Financial Time Series Models and Applications

by

Hu Mingming

A Thesis submitted to

the Faculty of Graduate Studies

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Statistics

University of Manitoba

Winnipeg, Manitoba

Abstract

Duration models are often concerned with time intervals between trades, longer durations indicating a lack of trading activities. In this thesis, we study parameter estimation for the Autoregressive Conditional Duration (ACD) and Stochastic Conditional Duration (SCD) models. Maximum likelihood methods can usually be used in the case of ACD models. However, the SCD models are based on the assumption that durations are generated by a dynamic stochastic latent variable which is often perturbed by Exponential, Weibull, Gamma or Log-Normal distributed innovations. This makes the use of maximum likelihood methods difficult.

One alternative method of parameter estimation, in this case, consists in using quasi-maximum likelihood after transforming the original nonlinear model into a state-space model and using the Kalman filter, a similar filtering scheme or the Generalized Method of Moments (GMM). We use the nonlinear filter and GMM method to analyze the Quadratic Stochastic Conditional duration model as well.

Acknowledgments

I would like to express my gratitude to my supervisors, Dr. Alexandre Leblanc and Dr. Thavaneswaran. Their expertise, understanding, and patience have made the deep impression to me. I appreciate their vast knowledge and skill in many areas and introduce to me an interesting area in statistics. I thank Dr. Alexandre Leblanc for his financial support and useful suggestions on all matters on statistical during my study. Under his supervision, I learn a lot. I thank Dr. Thavaneswaran introduce me to this useful area in statistics. I feel very fortunate to have had Dr. A. Leblanc and A. Thavaneswaran as my Supervisors.

I also will thank to the Committee member, Dr. Melody, for her helpful suggestions and statistical research experience in the summer of 2009.

Thanks also to the departmental IT support staff, Dave Gabrielson for helping me to deal with technical computing issues.

I would also like to thank my family for the support they have provided through my entire life. In as last, but not the least, I give thanks to many of my friends, one in particular is Yi Hua. Without her encouragement and spiritual support, I would not have finished this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Finance, perhaps the most popular area of Economics, is concerned with resource allocation, as well as resource management, acquisition and investment. In particular, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are among the most popular tools in financial market analysis because they are intuitive and easy to estimate. Furthermore, they capture the most important stylized facts of volatility. The following are key features of volatility:

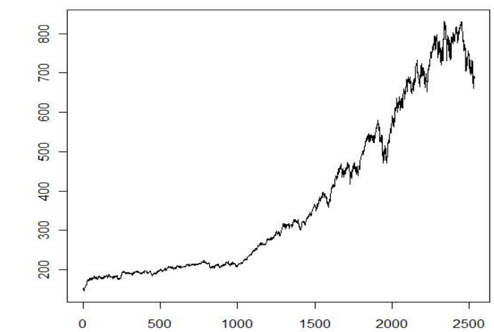
- volatility clustering: there exists clustering of changes in returns i.e. small changes tend to be followed by small changes and or large changes by large changes.
- volatility is leptokurtic: the distribution of returns has a higher probability

mass in the tail area ("fat tails") and a higher peak at the mean than that of a standard normal distribution;

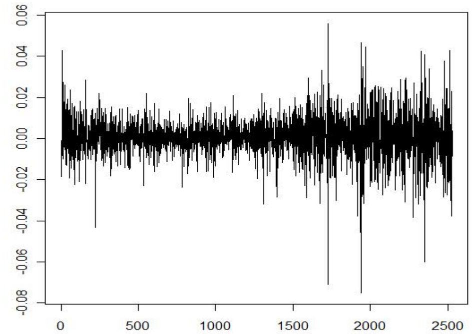
- heteroscedasticity: volatility is time-varying and non-constant;
- The squared values of the returns exhibit a high level of correlation whereas the values of the returns do not have much correlation.

As an example, consider the daily closing prices and returns for Standard & Poor's 100-share index ($S\&P100$), recorded from January 2, 1991 to December 29, 2000. The $S\&P100$ data exhibits the above 4 properties.

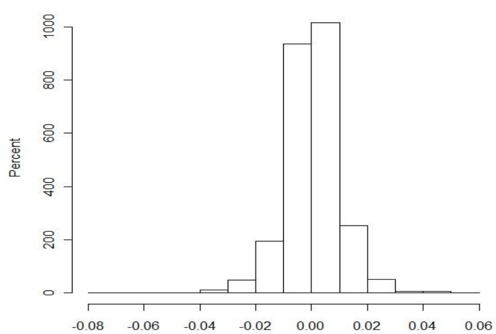
Indeed, Figure 1.1a shows the daily closing prices p_t and Figure 1.1b shows the returns y_t, where $y_t = \ln(\frac{p_t}{p_{t-1}})$. From the both figures, the variability of closing prices is seen to vary with low and high price fluctuations, indicating volatility clustering. Figure 1.1c shows the observed distribution of y_t. It exhibits a leptokurtic character with a high peak at the mean, a thin midrange and "fat tails". The Figure 1.1d (QQ plot) agrees as well. Figure 1.1e and Figure 1.1f respectively display the sample autocorrelation functions (SACF) of returns y_t and squared returns y_t^2. From these graphs, the majority of the sample autocorrelations for y_t^2 are significant, while those of y_t mostly are not. The next example will show that the simple model $y_t = \epsilon_{t-1}^2\epsilon_t$ (ϵ_t is a Gaussian white noise with variance σ_ϵ^2) can generate high peakedness. This model was considered by Gouriéroux (1997). The resulting process y_t is stationary having variance $\mathrm{Var}(y_t) = 3\sigma_\epsilon^6$ and conditional variance $\mathrm{Var}(y_t|\epsilon_{t-1}) = \sigma_\epsilon^2\epsilon_{t-1}^4$, which



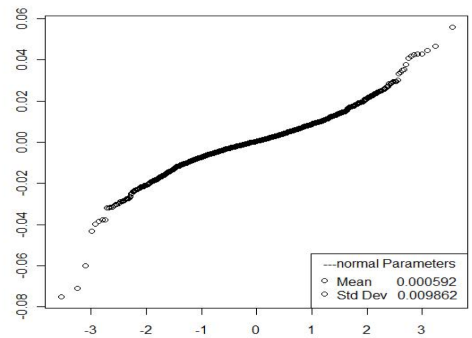
(a) Weekly closing prices p_t



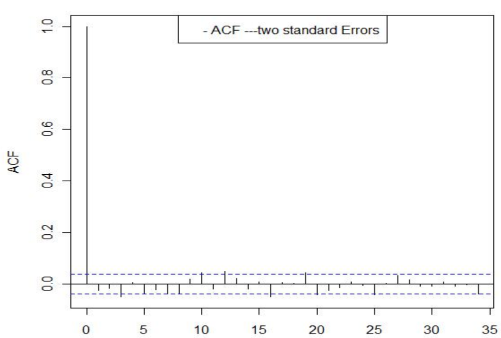
(b) Returns y_t



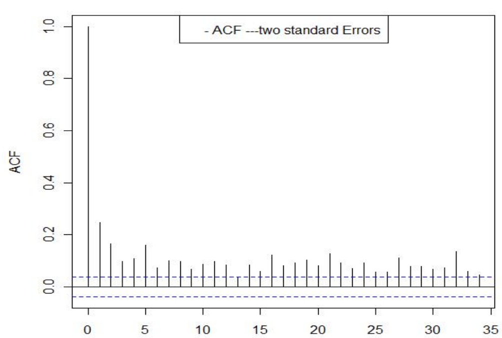
(c) Probability distribution curve of y_t



(d) QQ-plot of y_t



(e) SACF of return y_t



(f) SACF of return y_t^2

Figure 1.1: Sample Autocorrelation Function for the S&P 100 dataset

depends on lagged residuals. According to

$$E(\epsilon_t^{2n}) = \sigma_\epsilon^{2n}\frac{(2n)!}{2^n n!},$$

we can get that $E(y_t^4) = 315\sigma_\epsilon^{12}$ so that the kurtosis of the process y_t is $K^{(y)} = 35$.

1.2 Some Time Series Models

The following time series models are introduced. Suppose that a_t is a sequence of uncorrelated random variables having mean zero and constant variance σ_a^2.

- *Autoregressive (AR) model*

  The y_t process is called Autoregressive of order of p when it satisfies

$$y_t - \mu = \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + a_t,$$

under the stationary condition the solutions about the c for $\phi(c) = 0$ should lie outside the unicircle.

- *Moving Average (MA) model*

  A moving average process of order q is defined by

$$y_t = \mu + a_t - \sum_{i=1}^{q} \theta_i a_{t-i},$$

where the parameter μ is the mean of the process and θ_i are the coefficients for the process. These processes are well know in the time series literature and have been used successfully in recent decades.

- *Autoregressive Moving process (ARMA) model*

  The y_t process is called ARMA (p,q) when it satisfied

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(y_t - \mu) = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t,$$

  or, in other words,

$$y_t - \mu = \sum_{i=1}^{p} B^i \phi_i (y_{t-i} - \mu) + a_t - \sum_{j=1}^{q} B^j \theta_j a_{t-j}$$

where B denotes the backward shift operator, and μ is the mean of the process. The ϕ_i and θ_j are the coefficients for the process.

- *Autoregressive Conditional Heteroskedasticity (ARCH) model*

The ARCH model, introduced by Engle (1982), was the first model to provide a systematic framework for volatility modeling. In order to understand the ARCH model well, the structure of the model ARCH(p) is first introduced. It is given by

$$y_t | \mathcal{F}_{t-1}^y = \sqrt{h_t} Z_t,$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2,$$

where $p > 0$, $\omega > 0$, $\alpha_i \geq 0$, and $\mathcal{F}_{t-1}$ is the past information set available at time t. Note that the structure of the model, for h_t, is similar to an MA model, where y_t^2 plays the role of the innovations. To make σ_t^2 finite we assume that, $0 \leq \alpha_i < 1$. In additional, Z_t is taken to be a sequence of independent and identically distributed (i.i.d.) random variables with zero mean and unit variance (we write $Z_t \sim (0, 1)$ for this) that is uncorrelated with h_t. Then, it can be seen that

$$E(y_t) = E[E(y_t|\mathcal{F}_{t-1}^y)] = E[\sqrt{h_t}E(Z_t)] = 0,$$

and

$$\begin{aligned} Var(y_t) &= E(y_t^2) = E[E[y_t^2|\mathcal{F}_{t-1}^y]] = E[h_t E(Z_t^2)] \\ &= E[\omega + \textstyle\sum_{i=1}^p \alpha_i y_{t-i}^2] E[Z_t^2] = \omega + \textstyle\sum_{i=1}^p \alpha_i E[y_{t-i}^2]. \end{aligned}$$

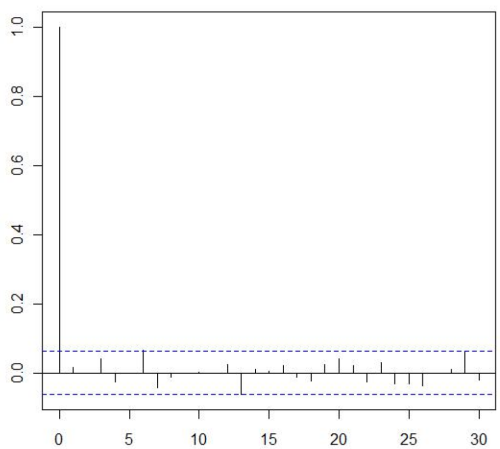
The stationarity of the process and the fact that $E[y_t] = 0$ implies that

$$Var(y_t) = Var(y_{t-i}) = E(y_{t-i}^2),$$

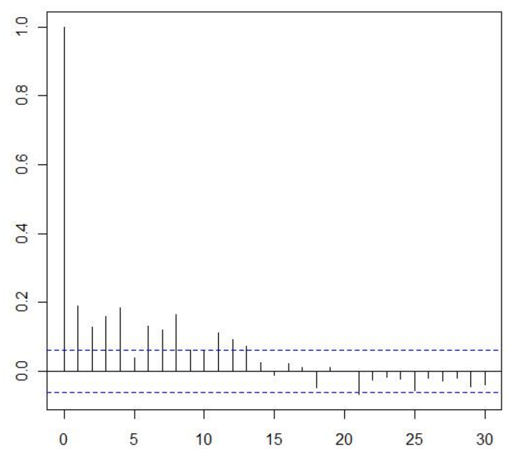
so that

$$Var(y_t) = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i}.$$

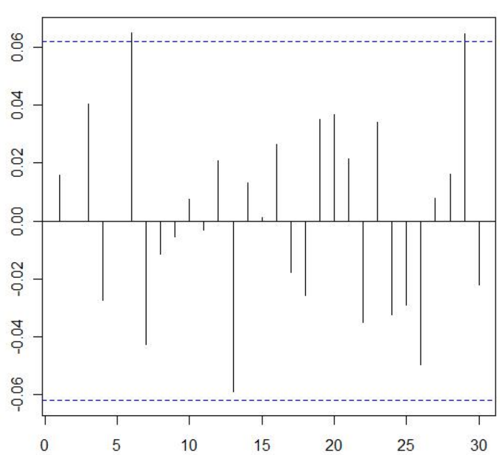
Details will be given in Section 2.2.1. The Autocorrelation function and Partial Autocorrelation function of the ARCH model will be shown in Figure 1.2.



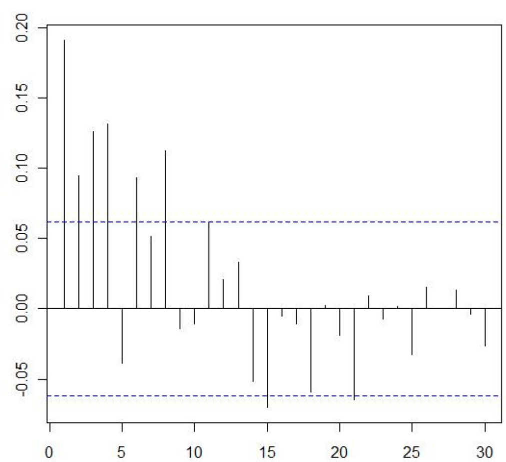
(a) Autocorrelation of Residuals y



(b) Autocorrelation of Residuals of y^2



(c) Partial Autocorrelation of Residuals y



(d) Partial Autocorrelation of Residuals y^2

Figure 1.2: Sample ACF and PACF of ARCH(1) model y_t and y_t^2, when ω=2.5, σ=0.2 and α_1=0

- *Normal GARCH Model*

Bollerslev (1986) introduced a useful extension known as the generalized ARCH (GARCH) model. This GARCH(P,Q) time series model is given by

$$y_t = \sqrt{h_t} Z_t,$$

$$h_t = \omega + \sum_{i=1}^{P} \alpha_i y_{t-i}^2 + \sum_{j=1}^{Q} \beta_j h_{t-j},$$

where Z_t is i.i.d. N(0,1), $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $P \geq 0$, and $Q \geq 0$. Note that the assumption of normality of Z_t implies that $y_t | \mathcal{F}_{t-1}^y \sim N(0, h_t)$. Comparing this to the ARMA model, we note the structure of the h_t process is essentially ARMA with, as before, y_t^2 playing the role of the innovations. Details will be given in Section 2.2.2.

- *GARCH Model with conditional t-distribution*

Originally, the GARCH model was introduced by Zakoïan (1994), with the normality assumption, that implied that $y_t | \mathcal{F}_{t-1}^y \sim N(0, h_t)$. Alternatively, the conditional density of y_t can be a student-t distribution with a degree of freedom, given by

$$f(y_t | \mathcal{F}_{t-1}^y) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2}) \sqrt{\pi (v-2) h_t}} \left[1 + \frac{y_t^2}{v-2} \right]^{-\dfrac{v+1}{2}},$$

where the gamma function is defined as

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx,$$

and where the degree of freedom, $v > 2$. It is a well-known property of the Student's t-distribution that, as the number of the degrees of freedom increases without bound, the t-distribution approaches a standard normal distribution. For finite v, however, the t-distribution has fatter tails than the corresponding normal distribution.

In Chapter 2, we mainly introduce volatility models. Volatility is a measure of price variability over some period of time. It typically describes the standard deviation of returns. The greater the volatility, the higher the risk. The volatility of price returns will be mainly described by GARCH and Stochastic volatility models. Then, some nonlinear time series models are introduced. In Chapter 3, Autoregressive Conditional Duration (ACD) models are introduced and the estimation of parameters are discussed in detail. In Chapter 4, linear SCD model are introduced and we discuss parameter estimation using Quasi-maximum likelihood method (QML) based on Kalman filtering and GMM methods. In Chapter 5, the new class of quadratic SCD models is presented along with parameter estimation based on QML and GMM methods.

Chapter 2

Nonlinear Models

2.1 Introduction

In the early 1990s, many nonlinear time series models have been introduced and studied in different areas. In this section, we begin by describing some of the important nonlinear models that can be found in the literature.

2.1.1 Random Coefficient AR Models

In engineering, econometrics, hydrology, meteorology and biology, the literature has demonstrated that many data sets can't be adequately modeled by classical linear time series models. For this reason, Nicholls & Quinn (1982) introduced the random coefficient autoregressive (RCAR) models, a family of models where coefficients have random characteristics. Since then, RCAR models attracted a lot of attention. The

RCAR models will be described in what follows. A p^{th} order RCAR, or RCAR(p), model is defined by

$$Z_t = \sum_{i=1}^{p} (\alpha_i + b_t^{(i)}) Z_{t-i} + a_t, \tag{2.1}$$

where the following conditions are met for $\{Z_t\}$ to be a strictly stationary process

(i) $\{a_t\}$ is a sequence of i.i.d. r.v.'s with mean 0 and variance $\sigma_a^2 < \infty$;

(ii) $\alpha_1, \alpha_2, \ldots \alpha_p$ are constants;

(iii) $\{\mathbf{b}_t = (b_t^{(1)}, b_t^{(2)}, \ldots, b_t^{(p)})'\}$ is a sequence of i.i.d. random vectors with zero mean vector and dispersion matrix $E(\mathbf{b}_t \mathbf{b}_t') = \Sigma$;

(iv) $\{\mathbf{b}_t\}$ and $\{a_t\}$ are mutually independent.

Taking $p = 1$, the RCAR(1) model is reduces to

$$Z_t = (\alpha + b_t) Z_{t-1} + a_t. \tag{2.2}$$

When $b_t = 0$ with probability 1, (2.2) reduces to an ordinary AR(1) model. As an example where such a model is valid, let Z_{t-1} be the amount of a given substance present in a system at the end of epoch $t-1, t = 1, 2, \ldots$ with $Z_0 = 0$. Suppose an amount a_t of this substance is introduced during the time interval $(t-1, t]$ and during the same interval a modification of the amount Z_{t-1} to $(\alpha + b_t)Z_{t-1}$ takes place. Then the total substance present at epoch t is described by model (2.2).

Assuming that $\{a_t\}$ and $\{b_t\}$ are mutually independent and normally distributed sequences, Nicholls & Quinn (1982) carried out the statistical analysis of model

(2.1). Specifically, they discuss maximum likelihood estimation of $\alpha_1, \alpha_2, \ldots, \alpha_p$. For the RCAR(1) model, they also obtain the conditional least squares (CLS) estimator of α, given by

$$\hat{\alpha} = (\sum_{t=2}^{N} Z_t Z_{t-1})/(\sum_{t=2}^{N} Z_{t-1}^2).$$

They show it is consistent and asymptotically normal. Tjøstheim (1986) also studies the problem of estimation in RCAR(p) models by using a maximum likelihood type penalty function. Finally, parameter estimation in RCAR(p) models using the theory of estimating functions is discussed by Thavaneswaran & Abraham (1988) and Chandra & Taniguchi (2001).

2.1.2 Bilinear Time Series

Bilinear time series, which were introduced by Granger & Anderson (1978), are another useful tool suitable for describing a nonlinear model. The general form of a bilinear time series model is given by

$$Z_t + \sum_{i=1}^{p} \alpha_i Z_{t-i} = a_t + \sum_{j=1}^{q} \beta_j a_{t-j} + \sum_{m=1}^{P} \sum_{m=1}^{Q} \gamma_{mn} a_{t-m} Z_{t-n}, \qquad (2.3)$$

where $\{a_t\}$ is a sequence of i.i.d. r.v.'s with mean zero and finite variance. Furthermore, for $s < t$, a_t and Z_s are independent. We can denote the model (2.3) by BL(p,q,P,Q). Though the model is nonlinear, it is linear in $\{a_t\}$ and $\{Z_t\}$ when they are considered separately. If we set $P = Q = 0$ then (2.3) reduces to an

ARMA(p,q) model and hence, can be considered as a generalization of an ARMA model. Statistical analysis of this model in its general form is very difficult. One of the simplest form of the model is given by

$$Z_t = \gamma Z_{t-k} a_{t-l} + a_t \tag{2.4}$$

where, as before, $\{a_t\}$ is a sequence of i.i.d. random variables with zero mean and finite variance. When $0 \leq \gamma < 1$, $\{Z_t\}$ is a stationary process.

2.1.3 Threshold Models

When a time series shows sudden changes at certain time points, threshold models can often be used. The basic idea here is that a nonlinear model can sometimes be described as having different regimes over its state space. Then within each regime, the model could be linear. The first model of this type was introduced by Tong (1983).

In what follows, we briefly present different families of threshold models. Let $\{c_0, c_1 \ldots, c_m\}$ denote a ordered subset of real numbers such that $c_0 < c_1 \cdots < c_m$ where $c_0 = -\infty$ and $c_m = +\infty$. We denote the state-space as R and let $R_j = (c_{j-1}, c_j]$ so that $\{R_1, R_2, \ldots, R_m\}$ forms a partition of the real line.

The Threshold Autoregressive (TAR) model of order p is given by

$$Z_t = \alpha_0^{(j)} + \sum_{i=1}^{p_j} \alpha_i^{(j)} Z_{t-i} + \alpha_t \quad \text{if } Z_{t-d} \in R_j, \quad j = 1, 2, \ldots, m,$$

where $\{a_t\}$ is an i.i.d. innovation sequence with mean 0 and variance σ^2, and d is called the delay parameter, which is integer. If $Z_{t-d} \in R_j$ then the model is said to be in regime j at time t. Within each regime, Z_t follows an autoregressive model, but not necessarily of the same order. The parameters involved in the model are $\sigma^2, d, \alpha_i^{(j)}, c_j, i = 1, 2, \ldots, p_j, j = 1, 2, \ldots, m-1$. Note that the order $p_1, \ldots, p_m$ of the model in the different regimes are assumed known.

As an example, consider a self-exciting Threshold Autoregressive (SETAR) model with two regimes:

$$Z_t = \begin{cases} \alpha_0^{(1)} + \sum_{i=1}^{p_1} \alpha_i^{(1)} Z_{t-i} + a_t & \text{if } Z_{t-d} \leq c \\ \alpha_0^{(2)} + \sum_{i=1}^{p_2} \alpha_i^{(2)} Z_{t-i} + a_t & \text{if } Z_{t-d} > c \end{cases}$$

where $0 < d < c$ and c is the threshold parameter. This model was introduced by Tong (1983). Petruccelli (1986) studied the properties of the least squares estimator of the parameters in this model when $\alpha_0^{(1)} = \alpha_0^{(2)} = 0$ and $p_1 = p_2 = 1$.

A second example of a slightly different nature is the smooth Threshold Autoregressive (STAR) model (also with two regimes), introduced by Chan & Tong (1986) and given by

$$Z_t = \alpha_0^{(1)} + \sum_{i=1}^{p} \alpha_i^{(1)} Z_{t-i} + \left[\alpha_0^{(2)} + \sum_{i=1}^{p} \alpha_i^{(2)} Z_{t-i}\right] F(\frac{Z_{t-d} - \gamma}{s}) + a_t,$$

where s acts as a smoothing parameter and $F(.)$ is the distribution function of a standard normal variate.

A threshold moving average (TMA) model with m regimes can been defined as

$$Z_t = \theta_0^{(j)} + a_t^{(j)} + \sum_{i=1}^{q_j} \theta_i^{(j)} a_{t-i}^{(j)} \quad \text{if } Z_{t-d} \in R_j, \quad j = 1, 2, \cdots, m,$$

where a $\{a_t^{(j)}\}$ is a sequence of independent Gaussian random variables with zero mean and variance σ_j^2. Finally, a threshold ARMA model has also been introduced by Brockwell, Liu and Tweedie (1992) and is given by

$$Z_t = \sum_{j=1}^{m} \left[\alpha_0^{(j)} + \sum_{i=0}^{p} \alpha_i^{(j)} Z_{t-i} + \sum_{i=1}^{q} \theta_i^{(j)} a_{t-i} \right] I_{\{Z_{t-d} \in R_j\}} + a_t$$

where I_A denotes the indicator function of a set A and the other variables and parameters are defined as before.

2.2 Modelling Variance: Conditional Heteroscedasticity

Changes in variance, or volatility, over time can be modelled using an approach based on autoregressive conditional heteroscedasticity. In the real word, share prices and foreign exchange rates are known to exhibit such conditional heteroscedasticity, which is detectable through the following phenomena.

1. Large and small values frequently show in the data, suggesting a non-normal heavy-tailed distribution.

2. Sample autocorrelations of the observed process are small whereas sample autocorrelations of the absolute and squared process are significantly different from zero even for large lags.

Such phenomena, including the important concept of conditional heteroscedasticity, may be simply modeled by assuming that

$$y_t = Z_t \sqrt{h_t}, \tag{2.5}$$

where $\{Z_t\}$ is a sequence of i.i.d. symmetric random variables with zero mean and unit variance. Further, $\{h_t\}$ is a sequence of non-negative random variables such that Z_t and h_t are independent for every fixed t. In this model, the conditional distribution of $y_t|h_t$ has variance h_t and we refer to $\{h_t\}$ as the time varying stochastic volatility of $\{y_t\}$. Now, some features of model (2.5) will be studied according to how the process $\{h_t\}$ is specified.

