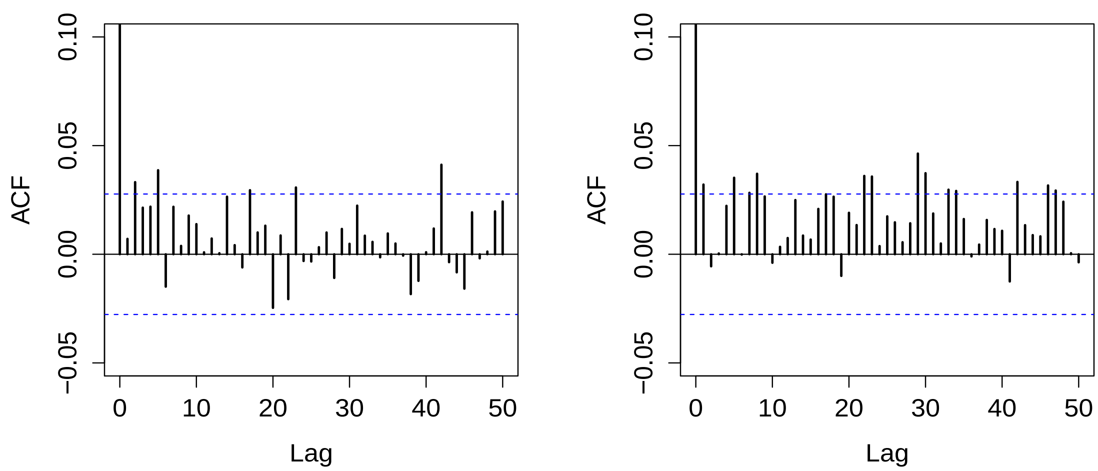

Figure 4.3: ACF plots for the simulated bivariate dataset of Series 1 (left) and Series 2 (right)

4.9.2 Estimation results

We employ the method described earlier in Section 4.5 to estimate the BVBSt-ACD model with the simulated bivariate dataset. The estimation results of both BVBS-ACD and BVBSt-ACD models are presented in Table 4.20. All estimates are statistically significant at 1% level. The BVBSt-ACD model is seen to provide much better fit to the data. The difference in AIC and BIC values between the two models are -265.57 and -250.535, respectively, which indicates that the AIC and BIC values of the BVBSt-ACD model are substantially smaller than those of the BVBS-ACD model.

Table 4.20: Estimation results for two bivariate ACD models

	BVBSt-ACD model												
	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$	$\hat{\nu}$	maxlnL	AIC	BIC
MLE	-0.004	0.997	0.003	-0.009	0.990	0.007	1.036	1.097	0.213	9.107	1560.22	-3100.44	-2950.096
SE	0.001	0.002	0.001	0.002	0.004	0.002	0.014	0.015	0.015	0.720			
	BVBS-ACD model												
	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\gamma}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\rho}$		maxlnL	AIC	BIC
MLE	-0.005	0.998	0.003	-0.010	0.991	0.008	1.173	1.240	0.215		1426.435	-2834.87	-2699.561
SE	0.001	0.002	0.001	0.002	0.003	0.001	0.012	0.012	0.013				

Then, the in-sample predictive model of the BVBSt-ACD model is given by

$$\hat{\sigma}_{1i} = \exp\left(-0.004 + 0.997 \ln \hat{\sigma}_{1,i-1} + 0.003 \left[\frac{X_{1,i-1}}{\hat{\sigma}_{1,i-1}}\right]\right),$$

$$\hat{\sigma}_{2i} = \exp\left(-0.009 + 0.990 \ln \hat{\sigma}_{2,i-1} + 0.007 \left[\frac{X_{2,i-1}}{\hat{\sigma}_{2,i-1}}\right]\right)$$

while that of the BVBS-ACD model is given by

$$\hat{\sigma}_{1i} = \exp\left(-0.005 + 0.998 \ln \hat{\sigma}_{1,i-1} + 0.003 \left[\frac{X_{1,i-1}}{\hat{\sigma}_{1,i-1}}\right]\right),$$

$$\hat{\sigma}_{2i} = \exp\left(-0.010 + 0.991 \ln \hat{\sigma}_{2,i-1} + 0.008 \left[\frac{X_{2,i-1}}{\hat{\sigma}_{2,i-1}}\right]\right).$$

4.9.3 Goodness-of-fit

To evaluate the goodness-of-fit of the BVBSt and BVBS-ACD models, we investigate the in-sample one-step-ahead density forecasts implied by the predictive models. We examine the uniformity of the z series by Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests and also check the independence of the z series by Ljung-Box (LB) tests. Results of these tests are presented in Tables 4.21 and 4.22.

