Identification of stochastic systems driven by Lévy processes

by

Máté Mánfay

Submitted to

Central European University

Department of Mathematics and its Applications

In partial fulfillment of the requirements for the degree of Doctor of Philosophy

Supervisor: Professor László Gerencsér

Budapest, Hungary

2014

Declaration

I, candidate for the degree of Doctor of Philosophy at the Central European University Department of Mathematics and its Applications, declare herewith that the present thesis is based on my research and only such external information as properly credited in notes and bibliography. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Budapest, 7 March 2014

Máté Mánfay

Declaration

I hereby declare that Máté Mánfay took an equal part in our joint papers, the materials of which have been included in this thesis.

Budapest, 7 March 2014

Dr. László Gerencsér

Supervisor

Acknowledgments

First, I would like to express my special appreciation and thanks to my advisor Dr. László Gerencsér. His support and advice have a fundamental contribution to this dissertation. Without his supervision and constant guidance this dissertation would not have born.

Both the Central European University and the Institute for Computer Science and Control, Hungarian Academy of Sciences (MTA SZTAKI), Systems and Control Lab are greatly acknowledged for supporting my work in the past few years.

From the beginning of my studies all of my teachers at the Noszlopy Gáspár Primary School, the Fazekas Mihály Secondary School, the Budapest University of Technology and Economics and the Central European University have been contributed to the development of my interest and knowledge in mathematics. It would be a nearly impossible task to mention all of them, but I cannot express how grateful I am to them.

Finally, a special thanks should go to my loved ones who have supported me throughout the entire process. Words cannot express how thankful I am to my mother, my father, my sister for all of the sacrifices that they have made for me. I would also like to thank my beloved Ildi for helping me through tough times.

Abstract

The research carried out in this thesis is motivated by my interest in the analysis of financial time series. From the technical point of view we study the identification of discrete time stochastic systems driven by the increments of Lévy processes. As an alternative to the maximum likelihood method we develop and analyze a novel identification method by adapting the so-called empirical characteristic function method originally devised for estimating parameters of characteristic functions from i.i.d. samples. First of all, we present an essentially asymptotically efficient three-stage identification method for the system and noise parameters of stable and inverse stable linear systems. Then we present an alternative extension of the empirical characteristic function (ECF) method applicable for stable, but possibly not inverse stable linear stochastic systems. Thirdly, we propose an essentially asymptotically efficient estimation method for the system parameters of general autoregressive conditional heteroscedasticity (GARCH) processes. For each of the above problems we precisely characterize the estimation error in the form of martingale representation theorems. After that we develop recursive estimation methods for stable and inverse stable linear systems along the line of arguments applied for the off-line identification of linear systems. Finally, we discuss a particular technical problem, the stability of time-varying stochastic systems driven or modulated by a Lévy process with discrete time interventions, such as parameter resettings.

Table of Contents

Declaration of authorship iii

Acknowledgments iv

Abstract v

Table of contents vi

1 Introduction 2

1.1 Notations 7

2 Lévy Processes 8

2.1 Motivation, basic properties 8

2.2 Semimartingales 10

2.3 Lévy processes in finance 12

3 ECF method for i.i.d. data 19

3.1 ECF with finite set of moment conditions 19

3.2 ECF with full continuum of moment conditions 24

3.3 Discussion 30

4 ECF method for linear Lévy systems **32**

4.1 Discrete-time Lévy-systems 33

4.2 A three-stage method 36

4.3 A summary of results on the PE method 39

4.4 The ECF method for estimating the noise parameters 41

4.5 Re-estimation of θ^* by the ECF method 46

4.6 Efficiency of the ECF method for θ^* 52

4.7 Discussion 57

5 ECF identification of linear Lévy systems with possibly unstable zeros **59**

5.1 ECF method for filtered data 60

5.2 Estimating the system dynamics 66

5.3 ECF for i.i.d. data revisited 70

5.4 Factor models 71

6 ECF estimation of the parameters of GARCH processes **74**

6.1 Basic properties of GARCH processes 74

6.2 ML method for GARCH processes 79

6.3 ECF method for GARCH processes 83

6.4 Analysis of the ECF method for GARCH processes 87

6.5 Efficiency of the ECF method for GARCH processes 93

6.6 Discussion 96

7 Recursive ECF estimation **99**

7.1 General recursive estimation scheme 100

7.2 Recursive ECF for i.i.d. sample 105

7.3 Recursive ECF for linear Lévy systems with known noise characteristics 108

7.4 Recursive ECF for linear Lévy systems 113

8 Stability of hybrid Lévy systems **120**

8.1 Motivation 120

8.2 Modified geometric drift condition for time-invariant systems 122

8.3 A stochastic Desoer's Theorem 125

8.4 Time-varying systems with parameter resettings 128

8.5 Discussion: state resetting for jump processes 129

8.6 Proofs 129

9 Simulation results **136**

9.1 Simulating CGMY processes 136

9.2 Re-estimation of ARMA system paramters by ECF method 137

9.3 Re-estimation of ARMA system parameters by recursive ECF method 140

9.4 ECF estimate of GARCH processes 140

10 Appendix **142**

10.1 Gamma processes 142

10.2 L-mixing processes 145

11 Conclusions **150**

Bibliography **152**

Chapter 1

Introduction

Lévy processes are widely used to model phenomena arising in natural sciences, economics, financial mathematics, queueing theory, telecommunication, robotics, mechanical systems and biology, see [18],[59],[5],[52],[24],[19].

The classical model for modelling market dynamics, namely geometric Brownian motion, was proposed by Louis Bacehelier [4]. This model is still the accepted core model despite the fact that empirical studies revealed that its assumptions are not realistic. For instance, since price movements are induced by transactions which can be unevenly distributed in real time, it would be more natural to use a time changed Brownian motion to model price dynamics. If the time change is defined by a gamma process, we obtain the so-called VG (shorthand for Variance Gamma) process. The VG processes reproduce a number of stylized facts of real price processes, such as fat tails and large kurtosis. It can be shown that this time changed Brownian process itself is a Lévy process. Extending the above construction novel price dynamics have been proposed by a variety of authors, called the geometric Lévy processes obtained by exponentiating a Lévy process (L_t).

The standard model of a price process within this framework is then

$$S_t = S_0 \exp L_t, \tag{1.0.1}$$

and (S_t) is called a geometric Lévy process. Hence, the Brownian motion with drift B_t is replaced by a general Lévy process in the classical model

$$S_t = S_0 \exp B_t. \tag{1.0.2}$$

This exponential Brownian motion model has been extensively studied in the literature, but recently more and more attention have been paid to the study of geometric Lévy processes. A variety of choices for (L_t) has been proposed in the literature. The main examples are: compound poisson process, stable process, variance Gamma (VG) process, tempered stable (CGMY) process, hypergeometric process and normal inverse Gauss (NIG) process, see [13], [48], [52], [19]. We give a short introduction for the theory of Lévy processes in Chapter 2. Additionally we present some Lévy processes of special form that are frequently used in financial modelling.

The motivation behind these models is that the returns of the stock process, say $(S_{t+h} - S_t)/S_t$ are assumed to be independent and stationary. While this is an attractive assumption, its consequences are less attractive. In particular it follows that the variance of the price process tends to infinity. A closer look at data in fact reveals that there is a weak correlation between, say, daily returns $(S_{t+1} - S_t)/S_t$.

In Chapter 4 we propose to introduce a new class of models that allows decaying memory and friction. For this purpose we first define a process Y_t which is the output of a finite dimensional stable linear SISO system, driven by a Lévy process:

$$\Delta Y = A \Delta L,$$

where A is a transfer function and (L_t) is a Lévy process. Such a system will be called a linear Lévy system. Then we define the price process according to (1.0.1):

$$S_n = S_0 \exp Y_n.$$

The application of the ML method would solve the full identification problem along standard lines, assuming that the density function of ΔL_n is known, which is unfortunately not the case. Hence we present a combination of advanced techniques in systems identification with a specific statistical technique, widely used in the context in finance, called the ECF (shorthand for empirical characteristic function) method. The ECF method was originally designed for i.i.d. samples and A. Feuerverger and P. McDunnogh [28] showed that it can be interpreted as the Fourier transform of the ML method. We will show that the proposed three-staged algorithm gives an essentially efficient estimate of both the system and the noise parameters. Furthermore, we prove martingale representation theorems for the estimation error.

Since in this thesis we focus on the applicability of the ECF method for linear and non-linear stochastic systems it is worth presenting the origins of the ECF method. In Chapter 3 we study the ECF method for i.i.d. data and give the efficiency result in the case when the full continuum of moment conditions is used.

In Chapter 5 we extend the results presented in Chapter 4 to finite dimensional stochastic Lévy systems with possibly unstable zeros. Recall that both the PE method and the ML method as presented in [36] assume that the system is non-minimum phase, i.e. it has an exponentially stable inverse. Our three-staged identification method in Chapter 4 is developed under the same assumption. In fact, the identification of finite dimensional linear stochastic systems with unstable zeros is barely discussed in the literature. We adapt the ECF method for linear stochastic systems with possibly unstable zeros. A novel challenge of this approach is that the exact characteristic function (c.f.) cannot be computed explicitly (which is the key assumption

for the ECF methods).

In Chapter 6 we investigate the possibility of adapting the ECF method for general autoregressive conditional heteroscedasticity (GARCH) processes. While applying the ECF method for such systems new problems arise as the GARCH processes are non-linear. Following the structure of the other Lévy driven systems studied in this thesis we consider a special type of GARCH model. GARCH processes were introduced by R. Engle and T. Bollerslev in 1982 and 1986 to model the price dynamics of highly volatile financial instruments, and became widely accepted in finance, see [25], [8]. If the driving noise process is formed by the increments of a Lévy process then the standard quasi-maximum likelihood method to identify a GARCH process can be replaced by an adaptation of the ECF method. We show that with an appropriately chosen score function the parameters of a Lévy driven GARCH process can be estimated in an essentially efficient way.

In the literature the empirical characteristic function method is presented as an off-line identification method. The framework of the individual steps of the three-stage method is familiar from system identification. Hence, following the approach of Ljung and Söderström, see [47], recursive identification methods can be developed in a natural manner. The resulting on-line estimation procedures can be mathematically rigourously analyzed either using the theory of L-mixing processes, see [32], or using the Markovian approach, see [6].

In Chapter 8 we study the stability of time-varying stochastic systems driven or modulated by a Lévy process with discrete time interventions, such as parameter or state resetting. We hope that the technical results presented in that chapter may contribute to the development of a continuous-time recursive maximum likelihood method for finite dimensional linear stochastic Lévy systems, along the lines of [38].

The objective of Chapter 9 is to present some numerical results. We apply both the off-line and the on-line identification method for simulated ARMA processes. The system parameters of a simulated GARCH process will be estimated, too. Our aim is to demonstrate that the

proposed algorithms are indeed able to give an adequate estimate of the system parameters. Finally, the last chapter gives a summary of the research.

1.1 Notations

$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex numbers
$\mathbb{R}^d, \mathbb{R}^{d\times d}$	set of real d-dimensional vectors, set of real $d \times d$ matrices
$\mathbb{C}^d, \mathbb{C}^{d\times d}$	set of complex d-dimensional vectors, set of complex $d \times d$ matrices
$\operatorname{int} S$	interior of the set S
$\lvert v\rvert$	norm of vector v
$\sigma(M)$	spectral radius of matrix M
$\mathbf{1}_S$	indicator function of the set or event S
q^{-1}	backward shift operator
$\to$	convergence
$\mathscr{F}^X$	natural filtration of the stochastic process $(X)_t$
$\mathbb{E}[X]$	expectation of X
$\operatorname{Var}(X)$	variance of X
$\operatorname{Cov}(X, Y)$	covariance of X and Y
$\langle X, Y\rangle$	predictable quadratic covariation of X and Y
$[X, Y]$	quadratic covariation of X and Y
$[X, X]$ or $[X]$	quadratic variation of X
$A \otimes B$	Kronecker product of A and B
$\langle f, g\rangle_H$	inner product on H
$C^k(S)$	set of k-times continuously differentiable functions of S

Chapter 2

Lévy Processes

2.1 Motivation, basic properties

Lévy processes have become a widely used tool in modeling price processes of financial instruments, such as stock prices or indices [52]. A Lévy process (L_t) is much like a Wiener process: a process with stationary an independent increments, but discontinuities or jumps are allowed. Hence, Lévy processes can be used to model shocks in financial markets. In chapter we follow the excellent introduction to Lévy processes of Jacod and Shiryaev, see [42].

Suppose that we are given a probability space $(\Omega, \mathscr{F}, P)$. In general, a Lévy process is a stochastic process defined as follows:

Definition 2.1. *We say that $(L_t), t \geq 0$ is a Lévy process if*

1. $L_0 = 0$,
2. *for any given $0 \leq t_1 < \ldots < t_n$, the random variables $L_{t_2} - L_{t_1}, L_{t_3} - L_{t_2}, \ldots,$ $L_{t_n} - L_{t_{n-1}}$ are independent.*
3. *for any $0 \leq s < t$, the distribution of $L_t - L_s$ and L_{t-s} are the same.*

A key building block in the theory of Lévy processes is the compound Poisson process. A more general class of pure jump Lévy process is formally obtained via

$$L_t = \int_0^t \int_{\mathbb{R}} xN(ds, dx), \tag{2.1.1}$$

where $N(dt, dx)$ is a time-homogeneous, space-time Poisson point process, counting the number of jumps of size x at time dt.