2.2.1 Autoregressive Conditional Heteroscedasticity

In 1982, Engle introduced ARCH models which perhaps now form the most popular family of models used to describe changing volatility. ARCH models allow for varying volatility by letting h_t depend on the past values of the process y_t. Specifically, the order p ARCH model is defined by

$$\begin{aligned} y_t &= Z_t\sqrt{h_t}, \\ h_t &= \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2, \end{aligned} \tag{2.6}$$

where $p > 0$, $\omega > 0$, $\alpha_i \geq 0$, and Z_t is a sequence of independent and identically distributed random variables with zero mean and unit variance (i.e. $Z_t \sim (0, 1)$). In order to make σ_t^2 finite, it is usually assumed that $0 \leq \alpha_i < 1$. Milhøj (1985) proved that the ARCH(p) process is stationary if and only if

$$\alpha_1 + \alpha_2 + \cdots + \alpha_p < 1. \tag{2.7}$$

We now briefly discuss conditions that guarantee the existence of higher order moments. First, from the previous assumptions on $\{Z_t\}$, it follows that

$$E(y_t) = E(E(y_t|\mathcal{F}_{t-1})) = E[\sqrt{h_t}E(Z_t)] = 0,$$

and

$$Cov(y_t, y_s) = Cov(Z_t\sqrt{h_t}, Z_s\sqrt{h_s}) = 0,$$

since for $s \neq t$, Z_t is independent of Z_s and $\{Z_t\}$ is independent of $\{h_t\}$. Under condition (2.7) we have seen that

$$E(y_t^2) \;\; = \;\; \omega + \textstyle\sum_{i=1}^{p} \alpha_i E(y_{t-1}^2),$$

so that, as previously claimed, we see that

$$Var(y_t) = E(y_t^2) = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i} < \infty,$$

implying that, $\{y_t\}$ is a white noise process.

In order to understand the ARCH model well, the structure of the ARCH(1) model is examined more thoroughly. First, in this case model (2.6) simplifies to

$$\begin{aligned} y_t &= Z_t \sqrt{h_t}, \\ h_t &= \omega + \alpha_1 y_{t-1}^2. \end{aligned} \tag{2.8}$$

From the previous calculations, under the normality assumption, $\{y_t\}$ is such that $E[y_t^{2k+1}] = 0$ for any integer $k > 0$, and $E(y_t^2) = 1$. In order to simplify, let $E(Z_t^4) = \lambda$. In particular, if Z_t is a standard normal variate, $\lambda = 3$ according to the formula

$$E(y_t^4) = \sigma_y^4 \frac{(4)!}{2^2 2!} = 3.$$

Thus, in the ARCH(1) model, $\{y_t\}$ has the following properties. First, we have

$$E(y_t) = E(y_t^3) = 0, \quad E(y_t^2) = E(h_t) = \frac{\omega}{1 - \alpha_1}, \quad \text{if } \ 0 \leq \alpha_1 < 1.$$

The autocorrelation function of $\{y_t^2\}$ is $Corr(y_t^2, y_{t-k}^2) = \alpha_1^k$, which is always non-negative. The fourth moment of y_t is given by

$$E(y_t^4) = \frac{\lambda\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-\lambda\alpha_1^2)}, \qquad \text{if} \;\; \lambda\alpha_1^2 < 1.$$

Indeed, this is obtained using the independence of Z_t and h_t, since

$$\begin{aligned} E(y_t^4) &= E(Z_t^4 h_t^2) = E(Z_t^4)E(h_t^2) \\ &= \lambda E(\omega + \alpha_1 y_{t-i}^2)^2 \\ &= \lambda\omega^2 + \alpha_1^2 E(y_{t-1}^4) + 2\omega\alpha_1 E(y_{t-1}^2) \\ &= \lambda\omega^2 + \alpha_1^2 E(y_{t-1}^4) + 2\omega^2 \frac{\alpha_1}{1-\alpha_1}. \end{aligned}$$

Using the stationary of $\{y_t\}$, we can now solve the previous equation, which leads to

$$E(y_t^4) = \frac{\lambda\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-\lambda\alpha_1^2)},$$

as was claimed. The coefficient of kurtosis follows easily and is given by

$$K_y = \frac{E(y_t^4)}{\{E(y_t^2)\}^2} = \frac{\lambda(1-\alpha_1^2)}{(1-\lambda\alpha_1^2)}, \qquad \text{if} \;\; \lambda\alpha_1^2 < 1. \tag{2.9}$$

We may rewrite (2.8) as

$$
\begin{aligned}
y_t^2 &= h_t + y_t^2 - h_t \\
&= h_t + h_t(Z_t^2 - 1) \\
&= \omega + \alpha_1 y_{t-1}^2 + \mu_t, \qquad (2.10)
\end{aligned}
$$

where $\{\mu_t = h_t(Z_t^2 - 1)\}$ is a sequence of uncorrelated r.v.'s with zero mean, implying that (2.10) defines an AR(1) model in y_t^2 with $\{\mu_t\}$ as an innovation sequence.

The definition of the ARCH(1) model implies that the kurtosis of the conditional distribution of y_t given h_t is the same as that of Z_t denoted by λ. However if $0 \leq \alpha_1 < 1$ and $\lambda > 1$, from (2.9) it is clear that the kurtosis of the marginal distribution of $\{y_t\}$ exceeds λ. For example, if Z_t is a standard normal variate then $\lambda = 3$ and K exceeds 3. Thus the unconditional distribution of y_t is leptokurtic and model (2.8) could be suitable for modeling heavy-tailed financial series.

Suppose that Z_t in (2.6) has a standard normal distribution. Then, it follows that the conditional distribution of y_t given $\{y_s, s \leq t-1\}$ is normal with mean zero and variance h_t. The unknown parameter vector to be estimated is $\theta = (\omega, \alpha_1, \cdots, \alpha_p)$. The log-likelihood function of θ, conditional on the observed data, is given by

$$
\begin{aligned}
\text{Ł}(\theta) &= \sum_{t=1}^{N} \log f(y_t | y_{t-1}, y_{t-2}, \cdots, y_{t-p}; \theta) \\
&= -(\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\log h_t - \frac{1}{2}\sum_{t=1}^{N}(y_t^2/h_t)), \qquad (2.11)
\end{aligned}
$$

with $h_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-1}^2$.

The MLE of θ can be obtained by maximizing Ł(θ). This is possible only by numerical methods. Hamilton (1994) describes a number of methods for obtaining the MLE. In certain practical situations, the normality assumption on Z_t may not be applicable. Bollerslev (1987) discussed the likelihood analysis of ARCH models when Z_t has a student t-distribution and Nelson (1991) studied such problems by assuming a generalized error distribution for Z_t. Next, we consider a generalized version of the ARCH(p) model.

2.2.2 Generalized ARCH Models

In 1986, Bollerslev introduced a useful extension known as generalized ARCH, or GARCH models. The difference between GARCH models and ARCH models is that a linear combination of lagged values of h_t are also added in the equation for conditional variance.

Specifically, the GARCH (p, q) model is given by

$$\begin{gathered} y_t = Z_t \sqrt{h_t}, \\ h_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \end{gathered} \tag{2.12}$$

where $\omega > 0, \alpha_i \geq 0, \beta_j \geq 0$ and $\{Z_t\}$ is a process of i.i.d. r.v.'s with mean zero and unit variance. In this case, the sequence $\{y_t\}$ is covariance stationary if and only if

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1.$$

In the original formulation of GARCH models, as they were introduced by Bollerslev, it is assumed that Z_t is a standard normal variate. Obviously, other distributions for Z_t could also be considered.

The above restrictions imposed on the parameters ensure that the conditional variance h_t is non-negative. Other authors observed, in empirical studies, that the variance remains non-negative and finite even if some of the constraints on the parameters are relaxed.

The simplest version of a GARCH model is the GARCH(1,1) model, given by

$$\begin{aligned} y_t &= Z_t\sqrt{h_t}, \\ h_t &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}. \end{aligned} \tag{2.13}$$

This may be rewritten as a non-Gaussian ARMA(1,1) model in terms of y_t^2 as follows,

$$\begin{aligned} y_t^2 &= h_t + y_t^2 - h_t \\ &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} + \mu_t, \end{aligned}$$

where $\mu_t = (y_t^2 - 1)h_t$. This can be further simplified to

$$y_t^2 = \omega + (\alpha_1 + \beta_1)y_{t-1}^2 - \beta_1\mu_{t-1} + \mu_t, \tag{2.14}$$

which is in the form of an ARMA(1,1) model. Note that model (2.13) is stationary if $\alpha_1 + \beta_1 < 1$.

It is readily verified that, under the normality assumption, for y_t defined by (2.13),

$$\begin{aligned} E(y_t) &= E(y_t^3) = 0, \\ E(y_t^2) &= E(h_t) = \omega/(1 - \alpha_1 - \beta_1), \quad \text{if } 0 \leq \alpha_1 + \beta_1 < 1. \end{aligned}$$

The coefficient of kurtosis can also be shown to be

$$K = \frac{\lambda(1 + \alpha_1 + \beta_1)(1 - \alpha_1 - \beta_1)}{1 - \lambda\alpha_1^2 - \beta_1^2 - 2\alpha_1\beta_1} \quad \text{if } \lambda\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1 < 1,$$

and the autocorrelation function of $\{y_t^2\}$ is

$$Corr(y_t^2, y_{t-k}^2) = \frac{\alpha_1(\alpha_1 + \beta_1)^{k-1} - \alpha_1(\alpha_1 + \beta_1)^{k+1} + \alpha_1^2(\alpha_1 + \beta_1)^k}{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2}.$$

Note that the GARCH model introduces flexibility in the structure of kurtosis and autocorrelations, when compared with the ARCH specification. Assuming that Z_t is a standard normal variate, the likelihood function of $\theta = (\omega, \alpha_1, \cdots, \alpha_p, \beta_1, \beta_2, \cdots, \beta_q)$ can be written as (2.11) with h_t given by (2.12).

Bayesian analysis of ARCH and GARCH models is discussed in Bauwens, Lubrano & Richard (1999). The unconditional marginal distributions of GARCH processes have tails that are fatter than those of the normal distribution. Pawlak & Schmidt (2001) studied the tail behavior of y_t^2 in GARCH models and obtained some bounds for the distribution function.

Abraham & Balakrishnan (2001) mentioned that the ARMA(1,1) representation of the GARCH(1,1) model given by (2.14) reveals that if $\alpha_1 + \beta_1 = 1$ then y_t^2 will have a unit root. In this case, the model (2.13) is referred to as an Integrated GARCH or IGARCH model.

2.3 Stochastic Volatility Models

As noted before, the volatility of financial series tends to change over time. The ARCH/GARCH models described earlier serve as tools for modeling and estimating the time-varying conditional variance. These models assume that volatility is driven by past observations. Alternatively, Taylor (1986) argued that the volatility process should be driven by some unobservable or latent economic force rather than by the movement of prices. The class of models that is formulated under this kind of belief is generally referred to as stochastic volatility (SV) models.

2.3.1 Linear stochastic Volatility Models

Stochastic volatility (SV) models were introduced by modeling the logarithm of volatility by a stochastic process. Stochastic Volatility models differ from ARCH models which can specify a process for the conditional variance of return. Important contributions to the Stochastic Volatility literature were made by Clark (1973), Taylor (1982), and Tauchen & Pitts (1983). But the ARCH/GARCH literature evolved much more rapidly than the one devoted to Stochastic Volatility models. One main advantage of ARCH/GARCH models is that they can easily be estimated using maximum likelihood methods, while this methodology is difficult to use for Stochastic Volatility models. A state space model will be used to present the SV model. One of the possible specifications of this type for model is given by

$$y_t = Z_t \exp(h_t/2), \tag{2.15}$$

where

$$h_{t+1} = \omega + \alpha h_t + \eta_{t+1}, \tag{2.16}$$

and $\{Z_t\}$ is a sequence of i.i.d. symmetric r.v's with mean zero and unit variance. Further, $\{Z_t\}$ and $\{\eta_t\}$ are assumed independent for each t. In this context, h_t can be interpreted as information which is very difficult to model directly into financial markets (see Tauchen & Pitts, 1983). In the initial developments of this model, it was assumed that $\{Z_t\}$ and $\{\eta_t\}$ are independent i.i.d. Gaussian sequences with mean 0 and variances 1 and σ_η^2 respectively. This model is referred to as a log-normal SV model.

Observe that the sequence $\{Z_t\}$ is always stationary and hence $\{y_t\}$ is stationary whenever $\{h_t\}$ is. If $|\alpha| < 1$ then $\{h_t\}$ defines a stationary Gaussian AR(1) process with

$$E(h_t) = \frac{\omega}{1-\alpha} = \mu_h \quad \text{and} \quad Var(h_t) = \frac{\sigma_\eta^2}{1-\alpha^2} = \sigma_h^2.$$

The distribution of y_t is symmetric about zero and hence $E(y_t^k) = 0$ for odd k. When k is even

$$\begin{aligned} E(y_t^k) &= E(Z_t^k)E(e^{\frac{kh_t}{2}}) \\ &= \frac{k!}{2^{k/2}(k/2)!}\exp\{\frac{k}{2}\mu_h + \frac{k^2\sigma_h^2}{8}\}. \end{aligned}$$

The kurtosis of y_t is then given by (see also (2.9))

$$K = 3\exp(\sigma_h^2) \geq 3,$$

which shows that, under this SV model, the marginal distribution of y_t has fatter tails than the corresponding normal distribution. The autocorrelation function of $\{y_t\}$ vanishes while that of $\{y_t^2\}$ is given by

$$corr(y_t^2, y_{t-k}^2) = \frac{\exp(\sigma_h^2\alpha^k) - 1}{3\exp(\sigma_h^2) - 1},$$

which can be negative if $\alpha < 0$, contrary to the case of ARCH models.

The dynamic properties of this SV model can also be revealed by using logarithms. Indeed, model (2.15) implies that

$$\log y_t^2 = h_t + \log Z_t^2$$

with h_t defined by (2.16). If Z_t is a standard normal variate, then $\log Z_t^2$ has mean -1.27 and variance 4.93. The autocorrelation function of $\{\log y_t^2\}$ is given by (*c.f.* Shephard, 1996).

$$\rho^k_{\log y_t^2} = \frac{1}{(1 + 4.93/\sigma_h^2)}\alpha_1^k.$$

The SV model (2.15)- (2.16) may also be defined by

$$y_t = Z_t\sqrt{h_t} \tag{2.17}$$

where

$$\log h_{t+1} = \omega + \alpha_1 \log h_t + \eta_{t+1}. \tag{2.18}$$

This form of the model is discussed in Jacquier *et al.* (1994). If Z_t and η_t are normal r.v.'s as before, then all the properties discussed above can also be shown to follow similarly from (2.17) and (2.18).

2.3.2 Inference for SV Models

The following lemma and models were originally discussed by H. Gong & A. Thavaneswaran (2008) . Consider

$$\begin{aligned}\theta_{t+1} &= a\theta_t + (b + \theta_t)\epsilon_{t+1} + c,\\ y_{t+1} &= A\theta_t + Be_{t+1} + C,\end{aligned}$$

where $\{\epsilon_{t+1}\}$ and $\{e_{t+1}\}$ are correlated Gaussian sequences having mean zero and variances σ_ϵ^2 and σ_e^2 respectively, and $cov(e_t, \epsilon_t) = \rho\sigma_e\sigma_\epsilon$. Here, $\{\theta_t\}$ is an unobserved process and $\{y_t\}$ is the observed process. Now, define $b_1 = \sqrt{E[(b+\theta_t)^2]}$ and $b_2 = E[b + \theta_t]$. After some transformation, the nonlinear time series model can be written in linear form as

$$
\begin{aligned}
\theta_{t+1} &= a\theta_t + b_1\tilde{\epsilon}_{t+1} + c, \\
y_{t+1} &= A\theta_t + Be_{t+1} + C,
\end{aligned}
$$

where $E[\tilde{\epsilon}_t] = 0$, $E[\tilde{\epsilon}_t\tilde{\epsilon}_{t+1}] = 0$, and $\{\tilde{\epsilon}_t\}$ is sequence of uncorrelated random variables. Before introducing different models, we first present a Lemma that is useful to study volatility.

Lemma 2.3.1. *The conditional mean of the filtered estimate of θ_t given F_t^y, $\hat{\theta}_t = E[\theta_t|F_t^y]$ is*

$$\hat{\theta}_{t+1} = a\hat{\theta}_t + c + (\frac{aA\gamma_t + Bb_1\rho\sigma_e\sigma_\varepsilon}{A^2\gamma_t + B^2\sigma_\varepsilon^2})[y_{t+1} - A\hat{\theta}_t - C]$$

and its conditional variance is $\gamma_t = E[(\theta_t - \hat{\theta}_t)^2|F_t^y]$, is given by

$$
\begin{aligned}
\gamma_{t+1} &= E[(\theta_{t+1} - \hat{\theta}_{t+1})^2|\mathcal{F}_{t+1}^y] \\
&= a^2\gamma_t + b^2\sigma_\varepsilon^2 - \frac{(aA\gamma_t + Bb_1\rho\sigma_e\sigma_\varepsilon)^2}{A^2\gamma_t + B^2\sigma_e^2}
\end{aligned}
$$

According to the Lemma 2.3.1, the recursive estimation will be given and the model parameters will be estimated. A proof is given in Appendix A.

2.3.3 Taylor's SV model

In 1982, Taylor introduced a Stochastic Volatility model with a lognormal (LN) specification for volatility. Wiggins (1987) and Scott (1987) analyzed prices of stock options based on Taylor's SV model. Chesney & Scott (1989) analyzed currency options. Specifically, Taylor's model specifies that logarithmic volatility follows an AR(1) process according to

$$\begin{aligned} y_{t+1} &= \sigma_t z_{t+1}, \\ \log \sigma_{t+1} &= \phi \log \sigma_t + \eta_{t+1}, \end{aligned} \tag{2.19}$$

where $|\phi| < 1$ for stationarity, z_t are $i.i.d$ $N(0,1)$, η_t are $i.i.d$ $N(0,\sigma_\eta^2)$, and $\{z_t\}$ and $\{\eta_t\}$ are independent processes. Note that this implies that $\log_t \sigma \sim N(0, \sigma_\eta^2/(1-\phi^2))$ for all t. Using the transformations

$$\log \sigma_t = \theta_t \quad \text{and} \quad \tilde{z}_{t+1} = \frac{e^{\theta_t}}{\sqrt{E[(e_t^\theta)^2]}} z_{t+1},$$

where $E\tilde{z} = 0$, $E[\tilde{z}_t \tilde{z}_{t+1}] = 0$, and $E\tilde{z}_t^2 = 1$. The model (2.19) becomes

$$
\begin{aligned}
y_{t+1} &= \sigma_t z_{t+1} = e^{\log(\sigma_t)} z_{t+1} = e^{\theta_t} z_{t+1} = \beta \tilde{z}_{t+1}, \\
\theta_{t+1} &= \phi \theta_t + \eta_{t+1},
\end{aligned}
$$

where $\beta = \sqrt{E[e^{\theta_t}]^2} = \sqrt{\exp(2\sigma_\eta^2/(1-\phi^2))}$. Now, from Lemma 2.3.1 the recursive estimates for the conditional mean of θ_t given $\mathcal{F}_t^y$, $\hat{\theta}_t = E[\theta_t | \mathcal{F}_t^y]$, and their conditional mean squared error, $\gamma_t = E[(\theta_t - \hat{\theta}_t)^2 | \mathcal{F}_t^y]$, are given by

$$
\begin{aligned}
\hat{\theta}_t &= \phi \hat{\theta}_{t-1}, \\
\gamma_t &= \phi^2 \gamma_{t-1} + \sigma_\eta^2,
\end{aligned}
$$

from using $A = 0$, $B = \beta$ and $C = 0$, $a = \phi$, $b_1 = 1$, $c = 0$, and $\rho = 0$ due to independence.

2.3.4 Quadratic SV model

Kawakatsu (2007) proposed a new stochastic volatility model which involves a quadratic term. Specifically, he used

$$
\begin{aligned}
y_{t+1} &= \theta_t + \exp(a + b\theta_t + c\theta_t^2) z_{t+1}, \\
\theta_{t+1} &= \phi \theta_t + \eta_{t+1},
\end{aligned}
$$

where $\{z_t\}$ and $\{\eta_t\}$ are two uncorrelated sequences of $i.i.d$ random variables with zero mean and unit variance. Since this is not a linear system, the following transformation is necessary:

$$\tilde{z}_{t+1} = \frac{\exp(a + b\theta_t + c\theta_t^2)}{\sqrt{E[(\exp(a + b\theta_t + c\theta_t^2))^2]}} z_{t+1},$$

where $b_1 = \sqrt{E[(\exp(a + b\theta_t + c\theta_t^2))^2]}$. By applying

$$E[\exp(a(z+b)^2)] = \exp(\frac{1}{2}\ln(1-2a) + \frac{ab^2}{1-2a}),$$

then

$$\begin{aligned} b_1 &= \sqrt{c(\theta + \tfrac{b}{2c})^2 + a - \tfrac{b^2}{4c}} \\ &= \sqrt{\exp(-\tfrac{1}{2}\ln(1-4c) + \tfrac{b^2}{4c(1-2c)} + a - \tfrac{b^2}{4c})}, \end{aligned}$$

and filtered estimate of θ_t and its conditional Mean squared error follow from Lemma 2.3.1.

Chapter 3

Conditional Duration Models

3.1 Introduction

In financial markets, high-frequency data are often observed daily or at a finer time scale. These data have recently attracted more attention because, in empirical studies of market microstructures, they provide significant information. Transaction-by-transaction or trade-by-trade data in security markets represent the ultimate high-frequency data, with time often being measured in seconds. Tsay (2005) mentioned several reasons for the usefulness of high-frequency financial data. First, they can be used to compare the efficiency of different trading systems in price discovery, which is defined as the process of establishing a market price at which demand and supply for an item are matched. Examples of trading systems that one could be interested in comparing are the open out-cry system of the NYSE and the computer trading system of NASDAQ. The second reason for the usefulness of high-frequency data is that they can be used to study the price discovery, liquidity, and volatility of stock markets.

As mentioned above, empirical studies have been very popular in finance, especially in the last 20 years. In particular, for studying stock markets, closing price is an interesting variable. However, if focusing solely on closing price, intraday events will be neglected. Because of the developments in computer technology and increased automation of financial markets, intraday data are now often available in the form of price, volumes and other factors. The analysis of high-frequency data (HFD) can deeply facilitate our understanding of market activity. There are several types of HFD which researchers are interested in, for example, transaction data or tick-by-tick data, 5-minute returns in financial exchanges and 1-minute returns on the cash market. These types of data have several important features that make their analysis quite challenging for financial economists and statisticians. Some of these characteristics are:

1. irregular time intervals,
2. discrete values, e.g. prices in multiples of tick size,
3. very large sample size,
4. multi-dimensional variables, e.g. price, volume, quotes, etc.

To highlight these characteristics, we briefly study the IBM transaction data recorded from November 1, 1990 to January 31, 1991. There are a total of 60328 trades done over a period of 63 days, 60265 of which are intraday trades. Table 3.1 presents a summary of these data. The following points are worth noting.

Table 3.1: Frequencies of price change in Multiples of Tick Size for IBM Data

Number(tick)	$\leq$ -3	-2	-1	0	1	2	$\geq$ 3
percentage	0.66	1.33	14.53	67.06	14.53	1.27	0.63

- About 2/3 of the intraday transactions are associated with no price change.
- Approximately 29% of the intraday transactions are associated with a change in price of one-tick.
- Less than 4% of the transactions are associated with a change of two ticks or more.
- The distribution of changes is almost perfectly symmetric about zero.

3.2 Autoregressive Conditional Duration Models

As was mentioned above, analyzing high frequency data is a very important part of many financial modelling problems. While some high frequency data are recorded at fixed time intervals, most of the transactions occur at irregular time points. In such a case, we lose part of the useful information if the data are recorded on a fixed interval. In order to avoid that loss, Engle & Russell (1998) introduced the autoregressive conditional duration (ACD) model. The modelling of time between events is the objective of ACD models. In particular, ACD models and their extensions can be used to model the behavior of irregularly spaced financial data. Engle (2000,2002) illustrated that ACD models share many features with GARCH models and that having irregular time-intervals is the main characteristic of HFD. In order to model durations, we let the conditional expected duration between the $(t-1)^{\text{th}}$ and t^{th} trades be

$$\psi_t = E[y_t|\mathcal{F}_{t-1}^y],$$

where $\mathcal{F}_{t-1}^y$ is past information, in this case past durations, up to and including the $(t-1)^{th}$ trade. The basic ACD model is defined as

$$\begin{aligned} y_t &= \psi_t \epsilon_t, \\ \psi_t &= \omega + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{j=1}^{q} \beta_j \psi_{t-j}, \end{aligned} \tag{3.1}$$

where $\omega > 0$, $\alpha_j > 0$, $\beta_j > 0$ and $\{\epsilon_t\}$ is a sequence of independent identically distributed random variables such that $E(\epsilon_t) = 1$. In order for this condition to be satisfied, we standardized distributions for the innovation process. Typically, ϵ_t follows a standard exponential or a standardized Weibull distribution. Note that this model structure implies $Var(y_t|\mathcal{F}_{t-1}^y) = \psi_t^2 Var(\epsilon_t)$. This is referred to as the ACD (p,q) model. In order for y_t to be covariance-stationary, it is sufficient for the parameters of model (3.1) to satisfy

$$\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1.$$

Moreover, when the roots of [1-$\sum_{j=1}^{p} \alpha_j B^j$-$\sum_{j=1}^{q} \beta_j B^j$] and [1-$\sum_{j=1}^{q} \beta_j B^j$] lie outside the unit circle, both the stationary and invertibility conditions are satisfied. In this case the conditional mean of y_t is, by definition, equal to ψ_t, but the unconditional mean of y_t can be shown to be

$$E(y_t) = \frac{\omega}{(1 - \sum_{j=1}^{p} \alpha_j - \sum_{j=1}^{q} \beta_j)}.$$

In order to see the impact of the distribution of the errors ϵ_t in the ACD model, we simulated series of 5000 observations from the ACD(1,1) model as given in (3.1). For this, we used four different innovational distributions for ϵ_t.