As seen in Table 4.21, at 10% significance level, the null hypothesises that the series z_1, z_2, $z_{1|2}$, $z_{2|1}$ and the pooled series $\{z_1, z_{2|1}\}$ and $\{z_2, z_{1|2}\}$ are i.i.d. $U(0, 1)$ can not be rejected. The LB tests show the absence of serial correlation in the associated z and z^2 series. Hence the BVBSt-ACD model provides a good fit to the data. As expected, the BVBS-ACD model doesn't fit data well. the null hypothesis that the z series should each be i.i.d. $U(0, 1)$ are rejected.

4.9.4 Out-of-sample forecast evaluation

We use the first three fifth of the simulated data, "in-sample" observations 1 to 3000, to estimate the model parameters and then employ the resulting predictive model to the rest of the data, "out-of-sample" observations 3001 to 5000, to form density forecasts. These evaluation results are presented in Tables 4.23 and 4.24. A seen in Table 4.23, at 10% significance level, the null hypothesis that the series z_1, z_2, $z_{1|2}$, $z_{2|1}$ and the pooled series $\{z_1, z_{2|1}\}$ and $\{z_2, z_{1|2}\}$ are i.i.d. $U(0, 1)$ can not be rejected. The KS and AD tests both seem to support the uniformity of the corresponding z series. The LB tests indicate no serial correlation in the associated z and z^2 series. As shown in Table 4.24, at roughly 1% significance level, the null hypothesis that the series z_1, z_2, $z_{1|2}$, $z_{2|1}$ and the pooled series $\{z_1, z_{2|1}\}$ and $\{z_2, z_{1|2}\}$ are i.i.d. $U(0, 1)$ can not be rejected. Therefore, these results all reveal that the BVBSt-ACD model yields better bivariate forecasts than the BVBS-ACD model.

Table 4.21: P-values for in-sample goodness-of-fit tests for the BVBSt-ACD model based on density forecasts

z series	KS	AD	LB(10)	LB(15)	LB(20)	$\text{LB}^2(10)$	$\text{LB}^2(15)$	$\text{LB}^2(20)$
z_1	0.938	0.993	0.798	0.688	0.787	0.808	0.756	0.768
z_2	0.914	0.894	0.522	0.602	0.546	0.610	0.781	0.680
$z_{1\|2}$	0.957	0.987	0.683	0.561	0.698	0.709	0.653	0.685
$z_{2\|1}$	0.824	0.962	0.741	0.718	0.657	0.782	0.858	0.808
$z_1, z_{2\|1}$	0.913	0.975	0.918	0.560	0.693	0.918	0.560	0.693
$z_2, z_{1\|2}$	0.935	0.960	0.778	0.477	0.637	0.760	0.610	0.748

Note: LB(l) represents LB test for z series over l lags, where $l = 10, 15$ and 20. $\text{LB}^2(l)$ represents LB test for z^2 series over l lags.

Table 4.22: P-values for in-sample goodness-of-fit tests for the BVBS-ACD model based on density forecasts

z series	KS	AD	LB(10)	LB(15)	LB(20)	$\text{LB}^2(10)$	$\text{LB}^2(15)$	$\text{LB}^2(20)$
z_1	0.007	0.001	0.819	0.715	0.791	0.828	0.782	0.772
z_2	0.052	0.004	0.524	0.597	0.517	0.590	0.763	0.647
$z_{1\|2}$	0.006	0.001	0.725	0.601	0.737	0.739	0.667	0.734
$z_{2\|1}$	0.022	0.005	0.749	0.724	0.655	0.758	0.858	0.811
$z_1, z_{2\|1}$	0.001	< 0.001	0.925	0.573	0.699	0.936	0.766	0.863
$z_2, z_{1\|2}$	0.002	< 0.001	0.813	0.492	0.633	0.809	0.631	0.762

Note: LB(l) represents LB test for z series over l lags, where $l = 10, 15$ and 20. $\text{LB}^2(l)$ represents LB test for z^2 series over l lags.