The intensity of $N(dt, dx)$ is defined by

$$\mathbb{E}\left[N(dt, dx)\right] = dt \cdot \nu(dx),$$

where $\nu(dx)$ is the Lévy-measure. Intuitively, $\nu(x)$ can be interpreted as the rate of jumps with size of x. Its simplest form is $\nu = \lambda\delta_h$, where δ_h denotes the Dirac measure concentrated at h. In this case (L_t) is a Poisson process of intensity λ and with jumps of size h. If $\nu = \sum_{i=1}^{n} \lambda_i \delta_{h_i}$, then (L_t) is a sum of independent Poisson processes. When ν is finite we can pass to the limit to obtain a process that has jumps with sizes in the full continuum. All other processes can be obtained by a limiting procedure in L^2, leading to the condition:

$$\int_{\mathbb{R}\setminus 0} \min(|x|^2, 1)\nu(dx) < \infty.$$

The above representation in (2.1.1) is mathematically rigorous if

$$\int_{\mathbb{R}} \min(|x|, 1)\nu(dx) < \infty. \tag{2.1.2}$$

Under this condition the sample paths of (L_t) are of *finite variation*, a property supported by empirical evidence for most indices.

Note that the logarithm of the characteristic function is linear in t, which is implied by the

fact that (L_t) has independent and stationary increments. The characteristic function plays a key role in the study of Lévy processes, because unlike the density function of L_t it typically has a closed form. The c.f. of a Lévy process is given by the following celebrated Lévy-Khintchine formula:

Theorem 2.1. *Let (L_t) be a Lévy process. Then there exist a triplet (b, c, ν), with $b, c \in \mathbb{R}, c \geq 0$, and ν being a Lévy measure satisfying $\nu(0) = 0$ and $\int_{\mathbb{R}} \min(|x|^2, 1)\nu(dx) < \infty$, such that*

$$\mathbb{E}\left[e^{iuL_t}\right] = \exp\left[t\left(ibu - \frac{u^2c}{2} + \int_{\mathbb{R}}\left(e^{iux} - 1 - iux\mathbf{1}_{|x|<1}\right)\nu(dx)\right)\right].$$

For pure-jump Lévy processes with finite variation trajectories we have the following simplified form of the Lévy-Khintchine formula:

Theorem 2.2. *Let (L_t) be a pure-jump Lévy process (having no Brownian motion component), with finite variation trajectories, then*

$$\mathbb{E}\left[e^{iuL_t}\right] = \exp\left[t\left(ibu + \int_{\mathbf{R}}\left(e^{iux} - 1\right)\nu(dx)\right)\right].$$

2.2 Semimartingales

In this section we present a basic background for the study of semimartingales following [42]. Let us use the notation $X_{t-} = \lim_{s\to t^-} X_s$. We first define the quadratic co-variation.

Definition 2.2. *The quadratic co-variation of two semimartingales X and Y is defined by $[X, Y]_t = X_tY_t - X_0Y_0 - \int_0^t X_{s-}dY_s - \int_0^t Y_{s-}dX_s$. If $X = Y$, then we get the quadratic variation of X.*

Proposition 2.1. *Let (X_t) be a semimartingale and (Y_t) be an adapted cádlág process with finite variation and $Y_0 = 0$. Then the $[X, Y]_t$ has the following properties*

- $[X,Y]_t = \int_0^t \Delta X_s dY_s$
- *If Y is predictable, then $[X,Y]_t = \int_0^t \Delta Y_s dX_s$ and $X_t Y_t = \int_0^t Y_s dX_s + \int_0^t X_{s-} dY_s$*
- *If Y is predictable and X is a local martingale, then $[X,Y]$ is a local martingale.*
- *If X or Y is continuous then $[X,Y] = 0$.*

The quadratic variation of purely discontinuous square integrable martingales can be written in a compact form.

Lemma 2.1. *Let X be a purely discontinuous square integrable martingale, then $[X,X]_t = \sum_{s\leq t}(\Delta X_s)^2$.*

The notation $\langle X,Y\rangle$ stands for the predictable quadratic variation of X and Y.

Theorem 2.3. *If X, Y are semimartingales with X^c, Y^c denoting their continuous martingale parts, respectively, then*

$$[X,Y]_t = \langle X^c, Y^c\rangle_t + \sum_{s\leq t} \Delta X_s \Delta Y_s.$$

Now we state the well-known Itô formula for semimartingales.

Theorem 2.4. *(Itô formula) Let $X = (X^1, \ldots, X^d)$ be a d-dimensional semimartingale and $f \in C^2(\mathbb{R}^d)$, then $f(X)$ is a semimartingale and*

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t D_i f(X_{s-}) dX_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t D_{i,j} f(X_{s-}) d < X^{i,c}, X^{j,c} > + \sum_{0\leq s\leq t} \left[f(X_s) - f(X_{s-}) - \sum_{i=1}^d D_i f(X_{s-}) \Delta X_s^i \right], \tag{2.2.1}$$

where D_i is the differential operator w.r.t. the i^{th} variable, and $D_{i,j}$ is the differential operator w.r.t. the i^{th} then the j^{th} variable.

The next results gives the Doléan-Dade formula stochastic exponential of a semimartingale.

Theorem 2.5. *Let X_t be a real-valued semimartingale. Then the stochastic differential equation $dY_t = Y_{t-}dX_t, Y_0 = 1$ has a one and only one (up to indistinguishability) cádlág solution. The solution is a semimartingale, denoted by $\varepsilon(X)$, and has the form of:*

$$\begin{aligned}\varepsilon(X)_t = \exp\left\{X_t - X_0 - \frac{1}{2}\langle X^c, X^c\rangle_t\right\} \\ \times \prod_{s\leq t}\left[(1+\Delta X_s)e^{-\Delta X_s}\right],\end{aligned} \tag{2.2.2}$$

where everything is well-defined and convergent. Moreover, if X_t has finite variation then so has $\varepsilon(X)$, and if X_t is a local martingale, then so is $\varepsilon(X)$.

The previous two results can be written in a more compact form for processes with finite variation.

Theorem 2.6. *(Itô formula for processes with finite variation) Let X be a semimartingale with finite variation and $f \in C^1$, then $f(X)$ is a semimartingale and*

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})dX_s^c + \sum_{0\leq s\leq t}\left[f(X_s) - f(X_{s-})\right]. \tag{2.2.3}$$

Note that if the continuous part X^c of X is zero, then the Itô formula gives f as sum of its jumps. As for the stochastic exponential, if X is a real-valued semimartingale with finite variation, then

$$\varepsilon(X)_t = e^{X_t - X_0}\prod_{s\leq t}\left[(1+\Delta X_s)e^{-\Delta X_s}\right] = e^{X_t^c}\prod_{s\leq t}\left[(1+\Delta X_s)\right].$$

2.3 Lévy processes in finance

To model the increments of the logarithm of a price process a wide range of geometric Lévy processes has been proposed by a variety of authors. In this section we present some Lévy

Figure 2.1: IBM stock price

processes that are widely used in financial modelling. In Figure 2.1, 2.2 and 2.3 real word data can be seen for IBM, Coke and MSFT, respectively[1]. Observe that shocks indeed occur in financial markets, this fact motivates the usage of the exponential Lévy model.

Compound Poisson process is defined by a rate λ and a jump size distribution F via

$$L_t = \sum_{i=1}^{N_t} X_i,$$

where N_t is a Poisson process with rate λ, and X_i-s are i.i.d. random variables with distribution F. The first model that considered taking the exponential of a compound Poisson process was introduced by Merton in [51]. Compound Poisson processes are also widely used for modeling in queueing theory. For example in [23] a generalized multi-server queue is used to model telecommunication networks. Among several properties of the model, customer arrivals, server failures and packet losses are modeled with compound Poisson processes.

[1]The source of the real word data and the figures is www.nasdaq.com

Figure 2.2: Coca-Cola stock price

Figure 2.3: Microsoft stock price

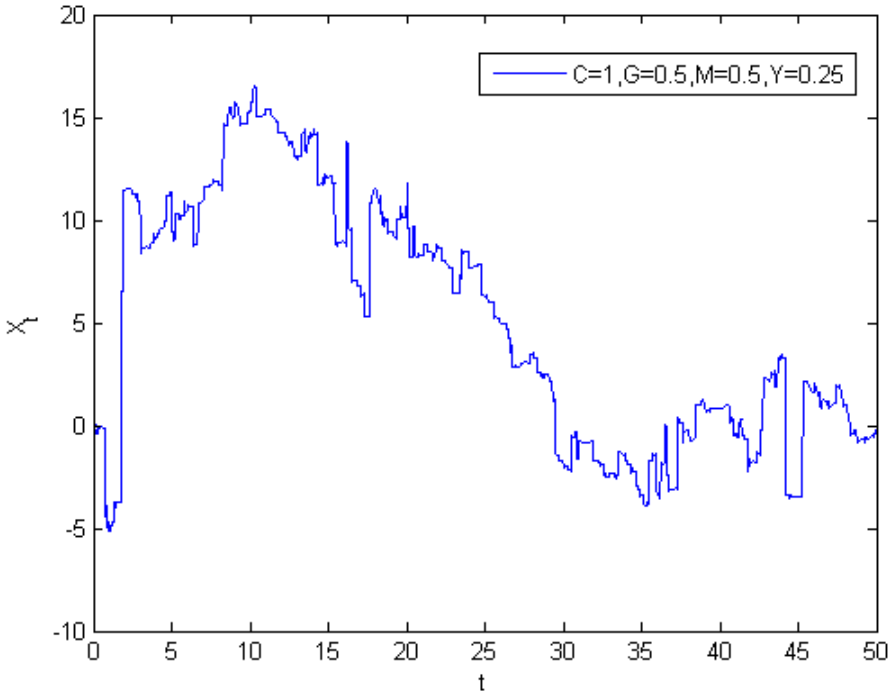

Figure 2.4: CGMY sample path

Mandelbrot suggested to use α-stable process to model the price dynamics of wool, see [50]. An α-stable process with $0 < \alpha < 2$ is defined via the Lévy measure

$$\nu(dx) = C^-|x|^{-1-\alpha}\mathbf{1}_{x<0}dx + C^+|x|^{-1-\alpha}\mathbf{1}_{x>0}dx.$$

A recently widely studied class of Lévy processes is the CGMY process due to Carr, Geman, Madan and Yor [13], see also [55], [2] and [1]. It is obtained by setting $C^- = C^+$, and then, separately for $x > 0$ and $x < 0$, multiplying the Lévy-density of the original symmetric stable process with a decreasing exponential. The corresponding Lévy-measure, using standard parametrization, is of the form:

$$\nu(dx) = \frac{Ce^{-G|x|}}{|x|^{1+Y}}\mathbf{1}_{x<0}dx + \frac{Ce^{-Mx}}{|x|^{1+Y}}\mathbf{1}_{x>0}dx,$$

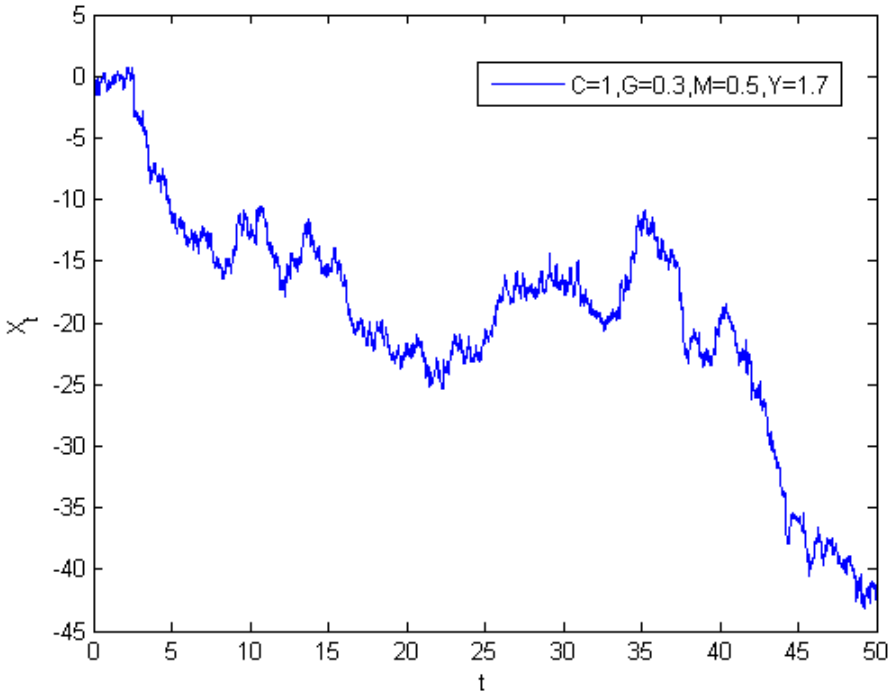

Figure 2.5: CGMY sample path

where $C, G, M > 0$, and $0 < Y < 2$. Intuitively, C controls the level of activity, G and M together control skewness. Typically $G > M$ reflects the fact that prices tend to increase rather than decrease. Y controls the density of small jumps, i.e. the fine structure. For $Y < 1$ the integrability condition (2.1.2) is satisfied, thus corresponding Lévy process is of finite variation. The characteristic exponent $\psi(u)$ of the CGMY process is given by

$$C\Gamma(-Y)\left((M-iu)^Y - M^Y + (G+iu)^Y - G^Y\right),$$

where Γ denotes the gamma-function. In Figure 2.4 and 2.5 sample paths of CGMY processes can be seen. A more general class of tempered stable distributions is studied in [62], Terdik and Woyczński obtain analytic formulas for the Rosiński measure of tempered processes.