For the first example, the error term was simulated using the standard exponential distribution with the parameter $\theta = 1$. For the second, ϵ_t follows the standard Weibull distribution with parameter $\gamma = 1.5$, the scale parameter being set to $\theta = \dfrac{1}{\Gamma(1 + \frac{1}{\gamma})}$, so that $E[\epsilon_t] = 1$ (see the Appendix C for the details). In the third case, the gamma distribution with parameter $\kappa = 1.5$ and scale parameter $\theta = \dfrac{1}{\kappa}$ was used, again to ensure that $E[\epsilon_t] = 1$. Finally, the error term follows a standard log-normal distribution with parameters $\mu = -\frac{\sigma^2}{2}$ and $\sigma = 1.5$. For the remainder of this section, the four cases mentioned above are briefly examined. Parameter estimation will be discussed in the next section. In all cases, the parameter values used for simulations were $\omega = 0.3$, $\alpha = 0.2$, and $\beta = 0.7$.

As a first example, consider the EACD(1,1) model given by (3.1), where $p = q = 1$ and ϵ_t follows the standard exponential distribution. From the properties of exponential random variables, we have $E[\epsilon_t] = 1$ and $Var[\epsilon_t] = 1$, implying that the conditional moments of y_t are $E(y_t|\mathcal{F}_{t-1}^y) = \psi_t$ and $Var(y_t|\mathcal{F}_{t-1}^y) = \psi_t^2$. For the first marginal (unconditional) moment of y_t, it follows that

$$E(y_t) = E[E[\psi_t \epsilon_t | \mathcal{F}^y_{t-1}]] = E(\psi_t),$$

and

$$E(\psi_t) = \omega + \alpha E(y_{t-1}) + \beta E(\psi_{t-1}).$$

Assuming the stationary condition holds, we know that

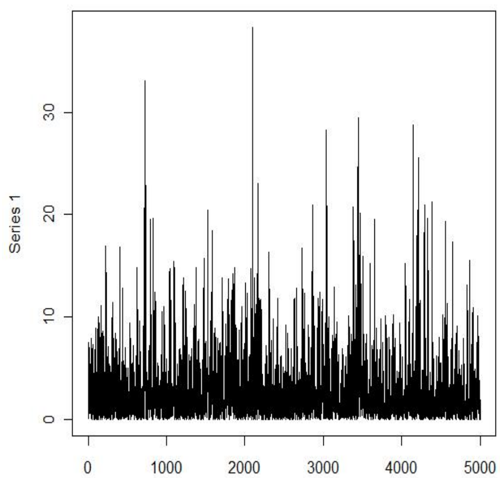
$$\begin{aligned} E(\psi_t) &= E(\psi_{t-1}) \\ E(y_{t-1}) &= E(y_t) \end{aligned}$$

so that the mean of y_t is given by

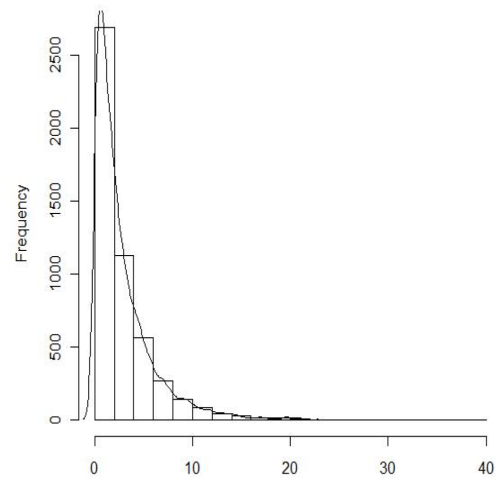
$$E[y_t] = E[\psi_t] = \frac{\omega}{1 - \alpha - \beta}.$$

Time plots, histograms and graphs of sample autocorrelation function are provided in Figures 3.1 to 3.3 for simulation of EACD(1,1), EACD(1,2) with $\alpha_1 = 0.2$, $\beta_1 = 0.3$, $\beta_2 = 0.4$ and EACD(2,1) with $\alpha_1 = 0.2$, $\alpha_2 = 0.3$ and $\beta_1 = 0.4$.

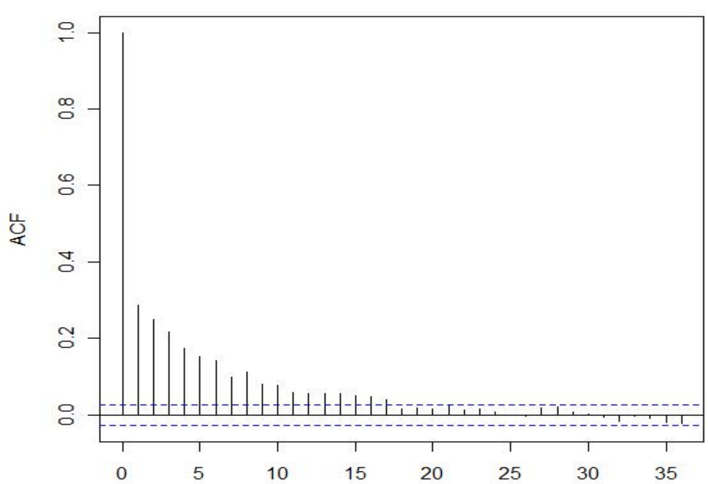
As another example, consider now the GACD(1,1) model given by (3.1), where the error term follows a gamma distribution $\mathrm{G}(\kappa, \theta)$ with $\theta = \frac{1}{\kappa}$. According to the properties of the gamma function, we know that $E(\epsilon_t) = \theta\kappa = 1$ and $\mathrm{Var}(\epsilon_t) =$



(a) Time plot for EACD (1,1)

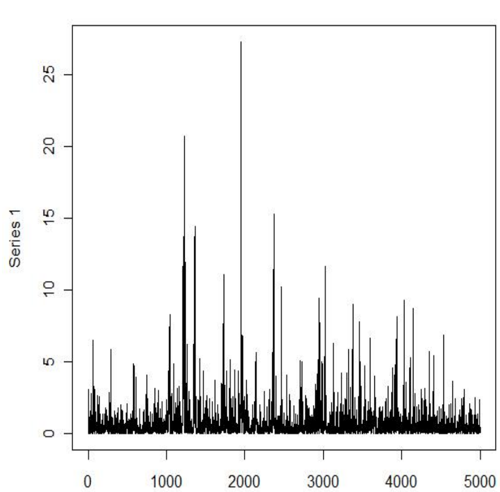


(b) Histogram for EACD (1,1)

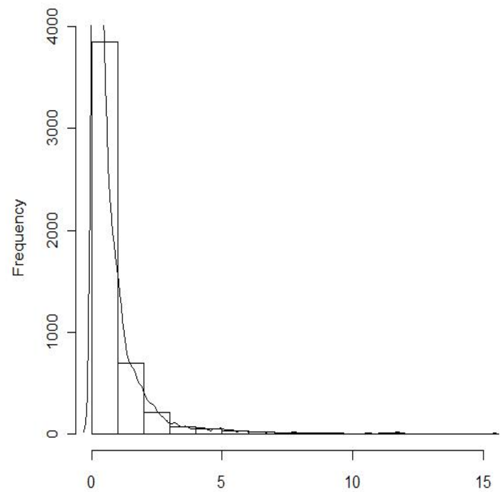


(c) Autocorrelation function

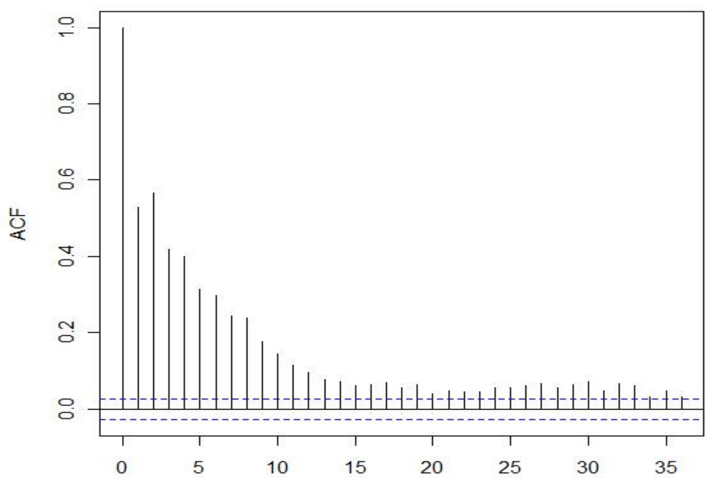
Figure 3.1: Simulated EACD (1,1) series with 5000 observations.



(a) Time plot for EACD (2,1)

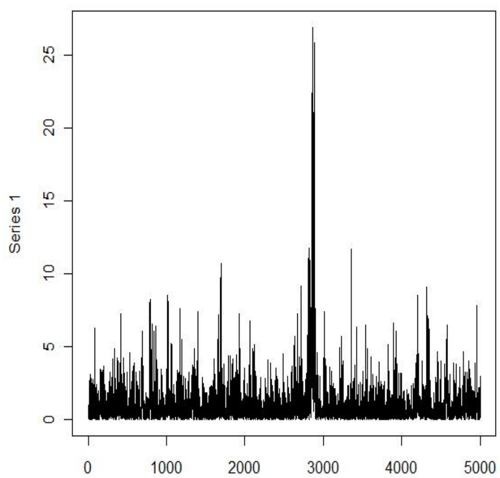


(b) Histogram for EACD (2,1)

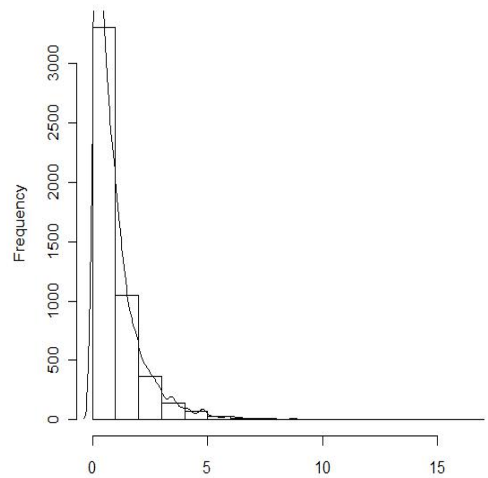


(c) Autocorrelation function for a EACD (2,1)

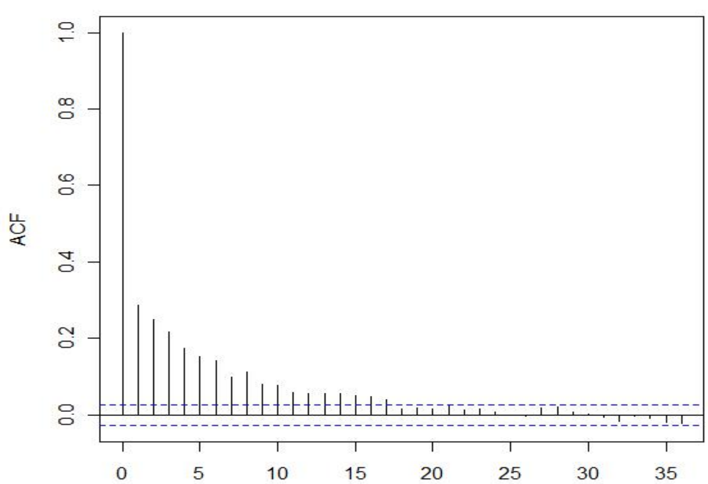
Figure 3.2: Simulated EACD(2,1) series with 5000 observations.



(a) Time plot for EACD (1,2)



(b) Histogram for EACD (1,2)



(c) Autocorrelation function for a EACD (1,2)

Figure 3.3: Simulated EACD(1,2) series with 5000 observations.

$\theta^2\kappa = \frac{1}{\kappa}$. If the stationary condition $\alpha + \beta < 1$ is satisfied, we know that

$$E(y_t) = E[E[\psi_t \epsilon_t | \mathcal{F}_{t-1}^y]] = E(\psi_t).$$

so that

$$E[y_t] = \theta\kappa E[\psi_t] = \theta\kappa \frac{\omega}{1-\alpha-\beta} = \frac{\omega}{1-\alpha-\beta},$$

and the conditional variance of y_t is

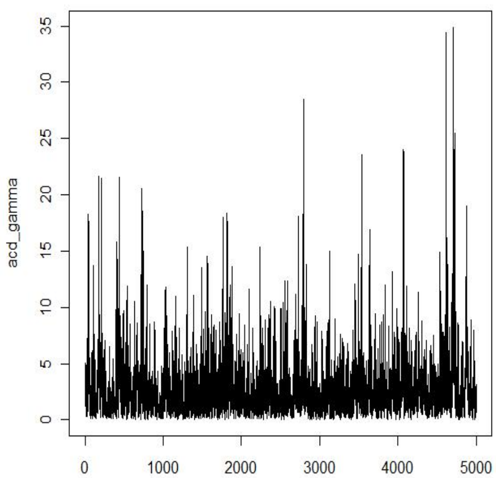
$$Var(y_t|\mathcal{F}_{t-1}^y) = \psi_t^2 Var(\epsilon_t) = \psi_t^2\theta^2\kappa = \frac{\psi_t^2}{\kappa}.$$

In our simulations, we considered a standard Gamma-distributed ACD model with parameters $\kappa = 1.5$ and $\theta = \frac{2}{3}$. A time plot, histogram and graph of the sample autocorrelation function obtained from the simulated data are provided in Figure 3.4.

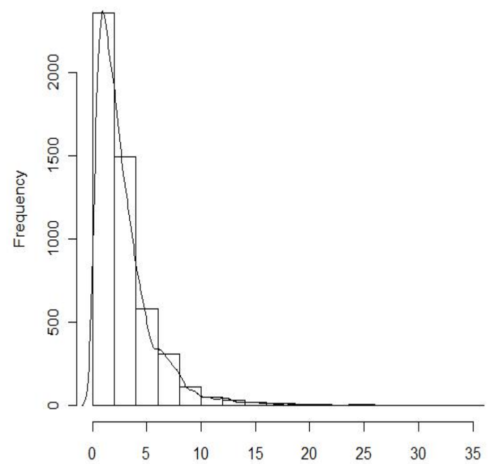
For the standard Weibull distributed ACD model with parameter $\gamma = 1.5$ and $\theta = \frac{1}{\Gamma(1+\frac{1}{1.5})}$, data were simulated and we produced a time plot, histogram and a graph of the autocorrelation function which are given in Figure 3.5.

For the standard Log-normal distribution with parameters $\mu = -\frac{\sigma^2}{2}$ and $\sigma = 1.5$, the Lognormal-ACD(1,1) was simulated. A time plot, histogram and the sample autocorrelation function of the simulated data are provided in Figure 3.6.

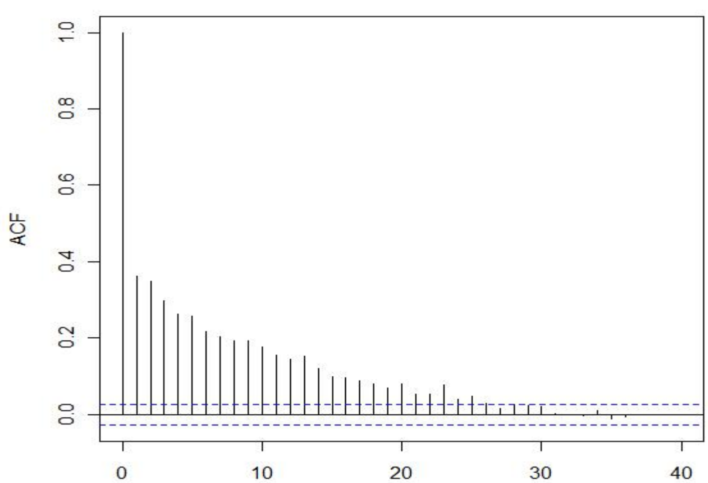
Note that all the histograms are skewed to the right and most of the data are between 0 and 10. From the plots of the ACF function, it seems that the autocor-



(a) Time plot for GACD (1,1)

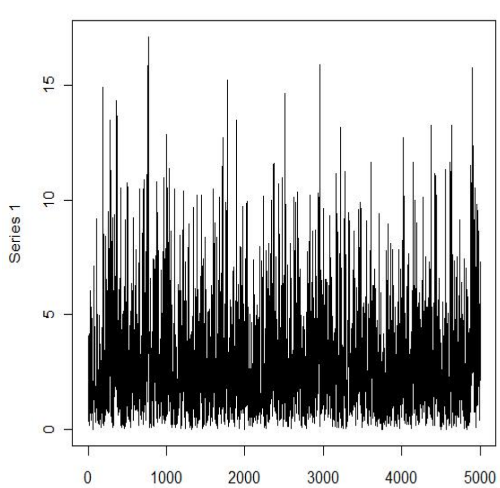


(b) Histogram for GACD (1,1)

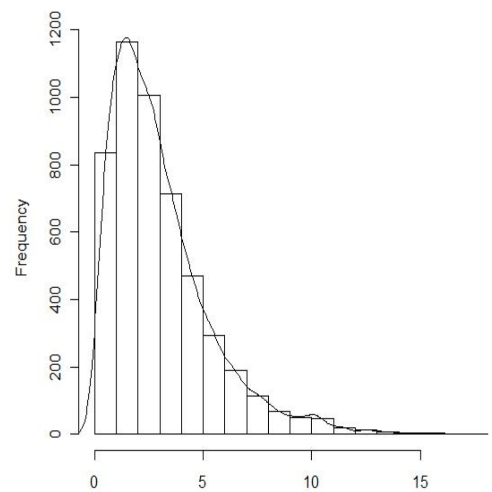


(c) Autocorrelation function for GACD (1,1)

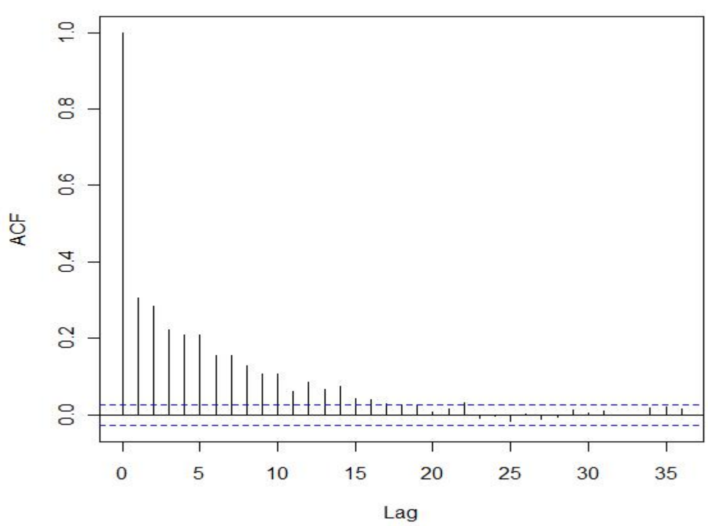
Figure 3.4: Simulated GACD(1,1) series with 5000 observations.



(a) Time plot for Weibull ACD (1,1)

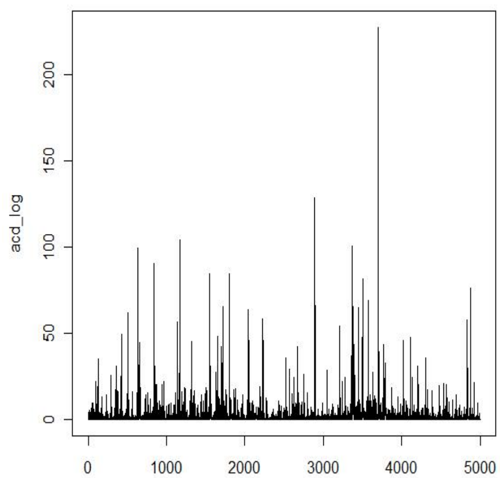


(b) Histogram for Weibull ACD (1,1)

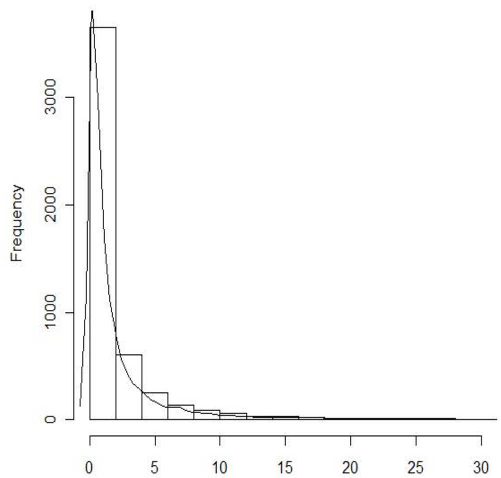


(c) Autocorrelation function for Weibull ACD (1,1)

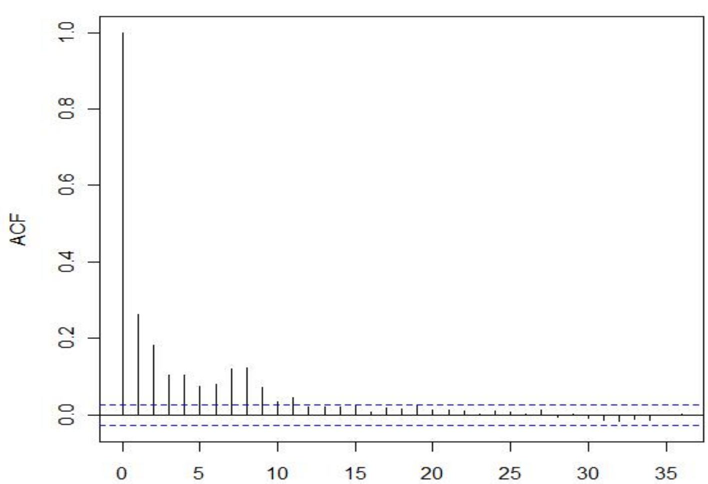
Figure 3.5: Simulated WACD(1,1) series with 5000 observations.



(a) Time plot for Log-normal ACD (1,1)



(b) Histogram for Log-normal ACD (1,1)



(c) Autocorrelation function for Log-normal ACD (1,1)

Figure 3.6: Simulated log-normal ACD(1,1) series with 5000 observations.

relation function is heavily influenced by the distribution of the innovations.

3.2.1 Parameter Estimation in ACD Models

Pathmanathan & Peiris (2009) mentioned ACD(p,q) models, for which the joint density of the durations $y_1, \ldots, y_T$ is

$$f(y_t|\Theta) = [\prod_{t=t_0+1}^{T} f(y_t|\mathcal{F}_{T-1}^{y}, \Theta)] \times f(y_{t_0}|\Theta),$$

where $t_0 = \max(p, q)$, Θ is the parameter vector and T is the sample size. As the sample size increases, the impact of the marginal density function, $f(y_{t_0}|\Theta)$ on the likelihood function will be ignored. Thus, the considered (approximate) likelihood function is written as

$$L(\Theta|\mathcal{F}_{T}^{y}) = \prod_{t=t_0+1}^{T} f(y_t|\mathcal{F}_{t-1}^{y}, \Theta).$$

Note that this can also be thought of as a conditional likelihood obtained by conditioning on the first t_0 observations. For an EACD model, the conditional log likelihood function is then

$$\begin{aligned}\ell(\Theta|\mathcal{F}_T^y) &= \sum_{t=t_0+1}^{T}\Big[\ln(\frac{1}{y_t}) + \ln(\frac{y_t}{\psi_t}) - \frac{y_t}{\psi_t}\Big],\\ &= \sum_{t=t_0+1}^{T}\Big[-\ln\psi_t - \frac{y_t}{\psi_t}\Big].\end{aligned}$$

For the GACD model, the conditional log likelihood function is

$$\ell(\Theta|\mathcal{F}_T^y) = \sum_{t=t_0+1}^{T}\Big[\kappa\ln(\kappa) - \ln(\Gamma(\kappa)) + (\kappa-1)\ln(\frac{y_t}{\psi_t}) - \frac{y_t\kappa}{\psi_t} - \ln(\psi_t)\Big]$$

For a WACD model, the conditional log likelihood function is

$$\begin{aligned}\ell(\Theta|\mathcal{F}_T^y) &= \Big\{\sum_{t=t_0+1}^{T}\gamma\ln[\Gamma(1+1/\gamma)] + \ln(\gamma/y_t) + \gamma\ln(y_t/\psi_t)\\ &\quad - \Big[\frac{\Gamma(1+1/\gamma)y_t}{\psi_t}\Big]^{\gamma}\Big\},\end{aligned}$$

when $\gamma = 1$, the conditional log likelihood function of the Weibull distribution can be reduced to exponential distribution.

Finally, for the log-normal ACD model, the conditional log likelihood function is instead

$$\ell(\Theta|\mathcal{F}_T^y) = \sum_{t=t_0+1}^{T}\Big[-\ln(\sqrt{2\pi}\sigma) - \frac{(\frac{y_t}{\psi_t} + \frac{1}{2}\sigma^2)^2}{2\sigma^2} - \ln(y_t)\Big].$$

In what follows, we consider simulated data generated from the EACD(1,1), EACD(1,2), EACD(2,1), GACD(1,1),WACD(1,1), and Log-normal ACD(1,1) models. For each model, we generated 100 samples, each with 5000 observations, and used maximum likelihood estimation based on the log-likelihood functions given above to estimate the parameter vector Θ. The following is a summary of our estimation results.

Table 3.2: Estimation results for simulated EACD(1,1) series.

EACD(1,1) Model			
Parameter	ω	α_1	β_1
True	0.3	0.2	0.7
Estimate	0.3023565	0.2007812	0.6981083
Standard Error	0.03719354	0.01580355	0.02268156

Table 3.3: Estimation results for simulated EACD(2,1) series.