Table 4.23: P-values for out-of-sample tests for the of BVBSt-ACD model based on density forecasts

z series	KS	AD	LB(10)	LB(15)	LB(20)	LB^2(10)	LB^2(15)	LB^2(20)
z_1	0.560	0.435	0.400	0.562	0.315	0.411	0.556	0.289
z_2	0.164	0.145	0.773	0.757	0.535	0.938	0.857	0.690
$z_{1\|2}$	0.609	0.357	0.260	0.468	0.281	0.273	0.470	0.257
$z_{2\|1}$	0.055	0.096	0.759	0.853	0.670	0.889	0.942	0.842
$z_1, z_{2\|1}$	0.444	0.572	0.562	0.698	0.313	0.560	0.669	0.396
$z_2, z_{1\|2}$	0.971	0.985	0.540	0.662	0.364	0.632	0.630	0.387

Note: LB(l) represents LB test for z series over l lags, where $l = 10, 15$ and 20. LB^2(l) represents LB test for z^2 series over l lags.

Table 4.24: P-values for out-of-sample tests for the of BVBS-ACD model based on density forecasts

z series	KS	AD	LB(10)	LB(15)	LB(20)	LB^2(10)	LB^2(15)	LB^2(20)
z_1	0.064	0.028	0.633	0.746	0.632	0.750	0.835	0.706
z_2	0.195	0.062	0.767	0.763	0.547	0.944	0.865	0.706
$z_{1\|2}$	0.035	0.037	0.445	0.699	0.675	0.568	0.833	0.774
$z_{2\|1}$	0.082	0.035	0.700	0.825	0.661	0.857	0.936	0.857
$z_1, z_{2\|1}$	0.013	0.002	0.594	0.729	0.407	0.725	0.796	0.592
$z_2, z_{1\|2}$	0.016	0.007	0.573	0.721	0.466	0.573	0.721	0.466

Note: LB(l) represents LB test for z series over l lags, where $l = 10, 15$ and 20. LB^2(l) represents LB test for z^2 series over l lags.

Chapter 5

Concluding Remarks

In this chapter, we first review the research contributions of this dissertation, and then point out some directions for future research.

5.1 Summary of Research

Replacing the conditional mean in the original ACD model with the conditional median, Bhatti (2010) developed the BS-ACD model which allows an unimodal hazard rate. A recent extension of this conditional median-type ACD model has been given by Leiva *et al.* (2014), which considered the BS, BS power-exponential and Student-t BS distributions, so called the scale-mixture BS distribution family; see Balakrishnan *et al.* (2009) and Díaz-García and Leiva (2005) for details on SBS distributions. Saulo and Leão (2017) discussed the conditional median-type ACD model based on log-symmetric distributions. Saulo *et al.* (2017a) compared the mean-based and median-based BS-ACD models for model-fitting, forecasting and influence analysis. Moreover, their study confirmed that the conditional median-type ACD models based on BS and associated distributions are superior to the existing ACD model (the Generalized Gamma ACD model, see Lunde (1999)) in terms of model-fitting and forecasting.

In this thesis, we have first considered the univariate BSt-ACD model and then introduced bivariate ACD models based on BVBS and BVBSt distributions. The main contributions include three parts. First, we have developed the maximum likelihood estimation method for the BSt-ACD model. Next, in spite of the vast literature on duration series of single assets, the modeling of bivariate duration series has not yet received much attention due to non-synchronosity. With the univariate ACD model, we only can analyze one stock at a time and the other stock at another time. We can not do the analysis jointly. So, the natural question is how to model them together. For this purpose, we proposed the BVBS-ACD and BVBSt-ACD models based on matched trade durations. The maximum likelihood estimation of model parameters based on a two-step procedure, NM followed by BFGS, and associated inferential methods have been developed. The goodness-of-fit and predictive performance of the models have been discussed.