Formally setting $Y = 0$ we get the Lévy density of the so-called Variance Gamma process that has been proposed by Madan and Senata in [49], see also [48] and [41]. A VG-process

is a time changed Brownian motion with drift, where the time change is a so-called gamma process, which is essentially the continuous time extension of the inverse of a Poisson process.

A gamma process with mean μ and variance ν, denoted by $(\gamma_t(\mu,\nu))$, is a stochastic process with independent increments, such that the distribution of the increment $\gamma_{t+h}(\mu,\nu)-\gamma_t(\mu,\nu)$ is a gamma distribution with mean μh and variance νh. The characteristic function of $\gamma_t(\mu,\nu)$ is given by

$$\varphi_t(u;\mu,\nu)=\left(\frac{1}{1-i\frac{\nu}{\mu}u}\right)^{\mu^2\frac{t}{\nu}}.$$

Note that the $\varphi_t(u;\mu,\nu)$ is a fractional power of the c.f. of an exponential distribution. Its worth mentioning that by scaling the process $\mu=1$ can be achieved. The Lévy measure of the gamma process is given by

$$\nu(dx)=\frac{\mu^2}{\nu}\frac{\exp(-\frac{\mu}{\nu}x)}{x}dx \quad \text{if } x>0,$$

and 0 otherwise. Note that the integral of $\nu(dx)$ is infinite, hence the gamma process has an infinite number of jumps in any finite interval. It is also said that the gamma process is an infinite activity process. Clearly, most of these jumps are very small as the Lévy measure is concentrated at the origin.

VG is a three-parameter class of processes, with explicit characteristic function and Lévy measure obtained as follows. Let $(B_t(\theta,\sigma))$ be a Brownian motion with drift θ and volatility σ, i.e.:

$$B_t(\theta,\sigma)=\theta t+\sigma B_t,$$

where (B_t) is a standard Brownian motion. The VG process $(X_t(\sigma,\nu,\theta))$ is then defined as

$$X_t(\sigma,\nu,\theta)=B_{\gamma_t(1,\nu)}(\theta,\sigma).$$

Hence, the VG process is a time-changed Brownian motion. Note that the mean μ is fixed to be 1, to avoid the ambiguity of parametrization due to re-scaling. According to [48] θ controls the skewness and ν controls the kurtosis of the process.

The Lévy-measure of a VG process can be obtained by first computing its characteristic function and then applying Lévy-Khintchine's formula in the inverse direction. Thus we get :

$$\nu(dx) = \begin{cases} \frac{\mu_n^2}{\nu_n} \frac{\exp(-\frac{\mu_n}{\nu_n}|x|)}{|x|} dx & \text{if } x < 0 \\ \frac{\mu_p^2}{\nu_p} \frac{\exp(-\frac{\mu_p}{\nu_p}x)}{x} dx & \text{if } x > 0, \end{cases}$$

where some parameters $\mu_p, \nu_p, \mu_n, \nu_n$, which is indeed formally the Lévy measure of a CGMY process with $Y = 0$. The parameters $\mu_p, \nu_p, \mu_n, \nu_n$, are obtained in terms of the original parameters as follows:

$$\begin{aligned} \mu_p &= \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} + \frac{\theta}{2}, \qquad \nu_p = \mu_p^2 \nu \\ \mu_n &= \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} - \frac{\theta}{2}, \qquad \nu_n = \mu_n^2 \nu \end{aligned}$$

Note that on the right hand side we have the difference of the Lévy measures of two gamma processes. Thus we get the following remarkable property of VG processes: a VG process $(X_t(\sigma, \nu, \theta))$ can be written as the difference of two gamma processes $(\gamma_{p,t})$ and $(\gamma_{n,t})$:

$$X_t(\sigma, \nu, \theta) = \gamma_{p,t}(\mu_p, \nu_p) - \gamma_{n,t}(\mu_n, \nu_n).$$

In particular, it follows that a VG process is of finite variation.

Chapter 3

ECF method for i.i.d. data

3.1 ECF with finite set of moment conditions

In this section we present the ECF method for i.i.d. data when a finite set of moment conditions is available. Although henceforth we will adapt the ECF method for more complex problems, the basic idea of the ECF method can be nicely presented even for i.i.d. data.

In this section we briefly describe the ECF method for i.i.d. samples, see [16]. A remarkable property of the ECF method is that, under idealistic conditions, it gives an efficient estimate of the unknown parameters of a given family of distributions [16]. A nice heuristic justification for this has been given by A. Feuerverger and P. McDunnogh in [28], showing that the equations defining the ECF method for i.i.d samples can be obtained as the Fourier transform of the likelihood equations.

Let $(r_1, r_2, \dots r_N)$ be i.i.d. observations, and let a closed form of the characteristic function $\varphi(u, \eta)$ be known, with η being a p-dimensional parameter vector, and $u \in \mathbb{R}$. The true value of the parameter will be denoted by η^*. The idea is to estimate η^* by a value of η for which the characteristic function (c.f.) best matches the empirical characteristic function (ECF). For

any fixed $u \in \mathbb{R}$ define the primary error or score as the generalized moment function:

$$h_n(u, \eta) = e^{iur_n} - \varphi(u, \eta).$$

Averaging over the samples, i.e. for $n = 1, \ldots, N$ we define a mean-error or secondary score

$$\overline{h}_N(u, \eta) = \frac{1}{N} \sum_{n=1}^{N} h_n(u, \eta).$$

An important property of the primary score or generalized moment function is that for $\eta = \eta^*$ we have

$$\mathbb{E}\left[h_n(u, \eta^*)\right] = 0, \quad \text{for all} \quad u.$$

Taking a finite set of u-s say $u_1, \ldots, u_M$, with $M > p$, let

$$h_{k,n}(\eta) = h_n(u, \eta) = e^{iu_k r_n} - \varphi(u_k, \eta).$$

Define the vector-valued mean-error as

$$\overline{h}_N(\eta) = (\overline{h}_N(u_1, \eta), \ldots, \overline{h}_N(u_M, \eta))^T,$$

and let us denote its expectation by

$$g(\eta) = \mathbb{E}\left[\overline{h}_N(\eta)\right].$$

Here $\dim g(\eta) = M > p$. Since $g(\eta^*) = 0$, we could, in principle, obtain η^* by solving the over-determined system of equations

$$g(\eta) = 0. \tag{3.1.1}$$

In our case of i.i.d. samples the entries of g are of the form $\varphi(u_k, \eta^*) - \varphi(u_k, \eta)$. Note

that g is not computable since η^* is unknown. Therefore we replace g by $\overline{h}_N(\eta)$, and seek the least-squares solution of the set of over-determined system of equations

$$\overline{h}_N(\eta) = 0. \tag{3.1.2}$$

More precisely, we minimize the weighted quadratic cumulative error

$$V_N(\eta) = |K^{-\frac{1}{2}}\overline{h}_N(\eta)|^2, \tag{3.1.3}$$

where K is an appropriate, $m \times m$, symmetric, positive definite weighing matrix, as proposed in [15]. Then, in Theorem 2 in [15] it is proved that under certain conditions the asymptotic distribution of $\sqrt{N}(\hat{\eta}_N - \eta^*)$ is normal, and a nice expression for its asymptotic covariance matrix is given.

It should be noted though, that in the proof of Theorem 2 of [15] $\hat{\eta}_N$ is assumed to be defined as the solution of the non-linear algebraic equation

$$\overline{h}^*_{\eta N}(\hat{\eta}_N)K^{-1}\overline{h}_N(\hat{\eta}_N) = 0, \tag{3.1.4}$$

tacitly assuming that the l.h.s. is $1/2$-times the gradient of $V_N(\eta)$ at $\hat{\eta}_N$. Here * denotes the complex conjugate of a vector or of a matrix. However, since the scores are complex valued, the actual gradient of $V_N(\eta)$ at $\eta = \hat{\eta}_N$ is equal to the l.h.s. of (3.1.4) plus its conjugate. Therefore the l.h.s. side of (3.1.4) will be loosely called a half-gradient.

The ambiguity in defining $\hat{\eta}_N$ in [15] may be due to the fact that the ECF method was derived there as a modification of the generalized moment method (GMM) developed earlier by Hansen, see [40]. But the latter author restricted himself to real valued score functions, in which case the "half gradient" is indeed $1/2$-times the gradient of $V_N(\eta)$. Going into further details of the proof one notes that the reasoning can be extended to complex valued function

by using the mean value theorem in its integral form. Hence, in what follows, we propose to define $\hat{\eta}_N$ as the possibly complex-valued solution of

$$\overline{h}^*_{\eta N}(\eta) K^{-1} \overline{h}_N(\eta) = 0. \tag{3.1.5}$$

In defining "the" solution of (3.1.4) we proceed as in [31]. $\hat{\eta}_N$ is a random variable such that $\hat{\eta}_N \in D_\eta$ for all ω. If the equation has multiple solutions then $\hat{\eta}_N$ is defined as any of them, and if the equation has no solution then $\hat{\eta}_N$ is any $\eta \in D_\eta$. The measurable selection theorem implies that such a random variable exists. Along the lines of Lemma 2.3. in [31] it can be shown that for any $d > 0$ equation (3.1.5) has unique solution in $\{|\eta - \eta^*| < d\}$ with probability at least $1 - O(N^{-s})$ for any $s > 0$.

The half-gradient given on the l.h.s. can be considered as a new set of score functions. The choice of this score functions are very nicely motivated by the arguments of Feuerverger and McDunnogh, see [28], and the remark at the end of the section.

To compute the asymptotic covariance matrix of the estimated parameter $\hat{\eta}_N$ we follow standard procedures. Again, the 'half-gradient' of $V_N(\eta)$, w.r.t. η, set equal to zero gives the p equations:

$$\overline{h}^*_{\eta N}(\eta) K^{-1} \overline{h}_N(\eta) = 0.$$

Note that the expectation of the right hand side is not necessarily 0 for $\eta = \eta^*$, since the mean errors $\overline{h}^*_{\eta,N}(\eta^*)$ and $\overline{h}_N(\eta^*)$ are not necessarily independent. Approximate the first term by its expectation denoted by

$$G = g_\eta(\eta^*), \tag{3.1.6}$$

where G is an $M \times p$ matrix, and define an approximating problem as

$$G^* K^{-1} \overline{h}_N(\eta) = 0. \tag{3.1.7}$$

It can be shown that using this approximation we get an estimator with asymptotic covariance matrix identical to that of $\hat{\eta}_N$.

The left hand side can be considered as a new set of exactly p scores with expectation $G^* K^{-1} g(\eta)$, and the corresponding auxiliary equation is given by

$$G^* K^{-1} g(\eta) = 0. \tag{3.1.8}$$

The sensitivity matrix or derivative of the left hand side of (3.1.8) at η^* is given by

$$T = G^* K^{-1} G.$$

To get the asymptotic covariance of the new set of scores defined by the left hand side of (3.1.7) we define the $M \times M$ covariance matrix, for any n, as

$$C_{k,l} = \mathbb{E}\left[\ h_{k,n}(\eta^*) h^*_{l,n}(\eta^*)\right].$$

Note that $C_{k,l}$ is explicitly known:

$$C_{k,l} = \varphi(u_k - u_l, \eta^*) - \varphi(u_k, \eta^*)\ \varphi(-u_l, \eta^*).$$

Then the asymptotic covariance of the new set of scores can be written as

$$S = G^* K^{-1} C K^{-1} G.$$

Finally, under suitable technical conditions, the asymptotic covariance matrix of the estimator $\hat{\eta}_N$ is obtained by standard arguments as

$$\Sigma_{\eta\eta} = \lim_{N\to\infty} \mathbb{E}\left[N(\hat{\eta}_N - \eta^*)(\hat{\eta}_N - \eta^*)^*\right] = T^{-1} S T^{-1}.$$

Equivalently, we have

$$\Sigma_{\eta\eta} = (G^* K^{-1} G)^{-1} \; G^* K^{-1} C K^{-1} G \; (G^* K^{-1} G)^{-1}.$$

It is easy to see that $\Sigma_{\eta\eta}$ is minimized for

$$K = C,$$

yielding the optimal asymptotic covariance matrix for $\hat{\eta}_N$:

$$\Sigma_{\eta\eta} = (G^* C^{-1} G)^{-1},$$

or equivalently

$$\Sigma_{\eta\eta} = \left(\varphi_\eta^*(\eta^*) C^{-1} \varphi_\eta(\eta^*)\right)^{-1},$$

with

$$\varphi_\eta(\eta) = \left(\varphi_\eta(u_1, \eta), \ldots, \varphi_\eta(u_M, \eta)\right)^T.$$