EACD(2,1) Model				
Parameter	ω	α_1	α_2	β_1
True	0.1	0.2	0.3	0.4
Estimate	0.1011735	0.1999071	0.3016044	0.395243
Standard Error	0.00913526	0.02835351	0.02086202	0.02900106

Table 3.4: Estimation results for simulated EACD(1,2) series.

EACD(1,2) Model				
Parameter	ω	α_1	β_1	β_2
True	0.1	0.2	0.3	0.4
Estimate	0.1014653	0.1966242	0.289038	0.4121622
Standard Error	0.01466038	0.01976934	0.07921241	0.09425074

Table 3.5: Estimation results for simulated WACD(1,1) series.

WACD(1,1) Model				
Parameter	ω	α_1	β_1	γ
True	0.3	0.2	0.7	1.5
Estimate	0.2995685	0.1986205	0.7010674	1.500781
Standard Error	0.03336015	0.00996145	0.01711365	0.01442231

Table 3.6: Estimation results for simulated GACD(1,1) series.

GACD(1,1) Model				
Parameter	ω	α_1	β_1	κ
True	0.3	0.2	0.7	1.5
Estimate	0.3039000	0.1973360	0.7014792	1.505007
Standard Error	0.02297376	0.01207850	0.0151672	0.02646482

Table 3.7: Estimation results for simulated Log-normal ACD(1,1) series.

Log-normal ACD(1,1) Model				
Parameter	ω	α_1	β_1	σ
True	0.3	0.2	0.7	1.5
Estimate	0.3043936	0.2001154	0.6986063	1.500403
Standard Error	0.03138267	0.01681737	0.02030652	0.01569941

We used the conditional likelihood approach and obtained the results shown in Tables 3.2 to 3.6. It appears that the method we used to estimate the parameter is reasonable and leads to little bias and has small variaility.

3.3 ACD models in finance

In real financial analysis, ACD models are a useful and popular tool. In this section, we consider time between transactions of IBM stocks on five consecutive trading days from November 1, 1990 to November 7, 1990. Specifically, we analyze these data using maximum likelihood estimation, as outlined in Section 3.3. While there are 3534 observations in total for this time period, the positive transaction durations were what we focused on. Specifically, we fitted different ACD (1,1) models to this data by considering the four families of distributions discussed previously. The results are shown in Tables 3.8 to 3.11.

Table 3.8: EACD(1,1) model fitted to the data

	Maximum Likelihood Estimation
EACD(1,1)	$\psi_t = 0.1813522 + 0.06522795 y_{t-1} + 0.8805849 \psi_{t-1}$ (0.04874086) (0.00962901) (0.02088729)

Table 3.9: WACD(1,1) model fitted to the data

	Maximum Likelihood Estimation
WACD(1,1)	$\psi_t = 0.16954624 + 0.06421319 y_{t-1} + 0.88471054 \psi_{t-1}$ (0.05094409) (0.01052793) (0.02213189)
$\hat{\gamma}$	0.87883443 (0.01126365)

Table 3.10: GACD(1,1) model fitted to the data

	Maximum Likelihood Estimation
GACD(1,1)	$\psi_t = 0.18135869 + 0.06522923 x_{t-1} + 0.88057686 \psi_{t-1}$ (0.0529322) (0.01034544) (0.0226847)
$\hat{\kappa}$	0.84790936 (0.01750003)

From the estimation results, we observed that parameters based on different innovational distributions are quite similar, suggesting some kind of robustness against

Table 3.11: Lognormal ACD(1,1) model fitted to the data

	Maximum Likelihood Estimation
Lognormal ACD(1,1)	$\psi_t = 0.1481680 + 0.0684174 y_{t-1} + 0.9029417 \psi_{t-1}$ (0.04815235) (0.01156778) (0.01858372)
$\hat{\sigma}$	1.2962554 (0.01542123)

model miss-specification in this case. The estimates of ω suggest that the expectation of the ψ_t is not zero. All estimates of β are around 0.9 suggesting that the duration process is invertible. The obtained estimates also suggest that $\alpha + \beta < 1$, or that the process is stationary.

Chapter 4

Stochastic Conditional Duration Models

4.1 Definition and Properties

In order to analyze intraday market activity, Gouriéroux *et al.* (1999) introduced duration models. They defined new classes of durations, which help to illustrate some important features of market activity. Specifically, they analyzed a class of parametric models for durations which they referred to as stochastic conditional duration (SCD) models. Bauwens & Veredas (2004) proposed to use SCD models for the inter-event duration processes where the conditional duration is modelled as a latent variable. The motivation for using a latent variable is that it captures the random flow of information that, in the case of financial markets, is very difficult to observe directly.

Usually, SCD models are used for modelling sequential durations. This is based on the assumption that there exists a stochastic latent variable which generates the durations. The observed duration y_t is modelled as the product of a latent variable Ψ_t and a positive random error term ϵ_t, which is an independent identically distributed random variable. The model can be defined as

$$\begin{aligned} y_t &= \Psi_t \epsilon_t, \qquad \text{where} \quad \Psi_t = e^{\psi_t} \\ \psi_t &= \omega + \beta \psi_{t-1} + u_t \end{aligned} \tag{4.1}$$

where $|\beta| < 1$, $u_t|\mathcal{F}_{t-1}^y$ follows a $N(0, \sigma_u^2)$ distribution, $\epsilon_t|\mathcal{F}_{t-1}^y$ follows some distribution with positive support and u_t is independent of $\epsilon_t|\mathcal{F}_{t-1}^y$ for all t. Note that $\psi_t = \omega + \beta\psi_{t-1} + u_t$ is essentially imposing a Gaussian AR(1) structure on the logarithm of the latent variable Ψ_t. $\mathcal{F}_{t-1}^y$ denotes the information set based on past values of ψ_t and y_t up to time $t-1$. The marginal distribution of y_t is determined by the distribution of the error term ϵ_t and the lognormal distribution of Ψ_t which comes from the fact that ψ_t is normally distributed. The only assumption imposed on ϵ_t is the existence of its moments

$$\begin{aligned} m_1 &= E(\epsilon_1) \\ m_p &= E(\epsilon_t - m_1)^p, \quad p = 2, \ldots \end{aligned}$$

Finally, let $\vartheta = m_2/m_1^2$.

The following two important theorems give the form of the moments of the observations and latent variable up to the second order.

Theorem 4.1.1. *The durations and latent variables of model* (4.1) *have the following moments*

$$\begin{aligned}
\mu_\Psi &= e^{\frac{\omega}{1-\beta}+\frac{1}{2}\frac{\sigma_u^2}{1-\beta^2}}, \\
\mu_y &= m_1\mu_\Psi, \\
\sigma_\Psi^2 &= \mu_\Psi^2(e^{\frac{\sigma_u^2}{1-\beta^2}}-1), \\
\sigma_y^2 &= \mu_y^2(\vartheta e^{\frac{\sigma_u^2}{1-\beta^2}}-1).
\end{aligned}$$

Proof. According to (4.1), $\{\psi_t\}$ is a Gaussian Stationary AR(1) process. As we know, this implies $\psi_t \sim N(\frac{\omega}{1-\beta}, \frac{\sigma_u^2}{1-\beta^2})$, which in turn implies that Ψ_t follows a lognormal distribution, specifically

$$\Psi_t \sim LN(\frac{\omega}{1-\beta}, \frac{\sigma_u^2}{1-\beta^2}).$$

The expressions for μ_Ψ and σ_Ψ^2 follow from the properties of lognormal distributions. The mean and variance of durations are obtained using the independence between Ψ_t and ϵ_t. Indeed, for the mean, we have

$$\mu_y = E(y_t) = E(\Psi_t)E(\epsilon_t) = m_1\mu_\Psi.$$

and for the variance,

$$\begin{aligned}\sigma_y^2 &= E[y_t^2] - [E[y_t]]^2 \\ &= E[\Psi_t^2]E[\epsilon_t^2] - (m_1\mu_\Psi)^2 \\ &= \mu_y^2(\vartheta e^{\frac{\sigma_u^2}{1-\beta^2}} - 1),\end{aligned}$$

upon using the previous definition of ϑ. □

Theorem 4.1.2. *The autocorrelation function (ACF) of the durations in model (4.1) is given by*

$$\rho_s^y = \frac{e^{\frac{\sigma_u^2\beta^s}{1-\beta^2}} - 1}{\vartheta e^{\frac{\sigma_u^2}{1-\beta^2}} - 1} \quad \forall\ s \geq 1$$

Proof. We know that

$$\rho_s^y = [E(y_t y_{t-s}) - \mu_y^2]/\sigma_y^2, \tag{4.2}$$

so that the main purpose of this proof is to calculate the expectation of $y_t y_{t-s}$. For this, we write

$$E(y_t y_{t-s}) = m_1^2 E(e^{\psi_t + \psi_{t-s}}) = m_1^2 E(e^{\lambda_{t,s}}),$$

where $\lambda_{t,s} = \psi_t - \psi_{t-s}$. From the autoregressive equation of ψ_t, we can write

$$\begin{aligned}\lambda_{t,s} = \psi_t + \psi_{t-s} &= 2\omega + \beta(\psi_{t-1} + \psi_{t-1-s}) + u_t + u_{t-s} \\ &= 2\omega + \beta\lambda_{t-1,s} + u_t + u_{t-s},\end{aligned}$$

implying that $\{\lambda_{t,s}\}$ is a ARMA(1,s) process. Because $E(\lambda_{t,s}) = 2\omega + \beta E(\lambda_{t-1,s})$, then $\mu_s = \frac{2\omega}{1-\beta}$. The variance σ_s^2 of $\lambda_{t,s}$ can be given by

$$
\begin{aligned}
\sigma_s^2 &= \beta\gamma_{1,s} + \sigma_u^2 + (1+\beta^s)\sigma_u^2, \\
\gamma_{1,s} &= \beta\sigma_s^2 + \beta^{s-1}\sigma_u^2,
\end{aligned}
$$

where $\gamma_{1,s} = Cov(\lambda_{t,s}, \lambda_{t-1,s})$. Solving for σ_s^2, we get

$$
\sigma_s^2 = \frac{2\sigma_u^2(1+\beta^s)}{1-\beta^2}
$$

This, in turn, implies that

$$
\begin{aligned}
E(e^{\lambda_{i,s}}) &= e^{\mu_s} + 0.5\sigma_s^2 = \exp\{\tfrac{2\omega}{1-\beta} + \tfrac{\sigma_u^2(1+\beta^s)}{1-\beta^2}\} \\
&= u_y^2 e^{\frac{\sigma_u^2\beta^s}{1-\beta^2}},
\end{aligned}
$$

upon using the properties of lognormal distributions. Hence by (4.2) the result follows. $\square$

4.2 Quasi Maximum Likelihood Estimation and the Kalman Filter

In this section, we adopt a state-space approach based on Kalman filtering and the QML method. This kind of approach is has been considered by Harvey (1994) to model the stochastic volatility of stock prices. The main benefits of using the Kalman filter with QML for estimation is the relative simplicity of the approach and the consistency of the obtained estimates.

4.2.1 State-space models and the Kalman filter

All time series models can be represented in state-space form. The Kalman filter is one of the most important tool in the application of time series modelling based on state-space representations. It is a recursive procedure used to optimally estimate parameters of the state vector at each time point. Filters are mainly used for prediction and smoothing of time series, which plays a key role in optimization problems. Specifically, the Kalman filter is a popular tool used in likelihood function estimation through the prediction error decomposition optimization. By using the Kalman filter to minimize the mean square error (MSE), the complex modelling of time series becomes realistic.

Before formally introducing the Kalman filter, a brief general description of state-space models is in order. (For details, see Bauwens & Veredas 2004 and Harvey, 1994) Generally speaking, a state-space model can be written as

$$Y_t = Z_t\alpha_t + d_t + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where Y_t is a multivariate time series with N elements, α_t is an unobservable $m \times 1$ vector known as the state vector, Z_t is an $N \times m$ matrix, d_t is a $N \times 1$ vector and ε_t is an $N \times 1$ random vector. It is assumed that the error ε_t are uncorrelated with mean 0 and covariance matrix H_t, that is, $E(\varepsilon_t) = 0$ and $Cov(\varepsilon_t) = E(\varepsilon_t \varepsilon_t^{'}) = H_t$. In general, the elements of the unobservable α_t are assumed to be generated by

$$\alpha_t = B_t \alpha_{t-1} + c_t + R_t \eta_t, \qquad t = 1, \ldots, T,$$

where B_t is an $m \times m$ matrix, c_t is an $m \times 1$ vector, R_t is an $m \times g$ matrix and the $g \times 1$ innovation vectors η_t satisfy $E(\eta_t) = 0$, $Cov(\eta_t) = Q_t$ and are independent.

We refer to Z_t, d_t, H_t and B_t, c_t, R_t and Q_t as system matrices. If the system matrices do not change over time, the model is said to be time-homogenous. Although the class of time-homogenous models is much broader than the class of stationary models, many time-homogenous models have a stationary form which can be obtained by applying a transformation such as differentiating.

In what follows, we focus on the univariate model, where $N = 1$, so that the measurement equation simplifies to

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{4.3}$$

where y_t univariate and $Var(\varepsilon_t) = h_t$. The transition equation is written as

$$\alpha_t = B_t \alpha_{t-1} + c_t + R_t \eta_t, \quad t = 1, \ldots, T, \tag{4.4}$$

where α_t is unobservable, B_t is an $m \times m$ matrix, c_t is an $m \times 1$ vector, R_t is $m \times g$ matrix and $\{\eta_t\}$ is a sequence of $g \times 1$ uncorrelated vectors with mean 0 and covariance matrix Q_t, that is $E(\eta_t) = 0$ and $Cov(\eta_t) = Q_t$.

Two essential assumptions are made about the initial state vector α_0:

1. we fix $E(\alpha_0) = a_0$ and $Var(\alpha_0) = P_0$,

2. the random vectors ε_t and η_t are uncorrelated with each other in all time periods, and also uncorrelated with the initial state, that is

$$E(\varepsilon_t \eta_t^{'}) = 0,$$

and

$$E(\varepsilon_t \alpha_0^{'}) = E(\eta_t \alpha_0^{'}) = 0,$$

where for all $t = 1, \ldots, T$.

In what follows, the linearity in α of the state-space model, given by (4.3) and (4.4), is crucial. Now, let a_t be the estimator of α_t and let P_t denote the $m \times m$ covariance matrix of the estimation error. Then, we can write

$$P_t = E[(\alpha_t - a_t)(\alpha_t - a_t)^{'}]$$

and the optimal estimator of α_t, given a_{t-1} and P_{t-1}, is given by

$$a_{t|t-1} = B_t a_{t-1} + c_t, \tag{4.5}$$

which has conditional covariance, given the prior information,

$$P_{t|t-1} = B_t P_{t-1} B_t^{'} + R_t Q_t R_t^{'}, \tag{4.6}$$

for t=1,..., T. Equations (4.5) and (4.6) are called the *prediction equations*. Denoting $E[y_t|\mathcal{F}_{t-1}^y] = \hat{y}_{t|t-1}$, the *updating equations* can be written as

$$a_t = a_{t|t-1} + P_{t|t-1}Z_t^{'}F_t^{'}(y_t - Z_t a_{t|t-1} - d_t), \tag{4.7}$$

and

$$P_t = P_{t|t-1} - P_{t|t-1}Z_t F_t^{-1} Z_t^{'} P_{t|t-1}, \tag{4.8}$$

and

$$v_t = y_t - \hat{y}_{t|t-1} = y_t - Z_t a_{t|t-1} - d_t, \tag{4.9}$$

where

$$F_t = E[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})^{'}|\mathcal{F}_{t-1}^y] = Z_t P_{t|t-1} Z_t^{'} + H_t, \tag{4.10}$$

for $t = 1, \ldots, T$. Equations (4.5) to (4.8) together make up the Kalman filter. They are derived under the assumption that all innovations are normally distributed.

The Kalman filter's starting values can be specified in term of a_0 and P_0 or $a_{1|0}$ and $P_{1|0}$. Given these initial conditions, the Kalman filter delivers the optimal estimation of the state vector by sequential updating. This optimality is valid only under the normality and linearity mentioned previously. For the initial values, we first get from (4.4) that

$$P_0 \quad = \quad Cov(\alpha_0) = (R_0 Q_0 R_0^{'})(I - B_0 B_0^{'})^{-1},$$

and from (4.6) that

$$P_{1|0} = B_1 P_0 B_1^{'} + R_1 Q_1 R_1^{'},$$

so that

$$F_1 = Z_1 P_{1|0} Z_1^{'} + H_1.$$

Also, the initial value α_0 has expectation $E(\alpha_0) = c_1(I - B)^{-1}$. According to

$$\begin{aligned} v_t &= y_t - Z_t a_{t|t-1} - d_t, \\ a_{t|t-1} &= B_t a_{t-1} + c_t, \end{aligned}$$

we know that

$$v_1 = y_1 - Z_1 a_{1|0} - d_1 = y_1 - Z_1(B_1 c_1 (I - B_1)^{-1} + c_1) - d_1.$$

The filtering of the series is then completed by cycling through (4.5) to (4.8) to obtained the sequential estimates $a_1, a_2, \ldots, a_t$.

4.2.2 Maximum Likelihood Estimation and the Prediction Error Decomposition

In the current context, the observations are not independent, so that the likelihood is decomposed as the product of conditional distributions according to

$$L(y;\psi) = f(y_1, y_2, \ldots, y_t|\psi) = \prod_{t=1}^{T} f(y_t|\mathcal{F}_{t-1}^{y}), \tag{4.11}$$

where $f(y_t|\mathcal{F}_{t-1}^{y})$ is the distribution of y_t conditional on the information set at time $t-1$. Recalling the derivation of the Kalman filter, equation (4.11) can be rewritten as

$$\log L = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T} v_t' F_t^{-1} v_t, \tag{4.12}$$

when assuming normality of the innovations, where the residuals v_t are given by (4.9) and the covariance matrix F_t is given by (4.10).

4.2.3 QML Estimation based on the Kalman Filter

For estimating latent variable models, several estimation methods have been proposed. Estimating the parameters of this kind of unobservable variable model is generally difficult because the likelihood function can't be evaluated easily. Here, a logarithmic transformation is used to estimate the parameters of the SCD model.

Using such a transformation, the SCD model (4.1) can be rewritten as

$$\ln y_t = \psi_t + \tilde{\xi}_t + \mu, \tag{4.13}$$

and

$$\psi_t = \omega + \beta\psi_{t-1} + u_t, \tag{4.14}$$

where $\tilde{\xi}_t = \xi_t - \mu$, $\xi_t = \ln \epsilon_t$ and $\mu = E[\xi_t] = E[\ln \epsilon_t]$. Model (4.1) has thus been linearized and we can hope to use the Kalman filter successfully to estimate its parameters. Equation (4.13) is called the *measurement equation* and (4.14) is called the *updating equation* or *transition equation*.

In the case where ϵ_t follows a Weibull (α,1) distribution, $\xi_t = \ln \epsilon_t$ has probability density function

$$f(\xi) = \alpha e^{\alpha\xi} e^{-e^{\alpha\xi}}, \qquad \xi_t \in R.$$

In order to estimate the vector of parameters $(\omega, \beta, \alpha, \sigma_u)'$ we compute the log likelihood function of the above model by making use of (4.12) and of the Kalman filter to obtain values for $\psi_1, \psi_2, \ldots, \psi_T$ for given values of all other parameters. Note that doing this would lead to the exact likelihood function if ξ_t were normally distributed with mean 0 and variance σ_ξ^2 depending on α. Obviously, this is not the case and so an approximate value of the likelihood is obtained instead. This explains the "quasi" in the name quasi maximum likelihood. To use this approach, we need to calculate the moments of $\xi_t = \ln \epsilon_t$. We do this by deriving its distribution. First,

ϵ_t follows the Weibull$(\alpha, 1)$ distribution, that is

$$f(\epsilon_t|\alpha) = \alpha\epsilon_t^{\alpha-1}e^{-(\epsilon_t)^\alpha}.$$

Now, let $\xi_t = \ln \epsilon_t$ so that $\epsilon_t = e^{\xi_t}$. The Jacobian of this transformation is given by $J = e^{\xi_t}$, so that

$$\begin{aligned} f(\xi_t|\alpha) &= \alpha(e^{\xi_t})^{\alpha-1}e^{-(e^{\xi_t})^\alpha}e^{\xi_t} \\ &= \alpha e^{\alpha\xi_t}e^{-e^{\alpha\xi_t}}, \end{aligned} \tag{4.15}$$

for $\varepsilon_t \in R$, which corresponds to an Extreme Value distribution. Indeed, when a random variable X follows the Extreme value distribution with parameters θ and η, its pdf is given by

$$f(x|\theta, \eta) = \tfrac{1}{\theta}e^{(x-\eta)/\theta - e^{(x-\eta)/\theta}}.$$

According to the properties of Extreme Value distribution, we know that $E(X) = -\gamma\theta + \eta$ and $Var(X) = \frac{\pi^2}{6}\theta^2$, where $\gamma = -0.5772$ is the famous Euler-Mascheroni constant. Going back to the previous distribution, it is now clear that (4.15) implies $\xi_t \sim EV(1/\alpha, 0)$ so that $\mu = E(\xi_t) = -0.5772/\alpha$, and $\sigma_\xi^2 = Var(\xi_t) = \pi^2/6\alpha^2$. For the time-homogenous stationary model, ψ_0 is set to be $E[\psi_0] = \frac{\omega}{1-\beta}$. Finally, we can get

$$E(\ln y_t) = E(\psi_t + \tilde{\xi}_t + \mu) = \frac{\omega}{1-\beta} - \frac{0.5772}{\alpha},$$

and

$$Var(\ln y_t) = \frac{\sigma_u^2}{1-\beta^2} + \frac{\pi^2}{6\alpha^2}.$$

Now we can estimate the parameters by the quasi maximum likelihood method based on the previous discussion. For the prediction, we use the formulae

$$\begin{aligned} P_{t|t-1} &= B_t P_{t-1} B_t' + R_t Q_t R_t', \\ F_t &= Z_t P_{t|t-1} Z_t' + H_t, \\ P_t &= P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}. \end{aligned}$$

The prediction error decomposition of the quasi log-likelihood function is given by (ignoring the constant term)

$$\ln L(\theta) = -\frac{1}{2}\sum_{t=1}^{N} \ln F_t - \frac{1}{2}\sum_{t=1}^{N} \frac{v_t^2}{F_t}, \tag{4.16}$$

where $v_t = x_t - \widehat{x}_{t|t-1}$, with $x = \ln y_t$ and $\widehat{x}_{t|t-1} = \widehat{\ln y_{t|t-1}}$, so that v_t is the difference between the log-duration and its prediction, and F_t is an estimate the conditional variance of v_t. Hence,

$$\begin{aligned} v_t &= \ln y_t - E(\ln y_t|\mathcal{F}_{t-1}^y) = \ln y_t - [\beta\hat{\psi}_{t-1} + \omega + \mu] \\ &= \ln y_t - \beta\hat{\psi}_{t-1} - \omega + \tfrac{0.5772}{\alpha}, \end{aligned}$$

and F_t is given by (4.10). For the initial values, we use

$$v_0 = \psi_0 + \ln \epsilon_0 - \frac{\omega}{1-\beta} + \frac{0.5772}{\alpha},$$

and

$$F_0 = \frac{\sigma_u^2}{1-\beta^2} + \frac{\pi^2}{6\alpha^2}.$$

Approximate estimates of the parameters are obtained by optimizing the quasi log-likelihood function derived above.

4.3 Using a nonlinear filtering scheme

Thavaneswaran & Gong (2009) studied recursive estimation for a class of continuous time nonlinear non-Gaussian stochastic volatility models used for option pricing in finance. Comparing to the kalman filter, nonlinear filter can be used in non-Gaussian distribution. They suggested a filtering procedure for discrete time stochastic volatility models associated with a nonlinear state space model given by

$$
\begin{aligned}
y_t &= A\theta_{t-1} + z_t, \\
\theta_t &= a\theta_{t-1} + (1+\theta_{t-1})\eta_t,
\end{aligned}
$$

where $\{z_t\}$ is a sequence of independent and identically distributed innovations with mean 0 and variance σ_Z^2, $\{\eta_t\}$ is a sequence of independent and identical distribution random variables with mean 0 and variance σ_η^2, and the sequences $\{z_t\}$ and $\{\eta_t\}$ are independent. The $\{y_t\}$ process is observed and $\{\theta_t\}$ is a nonobservable random parameter process.

Lemma 4.3.1. *In the above context, the MSE optimal sequential estimate $\hat{\theta}_t$ of θ_t is*

$$
\hat{\theta}_t = a\hat{\theta}_{t-1} + \frac{Aa\gamma_{t-1}}{A^2\gamma_{t-1} + \sigma_z^2}(y_t - A\hat{\theta}_{t-1}),
$$

and the MSE $\gamma_t = E[(\theta_t - \hat{\theta}_t)^2 | \mathcal{F}_{t-1}^y]$ of this minimizing estimator is

$$
\begin{aligned}
\gamma_t &= a^2\gamma_{t-1} + b_1^2\sigma_\eta^2 - \tfrac{(aA\gamma_{t-1})^2}{A^2\gamma_{t-1}+\sigma_z^2} \\
&= \frac{a^2\sigma_z^2\gamma_{t-1}}{A^2\gamma_{t-1} + \sigma_z^2} + b_1^2\sigma_\eta^2,
\end{aligned}
$$

where $b_1 = \sqrt{E(1+\theta_{t-1})^2}$.

In what follows, we adapt the argument used by Thavaneswaran & Gong (2009) to the SCD model setup. First, the model can be rewritten as

$$\begin{aligned} \psi_t &= \beta\psi_{t-1} + \omega + u_t, \\ x_t &= \psi_{t-1} + \xi_t + c \end{aligned}$$

where $x_t = \ln y_t$, $\xi_t = \ln \epsilon_t - \mu$ and $c = \mu = E[\ln \epsilon_t]$. The following theorem establishes a nonlinear filtering scheme for the SCD model.