In Chapter 2, we have developed the maximum likelihood estimation method for the BSt autoregressive conditional duration model. We have proposed the following two-step approach for the parameter estimation. In the first step, we estimate the ACD parameters by using the NM Algorithm with κ and ν fixed at their initial values. In the second step, we estimate over the whole parameter space by the BFGS algorithm. We have derived and implemented analytical gradients in the second step. The standard errors of the MLEs of the model parameters have been calculated by inverting the observed information matrix evaluated at the MLEs. We have examined the properties of the MLEs of the BSt-ACD model parameters through a Monte

Carlo simulation study. The bias, MSE and RMSE are found to be consistently small and tend to 0 as sample size increases, showing that the MLEs are asymptotically unbiased and consistent. The standard errors of the means become smaller, and the sample skewness and kurtosis are found to close to 0 and 3, respectively, with increasing sample size. This suggests the asymptotic normal distribution-sample theory of MLEs. We have also evaluated the performance of the model by model discrimination using both likelihood-based and information-based method. The results of the likelihood-ratio test show that the chi-square distribution provides only a reasonable approximation to the null distribution of the likelihood-ratio test when testing for the BSt_{12}-ACD model. But, the chi-square distribution provides a good approximation to the null distribution of the likelihood ratio test when testing for the BSt_7-ACD and BSt_9-ACD models. However, when testing for the BS-ACD model, the mixture chi-square distribution didn't provide a good approximation to the null distribution of the likelihood ratio test. Furthermore, we have seen that when the true model is BSt_7-ACD and BS-ACD, the test has reasonable power to reject the other candidate models. However, when the true model is BSt_9-ACD and BSt_{12}-ACD, the test have low powers to reject BSt_7-ACD and BSt_9-ACD, respectively. When investigating the performance of AIC and BIC, we first note that both of them result in the same selection rates. Hence, in this case, inference can be based on either AIC or BIC. The results show that the information-based criteria perform well in discriminating between the BSt_7, BSt_9, BSt_{12} and BS-ACD models and show an improvement in performance with sample size. Finally, we have illustrated the proposed methodology using real high frequency data on two stocks from the New York Stock Exchange. In general, the BSt-ACD model outperform all the considered models and the BS-ACD model provides close values, but the GG-ACD model turns out to be worse in terms of AIC values, goodness-of-fit and out-of-sample density forecasts.

In Chapter3, we have proposed a bivariate autoregressive conditional duration model based on the bivariate BS distribution, allowing us to jointly analyze the dependence of two matched duration series. The maximum likelihood estimation of model parameters and associated inferential methods have been developed. We have done the estimation by using a hybrid of optimization algorithms, NM followed by BFGS. First, we have applied the NM Algorithm to estimate the ACD parameters by fixing κ_1, κ_2 and ρ at their initial values. We have then used the BFGS algorithm to estimate over the entire parameter space. We have derived and implemented analytical gradients in the second step. The standard errors have been estimated from the negative of the inverse Hessian matrix evaluated at the MLEs. We have conducted a simulation study to evaluate the properties of the model as well as the performance of the inferential methods developed here. For different correlation coefficients, all the estimators exhibit lower Bias and RMSEs which decrease with increased sample size, tending towards 0. This empirically supports the asymptotic unbiasedness, consistency and efficiency of the MLEs. Moreover, as n increases, the empirical distributions of all the estimators become close to normal distribution in terms of skewness and kurtosis. Finally, we have illustrated the proposed methodology using real high frequency data on two stocks from the New York Stock Exchange. The estimation results reveal a weak positive correlation between the two matched duration series. To justify that it is necessary to include the correlation parameter in the model, we conduct model comparison between the restricted model ($\rho = 0$) and the full model by using likelihood ratio test and AIC. The results suggest that the proposed

BVBS-ACD model is a much better choice than the restricted model. In-sample and out-of-sample density forecasts have been employed to assess the goodness-of-fit and predictive ability of the model, which supports the performance of the BVBS-ACD model developed here.