3.2 ECF with full continuum of moment conditions

In this section we summarize the results of Carasco and Florens [16]. The paper in question extends the general method of moments by using the full continuum of moment conditions given by the empirical characteristic function. Suppose we are given an i.i.d. realization $X_1, \ldots, X_N$ of the random variable X. The characteristic function of X is $\varphi(u, \eta^*)$, where η^* denotes the unknown parameter vector, that is to be estimated. Let $\varphi_N(u)$ denote the empirical characteristic function

$$\varphi_N(u) = \frac{1}{N} \sum_{j=1}^{N} e^{iuX_j}.$$

The moment conditions proposed by the authors are

$$h(u, X_j, \eta) = e^{iuX_j} - \varphi(u, \eta), \tag{3.2.1}$$

clearly

$$\mathbb{E}\left[h(u, X_j, \eta^*)\right] = 0$$

holds for all $u \in \mathbb{R}$ and for each j. Fix a probability density function π and define the Hilbert space $L^2(\pi)$ by

$$L^2(\pi) = \left\{ f : \mathbb{R} \to \mathbb{C} \,\middle|\, \int_{\mathbb{R}} |f(t)|^2 \pi(t) dt < \infty \right\}.$$

The natural definition of the the inner product on this Hilbert space is

$$\langle f, g \rangle_H = \int_{\mathbb{R}} f(t) g^*(t) \pi(t) dt.$$

Let B be a bounded linear operator defined on $L^2(\pi)$, or on a subspace of it, and let B_N be a sequence of linear operators converging to B. Define the averaged scores by

$$\overline{h}_N(u, \eta) = \frac{1}{N} \sum_{j=1}^{N} h(u, X_j, \eta).$$

In [15] authors define the estimated parameter vector via

$$\hat{\eta}_N = \arg\min_{\eta} ||B_N \overline{h}_N(\cdot, \eta)||. \tag{3.2.2}$$

They also note that for a finite, say M, number of moment condition the above norm in (3.2.2) can be written as follows: first define $u = (\frac{1}{M}, \frac{2}{M}, \ldots, 1)$, and stack $\overline{h}_N(u, \eta)$-s, $u = (\frac{1}{M}, \frac{2}{M}, \ldots, 1)$ into an M-dimensional column vector $\overline{h}_N(\eta)$. Then, for a given $M \times M$, positive

definite matrix A_M the estimate $\hat{\eta}_N$ is given by

$$\hat{\eta}_N = \arg\min_{\eta} \overline{h}_N^*(\eta) A_M \overline{h}_N(\eta). \tag{3.2.3}$$

Which leads us to the equation that we have mentioned in the previous section. Again, while this seems a reasonable definition of $\hat{\eta}_N$ it turns out that this method does not give an efficient estimate. Thus we define $\hat{\eta}_N$ as the solution of the 'half-gradient' equation

$$\int_{\mathbb{R}} \left(B_N \overline{h}^*_{\eta N}(t, \eta) \right) \left(B_N \overline{h}_N(t, \eta) \right) \pi(t) dt = 0. \tag{3.2.4}$$

Nevertheless the analysis in [16] suggests that the authors work with the latter definition of $\hat{\eta}_N$. It was also proved that the operator B that gives an estimator with minimal asymptotic variance is given by $B = K^{-1/2}$, where operator K is defined as follows:

$$f : L^2(\pi) \to L^2(\pi) \tag{3.2.5}$$

$$f(t) \to g(s) = \int_{\mathbb{R}} k(s,t) f(t) \pi(t) dt, \tag{3.2.6}$$

where

$$k(s,t) = \mathbb{E}\left[h(s, X, \eta^*) h^*(t, X, \eta^*) \right].$$

We note here that the domain of $K^{-1/2}$ is not the whole $L^2(\pi)$, but a subset of it, which corresponds to the reproducing kernel Hilbert space (RKHS) of K, denoted by $\mathscr{H}(K)$. The norm $||.||_K$ in $\mathscr{H}(K)$ is denoted by

$$||g||_K^2 = ||K^{-1/2} g||^2.$$

According to Parzen, see Chapter 3 in [54], in case of a covariance kernel K the elements of $\mathscr{H}(K)$ can be characterized by an integral representation theorem. For the present case the

theorem implies that $\mathscr{H}(K)$ consists of all functions f that can be written as

$$f(t) = \mathbb{E}\left[G(X)h^*(t, X, \eta^*)\right] \tag{3.2.7}$$

with some unique G, which is the element of the Hilbert subspace spanned by $h^*(t, \cdot\,, \eta^*)$.

It can be proved that

$$\hat{\eta}_N \to \eta^* \text{ in probability}$$

under certain technical conditions. It is also showed that

$$\sqrt{n}\left(\hat{\eta}_N - \eta^*\right) \to \mathscr{N}\left(0, \left(||\mathbb{E}_{\eta^*}\left[h_\eta\right]||_K^2\right)^{-1}\right) \tag{3.2.8}$$

The most interesting result of the paper is that the proposed method gives an asymptotically efficient estimator. Here, we sketch the proof of this last statement.

At first, we show that $g = \varphi_\eta^* \in \mathscr{H}(K)$. For simplicity we assume that η is scalar, the idea of the proof is essentially the same if φ_η^* is a vector. To this end we use the integral representation theorem for $\mathscr{H}(K)$. If there exists a function G in L^2 such that $\mathbb{E}\left[G(X)\right] = 0$ and

$$g(u) = \varphi_\eta^*(u) = \mathbb{E}_{\eta^*}\left[G(X)e^{-iuX}\right],$$

then $g \in \mathscr{H}(K)$ follows. Denote the set of such G-s by $C(g)$, i.e.

$$C(g) = \left\{G \,\middle|\, \mathbb{E}\left[G^2(X)\right] < \infty,\ \mathbb{E}\left[G(X)\right] = 0,\ g(t) = \mathbb{E}\left[G(X)h^*(t, X, \eta^*)\right]\right\}.$$

We show that $C(\varphi_\eta^*)$ is non-empty and what is more we give the elements of $C(\varphi_\eta^*)$.

$$\begin{aligned}\varphi_\eta(-u) &= \mathbb{E}_{\eta^*}\left[G(X)e^{-iuX}\right] = \int_{\mathbb{R}} G(x)e^{-iux}f(x,\eta^*)dx \\ G(x)f(x,\eta^*) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{iux}\varphi_\eta(u)f(x,\eta^*)du \\ G(x)f(x,\eta^*) &= f_\eta(x,\eta)\end{aligned}$$

Hence

$$G(x) = \frac{f_\eta(x,\eta)}{f(x,\eta^*)},$$

which means that $\varphi_\eta^* \in \mathscr{H}(K)$ indeed holds. Moreover, an other result of Parzen, which can be found in [54], reads as

Theorem 3.1. *For any $g \in \mathscr{H}(K)$, the norm of g can be characterized as*

$$||g||_K^2 = \min_{G \in C(g)} \mathbb{E}_{\eta^*}\left[G^2\right]. \tag{3.2.9}$$

Using this characterization with $g = \varphi_\eta^*$ we obtain that

$$||\varphi_\eta^*||_K^2 = \min_{G \in C(\overline{\varphi_\eta})} \mathbb{E}_{\eta^*}\left[G^2\right] = \min_\eta \mathbb{E}\left[\left(\frac{f_\eta(x,\eta)}{f(x,\eta^*)}\right)^2\right]. \tag{3.2.10}$$

It is relatively easy to show that the r.h.s. reaches its minimum value at $\eta = \eta^*$, so that we get

$$||\varphi_\eta^*||_K^2 = \mathbb{E}\left[\left(\frac{f_\eta(x,\eta^*)}{f(x,\eta^*)}\right)^2\right],$$

thus the asymptotic efficiency of the estimate follows.

A nice interpretation of the ECF method is obtained by considering the log-likelihood equation

$$\int_{-\infty}^{\infty} \frac{\partial \log f(x,\eta)}{\partial \eta}(f_N(x,\eta^*) - f(x,\eta))dx = 0, \tag{3.2.11}$$

where f_N and f stand for the empirical pdf and the pdf, respectively. Taking the Fourier transform of the pdf-s by using Parseval's theorem yields

$$\int_{-\infty}^{\infty} w(t,\eta)(\varphi_N(t,\eta^*) - \varphi(t,\eta))dt = 0, \tag{3.2.12}$$

where φ_N is the empirical characteristic function and

$$w(t,\eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\partial \log f(x,\eta)}{\partial \eta} e^{-itx} dx.$$

The l.h.s. of (3.2.12) can be interpreted as a weighted combination of the primary ECF scores. It can be shown that replacing η by η^* in $w(t,\eta)$ does not effect the asymptotic covariance of the modified estimate $\hat{\eta}_N$, obtained as the solution of the modified version of (3.2.12). Since the ML estimate is asymptotically optimal it follows that the weighting function $w(t,\eta^*)$ is optimal among all feasible weighting functions.

Comparing this observation with the result of the previous section gives that, if the full continuum of u-s is used, then the weighting function $w(t,\eta^*)$ is equivalent to the weighting G^*K^{-1} meaning that

$$MG^*K^{-1} = w(t,\eta^*),$$

where M is a non-singular transformation.

We can conclude that the ML method is equivalent to fitting $\varphi(t,\eta)$ to the empirical c.f. $\varphi_N(t)$ using appropriate weights. We note that the above computation can be carried out only formally because the score functions are not necessarily integrable w.r.t. the Lebegue measure.

Not surprisingly the characteristic function fitting equation of Feuerverger and McDunnough can be transformed to the 'half-derivative' equation of Carrasco and Florens as follows: for simplicity write K instead of K_N, and use that $h_{\eta,n} = -\varphi_\eta$. The latter method gives the

estimated parameter as the solution of

$$\langle K^{-1/2}\varphi_\eta, K^{-1/2}\varphi_\eta\rangle_H = 0.$$

If φ_η were in the range of K^{-1} then the last equation could be written as

$$\langle K^{-1}\varphi_\eta, \varphi_\eta\rangle_H = 0, \tag{3.2.13}$$

which has the same form as (3.2.12). However, these calculation can be carried out only formally as φ_η is not in the range of K^{-1} and the score function is not integrable with respect to the Lebesgue measure.

3.3 Discussion

Intuitively one could reason that using the full continuum of u-s would give a more efficient estimate than using only a finite set of moment conditions. Indeed, efficiency can be achieved only with the full continuum of moment conditions. Yet, by increasing M, the number of fixed u-s, we get more and more efficient estimates, as it is pointed out by Carrasco and Florens in [15]. Following their line of argument we define $H = L^2[0,1]$, then the asymptotic variance of the ECF estimate with the continuum of moment condition is

$$||\mathbb{E}[h_\eta]||_K^{-2}.$$

On the other hand, choose H_N to be the Hilbert space of 2^N dimensional vectors, and let $u_k = k/2^N$. In this case the optimal weighting matrix $K^{(N)} = (K^{(N)})_{k,l}$ is given by

$$K_{k,l} = \mathbb{E}\left[h_{\frac{k}{2^N},n}(\eta^*)h^*_{\frac{l}{2^N},n}(\eta^*)\right],$$

for any n. The asymptotic covariance matrix of the estimator is then

$$||\mathbb{E}[h_\eta]||^{-2}_{K^{(N)}}.$$

The above variances can be compared using the result of Parzen, see [54] page 316-318. The result reads as follows: if $f \in \mathscr{H}(K)$, then

$$||f||_K \geq ||f||_{K^{(N+1)}} \geq ||f||_{K^{(N)}}.$$

Moreover,

$$||f||_{K^{(N)}} \to ||f||_K \text{ as } N \to \infty,$$

which means that the variance of the estimate can be arbitrarily close to the Cramer-Rao bound if we use a sufficiently large number of moment conditions.

Chapter 4

ECF method for linear Lévy systems

The objective of this chapter is to present a combination of advanced techniques in systems identification with a specific statistical technique, widely used in the context in finance, called the ECF (shorthand for empirical characteristic function) method. The ECF method was originally designed for i.i.d. samples and A. Feuerverger and P. McDunnogh [28] showed that it can be interpreted as the Fourier transform of an ML method. Several papers study the problem of identifying the noise characteristics of a linear system, but only a few pays attention to the problem of identifying the system parameters as well. Brockwell and Schlemm [11] consider the parametric estimation of the driving Lévy process of a multivariate continuous-time ARMA processes, but the identification of system parameters is out of the scope of their paper. Calder and Davis [12] discuss the M-estimators of ARMA processes with a given distribution on the noise process. The quasi-maximum likelihood estimation of multivariate Lévy-driven continuous-time ARMA processes is studied by Schlemm and Stelzer in [61], the method presented there identify the system parameters and the covariance structure of the noise process, but further characteristics of the driving noise are not estimated.

In this chapter we present a three-stage identification method for single-input-single-output

(SISO) that estimates both the system parameters and the characteristics of the noise process. We give the precise characterization of the estimation error as well. We adapt the ECF method for linear systems and demonstrate that our method can outperform standard system identification methods such as prediction error method or quasi maximum likelihood estimation method and that it is essentially asymptotically efficient in the sense discussed in Section 4.6. The results of this chapter are based on the recently submitted joint paper [35].