Theorem 4.3.2. *For the SCD model, the MSE optimal linear sequential estimator* $\hat{\psi}_t$ *and its MSE* $\gamma_t = E[(\psi_t - \hat{\psi}_t)^2 | \mathcal{F}_t^y]$ *are given by*

$$\begin{aligned} \hat{\psi}_t &= \beta\hat{\psi}_{t-1} + \omega + \frac{\beta\gamma_{t-1}}{\gamma_{t-1} + \sigma_\xi^2}(x_t - \hat{\psi}_{t-1} - c), \\ \gamma_t &= \beta^2\gamma_{t-1} + \sigma_u^2 - \frac{\beta^2\gamma_{t-1}^2}{\gamma_{t-1} + \sigma_\xi^2} \\ &= \sigma_u^2 + \beta^2\gamma_{t-1}\left(1 - \frac{\gamma_{t-1}}{\gamma_{t-1} + \sigma_\xi^2}\right). \end{aligned}$$

Proof. First, proceeding like Gong & Thavaneswaran (2009), we consider conditional linear predictions of the form

$$\hat{\psi}_t = \beta\hat{\psi}_{t-1} + \omega + G_t(x_t - \hat{\psi}_{t-1} - c),$$

where the choice of G_t is made by minimizing the prediction MSE. Note that, we can write

$$
\begin{aligned}
\psi_t - \hat{\psi}_t &= \beta\psi_{t-1} + \omega + u_t - [\beta\hat{\psi}_{t-1} + \omega + G_t(x_t - \hat{\psi}_{t-1} - c)] \\
&= \beta[\psi_{t-1} - \hat{\psi}_{t-1}] + u_t - G_t(x_t - \hat{\psi}_{t-1} - c) \\
&= \beta[\psi_{t-1} - \hat{\psi}_{t-1}] + u_t - G_t(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t),
\end{aligned}
$$

since

$$
x_t - \hat{\psi}_{t-1} - c = \psi_{t-1} + \xi_t + c - \hat{\psi}_{t-1} - c = \psi_{t-1} - \hat{\psi}_{t-1} + \xi_t. \tag{4.17}
$$

By using (4.17), then we can write

$$
\begin{aligned}
\gamma_t &= E[(\psi_t - \hat{\psi}_t)^2 | \mathcal{F}^y_{t-1}] \\
&= E[(\beta(\psi_{t-1} - \hat{\psi}_{t-1}) + u_t - G_t(y_t - \hat{\psi}_{t-1} - c))^2 | \mathcal{F}^y_{t-1}] \\
&= E[(\beta(\psi_{t-1} - \hat{\psi}_{t-1}) + u_t - G_t(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t)^2 | \mathcal{F}^y_{t-1}] \\
&= E\Big[\beta^2(\psi_{t-1} - \hat{\psi}_{t-1})^2 + u_t^2 + G_t^2(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t))^2 + \\
&\quad 2\beta[\psi_{t-1} - \hat{\psi}_{t-1}]u_t - 2G_t(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t)(\beta(\psi_{t-1} - \hat{\psi}_{t-1}) + u_t)\Big] \\
&= \beta^2 E[(\psi_{t-1} - \hat{\psi}_{t-1})^2] + \sigma_u^2 + G_t^2 E[(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t)^2] - \\
&\quad 2G_t\beta E[(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t)(\psi_{t-1} - \hat{\psi}_{t-1})].
\end{aligned} \tag{4.18}
$$

Now, note that the independence of ψ_t and ξ_{t+1} allows us to write

$$\begin{aligned} E(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t)^2 &= E\{(\psi_{t-1} - \hat{\psi}_{t-1})^2 + \xi_t^2 + 2(\psi_{t-1} - \hat{\psi}_{t-1})\xi_t\} \\ &= E(\psi_{t-1} - \hat{\psi}_{t-1})^2 + E(\xi_t)^2 \\ &= \gamma_{t-1} + \sigma_\xi^2, \end{aligned}$$

and

$$\begin{aligned} E[(\psi_{t-1} - \hat{\psi}_{t-1} + \xi_t)(\psi_{t-1} - \hat{\psi}_{t-1})] &= E(\psi_{t-1} - \hat{\psi}_{t-1})^2 + E[(\psi_{t-1} - \hat{\psi}_{t-1})\xi_t] \\ &= E(\psi_{t-1} - \hat{\psi}_{t-1})^2 \\ &= \gamma_{t-1} \end{aligned}$$

Going back to (4.18), we get

$$\begin{aligned} \gamma_t &= \beta^2\gamma_{t-1} + \sigma_\mu^2 + G_t^2(\gamma_{t-1} + \sigma_\xi^2) - 2G_t\beta\gamma_{t-1} \\ &= G_t^2(\gamma_{t-1} + \sigma_\xi^2) - 2G_t\beta\gamma_{t-1} + \beta^2\gamma_{t-1} + \sigma_\mu^2 \\ &= (\gamma_{t-1} + \sigma_\xi^2)\left[G_t^2 - 2G_t\tfrac{\beta\gamma_{t-1}}{\gamma_{t-1}+\sigma_\xi^2} + (\tfrac{\beta\gamma_{t-1}}{\gamma_{t-1}+\sigma_\xi^2})^2 - (\tfrac{\beta\gamma_{t-1}}{\gamma_{t-1}+\sigma_\xi^2})^2\right] + \beta^2\gamma_{t-1} + \sigma_\mu^2 \\ &= (\gamma_{t-1} + \sigma_\xi^2)(G_t - \tfrac{\beta\gamma_{t-1}}{\gamma_{t-1}+\sigma_\xi^2})^2 - \tfrac{\beta^2\gamma_{t-1}^2}{\gamma_{t-1}+\sigma_\xi^2} + \beta^2\gamma_{t-1} + \sigma_\mu^2. \end{aligned}$$

Obviously, the best prediction of the form (5.2) is obtained when $G_t = \dfrac{\beta\gamma_{t-1}}{\gamma_{t-1} + \sigma_\xi^2}$.

The stated result follow directly. □

In order to estimate the model parameters, the filter we introduced here can also be applied to compute the likelihood function for an SCD model. For this, we still use the QML method to estimate the parameters based on the prediction decomposition given in (4.16). However, the v_t and F_t are now obtained using the second filtering scheme. Specifically, we know $v_t = x_t - \widehat{x}_{t|t-1}$, where $\widehat{x}_{t|t-1}$ is the conditional mean of $x_t = \ln y_t$, so that

$$v_t = x_t - [\beta\hat{\psi}_{t-1} + \omega + c].$$

and F_t is the conditional variance of v_t, that is

$$\begin{aligned} F_t &= Var(v_t|\mathcal{F}^y_{t-1}) = Var(\psi_t - \hat{\psi}_t + \xi_t) \\ &= \gamma_t + \tfrac{\pi^2}{6\alpha^2}. \end{aligned}$$

4.4 Estimation based on the Generalized Method of Moments

In this section we describe the generalized method of moments (GMM) and its use in estimation for SCD models with applications in finance. Hansen (1982) formalized the GMM approach to estimation and, since then, GMM has become one of the most widely used estimation method for statistical models in finance. The method has been used successfully to estimate stochastic volatility (SV) models by Ruiz (1994) and Andersen & Sorensen (1996). In our model, Bauwence & Vededas did suggest it may be a suitable estimator. The main problems for GMM are which moments to match and how many moments to include in the estimation. We here follow the suggestion of Andersen & Sorensen (1996). Having obtained all the moments, SCD models can be estimated following standard GMM procedure. Under the regularity conditions, the GMM estimators are consistent and asymptotically normal.

For the problem at hand, we select moments with the following considerations. Firstly, in determining the numbers of moments, we keep in mind the findings of Andersen & Sorensen (1996): using more moments improves the estimation performance but causes a deterioration in the estimation of the weighting matrix. One should try to achieve a reasonable trade-off between using too many or not enough moments. Secondly, since the autocorrelation is varying over different lags, we use cross moments to capture the dependence. Thirdly, the first four moments should be included to capture the mean, variance, skewness and kurtosis of the data. Consequently, we choose the first four univariate moments and the first ten cross moments, namely $E[y_t^m]$ for $m = 1, 2, 3, 4.$ and $E[y_t y_{t-r}]$ for $r = 1, \ldots, 10$ to estimate all the parameters.

The following theorem gives useful theoretical moments for the sequence of durations y_t. Knight & Ning (2008) originally derived these results for the Weibull SCD model.

Theorem 4.4.1. *For model (4.1), the moments and cross-moments of durations are given by*

$$E[y_t^m y_{t-r}^n] = \exp\left(\tfrac{\omega(m+n)}{1-\beta} + \tfrac{m^2+n^2+2mn\beta^r}{2(1-\beta^2)}\sigma_u^2\right)\Gamma(m/\gamma+1)\Gamma(n/\gamma+1), \tag{4.19}$$

for $m, n \geq 0$.

Proof. From model (4.1), we know that $\epsilon_t \sim i.i.d$ Weibull $(\gamma, 1)$, and $u_t \sim N(0, \sigma_\mu^2)$, so that $\psi_t \sim N(\frac{\omega}{1-\beta}, \frac{\sigma^2}{1-\beta^2})$. Now, note that

$$E[y_t^m y_{t-r}^n] = E[\exp(m\psi_t + n\psi_{t-r})\epsilon_t^m \epsilon_{t-r}^n].$$

On the other hand, we also have

$$\begin{aligned}
\psi_t &= \omega + \beta[\omega + \beta\psi_{t-2} + u_{t-1}] + u_t \\
&= \omega + \omega\beta + \beta^2\psi_{t-2} + \beta u_{t-1} + u_t \\
&\vdots \qquad \vdots \qquad \vdots \\
&= \frac{\omega(1-\beta^r)}{1-\beta} + \beta^r\psi_{t-r} + \sum_{j=0}^{r-1}\beta^j u_{t-j},
\end{aligned}$$

which allows us to write, when $m, n > 0$,

$$\begin{aligned}
E[y_t^m y_{t-r}^n] &= E[\exp(m\psi_t + n\psi_{t-r})\epsilon_t^m \epsilon_{t-r}^n] \\
&= E\Big\{\exp\Big[m(\omega\frac{1-\beta^r}{1-\beta} + \beta^r\psi_{t-r} + \sum_{j=0}^{r-1}\beta^j u_{t-j}) + n\psi_{t-r}\Big]\Big\}E[\epsilon_t^m \epsilon_{t-r}^n] \\
&= E\Big\{\exp(\tfrac{m\omega(1-\beta^r)}{1-\beta}) + (n + m\beta^r)\psi_{t-r} + m\sum_{j=0}^{r-1}\beta^j u_{t-j}\Big\}E[\epsilon_t^m \epsilon_{t-r}^n] \\
&= \exp\left(\tfrac{m\omega(1-\beta^r)}{1-\beta}\right)E\left[\exp(m\sum_{j=0}^{r-1}\beta^j u_{t-j})\right]E\Big[(n + m\beta^r)\psi_{t-r}\Big]E[\epsilon_t^m \epsilon_{t-r}^n] \\
&= \exp\left(\frac{m\omega(1-\beta^r)}{1-\beta}\right)\cdot A\cdot B\cdot C
\end{aligned}$$

where we let

$$
\begin{aligned}
A &= E\Bigg[\exp(m\sum_{j=0}^{r-1}\beta^j u_{t-j})\Bigg], \\
B &= E\Bigg[\exp[(n+m\beta^r)\psi_{t-r}]\Bigg], \\
C &= E\Bigg[\epsilon_t^m \epsilon_{t-r}^n\Bigg].
\end{aligned}
$$

Now, the value of A is,

$$
A = E\Bigg[\exp(m\sum_{j=0}^{r-1}\beta^j u_{t-j})\Bigg] = E\Bigg[\exp(mZ_r)\Bigg],
$$

where $Z_r = \sum_{j=0}^{r-1}\beta^j u_{t-j} \sim N(0, \dfrac{1-\beta^{2r}}{1-\beta^2}\sigma_u^2)$. A simple expression for A then easily follows by making use of the moment generating function of Z_r, that is

$$
A = \exp\left\{\frac{m^2\sigma_u^2}{2}\frac{(1-\beta^{2r})}{1-\beta^2}\right\}.
$$

Similarly, since $\psi_t \sim N(\frac{\omega}{1-\beta}, \frac{\sigma_u^2}{1-\beta^2})$ for all t, we have that

$$
\begin{aligned}
B &= E\Bigg\{\exp\Bigg[(n+m\beta^r)\psi_{t-r}\Bigg]\Bigg\} \\
&= \exp\left[\frac{(n+m\beta^r)\omega}{1-\beta} + \frac{(n+m\beta^r)^2\sigma_u^2}{2(1-\beta^2)}\right].
\end{aligned}
$$

Finally, we know that for $X \sim$ Weibull (α, β),

$$E(X^m) = \beta\Gamma[1 + \frac{m}{\alpha}].$$

This implies here that

$$\begin{aligned} C &= E[\epsilon_t^m \epsilon_{t-r}^n] \\ &= \Gamma(1 + \tfrac{m}{\gamma})\Gamma(1 + \tfrac{n}{\gamma}). \end{aligned}$$

To complete the proof, note that $n = 0$, $E[y_t^m y_{t-r}^n]$ reduces to

$$E[y_t^m] = E[\exp(m\psi_t)\epsilon_t^m] = \exp\left[\frac{m\omega}{1-\beta} + \frac{m^2\sigma_u^2}{2(1-\beta^2)}\right]\Gamma(1 + \frac{m}{\gamma}),$$

since $\psi_t \sim N(\frac{\omega}{1-\beta}, \frac{\sigma_u^2}{1-\beta^2})$. A similar result is easily obtained when $m = 0$. □

Note that in the case where $m = n = 1$, (4.19) is equivalent to

$$E[y_t y_{t-j}] = \exp\left[\frac{2\omega}{1-\beta} + \frac{(2 + 2\beta^j)\sigma_u^2}{2(1-\beta^2)}\right][\Gamma(1 + \frac{1}{\gamma})]^2,$$

where $j = 1, \ldots 10$. Finally, for $m = n = 2$,

$$E[y_t^2 y_{t-j}^2] = \exp[\frac{4\omega}{1-\beta} + \frac{(8 + 8\beta^j)\sigma_u^2}{2(1-\beta^2)}][\Gamma(1 + \frac{2}{\gamma})]^2,$$

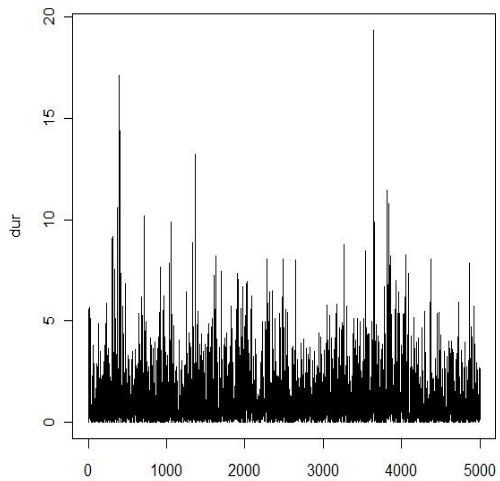
where $j = 1, \ldots 10$.

Since the all necessary theoretical moments are available, the parameters of the SCD model can be estimated using the GMM method. An R function can compute the necessary moments for $j = 1, \ldots 10$. and use these to obtained parameter estimates.

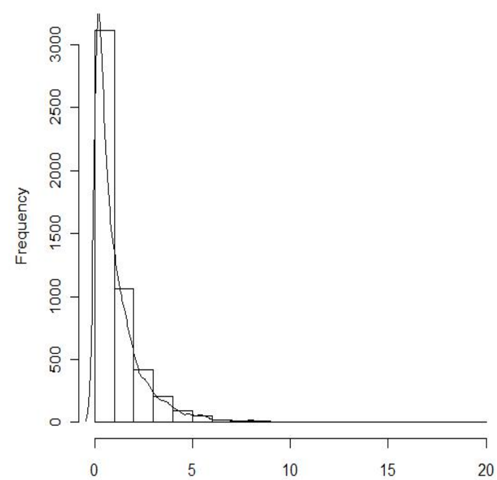
4.5 Simulation Study

For this simulation study, we simulated series of length 5000, which is considered to be representative of the typically large sample sizes that are associated with transaction data. Following the work of Knight & Ning (2008), we set $\Theta = (\omega,\ \beta,\ \gamma,\ \sigma_u)^t = (0.001\ 0.95\ 0.9\ 0.1)^t$. In order to make the model stationary and invertible, $|\beta|$ should be less than 1. The time plot and histogram of a typical observed Weibull SCD series is shown in Figure 4.1. In what follows, we refer to QML with Kalman filtering as method 1, QML with the filter introduced in Theorem 4.3.2 as method 2, GMM as method 3.

The histogram of the QML estimators using the Kalman filter obtained from 100 simulated series of length 5000 are shown in Figure 4.2. We did the same in Figure 4.3 and for the QML estimator using the filter introduced in Theorem 4.3.2, and in Figure 4.4 for the GMM estimator. From those histograms, it seems the sampling distributions are concentrated around the true values of the parameters, suggestting that the methods we used are reasonable. The result obtained in the GMM case are however suggesting the resulting estimators more biased. Comparing the three



(a) Time plot for Weibull SCD

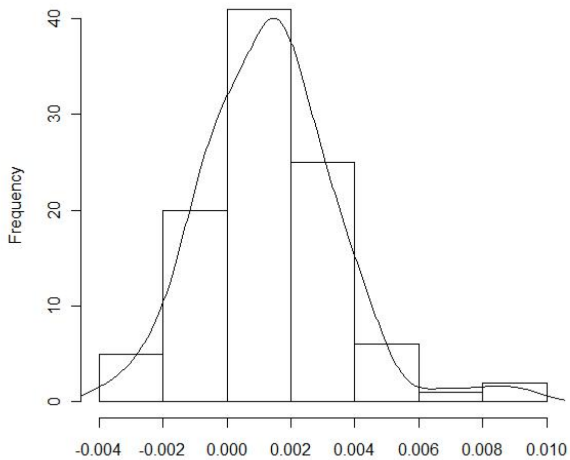


(b) Histogram for Weibull SCD

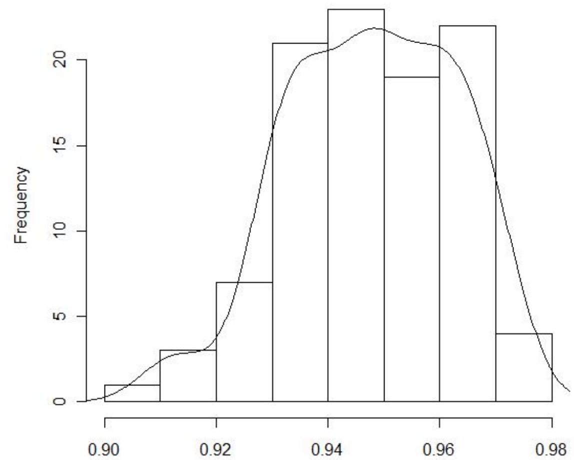
Figure 4.1: Simulated Weibull SCD series with 5000 observations.

methods based on these histograms, it seems that for Weillbull SCD model, method 2 is best for fitting the SCD model, and method 1 is better than method 3.

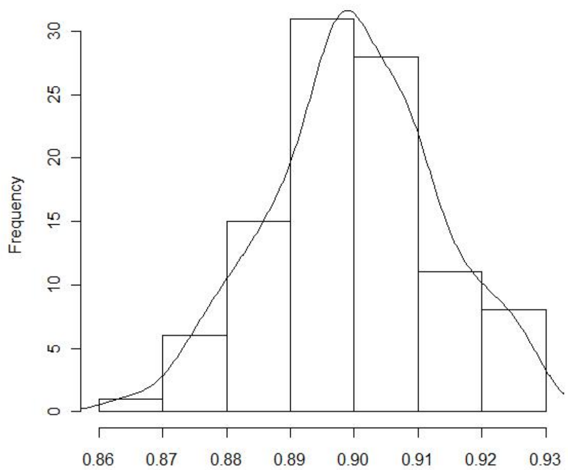
The estimation results based on the three methods discussed above are given in the Table 4.1. To help with the comparison, Table 4.2 gives the Average Squared-Error (ASE) (the sample equivalent of the traditional MSE) for the estimation of each parameter with each estimation method. From this, we see method 1 achieves lower ASE than all the other methods in the estimation of β, γ, and σ. Method 2 gives slightly smaller ASE than all the other method when estimating ω. Finally, note that for each parameter, the mean of all estimates fall within one standard error of the true parameter value.



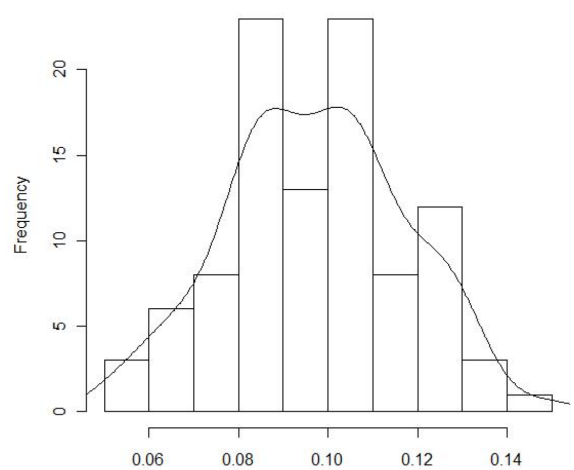
(a) Distribution of estimates ω=0.001



(b) Distribution of estimates β=0.95

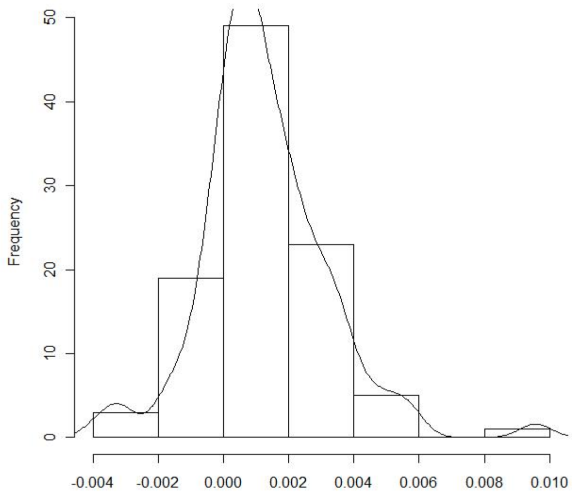


(c) Distribution of estimates γ=0.9

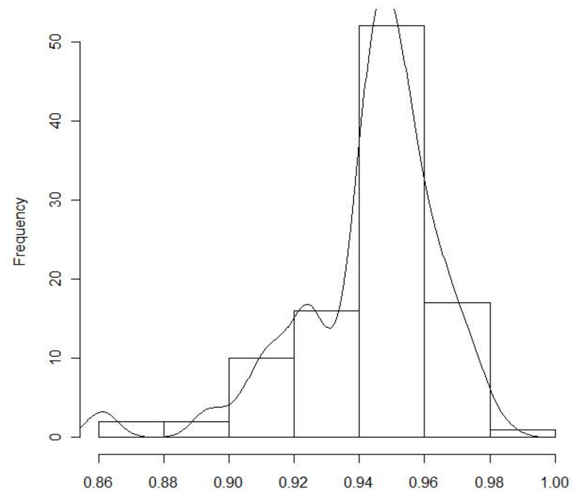


(d) Distribution of estimates σ=0.1

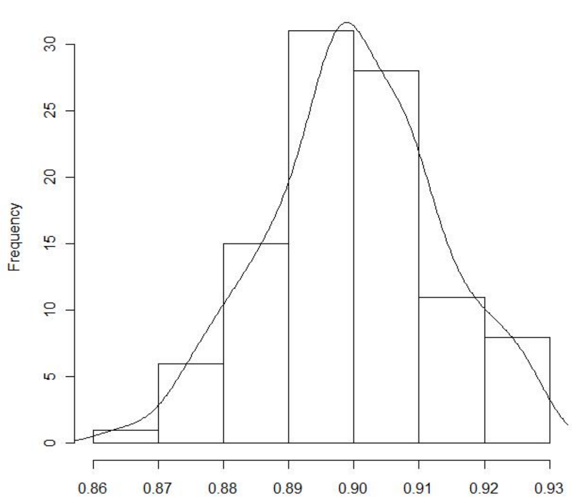
Figure 4.2: Histogram of QML estimators using the kalman filter.