In Chapter 4, we have proposed a bivariate autoregressive conditional duration model based on the bivariate Student-t BS distribution, which facilitates the joint modeling of trade durations and detect the dependence of two matched duration series. A further advantage of the proposed model is that it allows more flexibility in terms of the kurtosis and skewness through the inclusion of an additional shape parameter. The maximum likelihood estimation of model parameters and associated inferential methods have been developed. We have proposed a two-step estimation procedure. First, we fixed the distribution parameters, κ_1, κ_2, ρ and ν, at their initial values and maximized the log-likelihood function with respect to the ACD parameters through the NM Algorithm. Next, we estimated over the full parameter space by employing the BFGS algorithm. We have derived and implemented the analytical gradients in the second step. The standard errors have been obtained by inverting the negative Hessian evaluated at the MLEs. We have conducted a simulation study to evaluate the properties of the model as well as the performance of the inferential methods developed here. A number of different scenarios have been taken into account concerning the values of correlation coefficient and sample size. All estimators perform well across different levels of correlation, demonstrating the asymptotic unbiasedness and consistency properties of the MLEs. The closeness normal approximation for the estimates is seen through skewness and kurtosis when the sample size increases. Furthermore, we have evaluated the performance of the model by model discrimination based on information-based method. AIC/BIC can distinguish between the BVBSt_7, BVBSt_9,BVBSt_{12} and BVBS-ACD models with a relatively high selection rates for the true models. Finally, we have illustrated the proposed methodology by a simulated bivariate dataset and compared the results of the BVBSt-ACD model with those of the BVBS-ACD model. The suggested model provides better fit to the data in terms of AIC/BIC and in-sample density forecasts, and yields better bivariate forecast according to out-of-sample density forecasts. These show the main advantages of the BVBSt-ACD model developed here.

5.2 Future work

We now describe some research topics, that would be of interest to investigate further. First, as continuation of this work, one may develop Bayesian inference of univariate and bivariate ACD models based on BS and associated distributions, such as BS-ACD, BSt-ACD, BVBS-ACD and BVBSt-ACD models, discussed in this thesis. In Chapter 4, we proposed the BVBSt-ACD model and a two-step procedure to obtain the MLEs of model parameters. One may consider developing an EM algorithm, for this maximum likelihood estimation of model parameters.

Rahul *et al.* (2018) have proposed an univariate autoregressive moving average (ARMA) model based on BS distribution. One may further consider constructing a bivariate ARMA model based on BVBS and BVBSt distributions. It will then be of great interest to investigate model misspecification between ACD and ARMA models. Instead of modeling conditional expected duration in the ACD model, for bivariate or multivariate duration data, we may consider constructing the autoregressive conditional intensity Model based on BVBS, BVBSt and even multivariate BS distributions.

Another direction is could be to construct copula-based ACD Models based on BVBS, BVBSt and multivariate BS distributions.

In Chapter 3 and 4, we have proposed bivariate ACD models with a constant covariance structure. One may introduce a time-varying covariance/correlation into the current models and develop corresponding inferential methods.

Instead of modeling matched duration data, one may construct bivariate ACD model for the two full duration series by treating one of the duration series as censored.

In this thesis, our research focuses on ACD (1,1) instead of ACD (p, q) allowing $p > 1$ and $q > 1$. The general model can be defined and the structure of the model can be explained. However, the numbers of model parameters will increase. The model estimation and fitting to a given dataset will be more complicated. For example, if we take $p = 2$ and $q = 2$, i.e., BVBS-ACD (2,2), we will have 4 more ACD parameters due to the second order, in addition to 9 parameters in the BVBS-ACD (1,1) model. That is 13 model parameters in total. We need to consider how to come up with initial values and how stable the estimation procedure is, etc. Besides, we tried to apply the BVBS-ACD (2,2) model to real data. The model fitting did not improve substantially. That is why we concentrated on the BVBS-ACD (1,1) model.

Finally, the multivariate extension of the work in this thesis would pose a challenging task and would be of great interest as it would be of great practical value. When we move from a BVBS-ACD to MVBS-ACD (see Kundu, Balakrishnan, and Jamalizadeh (2013)), the model definition is not a problem. Suppose we take p variate. We have to take $x_{1i}, x_{2i}, ..., x_{pi}$. But now the number of model parameters would have exponentially increased. What would be the trivariate case? We will have 3 more ACD parameters from σ_{3i}, α_3,β_3 and γ_3. We will have 3 more distribution parameters, 1 shape parameter κ_3 and 2 correlation parameters ρ_{13} and ρ_{23}. 6 more parameters have been introduced simply because we went one dimension up. That is 15 model parameters in total. One can imagine what will happen if we go to higher dimensions. We need to come up with good initial values, a stable estimation procedure and appropriate tools to validate the model.