In [33] the same problem is tackled. Two methods are proposed. The first one is a two-step method that combines the prediction error method and the empirical characteristic function method for i.i.d. data. The second one estimates the the system parameters and the noise parameters simultaneously. It is proved that the second method may estimate the system parameters in a more efficient way than the first one, still it does not give an efficient estimator. Moreover, the method presented in [33] is applicable only if the driving noise is a zero mean process.

4.1 Discrete-time Lévy-systems

A natural object for study is a linear stochastic system driven by a Lévy-process. Since the study of continuous-time systems would lead to a number of technical difficulties, we restrict our attention to discrete-time, finite-dimensional linear stochastic systems driven by the increments of a Lévy-process:

$$\Delta y = A(\theta^*, q^{-1})\Delta L, \tag{4.1.1}$$

defined for the time range $-\infty < n < +\infty$, where ΔL_n is the increment of a Lévy process (L_t) with $-\infty < t < +\infty$, and $L_0 = 0$, over an interval $[(n-1)h, nh)$, with $h > 0$ being a fixed sampling interval, and $-\infty < n < +\infty$. The Lévy-measure of L will be denoted by $\nu(dx) = \nu(dx, \eta^*)$, where η^* denotes an unknown real parameter-vector with a known range, say $D_\eta \subset \mathbb{C}^r$. Note that the feasible range of η^* is allowed to be complex for technical reasons

inherent in the estimation procedure. The operator $A(\theta^*, q^{-1})$ is a rational, stable and inverse stable function of the backward shift operator q^{-1}, depending on some unknown real parameter-vector θ^*, taking its values from some known set $G_\theta \subset \mathbb{C}^p$. We suppose that A is monic, i.e. its zeroth coefficient is 1. The observed output process is then Δy. We call such systems briefly Lévy-systems.

The fundamental problem to be discussed in this chapter is the efficient identification of both the systems and the noise parameters. The ML method would be appropriate in solving the full identification problem (i.e. estimating both θ^* and η^*) along standard lines, if we knew the density function of ΔL_n is known, see [36]. Unfortunately, typically this is not the case with the mostly used Lévy processes.

Therefore we develop a new method, using a combination of the PE (prediction error) and an adapted version of the so-called ECF (empirical characteristic function) method, widely used in finance, to get a competitive alternative to the ML (maximum likelihood) method.

The ECF method was originally designed for i.i.d. samples. It has the remarkable property that under certain idealistic assumptions it is as efficient as the ML method. Certain extensions to dependent data are available in the literature at the cost of losing efficiency. Our main contribution is the development of a method for system identification using a suitably adapted ECF method, the efficiency of which is established solely relying on efficiency results for i.i.d. data.

Let us now describe a few additional technical details of our model. Let us assume that a state space representation in innovation form equation for this model is given by

$$\Delta X_{n+1} = H(\theta^*)\Delta X_n + K(\theta^*)\Delta L_n \tag{4.1.2}$$

$$\Delta Y_n = T(\theta^*)\Delta X_n + \Delta L_n. \tag{4.1.3}$$

Then stability and inverse stability of the system is then described by the following condition:

Condition 4.1. *It is assumed that the system matrices $H(\theta)$ and $H(\theta) - K(\theta)T(\theta)$ are stable for $\theta \in G_\theta$.*

To define the smooth dependence on θ suppose that $A(\theta, q^{-1})$ is three-times continuously differentiable w.r.t. θ for $\theta \in D_\theta$. Let $\mathscr{F}$ denote the natural filtration with $\mathscr{F}_{n-1}^{\Delta L} = \sigma\{\Delta L_k : k \leq n-1\}$. The system (7.3.1) is certainly well-defined if ΔL satisfies the integrability condition (2.1.2). Namely, then Δy_n can be written as a weighted sum of past values of ΔL, with exponentially decaying weights, converging in L_1. We will need the following technical condition:

Condition 4.2. *We assume that for all $q \geq 1$*

$$\int_{|x|\geq 1} |x|^q \nu(dx) < +\infty, \tag{4.1.4}$$

and that $\mathbb{E}[\Delta L_n] = 0$.

It follows, see [60], that for $q \geq 1$ and for all $h \geq 0$, the q-th moments of the increments of L are finite:

$$\mathbb{E}[|\Delta L_n|^q] < \infty. \tag{4.1.5}$$

We note here that Condition 5.2 holds in our benchmark examples to be presented in the next Section. Let $D_\theta \subset \mathbb{C}^p$ and $D_\theta^* \subset \mathbb{C}^p$ be compact domains such that

$$\theta^* \in D_\theta^* \subset \text{int } D_\theta \quad \text{and} \quad D_\theta \subset G_\theta.$$

For the q-dimensional η the domains D_η, D_η^* are defined analogously. Finally, let $\rho^* = (\theta^*, \eta^*)$ denote the joint parameter vector, and set

$$D_\rho = D_\theta \times D_\eta, \qquad D_\rho^* = D_\theta^* \times D_\eta^*, \qquad G_\rho = G_\theta \times G_\eta.$$

Here, $D_\rho \subset \mathbb{C}^{p+q}$ and $D_\rho^* \subset \mathbb{C}^{p+q}$. Before going into further details we present a few examples of Lévy processes used for modeling purposes.

4.2 A three-stage method

Now we turn to our main problem, the identification of Lévy systems, when both the system parameters and the noise parameters are unknown. A possible approach would be to adapt and apply the ECF method for the statistical analysis of dependent data. Preliminary results under restrictive conditions are available in the literature, see e.g. [44], [65]. We on this are available in the literature. The method proposed in the literature is based on the observation that, as an alternative to the joint probability density function, we can compute the joint characteristic function of blocks of unprocessed data, i.e. for blocks of the time series (y_n). Without going into further details we point out that the weakness of this approach is that the joint characteristic function is given in terms of an infinite product, and hence it is not clear how to use it in actual computations. Moreover, it is pointed out in the literature that the above ECF method for dynamic models may be less efficient than the ML method, see [14]. A novel idea is presented [34] for the case when the data is passed through a possibly non-FIR filter. Although the the exact c.f. cannot be computed explicitly it is found that an unbiased estimator for the exact c.f. can be obtained under the assumption that we can simulate the system with arbitrary feasible choice of the system parameters θ and noise parameter η.

In what follows we propose a completely different approach, combining the PE method with adapted versions of the ECF method for i.i.d. samples. Our novel method is a three-stage method, the first stage being a standard PE method for estimating the system parameters, taking into account only that the innovation process is i.i.d. having finite moments of all orders. Thus we get an estimate of the system parameters, say, $\hat{\theta}_N$.

In the second stage, using a certainty equivalence argument, pretending that $\hat{\theta}_N = \theta^*$, we

estimate the innovation process by inverting the system using the estimated system parameters $\hat{\theta}_N$. Then, the noise parameter η^* is estimated using the ECF method for i.i.d. sequences, resulting in an estimate $\hat{\eta}_N$. These procedures will be briefly described in Section 4.3. Finally, in the third stage, once again using a certainty equivalence argument, pretending that $\hat{\eta}_N = \eta^*$, we re-estimate the system parameters using a specific adaptation of the ECF method for systems-identification with i.i.d. innovation process, having a *known* characteristic function. This is a completely new and original step of our procedure, which deserves most of our attention.

The analysis of the effects of the estimation errors of $\hat{\theta}_N$ and $\hat{\eta}_N$ on subsequent steps are based on moment estimates of the estimation errors. The latter can be obtained by extending the techniques of [31], and exploiting the fact that all finite moments of the innovation process are finite.

To set the stage for the final, third step of our procedure we briefly summarize a simple known result on the ML estimate for the identification of a linear stochastic system with i.i.d. innovation with known characteristics, more accurately with known probability density function, say $f(., \eta^*)$, following [36]. In this case we can obtain the maximum likelihood estimate of the unknown system parameter θ^* via solving

$$\sum_{n=1}^{N} \frac{\partial}{\partial \theta} \log f\left(\varepsilon_n(\theta), \eta^*\right) = 0, \tag{4.2.1}$$

where

$$\varepsilon_n(\theta) = A^{-1}(\theta)\Delta y_n \tag{4.2.2}$$

is the estimated innovation process of the SISO system given under (7.3.1), using zero initial conditions for $n \leq 0$.

Then under certain technical conditions, in particular assuming that

$$\mathbb{E}[\Delta L_n] = 0, \tag{4.2.3}$$

the asymptotic covariance matrix of the ML estimate is given by

$$\Sigma_{\mathrm{ML}} = \mu^{-1} \left(R_P^*\right)^{-1}, \tag{4.2.4}$$

where

$$\mu = \mathbb{E}\left[\left(\frac{f'(\Delta L_n, \eta^*)}{f(\Delta L_n, \eta^*)}\right)^2\right],$$

with f' being the derivative of $f(., \eta^*)$ w.r.t. its first variable, and

$$R_P^* = \mathbb{E}\left[\varepsilon_{\theta n}^{(\mathrm{s})}(\theta^*)\varepsilon_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta^*)\right].$$

Note that μ can be interpreted as the Fisher information corresponding to the the location parameter m for the family of densities $f(\varepsilon - m, \eta^*)$. This interpretation of μ will exploited in Section 4.6 in proving the efficiency of our proposed three-stage method, and as a special case, in proving the efficiency of its third stage under the condition that the noise parameters are known.

The challenge we address in this chapter if we can achieve the same accuracy in estimating θ^* when we know the characteristic function of the innovation rather than its p.d.f. The surprising answer is a yes, or rather an almost yes. To achieve this we estimate θ^* by applying a suitable modification of the ECF method for i.i.d. data using θ-dependent residuals, and defining a set of initial scores as

$$h_{k,n}(\theta, \eta) = \left(e^{iu_k\varepsilon_n(\theta)} - \varphi(u_k, \eta^*)\right)\varepsilon_{\theta n}(\theta).$$

To summarize, our proposed three-stage method consists of the following steps:

1. Estimate θ^* by applying the PE method to obtain $\hat{\theta}_N$.

2. Invert the system with $\theta = \hat{\theta}_N$ to generate an estimated noise process or residual process, then estimate η^* by pretending that these residuals are i.i.d., and apply the ECF method

for for i.i.d. data to obtain $\hat{\eta}_N$.

3. Re-estimate θ^* by applying the ECF method for system identification, pretending that $\hat{\eta}_N = \eta^*$, to obtain an efficient estimate $\hat{\hat{\theta}}_N$ for the dynamics.

In the next two sections we briefly summarize some basic facts concerning the first two stages of our algorithm.

4.3 A summary of results on the PE method

In this section we briefly summarize some basic facts on the PE method for the case when the input noise has zero expectation, i.e. $\mathbb{E}[\Delta L_n] = 0$, see (4.2.3), and present a precise characterization of the error process that will be relevant later. Although general Lévy processes presented in Section 2.3 have non-zero mean, preprocessing our data, as is customary in classic time series analysis, we may achieve $\mathbb{E}[\Delta L_n] = 0$. First, we define the estimated innovation process $\varepsilon(\theta)$ as above, see (4.2.2). The prediction error estimator of parameter vector θ^* is then obtained by minimizing the cost function

$$V_{P,N}(\theta) = \frac{1}{2}\sum_{n=1}^{N} \varepsilon_n^2(\theta),$$

over G_θ, see [46]. An alternative, more convenient definition of the PE estimator $\hat{\theta}_N$ is obtained by setting the gradient of the cost functions equal to zero, and considering the equations:

$$\frac{\partial}{\partial\theta} V_{P,N}(\theta) = V_{\theta,P,N}(\theta) = \sum_{n=1}^{N} \varepsilon_{\theta n}(\theta)\ \varepsilon_n(\theta) = 0,$$

with $\varepsilon_{\theta n}(\theta)$ being a $p \times 1$ vector. For the precise definition of the random vector $\hat{\theta}_N$ see [31]. The asymptotic cost function associated with the PE method is defined as

$$W_{\theta,P} = \frac{1}{2}\mathbb{E}\left[\left(\varepsilon_n^{(\mathrm{s})}(\theta)\right)^2\right],$$

where $\varepsilon_n^{(\mathrm{s})}(\theta)$ denotes the stationary solution of (4.2.2) obtained by letting $-\infty < n < \infty$. (In general, the superscript $^{(\mathrm{s})}$ will be used throughout this chapter if the marked stochastic process is obtained by passing through a stationary process through an exponentially stable linear filter $-\infty < n < \infty$). We have

$$\begin{aligned}
&\frac{\partial}{\partial\theta}W_P(\theta^*) = 0 \text{ and} \\
&R_P^* := W_{\theta\theta,P}(\theta^*) = \mathbb{E}\left[\varepsilon_{\theta n}^{(\mathrm{s})}(\theta^*)\varepsilon_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta^*)\right].
\end{aligned}$$

Furthermore, $\theta = \theta^*$ is the unique solution of $W_{P,\theta}(\theta) = 0$ in G_θ, see [3]. The asymptotic covariance matrix of the PE estimate of θ^* is given by

$$\Sigma_P = \sigma^2\left(\mathbb{E}\left[\varepsilon_{\theta n}^{(\mathrm{s})}(\theta^*)\varepsilon_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta^*)\right]\right)^{-1}, \tag{4.3.1}$$

where σ^2 is the variance of ΔL_n. Since the ML method is efficient we have that $\mu^{-1} \leq \sigma^2$, we note that the accuracy of the ML method can significantly surpass that of the PE method, i.e. we can have $\mu^{-1} << \sigma^2$. Large difference between μ^{-1} and σ^2 can be achieved by taking the mixture of a mass like continuous pdf with and another continuous pdf.