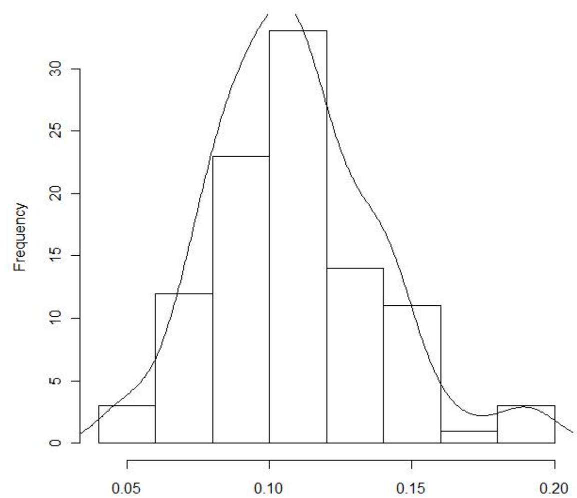
(a) Distribution of estimates ω=0.001



(b) Distribution of estimates β=0.95

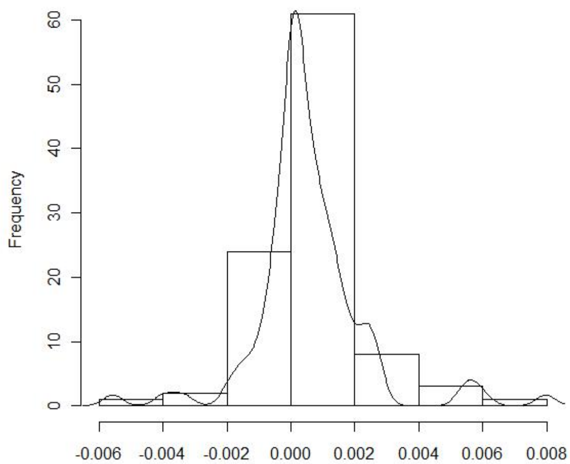


(c) Distribution of estimates γ=0.9

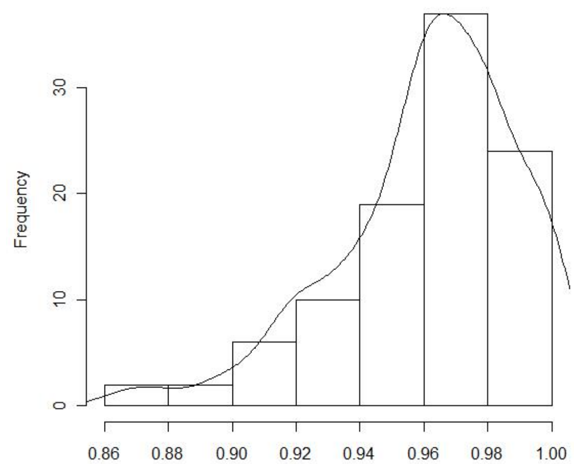


(d) Distribution of estimates σ=0.1

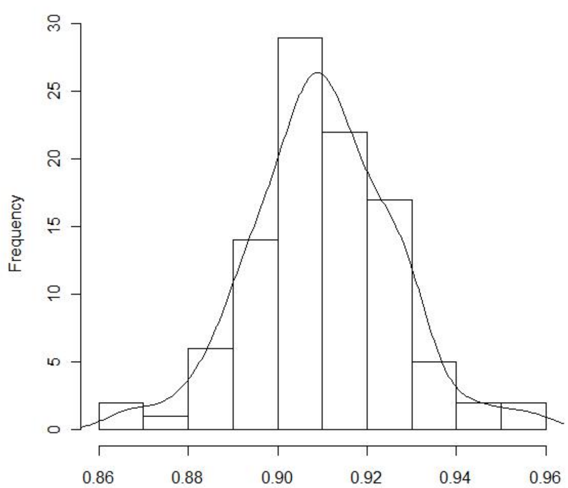
Figure 4.3: Histogram of QML estimators using nonlinear filter.



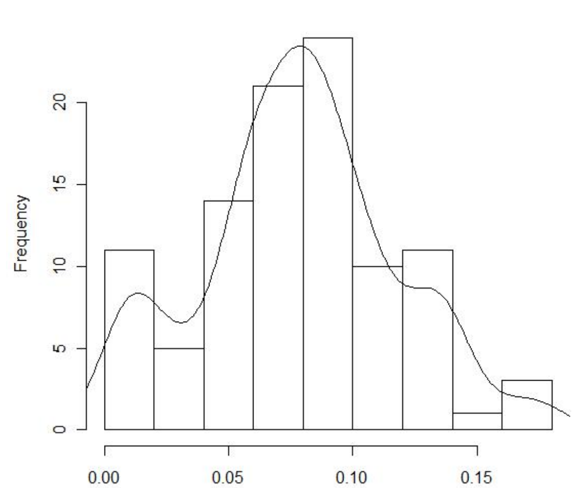
(a) Distribution of estimates ω=0.001



(b) Distribution of estimates β=0.95



(c) Distribution of estimates γ=0.9



(d) Distribution of estimates σ=0.1

Figure 4.4: Histogram of GMM estimators.

Table 4.1: Simulation Summary for the three Estimation Methods

parameters	True value	Method 1	Method 2	GMM
ω	0.001	0.001359561	0.001269947	0.0005527
		(0.002123128)	(0.001929450)	(0.001730760)
β	0.95	0.947831	0.9432045	0.9614191
		(0.01487210)	(0.02181952)	(0.02800412)
γ	0.9	0.9000105	0.9019865	0.9101544
		(0.01311689)	(0.01553524)	(0.01616926)
σ	0.1	0.09729156	0.1086845	0.05808528
		(0.01998328)	(0.02864012)	(0.04565775)

Table 4.2: Comparison of Average Squared Error

parameters	ASE		
	Method 1	Method 2	GMM
ω	0.002101897	0.001910156	0.001713452
β	0.01472338	0.02160132	0.02772408
γ	0.01298572	0.01537989	0.01600757
σ	0.01978345	0.02835372	0.04520117

4.6 Application of SCD Models in Finance

In this section, we will apply the SCD model to the IBM data. This data contains information on IBM trades from November 1, 1990 to January 31, 1991. Specifically, the data set includes data/time, volume, bid and ask price, and transaction data. We only use the data from the November 1, 1990 to December 21, 1990 in order to avoid holiday effects. This data set includes 35 trading days over 2 months, with 31668 observations. However, since IBM trading halted for over one hour and 15 minutes on November 23, the data from this day is deleted. Furthermore, we delete the trades that occurred before 9:50 am and after 4:00 pm in order to eliminate irregularities during opening and closing periods. This leads to a set of "working data" containing 25542 observations, with an observed sample mean of the adjusted duration of 2.888828 and a standard deviation of 3.644924. Note that the maximum observation is 47.31872 and the minimum is 0.079026. The three methods described previously were used to fit a Weibull SCD model to the IBM data. The resulting estimates of the model parameters and their standard error are shown in Table 4.3.

All estimates of β are close to one, showing high persistency of the duration process. Also, the fact that all estimates are less than one suggests the process is stationary. The estimates of σ_u are all significantly different from zero. According to Bauwens & Veredas (2004) and Knight & Ning (2008), the dispersion ratio, which we denote D_y, should help detect overdispersion in the observed data. This ratio is defined as $D_y = \frac{\sigma_y}{\mu_y}$, the standard deviation of durations divided by the mean duration. Based on the Weibull SCD model we have here that

$$D_y = \sqrt{\frac{\Gamma(1+\frac{2}{\gamma})}{(\Gamma(1+\frac{1}{\gamma}))^2}\exp(\frac{\sigma^2}{1-\beta^2}) - 1}.$$

We estimate D_y by substituting estimated parameters in the above formula. This "plug-in" approach leads to the model-based estimates of the dispersion ratio given in Table 4.3. From these results, it is reasonable to conclude that the true D_y is greater than 1, meaning that overdispersion is present here. Also, it seems trade durations exhibit persistency. Here, we use the Hessian matrix to obtain the standard errors of all estimates. It turns out that the inverse of the Hessian matrix approximates the variance/covariance matrix of parameter estimates. The Hessian matrix (and asymptotic standard errors for the parameters) can be computed via finite difference approximations. This procedure yields very precise asymptotic standard errors for all estimation methods. In the present case, where the sample size is finite, these are used as approximations to the unavailable finite sample standard errors.

Table 4.3: Estimates for both QML methods and GMM

parameter	QML based on Kalman filter	QML based on nonlinear filter	GMM Estimates
ω	0.02146828 (0.002598241)	0.02153400 (0.002611913)	0.0084164 (0.01554)
β	0.97592862 (0.002846242)	0.97594709 (0.002850144)	0.9908986 (0.01678)
γ	1.05149445 (0.005547853)	1.04415237 (0.005024234)	0.8917806 (0.00807)
σ	0.11526654 (0.007257183)	0.09865976 (0.006524264)	0.0467896 (0.04422)
Estimated Dispersion Ratio	1.232487	1.163427	1.246262

Chapter 5

Quadratic SCD Models

Kawakatu (2007) considered quadratic SV models in an attempt to capture the heavy-tailed property of asset returns. We do the same here for Conditional Durations.

5.1 Definition and Properties

Here, we introduce another type of nonlinear stochastic volatility models, the Quadratic Stochastic Conditional Duration (QSCD) model. The model is defined as

$$\begin{aligned} y_t &= \Psi_t \epsilon_t, \\ &= \exp(a\psi_t^2 + b\psi_t + c)\epsilon_t, \end{aligned} \tag{5.1}$$

where Ψ_t is again a latent variable, and

$$\psi_t = \beta\psi_{t-1} + \omega + u_t. \tag{5.2}$$

In general, $\epsilon_t > 0$, $u_t \in R$, are the two innovation sequences and are uncorrelated. In what follows, we focus on the case where $\epsilon_t \sim$ Weibull $(\alpha, 1)$ and $u_t \sim N(0, \sigma_u^2)$ and $\{\epsilon_t\}$ and $\{u_t\}$ are independent. Under the normality assumption on u_t, it follows that ψ_t is a stationary Gaussian AR(1) process whenever $|\beta| < 1$. More specifically, $\psi_t \sim N(\frac{\omega}{1-\beta}, \frac{\sigma_\mu^2}{1-\beta^2})$ and hence, the moments are given by

$$E(\psi_t) = \frac{\omega}{1-\beta},$$

$$E(\psi_t^2) = Var(\psi_t) + [E(\psi_t)]^2 = \frac{\sigma_u^2}{1-\beta^2} + \left(\frac{\omega}{1-\beta}\right)^2,$$

and

$$E(\psi_t^4) = 3\left(\frac{\sigma_u^2}{1-\beta^2}\right)^2.$$

The following two theorems are important in order to understand the properties of QSCD models.

Theorem 5.1.1. *The latent variable of model (5.1) has mean*

$$\mu_\Psi \quad = \quad \sqrt{\frac{1-\beta^2}{1-\beta^2-2a\sigma_u^2}} \exp\left\{\frac{(\omega+\beta\omega+b\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-2a\sigma_u^2)} - \frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + c\right\},$$

and variance

$$\sigma_\Psi^2 = \sqrt{\frac{1-\beta^2}{1-\beta^2-4a\sigma_u^2}} \exp\left\{ \frac{(\omega+\beta\omega+2b\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-4a\sigma_u^2)} - \frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + 2c \right\} - \mu_\Psi^2.$$

Theorem 5.1.2. *For the model given by (5.1) and (5.2), the moments of the duration y_t are*

$$\begin{aligned} E[y_t^m] &= E\{\exp[m(a\psi_t^2+b\psi_t+c)]\epsilon_t^m\} = E\{\exp[m(a\psi_t^2+b\psi_t+c)]\}E[\epsilon_t^m] \\ &= \sqrt{\tfrac{1-\beta^2}{1-\beta^2-2am\sigma_u^2}} \exp\left[\frac{(\omega+\beta\omega+mb\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-2am\sigma_u^2)} \right. \\ &\quad \left. -\frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + mc \right] \Gamma(1+m/\gamma). \end{aligned}$$

In addition, the cross-moments of durations y_t and y_{t-r} are

$$E(y_t y_{t-r}) = \frac{1}{\sqrt{1-2a\sigma_Z^2}} \frac{1}{\sqrt{(1-2A\sigma_\psi^2)}} \exp(B) \exp\left\{ \frac{C^2\sigma_\psi^2}{2(1-2A\sigma_\psi^2)} \right\} (\Gamma(1+1/\gamma))^2,$$

where

$$\begin{aligned} A &= a(\beta^{2r}+1) + 4\lambda a^2 \beta^{2r}, \\ B &= \frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2} + b\omega\frac{1-\beta^r}{1-\beta} + 2c + \lambda\left(b + \frac{2a\omega(1-\beta^r)}{1-\beta}\right)^2, \\ C &= 2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta} + b(\beta^r+1) + 4\lambda a\beta^r\left(b + \frac{2a\omega(1-\beta^r)}{1-\beta}\right), \end{aligned}$$

$\sigma_Z^2 = \dfrac{1-\beta^{2r-1}}{1-\beta^2}\sigma_u^2$ *and* $\sigma_\psi^2 = \dfrac{\sigma_\mu^2}{1-\beta^2}$. *In order for these moments and cross-moments to exist, the following conditions should be satisfied*

$$\frac{1-\beta^2}{1-\beta^2-8a\sigma_u^2} > 0, \tag{5.3}$$

$$1-2a\sigma_Z^2 > 0, \tag{5.4}$$

$$1-2A\sigma_\psi^2 > 0. \tag{5.5}$$

Proof. For the moments of the duration y_t, we use again the fact that ψ_t is normal when $|\beta| < 1$. Doing this, we have

$$\begin{aligned} E(y_t^m) &= E\{\exp[m(a\psi_t^2 + b\psi_t + c)]\} \\ &= \int_R \frac{1}{\sqrt{2\pi\frac{\sigma_u^2}{1-\beta^2}}}\exp\left[-\frac{(\psi_t - \frac{\omega}{1-\beta})^2}{\frac{2\sigma_u^2}{1-\beta^2}}\right]\exp[m(a\psi_t^2 + b\psi_t + c)]d\psi_t \\ &= \frac{1}{\sqrt{\frac{2\pi\sigma_u^2}{1-\beta^2}}}\int_R \exp\left[-\frac{\psi_t^2 - \frac{2\omega\psi_t}{1-\beta} + \frac{\omega^2}{(1-\beta)^2}}{\frac{2\sigma_u^2}{1-\beta^2}} + am\psi_t^2 + bm\psi_t + mc\right]d\psi_t \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\frac{2\pi\sigma_u^2}{1-\beta^2}}} \int_R \exp\Bigg[- \frac{(\psi_t - \frac{\omega+\omega\beta+bm\sigma_u^2}{1-\beta^2-2am\sigma_u^2})^2}{\frac{2\sigma_u^2}{1-\beta^2-2am\sigma_u^2}} \exp[\frac{(\omega+\beta\omega+bm\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-2am\sigma_u^2)} - \frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + mc]\Bigg] d\psi_t \\
&= \frac{\sqrt{\frac{2\pi\sigma_u^2}{1-\beta^2-2am\sigma_u^2}}}{\sqrt{\frac{2\pi\sigma_u^2}{1-\beta^2}}} \exp\Bigg[\frac{(\omega+\beta\omega+bm\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-2am\sigma_u^2)} - \frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + mc\Bigg] \\
&\quad \times \int_R \frac{1}{\sqrt{\frac{2\pi\sigma_u^2}{1-\beta^2-2am\sigma_u^2}}} \exp\Bigg[- \frac{(\psi_t - \frac{\omega+\omega\beta+bm\sigma_u^2}{1-\beta^2-2am\sigma_u^2})^2}{\frac{2\sigma_u^2}{1-\beta^2-2am\sigma_u^2}}\Bigg] d\psi_t \\
&= \sqrt{\frac{1-\beta^2}{1-\beta^2-2am\sigma_u^2}} \exp\Bigg[\frac{(\omega+\beta\omega+bm\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-2am\sigma_u^2)} - \frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + mc\Bigg].
\end{aligned}
$$

Now, recall that

$$
\begin{aligned}
\psi_t &= \beta^r \psi_{t-r} + \omega \tfrac{1-\beta^r}{1-\beta} + \sum_{j=0}^{r-1} \beta^j u_{t-j} \\
&= \beta^r \psi_{t-r} + \omega \tfrac{1-\beta^r}{1-\beta} + Z_r,
\end{aligned}
$$

where

$$
Z_r = \sum_{j=0}^{r-1} \beta^j u_{t-j} \sim N\bigg(0, \Big(\frac{1-\beta^{2r-1}}{1-\beta^2}\Big)\sigma_u^2\bigg),
$$

and the fact that Z_r is independent of $\{\epsilon_t\}$ and of ψ_{t-j} for $j = r, r+1, \ldots$ Similarly, we can also get

$$\begin{aligned}
\psi_t^2 &= \left\{\beta^r\psi_{t-r} + \omega\frac{1-\beta^r}{1-\beta} + Z_r\right\}^2 \\
&= \beta^{2r}\psi_{t-r}^2 + \frac{\omega^2(1-\beta)^2}{(1-\beta^2)} + Z_r^2 + 2\omega\frac{1-\beta^r}{1-\beta}\beta^r\psi_{t-r} + \\
&\quad 2\beta^r\psi_{t-r}Z_r + 2\omega\frac{1-\beta^r}{1-\beta}Z_r.
\end{aligned}$$

We can now calculate the cross effect between y_t and y_{t-r} where $r = 1, \ldots, 10$.

Using the two expressions obtained above, we can write

$$\begin{aligned}
\psi_t + \psi_{t-r} &= \beta^r\psi_{t-r} + \psi_{t-r} + \omega\frac{1-\beta^r}{1-\beta} + Z_r \\
&= (\beta^r + 1)\psi_{t-r} + \omega\frac{1-\beta^r}{1-\beta} + Z_r
\end{aligned} \tag{5.6}$$

and

$$\begin{aligned}
\psi_t^2 + \psi_{t-r}^2 &= \beta^{2r}\psi_{t-r}^2 + \psi_{t-r}^2 + \frac{\omega^2(1-\beta^r)^2}{(1-\beta^2)} + Z_r^2 \\
&\quad +2\frac{\omega\beta^r(1-\beta^r)}{1-\beta}\psi_{t-r} + 2\beta^r\psi_{t-r}Z_r + 2\frac{\omega(1-\beta^r)}{1-\beta}Z_r \\
&= (\beta^{2r}+1)\psi_{t-r}^2 + 2\frac{\omega\beta^r(1-\beta^r)}{1-\beta}\psi_{t-r} + \frac{\omega^2(1-\beta^r)^2}{(1-\beta^2)} \\
&\quad +Z_r^2 + \left(2\beta^r\psi_{t-r} + \frac{2\omega(1-\beta^r)}{1-\beta}\right)Z_r
\end{aligned} \tag{5.7}$$

Combining (5.6) and (5.7), we get

$$
\begin{aligned}
&a(\psi_t^2+\psi_{t-r}^2)+b(\psi_t+\psi_{t-r})+2c\\
&=a(\beta^{2r}+1)\psi_{t-r}^2+2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta}\psi_{t-r}+\frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2}\\
&\qquad+aZ_r^2+a\left(2\beta^r\psi_{t-r}+\frac{2\omega(1-\beta^r)}{1-\beta}\right)Z_r+b(\beta^r+1)\psi_{t-r}+b\omega\frac{1-\beta^r}{1-\beta}+bZ_r+2c\\
&=a(\beta^{2r}+1)\psi_{t-r}^2+\left[2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta}+b(\beta^r+1)\right]\psi_{t-r}+\frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2}\\
&\qquad+b\omega\frac{1-\beta^r}{1-\beta}+2c+aZ_r^2+\left[b+2a\beta^r\psi_{t-r}+\frac{2a\omega(1-\beta^r)}{1-\beta}\right]Z_r
\end{aligned}
$$

For the cross-moments of durations, we can then write

$$
\begin{aligned}
E(y_ty_{t-r}) &= E\{\exp[a(\psi_t^2+\psi_{t-r}^2)+b(\psi_t+\psi_{t-r})+2c]\}E(\epsilon_t)E(\epsilon_{t-r})\\
&= E\Bigg\{\exp\Bigg[a(\beta^{2r}+1)\psi_{t-r}^2+\left[2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta}+b(\beta^r+1)\right]\psi_{t-r}\\
&\quad+\frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2}+b\omega\frac{1-\beta^r}{1-\beta}+2c+aZ_r^2\\
&\quad+\left(b+2a\beta^r\psi_{t-r}+\frac{2a\omega(1-\beta^r)}{1-\beta}\right)Z_r\Bigg]\Bigg\}E(\epsilon_t)E(\epsilon_{t-r})
\end{aligned}
\tag{5.8}
$$

From Ghahramani & Thavaneswaran (2010), we know that

$$
E[e^{r(aY+bY^2)}]=\frac{1}{\sqrt{1-2br\sigma_Y^2}}\exp\frac{r^2a^2\sigma_Y^2}{2(1-2b\sigma_Y^2)}
$$

when $b < \dfrac{1}{2r\sigma_Y^2}$. From (5.8), we know

$$E\left\{ \exp[aZ_{t,r}^2 + (b + 2a\beta^r\psi_{t-r} + \frac{2a\omega(1-\beta^r)}{1-\beta})Z_{t,r}]|F_t^\psi \right\}$$

$$= \frac{1}{\sqrt{1-2a\sigma_Z^2}} \exp\left\{ \frac{\sigma_Z^2}{2(1-2a\sigma_Z^2)} \left(b + 2a\beta^r\psi_{t-r} + \frac{2a\omega(1-\beta^r)}{1-\beta} \right)^2 \right\}$$

and $1 > 2a\sigma_Z^2$ then $a < \dfrac{1}{\sigma_Z^2}$. Here we let $\lambda = \dfrac{1}{\sqrt{1-2a\sigma_Z^2}}$, then (5.8) can be written

$$\begin{aligned}
&\frac{1}{\sqrt{1-2a\sigma_Z^2}} \exp\left\{ \frac{\lambda^2\sigma_Z^2}{2} \left(b + 2a\beta^r\psi_{t-r} + \frac{2a\omega(1-\beta^r)}{1-\beta} \right)^2 \right\} \\
&= \frac{1}{\sqrt{1-2a\sigma_Z^2}} \exp\left\{ \frac{\lambda^2\sigma_Z^2}{2} \left(b + \frac{2a\omega(1-\beta^r)}{1-\beta} \right)^2 + 2\lambda^2 a^2 \sigma_Z^2 \beta^{2r} \psi_{t-r}^2 \right. \\
&\left. + 2a\lambda^2\sigma_Z^2\beta^r \left(b + \frac{2a\omega(1-\beta^r)}{1-\beta} \right) \psi_{t-r} \right\}
\end{aligned}$$

From (5.8), we know

$$\begin{aligned}
&E\left\{ E\left\{ \exp\left[a(\beta^{2r}+1)\psi_{t-r}^2 + \left(2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta} + b(\beta^r+1) \right)\psi_{t-r} + \right.\right.\right. \\
&\frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2} + b\omega\frac{1-\beta^r}{1-\beta} + 2c + a[Z_{t,r}]^2 + \\
&\left.\left.\left. \left(b + 2a\beta^r\psi_{t-r} + \frac{2a\omega(1-\beta^r)}{1-\beta} \right) Z_{t,r} \right] \right\} |F_t^\psi \right\} E(\epsilon_t)E(\epsilon_{t-r})
\end{aligned}$$

$$
\begin{aligned}
&= E\Bigg\{ \exp\Bigg[a(\beta^{2r}+1)\psi_{t-r}^2 \ + \Bigg(2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta} + b(\beta^r+1)\Bigg)\psi_{t-r} + \\
&\quad \frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2} + b\omega\frac{1-\beta^r}{1-\beta} + 2c\Bigg]\Bigg\}\times \\
&\quad \frac{1}{\sqrt{1-2a\sigma_Z^2}}\exp\left\{\lambda\Bigg(b+\frac{2a\omega(1-\beta^r)}{1-\beta}\Bigg)^2 + 4\lambda a^2\beta^{2r}\psi_{t-r}^2 + 4\lambda a\beta^r(b+\frac{2a\omega(1-\beta^r)}{1-\beta})\psi_{t-r}\right\} \\
&= \frac{1}{\sqrt{1-2a\sigma_Z^2}}E\Bigg\{ \exp\Bigg[a(\beta^{2r}+1)+4\lambda a^2\beta^{2r}\Bigg]\psi_{t-r}^2 + \\
&\quad \Bigg[2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta} + b(\beta^r+1) + 4\lambda a\beta^r\Bigg(b+\frac{2a\omega(1-\beta^r)}{1-\beta}\Bigg)\Bigg]\psi_{t-r} + \\
&\quad \frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2} + b\omega\frac{1-\beta^r}{1-\beta} + 2c + \lambda\Bigg(b+\frac{2a\omega(1-\beta^r)}{1-\beta}\Bigg)^2\Bigg\}
\end{aligned}
$$

Let

$$
\begin{aligned}
A &= a(\beta^{2r}+1)+4\lambda a^2\beta^{2r}, \\
B &= \frac{a\omega^2(1-\beta^r)^2}{(1-\beta)^2} + b\omega\frac{1-\beta^r}{1-\beta} + 2c + \lambda\Bigg(b+\frac{2a\omega(1-\beta^r)}{1-\beta}\Bigg)^2, \\
C &= 2\frac{a\omega\beta^r(1-\beta^r)}{1-\beta} + b(\beta^r+1) + 4\lambda a\beta^r\Bigg(b+\frac{2a\omega(1-\beta^r)}{1-\beta}\Bigg).
\end{aligned}
$$

Then, we can write

$$
\begin{aligned}
&E(y_t y_{t-r}) \\
&= \frac{1}{\sqrt{1-2a\sigma_Z^2}}\exp(B)\times E\Bigg[\exp(A\psi_{t-r}^2)+c\psi_{t-r}\Bigg]E(\epsilon_t)E(\epsilon_{t-r}) \\
&= \frac{1}{\sqrt{1-2a\sigma_Z^2}}\frac{1}{\sqrt{(1-2A\sigma_\psi^2)}}\exp(B)\exp\left[\frac{C^2\sigma_\psi^2}{2(1-2A\sigma_\psi^2)}\right]\Bigg(\Gamma(1+1/\gamma)\Bigg)^2,
\end{aligned}
$$

as claimed. $\square$

5.2 QML Estimation based on Nonlinear Filtering

As we have done in Chapter 4, we will use a logarithmic transformation to estimate the parameters of the QSCD model. For this, we rewrite (5.1) as

$$
\begin{aligned}
\ln y_t &= a\psi_t^2 + b\psi_t + c + \ln \epsilon_t \\
&= a\psi_t^2 + b\psi_t + c^* + \xi_t, \qquad (5.9)
\end{aligned}
$$

where as before, we use $x_t = \log y_t$, $\xi_t = \ln \epsilon_t - \mu$, $\mu = E(\ln \epsilon_t)$ and $c^* = c + \mu$. From previous result, we know that $\ln \epsilon_t$ follows the extreme value distribution $(0, \dfrac{\pi^2}{6\alpha^2})$. From (5.2), we can get

$$\psi_t^2 = (\beta\psi_{t-1} + \omega)^2 + e_t + \sigma_u^2,$$

where $e_t = 2(\beta\psi_{t-1} + \omega)u_t + u_t^2 - \sigma_u^2$ is such that $E(e_t) = 0$ and

$$
\begin{aligned}
E(e_t^2) &= E[u_t^4 + 4(\beta\psi_{t-1} + \omega)^2 u_t^2 + 4(\beta\psi_{t-1} + \omega)u_t^3] \\
&\quad + \sigma_u^4 - 2\sigma_u^2 E[2(\beta\psi_{t-1} + \omega)u_t + u_t^2]
\end{aligned}
$$

$$
\begin{aligned}
&= 3\sigma_u^4 + 4E[(\beta^2\psi_{t-1}^2 + \omega^2 + 2\beta\omega\psi_{t-1})u_t^2] + \sigma_u^4 - 2\sigma_u^4 \\
&= 2\sigma_u^4 + 4\sigma_u^2[\beta^2(\tfrac{\sigma_u^2}{1-\beta^2} + \tfrac{\omega^2}{(1-\beta)^2}) + \omega^2 + \tfrac{2\beta\omega^2}{1-\beta}] \\
&= \frac{2+2\beta^2}{1-\beta^2}\sigma_u^4 + \frac{4\omega^2\sigma_u^2}{(1-\beta)^2} = \tau,
\end{aligned}
$$

and is uncorrelated with ψ_{t-1}.