Bibliography

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, **36**(1), 106–108.

Anderson, T. W. (1971). *The statistical analysis of time series*. John Wiley & Sons, New York.

Azevedo, C., Leiva, V., Athayde, E., and Balakrishnan, N. (2012). Shape and change point analyses of the Birnbaum–Saunders-t hazard rate and associated estimation. *Computational Statistics & Data Analysis*, **56**(12), 3887–3897.

Balakrishnan, N. and Kundu, D. (2018). Birnbaum-saunders distribution: A review of models, analysis and applications. *Applied Stochastic Models in Business and Industry*. (to appear).

Balakrishnan, N. and Lai, C. D. (2009). *Continuous Bivariate Distributions*. Springer-Verlag, New York.

Balakrishnan, N. and Pal, S. (2013). Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on COM-Poisson family. *Computational Statistics & Data Analysis*, **67**, 41–67.

Balakrishnan, N. and Pal, S. (2016). Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. *Statistical Methods in Medical Research*, **25**(4), 1535–1563.

Balakrishnan, N. and Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, **25**(16), 2797–2816.

Balakrishnan, N. and Zhu, X. (2015). Inference for the bivariate Birnbaum–Saunders lifetime regression model and associated inference. *Metrika*, **78**(7), 853–872.

Balakrishnan, N., Leiva, V., and López, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum–Saunders distribution. *Communications in Statistics: Simulation and Computation*, **36**(3), 643–656.

Balakrishnan, N., Leiva, V., Sanhueza, A., and Vilca, F. (2009). Estimation in the Birnbaum–Saunders distribution based on scale-mixture of normals and the EM-algorithm. *SORT*, **33**(2), 171–192.

Balakrishnan, N., Barui, S., and Milienos, F. (2017). Proportional hazards under Conway–Maxwell-Poisson cure rate model and associated inference. *Statistical Methods in Medical Research*, **26**(5), 2055–2077.

Bard, Y. (1974). *Nonlinear Parameter Estimation*. Academic Press, New York.

Barros, M., Paula, G. A., and Leiva, V. (2009). An R implementation for generalized Birnbaum–Saunders distributions. *Computational Statistics & Data Analysis*, **53**(4), 1511–1528.

Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique*, **60**, 117–149.

Bauwens, L., Giot, P., Grammig, J., and Veredas, D. (2004). A comparison of financial duration models via density forecasts. *International Journal of Forecasting*, **20**(4), 589–609.

Bhatti, C. R. (2010). The Birnbaum–Saunders autoregressive conditional duration model. *Mathematics and Computers in Simulation*, **80**(10), 2062–2078.

Birnbaum, Z. and Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability*, **6**(2), 319–327.

Bolker, B. M. (2008). *Ecological Models and Data in R.* Princeton University Press, Princeton, NJ.

Collett, D. (2003). *Modelling Survival Data in Medical Research.* CRC Press, Boca Raton, FL.

Díaz-García, J. A. and Leiva, V. (2005). A new family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, **128**(2), 445–457.

Díaz-García, J. A. and Leiva, V. (2007). Erratum to "a new family of life distributions based on the elliptically contoured distributions". *Journal of Statistical Planning and Inference*, **137**, 1512–1513.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, **39**(4), 863–883.

Diebold, F. X., Hahn, J., and Tay, A. S. (1999). Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange. *Review of Economics and Statistics*, **81**(4), 661–673.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, **68**(1), 1–22.

Engle, R. F. and Lunde, A. (2003). Trades and quotes: a bivariate point process. *Journal of Financial Econometrics*, **1**(2), 159–188.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, **66**(5), 1127–1162.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer-Verlag, New York.

Fan, J. and Yao, Q. (2017). *The Elements of Financial Econometrics*. Cambridge University Press.

Grammig, J. and Maurer, K.-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal*, **3**(1), 16–38.

Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data*. Springer-Verlag, New York, NY.

Jin, X. and Kawczak, J. (2003). Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, **4**, 103–124.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Volume 2, Second edition. John Wiley & Sons, New York.

Kundu, D., Kannan, N., and Balakrishnan, N. (2008). On the hazard function of Birnbaum–Saunders distribution and associated inference. *Computational Statistics & Data Analysis*, **52**(5), 2692–2702.