To formulate our next result we need the following definition:

Definition 4.1. *Let $(X_n), n \geq 0$ be a stochastic process, and let*

$$f : \mathbb{Z}^+ \to \mathbb{R}^+$$

be a function. We write that

$$X_n = O_M(f(n))$$

if for all $q \geq 1$

$$\sup_n \frac{\mathbb{E}^{1/q} |X_n|^q}{f(n)} < \infty$$

holds.

The following theorem is proven in [31]:

Theorem 4.1. *Under Conditions 4.1, 5.2 we have*

$$\hat{\theta}_N - \theta^* = -(R_P^*)^{-1} V_{\theta,P,N}(\theta^*) + r_N,$$

with $r_N = O_M(N^{-1})$.

The key point of this result is that the error term r_N is under control. A direct consequence is the following result, which in fact is proved independently in [31] as an auxiliary result:

Corollary 4.1. *Under Conditions 4.1, 5.2 we have*

$$\hat{\theta}_N - \theta^* = O_M(N^{-1/2}).$$

4.4 The ECF method for estimating the noise parameters

The core of the second stage of our procedure is the following problem: estimate the unknown noise parameters are when we know the system parameters. This is solved by a simple adaptation of the ECF method for i.i.d. data. For this we note that

$$\varepsilon_n^{(\mathrm{s})}(\theta^*) = A^{-1}(\theta^*)\Delta y_n = A^{-1}(\theta^*)A(\theta^*)\Delta L_n = \Delta L_n, \tag{4.4.1}$$

for $-\infty < n < +\infty$. The score function for the ECF method to estimate η^* would be then defined by

$$h_{k,n}^{opt}(\theta^*, \eta) = h_n^{opt}(u_k, \theta^*, \eta) = e^{iu_k \varepsilon_n^{(\mathrm{s})}(\theta^*)} - \varphi(u_k, \eta).$$

In practice we can not reconstruct $\varepsilon_n^{(\mathrm{s})}(\theta^*)$ exactly, and rather work with the approximating process $\varepsilon_n(\theta^*)$ defined by (4.4.1) with zero initial conditions, i.e. $\Delta L_n = \varepsilon_n(\theta^*) = 0$ for $n \leq 0$.

We can now ask ourselves, what happens if the above straightforward procedure is applied in the case when when θ^* is unknown, and is replaced by an arbitrary, feasible θ. To see this, let us define

$$\varepsilon_n^{(\mathrm{s})}(\theta) = A^{-1}(\theta)\Delta y_n = A^{-1}(\theta)A(\theta^*)\Delta L_n \tag{4.4.2}$$

for $-\infty < n < +\infty$. The extended, θ-dependent primary score function would then be defined by

$$h_{k,n}^{(\mathrm{s})}(\theta, \eta) = h_n(u_k, \theta^*, \eta^*) = e^{iu_k \varepsilon_n^{(\mathrm{s})}(\theta)} - \varphi(u_k, \eta),$$

with a fixed set of real numbers $u_1, \cdots, u_M$, with $M \geq \dim \eta$. Define

$$h_{k,n}(\theta, \eta) = h_n(u_k, \theta^*, \eta^*) = e^{iu_k \varepsilon_n(\theta)} - \varphi(u_k, \eta),$$

which is computable. Note that it is no longer true that the primary score satisfies for any θ and for $\eta = \eta^*$ the equality

$$\mathbb{E}\left[h_n(u, \theta, \eta^*)\right] = 0, \quad \text{for all} \quad u.$$

Still, we proceed in computing a θ-dependent estimate $\hat{\eta}_N = \hat{\eta}_N(\theta)$ as if we had $\theta = \theta^*$.

Averaging over the samples, i.e. for $n = 1, \ldots, N$ we define a mean-error or secondary score as

$$\overline{h}_{k,N}(\theta, \eta) = \frac{1}{N} \sum_{n=1}^{N} h_{k,n}(\theta, \eta).$$

Then define the vector-valued mean-error as

$$\overline{h}_N(\theta, \eta) = (\overline{h}_{1,N}(\theta, \eta), \ldots, \overline{h}_{M,N}(\theta, \eta))^T.$$

and let us denote its expectation by

$$g_N(\theta, \eta) = \mathbb{E}\left[\overline{h}_N(\theta, \eta)\right].$$

Note that the expectation of the r.h.s. depends on N since $\varepsilon_n(\theta)$ is non-stationary, due to the 0 initialization. However, it is easily seen that the limit

$$g(\theta, \eta) = \lim_{N \to \infty} g_N(\theta, \eta)$$

exists; this is seen simply by approximating $\varepsilon_n(\theta)$ by $\varepsilon_n^{(\mathrm{s})}(\theta)$. Recall that here $\dim g(\theta, \eta) = M > p$, and that $g(\theta^*, \eta^*) = 0$. However, choosing an arbitrary θ, the equation $g(\theta, \eta) = 0$ may have no solution in η.

Now, we first define our estimator $\hat{\eta}_N = \hat{\eta}_N(\theta)$ as the possibly complex-valued solution of

$$V'_{E,N}(\theta, \eta) := \overline{h}^*_{\eta N}(\theta, \eta) K^{-1} \overline{h}_N(\theta, \eta) = 0. \tag{4.4.3}$$

The precise definition of the solution of (4.4.3) is analogous with that of $3.1.5$. The l.h.s. is then considered as a new set of score functions the asymptotic value of which is given by

$$W'_E(\theta, \eta) = g^*_\eta(\theta, \eta) K^{-1} g(\theta, \eta). \tag{4.4.4}$$

To formulate our results we need some technical conditions. Let ρ be the joint parameter i.e. $\rho = (\theta, \eta)$. Let $D_\rho \subset \mathbb{C}^{p+r}$ and $D_\rho^* \subset \mathbb{C}^{p+r}$ be compact domains such that $\rho^* \in D_\rho^* \subset$ int D_ρ and $D_\rho \subset G_\rho$.

Condition 4.3. *For each $\theta \in D_\theta$ the equation $W_E'(\theta, \eta) = 0$ has a unique solution $\eta^*(\theta)$ in D_η^*, i.e.*

$$W_E'(\theta, \eta^*(\theta)) = 0.$$

Note that the expectation of the l.h.s. of (4.4.3) is not necessarily 0 for $\eta = \eta^*$ and $\theta = \theta^*$, since the mean errors $\overline{h}_{\eta N}^*(\theta^*, \eta^*)$ and $\overline{h}_N(\theta^*, \eta^*)$ are not necessarily independent. Hence we define, as in the i.i.d. case, an approximating problem as follows: replace the first term of the l.h.s. of (4.4.3) by the asymptotic value of its expectation, evaluated at the solution $(\theta, \eta^*(\theta))$ to get

$$\bar{V}_{E,N}'(\theta, \eta) := G^*(\theta) K^{-1} \overline{h}_N(\theta, \eta) = 0, \tag{4.4.5}$$

where $G(\theta)$ is the $M \times p$ matrix

$$G(\theta) = g_\eta(\theta, \eta^*(\theta)).$$

It can be shown that using this approximation we get an estimator with asymptotic covariance matrix identical to that of $\hat{\eta}_N(\theta)$. The left hand side of $(4.4.5)$ can be considered as a new set of exactly p scores with asymptotic expectation

$$\bar{W}_E'(\theta, \eta) = g_\eta^*(\theta, \eta^*(\theta)) K^{-1} g(\theta, \eta) = G^*(\theta) K^{-1} g(\theta, \eta).$$

Following the arguments applied in the i.i.d. case, simple matrix inequalities yield that the optimal choice of $K(\theta)$ is $K^{(\mathrm{opt})} = K^{(\mathrm{opt})}(\theta, \eta^*(\theta)) = C(\theta, \eta^*(\theta))$, with $C(\theta, \eta)$ being the

$M \times M$ matrix with entries

$$C_{k,l}(\theta,\eta) = \mathbb{E}\left[\, h^{(\mathrm{s})}_{k,n}(\theta,\eta) h^{(\mathrm{s})*}_{l,n}(\theta,\eta)\right].$$

In actual computations we use $\hat{C}(\theta,\eta)$ with entries

$$\hat{C}_{k,l}(\theta,\eta) = \frac{1}{N}\sum_{n=1}^{N} h_{k,n}(\theta,\eta) h^{*}_{l,n}(\theta,\eta),$$

since the expected value is not computable. Moreover, since $\eta^*(\theta)$ is unknown we first take $K_1 = \hat{C}(\hat{\theta}_N,\eta)$, with any η to obtain the preliminary estimate $\hat{\eta}^{(\mathrm{pre})}_N(\hat{\theta}_N)$ of η^*, then re-estimate η^* using $K_2 = \hat{C}(\hat{\theta}_N,\hat{\eta}^{(\mathrm{pre})}_N)$ to get $\hat{\eta}_N(\hat{\theta}_N)$. It can be seen that the error we get by substituting $C(\hat{\theta}_N,\eta^*(\theta))$ with $\hat{C}(\hat{\theta}_N,\hat{\eta}_N(\hat{\theta}_N))$ is $O_M(N^{-1})$.

Let the Jacobian of $\bar{W}'_E(\theta,\eta)$ w.r.t. η at $\eta = \eta^*(\theta)$ be denoted by $R^*_E(\theta)$. Then it is easy to see that

$$R^*_E(\theta) = \bar{W}'_{E,\eta}(\theta,\eta^*(\theta)) = G^*(\theta)K^{-1}G(\theta).$$

Our next result characterizes the estimation error of the ECF method for the noise parameter η^* :

Theorem 4.2. *Under Conditions 4.1, 5.2 and 4.3 we have for any fixed feasible $\theta \in D_\theta$*

$$\hat{\eta}_N(\theta) - \eta^*(\theta) = -(R^*_E(\theta))^{-1}\bar{V}'_{E,N}(\theta,\eta^*(\theta)) + O_M(N^{-1}),$$

with an error term that is $O_M(N^{-1})$ uniformly in θ for $\theta \in D_\theta$.

The proof Theorem 4.2 is isomorphic to that of the martingale representation theorem in [31]. It is partially based on the following lemma, which itself is a direct corollary of the above theorem:

Lemma 4.1. *Under Conditions 4.1, 5.2 and 4.3 we have for any fixed $\theta \in D_\theta$*

$$\hat{\eta}_N(\theta) - \eta^*(\theta) = O_M(N^{-1/2}),$$

with an error term that is $O_M(N^{-1})$ uniformly in θ for $\theta \in D_\theta$.

Using the above theorem with $\theta = \hat{\theta}_N$, the estimation that we obtained by the PE method, and setting $\hat{\eta}_N = \eta^*(\hat{\theta}_N)$, according to the second stage of our three-stage procedure, we conclude that the following result holds:

Theorem 4.3. *Under Conditions 4.1, 5.2 and 4.3 we have*

$$\hat{\eta}_N - \eta^* = -(R_E^*(\theta^*))^{-1} G^*(\theta^*) K^{-1} \overline{h}_N(\theta^*, \eta^*) + O_M(N^{-1}).$$

The proof of the last result is partially based on the observations that $\hat{\eta}_N - \eta^* = \eta^*(\hat{\theta}_N) - \eta^*(\theta^*) = O_M(N^{-1/2})$, and that

$$\left\| W'_{E,\eta}(\hat{\theta}_N, \eta^*(\hat{\theta}_N)) - W'_{E,\eta}(\theta^*, \eta^*) \right\| = O_M(N^{-1/2}).$$

4.5 Re-estimation of θ^* by the ECF method

In this section we consider the problem of identifying the system parameters θ^*, using an appropriate version of the ECF method, under the assumption that the noise characteristics η^* is known. This is a simplified benchmark problem of interest of its own. The main innovation of our three-stage procedure is the development of the third stage and the proof that the proposed procedure is essentially asymptotically efficient, i.e. taking a sufficient large number of u-s, the asymptotic covariance matrix of the estimator of θ^* is close to optimal, given by the Cramer-Rao inequality. Following the philosophy of the ECF method take a fix set u_i-s,

$1 \leq i \leq M$. The first natural candidate for a score function would be

$$f_{k,n}(\theta) = e^{iu_k\varepsilon_n(\theta)} - \varphi(u_k, \eta^*),$$

see [33]. However, the corresponding sensitivity matrix would have zero expectation, as

$$\mathbb{E}\left[e^{iu_k\varepsilon_n(\theta^*)} iu_k \varepsilon_{\theta n}(\theta^*)\right] = 0.$$

The only natural way to avoid this is to make $\varepsilon_{\theta n}(\theta^*)\varepsilon_{\theta n}^T(\theta^*)$ appear in the sensitivity matrix. Hence, we define the score functions as

$$h_{k,n}(\theta) = \left(e^{iu_k\varepsilon_n(\theta)} - \varphi(u_k, \eta^*)\right) \varepsilon_{\theta n}(\theta), \tag{4.5.1}$$

with $h_{k,n}(\theta)$ being $p \times 1$ column vectors. Note that in this case at $\theta = \theta^*$ both $h_{k,n}^{(\mathrm{s})}(\theta)$-s and $\left(e^{iu_k\varepsilon_n^{(\mathrm{s})}(\theta)} - \varphi(u_k, \eta^*)\right)$ have zero expectation.