Using the same method of minimizing the prediction MSE, $\gamma_t = E[\psi_{t-1}^2 - \hat{\psi}_{t-1}^2]^2$ as that presented in Section 4.3, the optimal recursive filtering scheme is given in the next theorem.

Theorem 5.2.1. *In the current context, an MSE optimal sequential estimator of ψ_t^2 is*

$$
\hat{\psi}_t^2 = (\beta\hat{\psi}_{t-1} + \omega)^2 + G_t(x_t - a\hat{\psi}_{t-1}^2 - b\hat{\psi}_{t-1} - c^*),
$$

with conditional MSE

$$
\begin{aligned}
\gamma_t &= (\beta^2 - aG_t)^2\gamma_{t-1} + (2\omega\beta - bG_t)^2\left(\tfrac{\sigma_\mu^2}{1-\beta^2} + (\tfrac{\omega}{1-\beta})^2 - 2(\tfrac{\omega}{1-\beta})\hat{\psi}_{t-1} + \hat{\psi}_{t-1}^2\right) \\
&\quad + 2(\beta^2 - aG_t)(2\omega\beta - bG_t)\hat{\psi}_{t-1}\left(\hat{\psi}_{t-1}^2 - \tfrac{\omega}{1-\beta}\hat{\psi}_{t-1} - (\tfrac{\sigma_\mu^2}{1-\beta^2} + (\tfrac{\omega}{1-\beta})^2)\right) \\
&\quad + \tau + G_t^2\frac{\pi^2}{6\alpha^2},
\end{aligned}
$$

where $G_t = \dfrac{A_t}{B_t}$,

$$\begin{aligned} A_t &= 2\beta^2 a\gamma_{t-1} + 4\omega\beta b(\tfrac{\sigma_u^2}{1-\beta^2} + \tfrac{\omega^2}{(1-\beta)^2} - 2\tfrac{\omega}{1-\beta}\hat{\psi}_{t-1} + \hat{\psi}_{t-1}^2) \\ &\quad +2(\beta^2 b + 2a\beta\omega)\left(\hat{\psi}_{t-1}^3 - (\tfrac{\omega}{1-\beta})\hat{\psi}_{t-1}^2 - (\tfrac{\sigma_u^2}{1-\beta^2} + \tfrac{\omega^2}{(1-\beta)^2})\hat{\psi}_{t-1}\right), \end{aligned}$$

and

$$\begin{aligned} B_t &= 2a^2\gamma_{t-1} + 2b^2(\tfrac{\sigma_\mu^2}{1-\beta^2} + (\tfrac{\omega}{1-\beta})^2 - 2\tfrac{\omega}{1-\beta}\psi_{t-1} + \psi_{t-1}^2) \\ &\quad +2\tfrac{\pi^2}{6\alpha^2} + 4ab\hat{\psi}_{t-1}(\hat{\psi}_{t-1}^2 - \tfrac{\omega}{1-\beta}\hat{\psi}_{t-1} - (\tfrac{\sigma_\mu^2}{1-\beta^2} + (\tfrac{\omega}{1-\beta})^2)). \end{aligned}$$

Proof. First, proceeding as in Thavaneswaran & Gong (2009), we consider conditional predictions of the form

$$\hat{\psi}_t^2 = (\beta\hat{\psi}_{t-1} + \omega)^2 + G_t(x_t - a\hat{\psi}_{t-1}^2 - b\hat{\psi}_{t-1} - c^*)$$

where the choice of G_t is made by minimizing the prediction MSE.

Here, we write

$$\begin{aligned} (\psi_t^2 - \hat{\psi}_t^2)^2 &= \Big[(\beta\psi_{t-1} + \omega)^2 + e_t - (\beta\hat{\psi}_{t-1} + \omega)^2 + G_t(\log y_t - \hat{\psi}_{t-1} - c)\Big]^2 \\ &= \Big[(\beta\psi_{t-1} + \omega)^2 - (\beta\hat{\psi}_{t-1} + \omega)^2 + e_t \\ &\qquad +G_t(\xi_t + a\psi_{t-1}^2 + b\psi_{t-1} + c^* - a\hat{\psi}_{t-1}^2 - b\hat{\psi}_{t-1} - c^*)\Big]^2 \end{aligned}$$

$$
\begin{aligned}
&= \Big[\beta(\psi_{t-1} - \hat{\psi}_{t-1})(\beta(\psi_{t-1} + \hat{\psi}_{t-1}) + 2\omega) + e_t \\
&\quad + G_t(\xi_t + a(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2) + b(\psi_{t-1} - \hat{\psi}_{t-1}))\Big]^2 \\
&= \Big[(\beta^2 - aG_t)(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2) + (2\omega\beta - bG_t)(\psi_{t-1} - \hat{\psi}_{t-1}) + e_t - G_t\xi_t\Big]^2 \\
&= (\beta^2 - aG_t)^2(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)^2 + (2\omega\beta - bG_t)^2(\psi_{t-1} - \hat{\psi}_{t-1})^2 + e_t^2 + G_t^2\xi_t^2 \\
&\quad + 2(\beta^2 - aG_t)(2\beta\omega - bG_t)(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)(\psi_{t-1} - \hat{\psi}_{t-1}) + 2(\beta^2 - aG_t)(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)e_t \\
&\quad - 2(\beta^2 - aG_t)G_t\xi_t(\psi_{t-1}^2 - \hat{\psi}_{t-}^2) + 2(2\omega\beta - bG_t)(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)e_t \\
&\quad - 2(2\omega\beta - bG_t)(\psi_{t-1} - \hat{\psi}_{t-1})G_t\xi_t - 2G_t\xi_t e_t
\end{aligned}
$$

From this, we further have

$$
\begin{aligned}
\gamma_t &= E[(\psi_t^2 - \hat{\psi}_t^2)^2 | \mathcal{F}_{t-1}^y] \\
&= (\beta^2 - aG_t)^2\gamma_{t-1} + (2\omega\beta - bG_t)^2 E[\psi_{t-1} - \hat{\psi}_{t-1}]^2 + E(e_t^2) + G_t^2 E(\xi_t^2) \\
&\quad + 2(\beta^2 - aG_t)(2\beta\omega - bG_t)E[(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)(\psi_{t-1} - \hat{\psi}_{t-1})] \\
&\quad + 2(\beta^2 - aG_t)E[(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)e_t] \\
&\quad + 2(2\omega\beta - bG_t)E[(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)e_t] \\
&= A + B + C + D_1 + D_2
\end{aligned}
$$

where

$$
\begin{aligned}
A &= (\beta^2 - aG_t)^2\gamma_{t-1} + E(e_t^2) + G_t E(\xi_t^2) \\
&= (\beta^2 - aG_t)^2\gamma_{t-1} + \tau + G_t E(\xi_t^2),
\end{aligned}
$$

$$
\begin{aligned}
B &= (2\omega\beta - bG_t)^2 E[\psi_{t-1}^2 - 2\psi_{t-1}\hat{\psi}_{t-1} + \hat{\psi}_{t-1}^2] \\
&= (2\omega\beta - bG_t)^2 \left[\left(\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2}\right) - \frac{2\omega}{1-\beta}\hat{\psi}_{t-1} + \hat{\psi}_{t-1}^2\right], \\
&= (4\omega^2\beta^2 - 4\omega\beta bG_t + b^2G_t^2)\left[\left(\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2}\right) - \frac{2\omega}{1-\beta}\hat{\psi}_{t-1} + \hat{\psi}_{t-1}^2\right], \\
C &= 2(\beta^2 - aG_t)(2\beta\omega - bG_t)E[(\psi_{t-1}^3 - \psi_{t-1}^2\hat{\psi}_{t-1} - \psi_{t-1}\hat{\psi}_{t-1}^2 + \hat{\psi}_{t-1}^3)] \\
&= 2(\beta^2 - aG_t)(2\beta\omega - bG_t)\left[\hat{\psi}_{t-1}^3 - (\frac{2\omega}{1-\beta})\hat{\psi}_{t-1}^2 - (\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2})\hat{\psi}_{t-1}\right], \\
D_1 &= 2(\beta^2 - aG_t)E[(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)e_t] = 0, \\
D_2 &= 2(2\omega\beta - bG_t)E[(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)e_t] = 0,
\end{aligned}
$$

so that

$$
\begin{aligned}
\gamma_t &= (\beta^4 - 2\beta^2 aG_t + a^2G_t^2)\gamma_{t-1} + \tau + G_t^2 E(\xi_t^2) \\
&\quad + (4\omega^2\beta^2 - 4\omega\beta bG_t + b^2G_t^2)\left[\left(\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2}\right) - \frac{2\omega}{1-\beta}\hat{\psi}_{t-1} + \hat{\psi}_{t-1}^2\right] \\
&\quad + 2(\beta^2 - aG_t)(2\beta\omega - bG_t)\left[\hat{\psi}_{t-1}^3 - (\frac{2\omega}{1-\beta})\hat{\psi}_{t-1}^2 - (\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2})\hat{\psi}_{t-1}\right]
\end{aligned}
$$

It is now straightforward to minimize γ_t with respect to G_t since the former is a degree two polynomial in the latter. This leads to the stated result. □

In order to estimate the model parameters, the filter introduced above can be applied to compute a quasi-likelihood function for the model given by (5.2) and (5.9) as we did previously in Chapter 4 for the regular SCD model. Specifically, we find the value of θ maximizing the quasi-likelihood

$$\ln L(\theta) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{N}\ln F_t - \frac{1}{2}\sum_{t=1}^{N}\frac{v_t^2}{F_t},$$

where $v_t = x_t - \hat{x}_{t|t-1}$, $\hat{x}_{t|t-1}$ is the conditional mean $E[x_t|\mathcal{F}_{t-1}^y]$,

$$v_t = x_t - [a\hat{\psi}_{t-1}^2 + b\hat{\psi}_{t-1} + c^*]$$

and F_t is the conditional variance of v_t, that is

$$\begin{aligned} F_t &= Var(v_t|\mathcal{F}_{t-1}^y) \\ &= E[v_t^2|\mathcal{F}_{t-1}^y] - \left(E[v_t|\mathcal{F}_{t-1}^y]\right)^2. \end{aligned}$$

Note that (5.2) implies

$$\begin{aligned} v_t &= a\psi_{t-1}^2 + b\psi_{t-1} + c^* + \xi_t - \left(a\hat{\psi}_{t-1}^2 + b\hat{\psi}_{t-1} + c^*\right) \\ &= a(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2) + b(\psi_{t-1} - \hat{\psi}_{t-1}) + \xi_t. \end{aligned}$$

and

$$\begin{aligned} v_t^2 &= a^2(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)^2 + b^2(\psi_{t-1} - \hat{\psi}_{t-1})^2 + \xi_t^2 + 2ab(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)(\psi_{t-1} - \hat{\psi}_{t-1}) \\ &\quad +2a(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)\xi_t + 2b(\psi_{t-1} - \hat{\psi}_{t-1})\xi_t. \end{aligned}$$

From this, we have that

$$E(v_t^2|\mathcal{F}_{t-1}^y) = a^2\gamma + b^2E(\psi_{t-1} - \hat{\psi}_{t-1})^2 + \frac{\pi^2}{6\alpha^2} + 2abE(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2)(\psi_{t-1} - \hat{\psi}_{t-1}),$$

and

$$E(v_t|\mathcal{F}_{t-1}^y) = aE(\psi_{t-1}^2 - \hat{\psi}_{t-1}^2) + bE(\psi_{t-1} - \hat{\psi}_{t-1}),$$

implying(5.2) can be rewritten as

$$\begin{aligned} Var(v_t|\mathcal{F}_{t-1}^y) &= a^2\gamma + b^2\left(\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2} + \hat{\psi}_{t-1}^2 - 2\frac{\omega\hat{\psi}_{t-1}}{1-\beta}\right) \\ &+\frac{\pi^2}{6\alpha^2} - a^2\left(\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2} - \hat{\psi}_{t-1}^2\right)^2 \\ &-b^2\left(\frac{\omega}{1-\beta} - \hat{\psi}_{t-1}\right)^2 - 2ab\left(\frac{\sigma_u^2}{1-\beta^2} + \frac{\omega^2}{(1-\beta)^2}\right)\left(\frac{\omega}{1-\beta}\right). \end{aligned}$$

5.3 GMM Estimation

For the model defined by (5.1) and (5.2), estimating the parameters can also be done using the GMM method. Bauwens & Veredas (2004) suggested this method may lead to suitable estimators. Using the approach of Andersen & Sorensen (1996), we calculate the expected value of y_t^m, $m = 1, 2, 3, 4$, and $y_t y_{t-r}$, where $r = 1, \ldots, 10$. These moments and cross-moments of duration were derived in Theorem 5.1.2.

5.4 A Simulation Study

We simulated series of length 5000. To allow for the use of the GMM method, conditions (5.3) to (5.5) should be satisfied. From (5.3), it is required that $a < \dfrac{1-\beta^2}{8\sigma_u^2}$. In order to make the model stationary and invertible, $|\beta|$ should be less than 1. The parameters σ_u and α should both be greater than 0. For the simulation study, we let $\beta = 0.95$ and $\sigma_u = 0.1$ implying parameter a should be less than 1.21875, for condition (5.3) to be satisfied. Following the work of Kawakatsu (2007) we set $\Theta \in (\omega,\ \beta,\ \gamma,\ \sigma,\ a,\ b,\ c)^t = (0.001,\ 0.95,\ 0.9,\ 0.1,\ 0.1,\ 0.45,\ 0)$. For the other two conditions, we see from Table 5.1 that they are both satisfied in all considered cases. Note the dependence on r of σ_Z^2. Estimation results are shown in Table 5.2.

Table 5.1: Parameter Values used for simulation of QSCD data.

parameters	β	σ_u	a	σ_Z^2	$\frac{1-\beta^2}{1-\beta^2-8a\sigma_u^2}$
$r = 10$	0.95	0.1	0.1	0.063861175	1.089>0
$r = 1$	0.95	0.1	0.1	0.005128205	1.089>0
parameters	$1-2a\sigma_Z^2$	λ	A	$1-2A\sigma_\Psi^2$	
$r = 10$	0.9872278>0	0.03234369	0.1363124	0.972>0	
$r = 1$	0.9989744>0	0.002566735	0.1903427	0.9989>0	
parameters	β	σ_u	a	σ_Z^2	$\frac{1-\beta^2}{1-\beta^2-8a\sigma_u^2}$
$r = 10$	0.9	0.1	0.1	0.04552183	1.043956>0
$r = 1$	0.9	0.1	0.1	0.00526358	1.043956>0
parameters	$1-2a\sigma_Z^2$	λ	A	$1-2A\sigma_\Psi^2$	
$r = 10$	0.9908956>0	0.02297004	0.1122694	0.988>0	
$r = 1$	0.9989474>0	0.002634352	0.1810854	0.981>0	

Both methods discussed earlier were used: the QML optimization relying on the nonlinear filter and the GMM method. For comparison, in Table 5.3 we give the

ASE of the estimates obtained from both methods. The GMM estimates provide the smallest ASE for ω and γ, while the QML method gave a lower ASE for parameters β, σ, a, b, and c.

Table 5.2: Simulation Estimations of QML and GMM

parameters	True value	QML	GMM
ω	0.001	-0.01192816	-0.0031345
		(0.02647389)	(0.006553696)
β	0.95	0.9496733	0.9232761
		(0.02288866)	(0.07948605)
γ	0.9	0.8998673	0.9064824
		(0.01778818)	(0.01398802)
σ_u	0.1	0.08692612	0.061956
		(0.02190208)	(0.05289102)
a	0.1	0.1019410	0.1828681
		(0.02496394)	(0.2391935)
b	0.45	0.4467866	0.4329538
		(0.02681089)	(0.1866534)
c	0	-0.009498322	0.01661534
		(0.02144684)	(0.02972771)

Table 5.3: Average Squared Error (ASE) of parameter estimates for the QSCD model

parameters	ASE	
	QML	GMM
ω	0.02597438	0.006488808
β	0.0224568	0.07869906
γ	0.01745255	0.01384953
σ	0.02148883	0.05236735
a	0.02449292	0.2368252
b	0.02630502	0.1848054
c	0.02104218	0.02943337

5.5 Another look at the IBM data

We here take another look at the IBM data introduced in Section 4.6. In Table 5.4, we give the estimates of each parameter. Note that the standard errors of these estimates are not provided. This is because the standard errors provided by the standard optimization method that we used are unreliable here. This difficulty has been raised by others who have suggested that the complexity (and nonlinearity) of the model makes it extremely difficult to numerically approximate second order derivatives. They also suggested that the values of the estimates should still by reasonably accurate as they are obtained without making use of higher order derivatives.

Again, all estimates of β are close to one, and smaller than one, suggesting persistency of the duration process and stationarity. As before, we define the dispersion ratio $D_y = \frac{\sigma_y}{\mu_y}$, the standard deviation of durations divided by the mean duration. We can calculate this ratio, based on the Weibull QSCD model formulae given in Theorem 5.2.1, which leads to

$$D_y = \frac{\sqrt{\sqrt{\frac{1-\beta^2}{1-\beta^2-4a\sigma_u^2}}\exp\{\frac{(\omega+\beta\omega+2b\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-4a\sigma_u^2)} - \frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + 2c\}\Gamma(1+2/\gamma) - \mu_y^2}}{\sqrt{\frac{1-\beta^2}{1-\beta^2-2a\sigma_u^2}}\exp\{\frac{(\omega+\beta\omega+b\sigma_u^2)^2}{2\sigma_u^2(1-\beta^2-2a\sigma_u^2)} - \frac{\omega^2(1+\beta)}{2\sigma_u^2(1-\beta)} + c\}\Gamma(1+1/\gamma)}.$$

We obtain estimates of the dispersion ratio by substituting estimated parameters. These estimates are given in Table 5.4, along with all other parameter estimates obtained with both methods, and once again suggest that overdispersion is present here.

Table 5.4: Estimates of QSCD model parameters obtained with QML and GMM

parameter	QML based on Thava's filter	GMM Estimates
ω	0.05365050	-0.02669
β	0.97489427	0.50186
γ	0.92376087	0.83362
σ	0.13471760	0.1007
a	0.10553067	0.15909
b	0.44141709	1.61973
c	0.08332539	1.02253
Estimated Dispersion Ratio	1.473206	1.242109

Chapter 6

Conclusions

In this thesis, two estimation method were developed for the estimation of the parameters in SCD models and in quadratic SCD models. As there is no closed form likelihood for both models, maximum likelihood available estimation method can't be applied. However, it is shown that the analytical form of the filter and moments can be derived, therefore QML methods based on filtering methods and GMM are two reasonable approaches. For the linear SCD model, two different filtering schemes were used one of which is the popular Kalman filter. However, the complexity of quadratic the SCD model, which has a quadratic form for the latent variable, made it impossible to use the QML approach based on Kalman filtering. A nonlinear filter was used instead. In both cases, the GMM methodology was also successfully applied to estimate model parameters.

Appendix A

Proof of Lemma 2.3.1 in Chapter 2

Proof. Using the same approach as Gong & Thavaneswaran (2008) , we have

$$\hat{\theta}_{t+1} = a\hat{\theta}_t + c + G_t(y_{t+1} - A\hat{\theta}_t - c^*),$$

so that

$$\theta_{t+1} - \hat{\theta}_{t+1} = a(\theta_t - \hat{\theta}_t) + b_1\tilde{\epsilon}_{t+1} - G_t(y_{t+1} - A\hat{\theta}_t - c^*),$$

and

$$
\begin{aligned}
\gamma_{t+1} &= E[\theta_{t+1} - \hat{\theta}_{t+1}]^2 \\
&= E[a(\theta_t - \hat{\theta}_t) + b_1\tilde{\epsilon}_{t+1} - G_t(y_{t+1} - A\hat{\theta}_t - c^*)]^2 \\
&= a^2E[\theta_t - \hat{\theta}_t]^2 + b_1^2\sigma_\epsilon^2 + G_t^2(y_{t+1} - A\hat{\theta}_t - c^*)^2 + \\
&\quad 2ab_1(\theta_t - \hat{\theta}_t)E[\tilde{\epsilon}_{t+1}] - 2aE[G_t(\theta_t - \hat{\theta}_t)(y_{t+1} - A\hat{\theta}_t - c^*)] - \\
&\quad 2b_1G_tE[\tilde{\epsilon}_t(y_{t+1} - A\hat{\theta}_t - c^*)].
\end{aligned}
$$

Now, note that

$$E[y_{t+1} - A\hat{\theta}_t - c^*] = E[A(\theta_t - \hat{\theta}_t) + Be_{t+1}],$$

or

$$E[y_{t+1} - A\hat{\theta}_t - c^*]^2 = A^2E[(\theta_t - \hat{\theta}_t)^2] + B^2\sigma_e^2.$$

Also, we have that

$$
\begin{aligned}
E[(y_{t+1} - A\hat{\theta}_t - c^*)(\theta_t - \hat{\theta}_t)] &= E[(\theta_t - \hat{\theta}_t)(A(\theta_t - \hat{\theta}_t) + Be_{t+1})] \\
&= AE[(\theta_t - \hat{\theta}_t)^2],
\end{aligned}
$$

$$
\begin{aligned}
E[(y_{t+1} - A\hat{\theta}_t - c^*)\tilde{\epsilon}_{t-1} b_1] &= E[b_1[A(\theta_t - \hat{\theta}_t + Be_{t+1}]\tfrac{b+\theta_t}{b_1}\epsilon_{t+1}] \\
&= E[b_2(A(\theta_t - \hat{\theta}_t + Be_{t+1})\epsilon_{t+1}] \\
&= b_2 BE[e_t + 1\epsilon_{t+1}] \\
&= b_2 B\{Cov(e_{t+1}\epsilon_{t+1}) - E(e_{t+1})E(\epsilon_{t+1})\} \\
&= b_2 B\rho\sigma_e\sigma_\epsilon.
\end{aligned}
$$

These results allow us to write

$$
\begin{aligned}
\gamma_{t+1} &= a^2\gamma_t + b_1^2\sigma_\epsilon^2 + G_t^2(A^2\gamma_t + B^2\sigma_e^2) - 2aA\gamma_t G_t - 2b_2 G_t B\rho\sigma_e\sigma_\epsilon \\
&= (A^2\gamma_t + B^2\sigma_e^2)\left[G_t - \frac{aA\gamma_t + Bb_2\rho\sigma_e\sigma_\epsilon}{A^2\gamma_t + B^2\sigma_e^2}\right]^2 - \\
&\quad \frac{(aA\gamma_t + Bb_2\rho\sigma_e\sigma_\epsilon)^2}{A^2\gamma_t + B^2\sigma_e^2} + a^2\gamma_t + b_1^2\sigma_\epsilon^2,
\end{aligned}
$$

which is minimized by letting

$$
G_t = \frac{aA\gamma_t + Bb_2\rho\sigma_e\sigma_\epsilon}{A^2\gamma_t + B^2\sigma_e^2}.
$$

In this case, we get

$$
\gamma_{t+1} = a^2\gamma_t + b_1^2\sigma_\epsilon^2 - \frac{(aA\gamma_t + Bb_2\rho\sigma_e\sigma_\epsilon)^2}{A^2\gamma_t + B^2\sigma_e^2}.
$$