Kundu, D., Balakrishnan, N., and Jamalizadeh, A. (2010). Bivariate Birnbaum–Saunders distribution and associated inference. *Journal of Multivariate Analysis*, **101**(1), 113–125.

Kundu, D., Balakrishnan, N., and Jamalizadeh, A. (2013). Generalized multivariate Birnbaum–Saunders distributions and related inferential issues. *Journal of Multivariate Analysis*, **116**, 230–244.

Leiva, V. (2016). *The Birnbaum–Saunders Distribution*. Academic Press, New York.

Leiva, V., Riquelme, M., Balakrishnan, N., and Sanhueza, A. (2008). Lifetime analysis based on the generalized Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*, **52**(4), 2079–2097.

Leiva, V., Saulo, H., Leão, J., and Marchant, C. (2014). A family of autoregressive conditional duration models applied to financial data. *Computational Statistics & Data Analysis*, **79**, 175–191.

Lunde, A. (1999). A generalized gamma autoregressive conditional duration model. *Department of Economics, Politics and Public Administration, Aalbourg University, Denmark*.

Mayorov, K. (2011). *Modelling Trade Durations with the Birnbaum–Saunders Autoregressive Model*. Master's thesis, McMaster University, Hamilton, Ontario, Canada.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*. John Wiley, New York.

Mitchell, J. and Wallis, K. F. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, **26**(6), 1023–1040.

Mosconi, R. and Olivetti, F. (2005). Bivariate generalizations of the ACD models. In *Journal of Applied Econometrics Annual Conference, Venezia*.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, **7**(4), 308–313.

Pacurar, M. (2008). Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys*, **22**(4), 711–751.

Rahul, T., Balakrishnan, N., and Balakrishna, N. (2018). Time series with Birnbaum-Saunders marginal distributions. *Applied Stochastic Models in Business and Industry*. (to appear).

Rossi, B. and Sekhposyan, T. (2014). Evaluating predictive densities of us output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, **30**(3), 662–682.

Runde, R. (1997). The asymptotic null distribution of the box-pierce q-statistic for random variables with infinite variance an application to german stock returns. *Journal of Econometrics*, **78**(2), 205–216.

Sanhueza, A., Leiva, V., and Balakrishnan, N. (2008). The generalized Birnbaum–Saunders distribution and its theory, methodology, and application. *Communications in Statistics : Theory and Methods*, **37**(5), 645–670.

Saranjeet, K. B. and Ramanathan, T. (2018). Conditional duration models for high frequency data: A review on recent developments. *Journal of Economic Surveys*. (to appear).

Saulo, H. and Leão, J. (2017). On log-symmetric duration models applied to high frequency financial data. *Economics Bulletin*, **37**(2), 1089–1097.

Saulo, H., Leão, J., Leiva, V., and Aykroyd, R. G. (2017a). Birnbaum–Saunders autoregressive conditional duration models applied to high-frequency financial data. *Statistical Papers*, pages 1–25.

Saulo, H., Balakrishnan, N., Zhu, X., Gonzales, J. F., and Leão, J. (2017b). Estimation in generalized bivariate Birnbaum–Saunders models. *Metrika*, **80**(4), 427–453.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**(398), 605–610.

Simonsen, O. (2007). An empirical model for durations in stocks. *Annals of Finance*, **3**(2), 241–255.

Tay, A. S. and Wallis, K. F. (2000). Density forecasting: a survey. *Journal of forecasting*, **19**(4), 235–254.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. Third edition. John Wiley & Sons, Hoboken, NJ.

Vilca, F., Balakrishnan, N., and Zeller, C. B. (2014a). The bivariate sinh-elliptical distribution with applications to Birnbaum–Saunders distribution and associated regression and measurement error models. *Computational Statistics & Data Analysis*, **80**, 1–16.

Vilca, F., Balakrishnan, N., and Zeller, C. B. (2014b). A robust extension of the bivariate Birnbaum–Saunders distribution and associated inference. *Journal of Multivariate Analysis*, **124**, 418–435.

Wright, S. and Nocedal, J. (1999). Numerical optimization. *Springer Science*, **35**(67-68), 7.