For technical reasons we shall also consider a more general class of problems when the noise characteristics is misspecified, i.e. when η^* is unknown, and we apply the ECF method for the identification of the systems dynamics with an arbitrary feasible η, pretending that it is the true value. When η^* is misspecified we define the score functions as

$$h_{k,n}(\theta, \eta) = \left(e^{iu_k\varepsilon_n(\theta)} - \varphi(u_k, \eta)\right) \varepsilon_{\theta n}(\theta). \tag{4.5.2}$$

For the stationary approximation of $h_{k,n}(\theta, \eta)$ we obviously have

$$\mathbb{E}\left[h_{k,n}^{(\mathrm{s})}(\theta^*, \eta)\right] = 0,$$

because $\mathbb{E}\left[\varepsilon_{\theta n}(\theta^*)\right] = 0$, and the two terms in $h_{k,n}(\theta^*, \eta)$ are independent. The problem we address is to identify the system dynamics specified by θ^*. We note in passing that as a special

case of this technical problem for known η^* we will obtain the (essential) efficiency of the procedure. Surprisingly, the estimator, say $\hat{\hat{\theta}}_N(\eta)$, will be still consistent, just like in the case of the misspecified ML method, under the condition that the input noise has 0 mean. The third step of our three-stage procedure is then obtained by setting $\eta = \hat{\eta}_N$, with $\hat{\eta}_N$ being defined in the previous section. The (essential) efficiency of our procedure is then obtained by a simple comparison of $\hat{\hat{\theta}}_N(\hat{\eta}_N)$ and $\hat{\hat{\theta}}_N(\eta^*)$.

Define the pM-dimensional score column vector $h_n(\theta,\eta) = \left(h_{1,n}^T(\theta,\eta),\ldots,h_{M,n}^T(\theta,\eta)\right)^T$, and the sample mean of these score vectors:

$$\overline{h}_N(\theta,\eta) = \frac{1}{N}\sum_{n=1}^{N} h_n(\theta,\eta).$$

Let $K > 0$ be a fixed symmetric, $pM \times pM$, positive definite weighting matrix. Define the pM-dimensional column vectors

$$g_N(\theta,\eta) = \mathbb{E}\left[\overline{h}_N(\theta,\eta)\right] \text{ and } g(\theta,\eta) = \lim_{N\to\infty} g_N(\theta,\eta).$$

Note that $\theta = \theta^*$ is the solution of the over-determined set of non-linear algebraic equations

$$g(\theta^*,\eta) = 0. \tag{4.5.3}$$

Since g is not computable we approximate it by $\overline{h}_N$ and in analogy with (4.4.3) we seek a possibly complex-valued solution of the 'half-gradient' equation

$$V_N'(\theta,\eta) = \overline{h}_{\theta N}^*(\theta,\eta)K^{-1}\overline{h}_N^*(\theta,\eta) = 0 \tag{4.5.4}$$

to obtain $\hat{\hat{\theta}}_N(\eta)$. The precise definition of the random variable $\hat{\hat{\theta}}_N(\eta)$ can be given using the procedure seen in Section 3.1 after (3.1.5). The system of equations in (4.5.4) is no longer

over-determined because $\dim V'_N = p$. The asymptotic value of the new set of scores is then

$$W'(\theta,\eta) = \lim_{N\to\infty} \mathbb{E}[V'_N(\theta,\eta)] = g^*_\theta(\theta,\eta)K^{-1}g(\theta,\eta).$$

Condition 4.4. *For each $\eta \in D_\eta$ the equation $W'(\theta,\eta) = 0$ has a unique solution $\theta = \theta^*$ in D^*_θ.*

In view of (4.5.3) it is not surprising that the solution of $W'(\theta,\eta) = 0$ does not depend on η. The Jacobian of W' at $\theta = \theta^*$:

$$R^*(\eta) := G^*(\eta)K^{-1}G(\eta),$$

with the $pM \times p$ matrix $G(\eta) = g_\theta(\theta^*,\eta)$. In analogy with (4.4.5) the auxiliary problem that corresponds to equation (4.5.4) can be formulated as

$$\bar{V}'_N(\theta,\eta) := G^*(\eta)K^{-1}\overline{h}_N(\theta,\eta) = 0. \tag{4.5.5}$$

The following result, which can be proved using the reasoning seen in [31], is a martingale representation theorem for the η-dependent estimate of θ^*.

Theorem 4.4. *Under Conditions 4.1, 5.2 and 4.4 we have*

$$\hat{\bar{\theta}}_N(\eta) - \theta^* = -(R^*(\eta))^{-1}V'_N(\theta^*,\eta) + O_M(N^{-1}). \tag{4.5.6}$$

It is worth mentioning that the theorem remains valid if we write $\bar{V}'_N$ instead of V'_N, that is

$$\hat{\bar{\theta}}_N(\eta) - \theta^* = -(R^*(\eta))^{-1}\bar{V}'_N(\theta^*,\eta) + O_M(N^{-1}) \tag{4.5.7}$$

holds under the conditions of the above theorem. Specifying the r.h.s. we have that $\hat{\bar{\theta}}_N(\eta) - \theta^*$

equals to

$$-(R^*(\eta))^{-1}G^*(\eta)K^{-1}\overline{h}_N(\theta^*,\eta)+O_M(N^{-1}). \tag{4.5.8}$$

Sketch of the proof: Let us write $\hat{\hat{\theta}}_N$ instead of $\hat{\hat{\theta}}_N(\eta)$ to simplify notations. Using the fact that an exponentially stable filter with L-mixing input produces an L-mixing output, see [30], we get that Δy_n is an L-mixing process. The estimated innovation process and its derivatives with respect to θ can be written as

$$\begin{aligned}\varepsilon_n(\theta) &= A^{-1}(\theta)\Delta y_n\\ \varepsilon_{\theta n}(\theta) &= A_\theta^{-1}(\theta)\Delta y_n\\ \varepsilon_{\theta\theta n}(\theta) &= A_{\theta\theta}^{-1}(\theta)\Delta y_n.\end{aligned}$$

The notation $A^{-1}(\theta)\Delta y_n$ is understood as follows: take the state-space representation of $A^{-1}(\theta)$ and compute $\varepsilon_n(\theta)$ by substituting θ into the corresponding state transition matrices. The derivatives of the filter $A^{-1}(\theta)$ w.r.t. θ are defined accordingly. Again, since $A^{-1}(\theta)$ and its derivatives with respect to θ are exponentially stable and $A(\theta)$ is three-times continuously differentiable w.r.t. θ, and $\theta\in D_\theta$, where D_θ is compact, we conclude that the processes $\varepsilon_n(\theta)$, $\varepsilon_{\theta n}(\theta)$ and $\varepsilon_{\theta\theta n}(\theta)$ are L-mixing uniformly when $\theta\in D_\theta$.

One can show that for any given $d>0$ the equation $V_N'(\theta,\eta)=0$ has a unique solution in D_θ and it is in the sphere $S=\{\theta:|\theta-\theta^*|<d\}$ with probability at least $1-O(N^{-s})$ for any $s>0$. This result is the η-dependent version of Lemma 2.3. in [31]. Using this result write

$$0=V_N'\left(\hat{\hat{\theta}}_N,\eta\right)=V_N'\left(\theta^*,\eta\right)+\overline{V}_{\theta N}'(\eta)\left(\hat{\hat{\theta}}_N-\theta^*\right), \tag{4.5.9}$$

where

$$\overline{V}_{\theta N}'(\eta)=\int_0^1 V_{\theta N}'\left((1-\lambda)\,\theta^*+\lambda\hat{\hat{\theta}}_N,\eta\right)d\lambda.$$

Letting η vary in the compact domain D_η one may follow the line of arguments presented in the proof of Theorem 2.1. in [31] to conclude that

$$\left\| \overline{V}'^{-1}_{\theta N}(\eta) - W'^{-1}_{\theta}(\theta^*, \eta) \right\| = O_M(N^{-1/2}). \tag{4.5.10}$$

except for an event of probability $O_M(N^{-s})$ for any $s > 0$. Finally,

$$\begin{aligned} \hat{\hat{\theta}}_N - \theta^* = -\overline{V}'^{-1}_{\theta N}(\eta) V'_N(\theta^*, \eta) = \\ - \left(W'^{-1}_{\theta}(\theta^*, \eta) + O_M(N^{-1/2}) \right) V'_N(\theta^*, \eta) = \\ -W'^{-1}_{\theta}(\theta^*, \eta) V'_N(\theta^*, \eta) + O_M(N^{-1}) \end{aligned}$$

holds, again except from an event of probability $O_M(N^{-s})$ for any $s > 0$. Hence the last expression reads as

$$- \left(R^*(\eta) \right)^{-1} V'_N(\theta^*, \eta) + O_M(N^{-1}).$$

Which concludes the proof. □

Now set $\eta = \hat{\eta}_N$, where $\hat{\eta}_N$ is the estimate of the noise characteristics that we obtained at the second step of the procedure. Then the third step of our three-stage method is simply carried out by repeating the above procedure with the scores

$$h_{k,n}(\theta, \hat{\eta}_N) = \left(e^{iu_k \varepsilon_n(\theta)} - \varphi(u, \hat{\eta}_N) \right) \varepsilon_{\theta n}(\theta). \tag{4.5.11}$$

Applying Theorem 4.4 with the choice $\eta = \hat{\eta}_N$ and using that $R^*(\eta)$ and $\bar{V}'_N(\theta^*, \eta)$ are smooth enough in η and that $\hat{\eta}_N - \eta^* = O_M(N^{-1/2})$ we obtain the following result:

Theorem 4.5. *Suppose that $\hat{\eta}_N$ is obtained by the second step of the three-stage identification*

procedure. Then under Conditions 4.1, 5.2 and 4.4 we have

$$\hat{\hat{\theta}}_N(\hat{\eta}_N) - \theta^* = -(R^*(\eta^*))^{-1}\bar{V}'_N(\theta^*, \eta^*) + O_M(N^{-1}). \tag{4.5.12}$$

Specifying the r.h.s. we have that $\hat{\hat{\theta}}_N(\hat{\eta}_N) - \theta^*$ equals to

$$-(R^*(\eta^*))^{-1}G^*(\eta^*)K^{-1}\overline{h}_N(\theta^*, \eta^*) + O_M(N^{-1}). \tag{4.5.13}$$

Observe that (4.5.8) and (4.5.13) are the same if $\eta = \eta^*$, which together with Theorem 4.3 and the fact that $\varepsilon^{(s)}_{\theta n}(\theta^*)$ has zero expectation and is independent of ΔL_n imply the next corollary.

Corollary 4.2. *The three-stage method and the ECF identification of θ^* with known η^* yield the same asymptotic covariance for the estimate of θ^*. We also have that the estimates $\hat{\eta}_N$ and $\hat{\hat{\theta}}_N(\hat{\eta}_N)$ are asymptotically uncorrelated.*

4.6 Efficiency of the ECF method for θ^*

In this section we show that the three-stage method gives an essentially efficient estimate of θ^*. Recall the notations

$$R^*_P = \mathbb{E}\left[\varepsilon^{(\mathrm{s})}_{\theta n}(\theta^*)\varepsilon^{(\mathrm{s})\mathrm{T}}_{\theta n}(\theta^*)\right],$$

and C with entries

$$C_{k,l} = \varphi(u_k - u_l, \eta^*) - \varphi(u_k, \eta^*)\varphi(-u_l, \eta^*).$$

Theorem 4.6. *Let $K = C \otimes R^*_P$ and let $\hat{\hat{\theta}}_N(\hat{\eta}_N)$ be the estimate obtained by the third step of out three-stage procedure, then we have*

$$\mathbb{E}\left[N\left(\hat{\hat{\theta}}_N(\hat{\eta}_N) - \theta^*\right)\left(\hat{\hat{\theta}}_N(\hat{\eta}_N) - \theta^*\right)^*\right] = \Sigma_{\theta\theta} + O_M(N^{-1/2}),$$

where the asymptotic covariance matrix is given by

$$\Sigma_{\theta\theta} = \left(\psi^* C^{-1} \psi\right)^{-1} \left(R_P^*\right)^{-1},$$

with $\psi = \left(iu_1\varphi(u_1,\eta^*),\ldots,iu_M\varphi(u_M,\eta^*)\right)^T$.