$\square$

Appendix B

R code

USE KALMAN FILTER TO ESTIMATE THE SCD MODEL PARAMETERS

```
for (j in 1:100){
SCD_Simul<- function(n, Coeff){
# inputs: n  length of desired series. m the replication number
# parms      vector of parameters for SCD process; has the form:
#                              (omega,beta,gamma,sigma,)
#  model:  e_(t) is any special distribution for the error term
#         mu_(t) is normal distribution
#         d_t = exp(psi_t)* e_(t)
#         psi_t = omega + beta*psi_(t-1) + mu_t
# output: dur series
# usage: Coeff=c(0.001,0.95,0.9,0.1)
#        distrib= "weibull"
```

```
#          scd  = SCD_Simul(5000,Coeff,distrib,forma)
# error term
      er  <- rweibull(n, shape=Coeff[3], scale = 1)
     psi  <- rep(1,n)
     dur  <- rep(1,n)
     mu   <- rnorm(n,mean=0,sd=Coeff[4])
   psi[1] <- Coeff[1]/(1-Coeff[2])
            # initial value for the psi
   dur[1] <- exp(psi[1])*er[1]
            # initial value for the duration
 for (i in 2:n){
     psi[i] <- Coeff[1]+ Coeff[2]*psi[i-1]+ mu[i]
            # Eq psi_(i)=w + b*psi_(i-1)+u_(i)
     dur[i] <- exp(psi[i])*er[i]
            # Eq y_(i)=e^(psi_(i))*e_(i)
            # install each series duration into the column of sumdur
 }
return(dur)
}
Coeff=c(0.001,0.95,0.9,0.1)
n=5000
scd_wei=SCD_Simul(n,Coeff)
        # simulate data
 #For Estimation part
```

```
     scd <- scd_wei
      n  <- length(scd)
estimscd <- function(pars){
estimscd <-0
      H  <- pi^2/(6*pars[3]^2)
            # it is the variance about error term
      u  <- (-0.5772/pars[3])
            # it is constant d_i=-0.5772/alpha
      p  <- rep(1,n)
            # give the number for the P
     p[1]<- pars[4]^2/(1-pars[2]^2)
            # initial value for the P
    pcon <- rep(1,n)
            # give the number for the conditional P
  pcon[1]<- (pars[2]^2)*p[1]+pars[4]^2
            # intial value for the conditional P
      F  <- rep(2,n)
            # covaraince for the series duration Y
     F[1]<- pcon[1]+H
      a  <- rep(1,n)
    a[1] <- pars[1]/(1-pars[2])
    acon <- rep(1,n)
  acon[1]<- pars[2]*a[1]+pars[1]
      v  <- rep(1,n)
```

```
     v[1]<- log(scd[1])-acon[1]-u
 for (t in 2:n){
    acon[t] <- pars[2]*a[t-1]+pars[1]
             # the prediction equation for conditoanl a_i
    pcon[t] <- (pars[2]^2)*p[t-1]+pars[4]^2
             # the prediction equation for conditoanl p_i
      v[t]  <- log(scd[t])-acon[t]-u
             # Eq v_i=y_i-E(y_i)
      a[t]  <- acon[t]+(pcon[t]*(F[t]^(-1)))*v[t]
             # the updating equation for the a_i
      F[t]  <- pcon[t]+H
             # the updating equation for the covarance F_i
      p[t]  <- pcon[t]-pcon[t]^2/F[t]
             # the updating equation for the p_i
   estimscd <- estimscd+0.5*log(abs(F[t]))+0.5*(v[t]^2)/F[t]
             #qusi-maximum likehood function
  }
return(estimscd)
}
# Using the package to estimate the parameters
pars=c(0.001,0.95,0.9,0.1)
scd_par=optim(pars,estimscd,method="Nelder-Mead",hessian= TRUE)
cat(j,"\n")
cat(scd_par$par,"\n")
```

```
j=j+1
}
```

USE THAVA'S FILTER TO ESTIMATE THE SCD MODEL PARAMETERS

```
for (j in 1:100){
SCD_Simul<- function(n, Coeff){
# inputs: n  length of desired series.
# parms:     vector of parameters for SCD process; has the form:
#                               (omega,delta,gamma,sigma,)
#  model:  e_(t) is any special distribution for the error term
#         mu_(t) is normal distribution
#         d_t = exp(psi_t)* e_(t)
#         psi_t = omega + delta*psi_(t-1) + mu_t
# output: dur series
# usage: Coeff=c(0.001,0.95,0.9,0.1)
#        distrib= "weibull"
#          scd  = SCD_Simul(5000,Coeff,distrib,forma)
# error term
      er  <- rweibull(n, shape=Coeff[3], scale = 1)
     psi  <- rep(1,n)
     dur  <- rep(1,n)
     mu   <- rnorm(n,mean=0,sd=Coeff[4])
   psi[1] <- Coeff[1]/(1-Coeff[2])
             # initial value for the psi
```

```
   dur[1] <- exp(psi[1])*er[1]
            # initial value for the duration
 for (i in 2:n){
     psi[i] <- Coeff[1]+ Coeff[2]*psi[i-1]+ mu[i]
            # Eq psi_(i)=w + b*psi_(i-1)+u_(i)
     dur[i] <- exp(psi[i])*er[i]
            # Eq y_(i)=e^(psi_(i))*e_(i)
            # install each series duration into the column of sumdur }
return(dur)
}
Coeff=c(0.001,0.95,0.9,0.1)
n=5000
scd_wei=SCD_Simul(n,Coeff)
     scd <- scd_wei
      n  <- length(scd)
estimscd <- function(pars){
 estimscd<-0
      H  <- pi^2/(6*pars[3]^2)
           # give the initial value for the variance Q
      u  <- (-0.5772/pars[3])
    gamma<- rep(1,n)
 gamma[1]<- (pars[4]^2)/(1-pars[2]^2)
    theta<- rep(1,n)
 theta[1]<- pars[1]/(1-pars[2])
```

```
     F  <- rep(2,n)
    F[1]<- ((pars[2]^2)/(1-pars[2]^2)+1)*(pars[4]^2) + H
     v  <- rep(1,n)
    v[2]<- log(scd[2])-pars[2]*theta[1]-pars[1]-u
 for (t in 2:n){
     gamma[t] <- ((pars[2]^2)*gamma[t-1]) + (pars[4]^2) -
                  (pars[2]^2*gamma[t-1]^2)/(gamma[t-1]+H)
     theta[t] <- pars[2]*theta[t-1] + pars[1]+
                 ((pars[2]*gamma[t-1])/(gamma[t-1]+H))*
                 (log(scd[t])-pars[2]*theta[t-1]-pars[1]-u)
        v[t]  <- log(scd[t])-pars[2]*theta[t-1]-pars[1]-u
        F[t]  <- gamma[t] + H
estimscd <- estimscd+0.5*log(abs(F[t]))+0.5*(v[t]^2)/F[t]
                #weilk0=weilk0 + 0.5*abs(log(ft[i])) +0.5*v[i]^2/ft[i]
  }
return(estimscd)
}
pars=c(0.001,0.9,0.9,0.01)
scd_par=optim(pars,estimscd,method="BFGS",hessian= TRUE)
cat(j,"\n")
cat(scd_par$par,"\n")
j=j+1
}
```

USE GMM METHOD TO ESTIMATE THE SCD MODEL PARAMETERS

```
library("gmm")
library("fMultivar")
for (j in 1:100){
SCD_Simul<- function(n, Coeff){
# inputs: n  length of desired series. m the replication number
# parms      vector of parameters for SCD(1,1) process; has the form:
#                              (omega,beta,gamma,sigma,)
#  model:  e_(t) is any special distribution for the error term
#         mu_(t) is normal distribution
#         d_t = exp(psi_t)* e_(t)
#         psi_t = omega + beta*psi_(t-1) + mu_t
# output: dur series
# usage: Coeff=c(0.001,0.95,0.9,0.1)
#        distrib= "weibull"
#          scd  = SCD_Simul(5000,Coeff,distrib,forma)
# error term
      er  <- rweibull(n, shape=Coeff[3], scale = 1)
     psi  <- rep(1,n)
     dur  <- rep(1,n)
     mu   <- rnorm(n,mean=0,sd=Coeff[4])
   psi[1] <- Coeff[1]/(1-Coeff[2])
             # initial value for the psi
   dur[1] <- exp(psi[1])*er[1]
             # initial value for the duration
```

```
 for (i in 2:n){
     psi[i] <- Coeff[1]+ Coeff[2]*psi[i-1]+ mu[i]
             # Eq psi_(i)=w + b*psi_(i-1)+u_(i)
     dur[i] <- exp(psi[i])*er[i]
             # Eq y_(i)=e^(psi_(i))*e_(i)
             # install each series duration into the column of sumdur
 }
return(dur)
}
Coeff=c(0.001,0.95,0.9,0.1)
n=5000
scd_wei=SCD_Simul(n,Coeff)
scd.wei<- scd_wei
scd.moments <- function(parm,data1=NULL){
      omega <- parm[1]
       beta <- parm[2]
      alpha <- parm[3]
      sigma <- parm[4]
 exv.d <- c(exp(omega/(1-beta) + sigma^2/(2*(1-beta^2)))*gamma(1/alpha+1),
            exp(2*omega/(1-beta) + 2*sigma^2/(1-beta^2))*gamma(2/alpha+1),
            exp(3*omega/(1-beta) + 9*sigma^2/(2*(1-beta^2)))*gamma(3/alpha+1),
            exp(4*omega/(1-beta) + 16*sigma^2/(2*(1-beta^2)))*gamma(4/alpha+1))
exv.dd <- c(exp(2*omega/(1-beta) + (1+beta^(1:10))*sigma^2/(1-beta^2))*
                                                         (gamma(1/alpha+1))^2)
```

```
  gmat <- c(exv.d,exv.dd)
  t(t(data1)-gmat)
}
scd.four <- cbind(abs(scd.wei), scd.wei^2, abs(scd.wei)^3,  scd.wei^4)
initial.val <- scd.four[-(1:10),]
dcorre.val <- tslag(scd.wei, 1:10, trim=T) * as.vector(scd.wei[-(1:10)])
scd.data <- cbind(initial.val, abs(dcorre.val))
 parm  <-c(0.01,0.9,0.9,0.1)
result <- gmm(scd.moments, x=scd.data, t0=parm, type=c("iterative") )
cat(summary(result)$coefficients[1:4])
cat(j,"\n")
j=j+1
}
```

USE THAVA'S FILTER TO ESTIMATE THE QUADRATIC SCD MODEL PARAMETERS

```
for (j in 1:50){
SCD_Simul<- function(n, Coeff)
{
# inputs: n  length of desired series
# parms      vector of parameters for SCD(1,1) process; has the form:
#                             (omega,delta,alpha,sigma,a,b,c,)
#            Coeff[1]=omega, Coeff[2]=beta, Coeff[3]=alpha, Coeff[4]=sigma
#            Coeff[5]=a, Coeff[6]=b, Coeff[7]=c
```

```
#  model: e_(t) is any special distribution for the error term
#          mu_(t) is normal distribution
#          d_t = psi_(t)* e_(t)
#          Cpsi_t = exp(psi_t)
#          psi_t = omega + delta*psi_(t-1) + mu_t
# output: dur series
# usage: Coeff=c(0.001,0.95,0.9,0.1,0.1,0.45,0)
#         distrib="weibull"
#         scd=SCD_Simul(5000,Coeff)
      e  <-rweibull(n, Coeff[3], scale = 1)
     psi <- rep(1,n)
  Calpsi <- rep(1,n)
    dur  <- rep(1,n)
     mu  <- rnorm(n,mean=0,sd=Coeff[4])
  psi[1]<- Coeff[1]/(1-Coeff[2])
  dur[1]<- psi[1]*e[1]
for (i in 2:n){
   psi[i]<- Coeff[1]+ Coeff[2]*psi[i-1]+ mu[i]
   Calpsi[i]<- exp(Coeff[5]*psi[i]^2+Coeff[6]*psi[i]+Coeff[7])
   dur[i]<- Calpsi[i]*e[i]
}
 return(dur)
}
Coeff=c(0.001,0.95,0.9,0.1,0.1,0.45,0)
```

```
n=5000

scdq_wei=SCD_Simul(n,Coeff)

scd<-scdq_wei

 n  <- length(scd)

estimscd <- function(pars){

     estimscd=0

       mu <-rnorm(n,mean=0,sd=pars[4])

                # general n normal distribution in order to create the psi2

       H  <- pi^2/(6*pars[3]^2)

                # give the initial value for the variance Q

       u  <- (-0.5772/pars[3])

                # u=E(log(epsilon_(t+1)))=-0.5772/alpha

       c1 <- pars[4]^2/(1-pars[2]^2)

                # var(psi_(t))=sigma^2/(1-beta^2)

       c2 <- pars[1]/(1-pars[2])

                # E(psi_(t))=omega/(1-beta)

    lamda <- ((2+2*pars[2]^2)/(1-pars[2]^2))*pars[4]^2

              +(4*(pars[1]^2)*(pars[4]^2))/(1-pars[2])^2

                # E(e_(t+1)^2)

     psi2 <- rep(1,n)

   psi2[1]<- c1+c2^2

                # E(psi_(t+1)^2)=sigma^2/(1-beta^2)+(omega/(1-beta))^2

     psi1 <- rep(1,n)

   psi1[1]<- sqrt(psi2[1])
```

```
   gamma<- rep(1,n)
gamma[1]<- (pars[4]^2/(1-pars[2]^4))*(2*pars[4]^2+(4*pars[2]^2*pars[1]^2
            +4*pars[2]*pars[4]^2)/(1-pars[2]^2)+4*pars[1]^2)
              # var(psi_(t+1)^2)
     F  <- rep(2,n)
    F[1]<- pars[5]^2*gamma[1] + H + pars[6]^2*c1
              # f_(t)=a^2*var(pai_(t)^2)+b^2*var(par_(t))+var(epsilon_(t+1))
     v  <- rep(1,n)
    v[2]<- log(scd[2])-(pars[5]*psi2[1] + pars[6]*psi1[1]+ pars[7] + u)
     d  <-rep(1,n)
  numer <-rep(1,n)
     G  <-rep(1,n)
   for (t in 2:n){
     d[t-1] <- 2*pars[5]*(pars[2]^2)*gamma[t-1]+4*pars[1]*pars[2]*pars[6]
               *(c1+c2^2-2*c2*psi1[t-1]+psi2[t-1])+2*((pars[2]^2)*pars[6]
               +2*pars[5]*pars[1]*pars[2])*psi1[t-1]*(psi2[t-1]-c2*psi1[t-1]
               -(c1+c2^2))
     numer[t-1]<- 2*(pars[5]^2)*gamma[t-1]+2*(pars[6]^2)*(c1+c2^2-2*c2*psi1[t-1]
                   +psi2[t-1])+2*H+4*pars[5]*pars[6]*psi1[t-1]*(psi2[t-1]
                   -c2*psi1[t-1]-(c1+(c2^2)))
     G[t-1] <- d[t-1]/numer[t-1]
     gamma[t] <-((pars[2]^2-pars[5]*G[t-1])^2)*gamma[t-1] + lamda
                +H*G[t-1]^2 +((2*pars[1]*pars[2]-pars[6]*G[t-1])^2)
                *(c1+c2^2-2*c2*psi1[t-1]+psi2[t-1])+ 2*(pars[2]^2
```

```
                    -pars[5]*G[t-1])*(2*pars[2]*pars[1]-pars[6]*G[t-1])*psi1[t-1]
                     *(psi2[t-1]-c2*psi1[t-1]-(c1+c2^2))
        psi2[t] <- (pars[2]^2)*psi2[t-1]+ (pars[1]^2)+ mu[t]^2
                    +2*pars[1]*pars[2]*psi1[t-1]+2*pars[2]*psi1[t-1]*mu[t]
                    +2*pars[1]*mu[t]
        psi1[t] <- sqrt(psi2[t])
        v[t] <- log(scd[t])-pars[5]*psi2[t-1]-pars[6]*psi1[t-1]-u-pars[7]
        F[t] <- (pars[5]^2)*gamma[t]+(pars[6]^2)*(c2^2+c1+psi2[t-1]-2*c2*psi1[t-1])
                +H-pars[5]^2*(c2^2+c1-psi2[t-1])^2
                -pars[6]^2*(c2-psi2[t-1])^2-2*pars[5]*pars[6]*(c2^2+c1)*(c1)
estimscd <- estimscd+0.5*log(abs(F[t]))+0.5*(v[t]^2)/F[t]
  }
return(estimscd)
}
 pars=c(0.001,0.95,0.9,0.1,0.1,0.45,0)  # yong 0
 scdq_par=optim(pars,estimscd,method="CG",hessian= TRUE)
cat(j,"\n")
cat(scdq_par$par,"\n")
j=j+1
}
```

USE GMM METHOD TO ESTIMATE THE QUADRATIC SCD MODEL PARAMETERS

```
library("gmm")
```

```
library("fMultivar")
for (j in 1:50){
SCD_Simul<- function(n, Coeff)
{
# inputs: n  length of desired series
# parms      vector of parameters for SCD(1,1) process; has the form:
#                              (omega,delta,alpha,sigma,a,b,c,)
#            Coeff[1]=omega, Coeff[2]=beta, Coeff[3]=alpha, Coeff[4]=sigma
#            Coeff[5]=a, Coeff[6]=b, Coeff[7]=c
#  model: e_(t) is any special distribution for the error term
#         mu_(t) is normal distribution
#         d_t = psi_(t)* e_(t)
#         Cpsi_t = exp(psi_t)
#         psi_t = omega + delta*psi_(t-1) + mu_t
# output: dur series
# usage: Coeff=c(0.001,0.95,0.9,0.1,0.1,0.45,0)
#        distrib="weibull"
#        scd=SCD_Simul(5000,Coeff)
      e  <-rweibull(n, Coeff[3], scale = 1)
     psi <- rep(1,n)
  Calpsi <- rep(1,n)
    dur  <- rep(1,n)
     mu  <- rnorm(n,mean=0,sd=Coeff[4])
  psi[1]<- Coeff[1]/(1-Coeff[2])
```

```
Calpsi[1]<- exp(Coeff[5]*psi[1]^2+Coeff[6]*psi[1]+Coeff[7])
  dur[1]<- Calpsi[1]*e[1]
for (i in 2:n){
   psi[i]<- Coeff[1]+ Coeff[2]*psi[i-1]+ mu[i]
   Calpsi[i]<- exp(Coeff[5]*psi[i]^2+Coeff[6]*psi[i]+Coeff[7])
   dur[i]<- Calpsi[i]*e[i] }
 return(dur)
}
Coeff=c(0.001,0.95,0.9,0.1,0.1,0.45,0)
n=5000
scdq_wei=SCD_Simul(n,Coeff)
scd.wei<-scdq_wei
r=10
scd.moments <- function(parm,data1=NULL){
      omega <- parm[1]
       beta <- parm[2]
      alpha <- parm[3]
      sigma <- parm[4]
        a   <- parm[5]
        b   <- parm[6]
        c   <- parm[7]
 z.v  <- ((1-beta^(2*(1:r)-1))/(1-beta^2))*(sigma^2)
psi.v <- sigma^2/(1-beta^2)
lamda <- z.v/(2*(1-2*a*z.v))
```

```
A   <- a*(beta^(2*(1:r))+1)+4*lamda*a^2*beta^(2*(1:r))
B   <- a*omega^2*(1-beta^(1:r))^2/(1-beta)^2+b*omega*(1-beta^(1:r))/(1-beta)+
            2*c+lamda*(2*a*omega*(1-beta^(1:r))/(1-beta)+b)^2
C.1  <- (2*a*omega*beta^(1:r)*(1-beta^(1:r))/(1-beta)+b*(beta^(1:r)+1))
C.2  <- 4*lamda*a*(beta^(1:r))*(2*a*omega*(1-beta^(1:r)/(1-beta))+b)
C   <- C.1+C.2
G   <- 1/sqrt(abs(1-2*b*z.v))
T   <- 1/sqrt(abs(1-2*A*psi.v))
exv.d <-
c(sqrt((1-beta^2)/(1-beta^2-2*a*sigma^2))*
       exp((omega+beta*omega+b*sigma^2)^2/(2*sigma^2*(1-beta^2-2*a*sigma^2))-
           (omega^2*(1+beta))/(2*sigma^2*(1-beta))+c)*gamma(1+1/alpha),
  sqrt((1-beta^2)/(1-beta^2-4*a*sigma^2))*
       exp((omega+beta*omega+2*b*sigma^2)^2/(2*sigma^2*(1-beta^2-4*a*sigma^2))-
           (omega^2*(1+beta))/(2*sigma^2*(1-beta))+2*c)*gamma(1+2/alpha),
  sqrt((1-beta^2)/(1-beta^2-6*a*sigma^2))*
       exp((omega+beta*omega+3*b*sigma^2)^2/(2*sigma^2*(1-beta^2-6*a*sigma^2))-
           (omega^2*(1+beta))/(2*sigma^2*(1-beta))+3*c)*gamma(1+3/alpha),
  sqrt((1-beta^2)/(1-beta^2-8*a*sigma^2))*
       exp((omega+beta*omega+4*b*sigma^2)^2/(2*sigma^2*(1-beta^2-8*a*sigma^2))-
           (omega^2*(1+beta))/(2*sigma^2*(1-beta))+4*c)*gamma(1+4/alpha))
exv.dd <- c(G*T*exp(B)*exp(C^2*psi.v/(2*(1-2*A*psi.v)))*(gamma(1/alpha+1))^2)
gmat <- c(exv.d,exv.dd)
t(t(data1)-gmat)
```

```
}
scd.four <- cbind(abs(scd.wei), scd.wei^2, abs(scd.wei)^3,  scd.wei^4)
initial.val <- scd.four[-(1:10),]
dcorre.val <- tslag(scd.wei, 1:10, trim=T) * as.vector(scd.wei[-(1:10)])
scd.data <- cbind(initial.val, abs(dcorre.val))
parm=c(0.001,0.9,0.9,0.01,0.1,0.45,0)
res.gmm.simg <- gmm(scd.moments, x=scd.data, t0=parm, type=c("twoStep"))
cat(summary(res.gmm.simg)$coefficients[1:7])
cat(j,"\n")
j=j+1
}
```

Bibliography

Abraham, B. and Balakrishnan, N. (2001). Time series in industry and business. Technical report, IIQP Research Report.

Andersen, T.G. and Sorensen, B.E. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* **14**(3), 328–352.

Bauwens, L., Lubrano, M., and Richard, J.F. (1999). *Baysian Inference in Dynamic Econometric Models.* Oxford University Press.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent variable model for the analysis of financial durations. *Journal of Econometrics* **119**, 381–412.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (1987). A conditional heteroscedastic time series model for speculative prices and rates of return. *Review of Economics & Statistics* **63**, 542–547.

Brockwell, P.J., Liu, J., and Tweedie, R.L. (1992). On existence of stationary threshold autoregressive moving average processes. *Journal of Time Series Analysis* **13**, 95–113.

Chan, K.S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* **7**, 179–190.

Chandra, S. and Taniguchi, M. (2001). Estimation functioins for non-linear time series models. *Annals of Institute of Statistical Mathematics* **53**(1), 125–141.

Chesney, M. and Scott, L.O. (1989). Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model. *Journal of Financial and Quantitative Analysis* **24**, 267–284.

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41**, 135–155.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987–1008.

Engle, R.F. (2000). The econometrics of Ultra-High Frequency Data. *Econometrica* **68**, 1–22.

Engle, R.F. (2002). New frontiers of ARCH models. *Journal of Applied Econometrics* **17**, 425–446.

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

Gong, H. and Thavaneswaran, A. (2009). Recursive estimation for continuous time stochastic volatility models. *Applied Mathematics Letters* **22**, 1770–1774.

Gong, H., Thavaneswaran, A., and Singh, J. (2008). Filtering for some time series models by using transformation. *The Mathematical Scientist* **33**, 141–147.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications*. Springer, New York.

Gouriéroux, C., Jasiak, J., and Le Fol, G. (1999). Intra-day market activity. *Journal of Financial Markets* **2**, 193–226.

Granger, C.W.J. and Anderson, A.P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoek and Ruprecht, Göttingen.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4), 1029–1054.

Harvey, A.C. (1994). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A.C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics* **14**(4), 429–434.

Jacquier, E., Polson, N.G., and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economics* **12**, 371–417.

Kawakatsu, H. (2007). Specification and specification of discrete time quadratic stochastic volatility models. *Journal Empirical Finance* **14**, 424–442.

Knight, J. and Ning, C. (2008). Estimation of the stochastic conditional duration model via alternative methods. *Econometrics Journal* **11**, 593–616.

Milhøj (1985). The moment structure of ARCH processes. *Scandinavian Journal of Statistics* **12**, 281–292.

Nelson, D.B. (1991). Conditional heteroscedasticity in asset pricing: a new approach. *Econometrica* **59**, 347–370.

Nicholls, D.F. and Quinn, B.G. (1982). *Random Coefficients Autoregressive Models: An Introduction.* Springer Verlag, New York.

Pathmanathan, D., Ng, K., and Peiris, M. (2009). On estimation of ACD models with different error distributions. *Sri Lankan Journal of Applied Statistics* **10**, 251–269.

Pawlak, M. and Schmid, W. (2001). On the distributional properties of GARCH processes. *Journal of Time Series Analysis* **22**, 339–352.

Peiris, S. and Thavaneswaran, A. (2004). A note on the filtering for some time series models. *Journal of Time Series Analysis* **25**, 397–407.

Petruccelli, J.D. (1986). On the consistency of least squares estimators of threshold AR(1) model. *Journal of Time Series Analysis* **7**, 269–278.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* **63**, 289–306.

Scott, L.O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis* **22**, 419–438.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Time Series Models in Econometrics, Finance and other fields.* Eds. Cox, D.R., Hinkley, D.V. and Bardorff-Nielson, 1–67, Chapman & Hall.

Tauchen, G. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica* **51**, 485–505.

Taylor, S.J. (1982). Financial returns modeled by the product of two stochastic processes, a study of daily sugar prices 1961-79. In *Time Series Analysis: Theory and Practice,* ed. O.D. Anderson, **1**, 203–226, North–Holland, Amsterdam.

Taylor, S.J. (1986). *Modelling Financial Time Series.* John Wiley, Chichester.

Thavaneswaran, A. and Abraham, B. (1988). Estimation for non-linear time series models using estimating equations. *Journal of Time Series Analysis* **13**, 99–108.

Thavaneswaran, A. and Ghahramani, M. (2010). Stochastic conditional duration models. Submitted to the *Journal of Statistical Theory and Applications*.

Tjøstheim, D. (1986). Estimation in non-linear time series models. *Stochastic Processes & Their Applications* **21**, 251–273.

Tong, H. (1983). *Threshold models in non-linear time series analysis.* Lecture Notes in Statistics No 21. Springer-Verlag, New York.

Tsay, R.S. (2005). *Analysis of Financial Time Series.* John Wiley & Sons.

Wiggins, J.B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics* **19**, 351–372.

Zakoïan (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**, 931–955.