Proof: From now on we write $\hat{\hat{\theta}}_N$ for $\hat{\hat{\theta}}_N(\hat{\eta}_N)$ to simplify notations. Using Theorem 4.5 and Corollary 4.2 we get that the covariance matrix of the estimator is

$$\mathbb{E}\left[\left(\hat{\hat{\theta}}_N - \theta^*\right)\left(\hat{\hat{\theta}}_N - \theta^*\right)^*\right] = (R^*(\eta^*))^{-1}\mathbb{E}\left[\bar{V}'_N(\theta^*,\eta^*)\bar{V}'^*_N(\theta^*,\eta^*)\right](R^*(\eta^*))^{-1} + r_1,$$

with $r_1 = O_M(N^{-3/2})$. Now, we compute the middle term of the above formula. At first, it is easy to see that since $\varepsilon^{(\mathrm{s})}_{\theta n}(\theta^*)$ and $\varepsilon^{(\mathrm{s})}_n(\theta^*) = \Delta L_n$ are independent we have

$$\mathbb{E}\left[\overline{h}_N(\theta^*,\eta^*)\overline{h}^*_N(\theta^*,\eta^*)\right] = \frac{1}{N}C \otimes R_P^* + r_2 = \frac{1}{N}K + r_2,$$

where $r_2 = O_M(N^{-3/2})$. Hence,

$$\begin{aligned}
&\mathbb{E}\left[\bar{V}'_N(\theta^*,\eta^*)\bar{V}'^*_N(\theta^*,\eta^*)\right] = \\
&G^*(\theta^*)K^{-1}\left(\frac{1}{N}K + r_2\right)K^{-1}G(\theta^*) = \\
&\frac{1}{N}G^*(\theta^*)K^{-1}G(\theta^*) + r_3,
\end{aligned}$$

with $r_3 = O_M(N^{-3/2})$, because $G(\theta^*)$ is bounded. We have

$$G(\theta^*) = \psi \otimes R_P^*,$$

since for the k^{th} block of rows of $G(\theta^*)$ we have

$$\mathbb{E}\left[h_{\theta;k,n}(\theta^*,\eta^*)\right] = iu_k\varphi(u_k,\eta^*)\mathbb{E}\left[\varepsilon_{\theta n}(\theta^*)\varepsilon_{\theta n}^T(\theta^*)\right].$$

Using the mixed-product property and the inverse of a Kronecker product, reading as $(A\otimes B)(C\otimes D) = AC\otimes BD$ and $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$, the covariance can be further calculated as follows:

$$\begin{aligned}
&\mathbb{E}\left[\left(\hat{\hat{\theta}}_N-\theta^*\right)\left(\hat{\hat{\theta}}_N-\theta^*\right)^*\right] = \\
&\frac{1}{N}\left(\left(\psi^*\otimes R_P^*\right)\left(C\otimes R_P^*\right)^{-1}\left(\psi\otimes R_P^*\right)\right)^{-1} + r_4 = \\
&\frac{1}{N}\left(\left(\psi^*C^{-1}\psi\right)\otimes R_P^*\right)^{-1} + r_4 = \\
&\frac{1}{N}\left(\psi^*C^{-1}\psi\right)^{-1}\left(R_P^*\right)^{-1} + r_4,
\end{aligned}$$

with $r_4 = O_M(N^{-3/2})$, which concludes the proof. □

Theorem 4.7. *Under the conditions of Theorem 4.6 the estimate $\hat{\hat{\theta}}_N$ is essentially asymptotically efficient.*

Suppose now that instead of a finite number of moment conditions we exploit the full continuum of moment conditions. In this case the continuous version of the score vector would be defined as

$$h_n(u,\theta) = \left(e^{iu\varepsilon_n(\theta)} - \varphi(u,\eta^*)\right)\varepsilon_{\theta n}(\theta) \text{ for each } u\in\mathbb{R},$$

and that of the 'half-gradient' equation (4.5.5) would read as

$$\langle \overline{h}_N(u,\theta), K^{-1}G\rangle = 0, \tag{4.6.1}$$

where $\langle\cdot\,,\cdot\rangle$ denotes an inner product that we need yet to specify. To define the inner product let us first introduce the Hilbert space $H = L^2(\pi) = \left\{f:\mathbb{R}\to\mathbb{C}\,\middle|\int|f(t)|^2\pi(t)dt<\infty\right\}$, with

π being a probability measure on $\mathbb{R}$. Then define the inner product on H as

$$\langle f, g\rangle = \int f(t)g^*(t)\pi(t)dt,$$

and the norm $||f||^2 = \langle f, f\rangle$. Let $k(s,t)$ be a symmetric, positive definite function, then define the operator K as

$$(Kf)(s) = \int k(s,t)f(t)\pi(t)dt. \tag{4.6.2}$$

We will use the notation $||f||_K^2 = ||K^{-1/2}f||^2$. As it is pointed out in [16] $K^{-1/2}f$ does not exists on the whole space H, the analysis of such operators is done by using the reproducing kernel Hilbert space (RKHS) of K. The results of Chapter 3 in [54] imply that $K^{-1/2}f \in H$ for each f in the RKHS of K, because for each f there exists an $F \in L^2(\pi)$ such that $||f||_K^2 = ||K^{-1/2}f||^2 = ||F||^2 < \infty$. It is worth mentioning that the norm $||f||_K^2$ does not depend on the choice of π, because π-s in the inner product and in K^{-1} cancel out each other, see equation (4.7) in [16]. For more details see the just cited paper.

Define C the π-dependent covariance operator as special case of K via

$$(Cf)(s) = \int c(s,t)f(t)\pi(t)dt, \tag{4.6.3}$$

with

$$c(s,t) = \mathbb{E}\left[h_{s,n}^{(s)}(\theta^*, \eta^*)h_{t,n}^{(s)*}(\theta^*, \eta^*)\right].$$

If the full continuum of u-s were defined via $u_s = s$ for all $s \in \mathbb{R}$, then the continuous version of Theorem 4.6 would give

$$\lim_{N\to\infty} \mathbb{E}\left[N\left(\hat{\hat{\theta}}_N - \theta^*\right)\left(\hat{\hat{\theta}}_N - \theta^*\right)^*\right] = \left(||iu\varphi(u, \eta^*)||_C^2\right)^{-1}(R_P^*)^{-1} \tag{4.6.4}$$

for the asymptotic covariance matrix of the estimate. Note that in the above formula the

asymptotic covariance matrix is decoupled, $||iu\varphi(u,\eta^*)||_C^2$ depends only on the noise characteristics and R_P^* depends on the derivative of the innovation process, hence on the parameters of the linear system.

Recall the notation μ used in connection with the ML method:

$$\mu = \mathbb{E}\left[\left(\frac{f'(\Delta L_n, \eta^*)}{f(\Delta L_n, \eta^*)}\right)^2\right],$$

and that the asymptotic covariance of the ML estimate is given by

$$\Sigma_{\mathrm{ML}} = \mu^{-1}\left(R_P^*\right)^{-1}.$$

Lemma 4.2. *Using the above notations we have*

$$\left(||iu\varphi(u,\eta^*)||_C^2\right)^{-1} = \mu^{-1}. \tag{4.6.5}$$

Proof: Recall that μ was shown to be equal to the Fisher of the location parameter. We do not prove (4.6.5) using direct computation, instead we reduce the problem to the efficiency result of the i.i.d. ECF method. More specifically, we show that $\left(||iu\varphi(u,\eta^*)||_C^2\right)^{-1}$ can be obtained as the asymptotic covariance of an efficient ECF method for the problem of estimating the location parameter of ΔL given an i.i.d. realization of ΔL.

To this end, consider the following identification problem: given a sequence of i.i.d. realizations of the distribution $\Delta L + \lambda^*$, where λ^* is a location parameter to be estimated, and ΔL is a random variable with known characteristic function φ.

According to [16] for an i.i.d. sample, which was generated by a random variable with a general characteristic function $\chi(u,\alpha^*)$, with α^* being an unknown parameter, the ECF method with the full continuum of u-s gives an asymptotically efficient estimate of α^* with asymptotic

covariance

$$(||\chi_\alpha(u,\alpha^*)||_C^2)^{-1}.$$

Apply the ECF method to identify λ^*, and compute the asymptotic covariance of the procedure. Let $\varphi_{\Delta L+\lambda}$ denote the c.f. of $\Delta L + \lambda$, then

$$\begin{aligned}\frac{\partial}{\partial\lambda}\varphi_{\Delta L+\lambda}(u,\eta) = \frac{\partial}{\partial\lambda}\mathbb{E}\left[e^{iu(\Delta L+\lambda)}\right] = \\ \frac{\partial}{\partial\lambda}\left(e^{iu\lambda}\mathbb{E}\left[e^{iu\Delta L}\right]\right) = iu\varphi_{\Delta L+\lambda}(u,\eta).\end{aligned} \tag{4.6.6}$$

Choosing $\chi = \varphi_{\Delta L+\lambda}$ and $\alpha = \lambda$ we have $\chi_\alpha(u,\lambda^*) = iu\varphi_{\Delta L+\lambda}(u,\eta^*)$ and

$$\left(||iu\varphi(u,\eta^*)||_C^2\right)^{-1} = \mu^{-1}$$

as stated in the Lemma. This also concludes the proof of Theorem 4.7.

4.7 Discussion

We have seen in the third step of our three-stage procedure the optimal choice of K is $C \otimes R_P^*$. Here both C and R_P^* are given by an expected value using the true value of parameters θ and η, hence they are not computable. To overcome this difficulty in practice we propose to approximate C and R_P^* is two steps. First, define the approximations $\hat{C}(\theta,\eta)$ and $\hat{R}_P^*(\theta)$ by

$$\hat{C}_{k,l}(\theta,\eta) = \frac{1}{N}\sum_{n=1}^{N} h_{k,n}(\theta,\eta)h_{l,n}^*(\theta,\eta),$$

and

$$\hat{R}_P^*(\theta) = \frac{1}{N}\sum_{n=1}^{N} \varepsilon_{\theta n}(\theta)\varepsilon_{\theta n}(\theta)^T.$$

We would like to use $\hat{C}(\theta^*, \eta^*)$ and $\hat{R}_P^*(\theta^*)$, but since θ^* and η^* are unknown we approximate them by $\hat{\theta}_N$ and $\hat{\eta}_N(\hat{\theta}_N)$, respectively. Recall that $\hat{\theta}_N$ is obtained in the first step and $\hat{\eta}_N(\hat{\theta}_N)$ is obtained in the second step of the three-stage procedure. Thus, we use

$$K = \hat{C}(\hat{\theta}_N, \hat{\eta}_N(\hat{\theta}_N)) \otimes \hat{R}_P^*(\hat{\theta}_N)$$

to get an approximation of $\hat{\hat{\theta}}_N(\hat{\eta}_N)$. It is relatively easy to see that Theorem 4.5 and 4.6 and Theorem 4.7 are valid for this approximation of $\hat{\hat{\theta}}_N(\hat{\eta}_N)$, too.

Throughout the chapter it was assumed that $\mathbb{E}[\Delta L_n] = 0$, arguing that this can be achieved by preprocessing our data. A more rigorous approach to handling the case $\mathbb{E}[\Delta L_n] \neq 0$ is to incorporate the unknown mean in the parameter-vector and estimate this extended vector.

Chapter 5

ECF identification of linear Lévy systems with possibly unstable zeros

In this chapter we investigate the possibility of implementing the ECF method for dependent data using blocks of unprocessed observations. The purpose of this chapter is to extend the results presented in Chapter 4 to finite dimensional stochastic Lévy systems with possibly unstable zeros. Recall that both the PE method and the ML method as presented in Chapter 4 assume that the system is non-minimum phase, i.e. it has a stable inverse. The same assumption is used in [46],[36] and [33]. In fact, the identification of finite dimensional linear stochastic systems with unstable zeros is barely discussed in the literature. A remarkable feature of the ECF method is that it is naturally applicable to the identification of finite dimensional stochastic systems with unstable zeros if properly adapted. Our starting point is the ECF method for dependent data, as presented in the literature, using blocks of data, see [44]. This idea is then extended by defining a c.f. in terms of data passed through a possibly non-FIR filter. The challenge of this approach is that the exact c.f. cannot be computed explicitly, which was a key assumption for the ECF methods for i.i.d. data and linear Lévy systems. However, it is

found that an unbiased estimator for the exact c.f. can be obtained under the assumption that we can simulate the system with arbitrary feasible choice of the system parameters θ and noise parameter η. The latter assumption is not unrealistic in view of the procedure presented in [21].

Thus we will finally arrive at a procedure which can be viewed as a statistical output error method. The actual data are compared to simulated data, and the parameters of the latter are adjusted so as to ensure a good fit in a statistical sense. The resulting method can be analyzed along the lines of the classic ECF, or rather GMM method. This chapter presents the results of the recently submitted joint paper of the author, see [34].

In retrospect, our method also extends the classic ECF method for i.i.d. data for situations when the c.f. is not available explicitly, but we do have an unbiased estimator of it in terms of a parameter-dependent random variable, say $\xi(\eta)$, which is computable via a mechanism of the form

$$\xi(\eta) = F(\rho, \eta), \tag{5.0.1}$$

where F is a fixed, known function of ρ and η, and ρ is a fixed random variable and η is allowed to vary. The data are generated via a true η^* and the problem is to identify η^*. The above problem formulation is perfectly in line with the problem of system identification with ρ denoting the input noise and η^* denoting the noise parameters.

5.1 ECF method for filtered data

In this section we extend the ECF method to dependent data obtained by taking an i.i.d. sequence and passing it through a stable finite dimensional linear system. A practically interesting object of study is a linear stochastic system driven by a Lévy-process, or rather the increments of a Lévy-process, see Chapter 4. Recall that we write the system in the form

$$\Delta y = A(\theta^*, q^{-1})\Delta L, \tag{5.1.1}$$

where the time range is $-\infty < n < +\infty$. Here ΔL_n, like before, denotes the increment of a zero mean Lévy process (L_t) over an interval $[(n-1)h, nh)$, with $h > 0$ being a fixed sampling interval. (L_t) itself is defined for $-\infty < t < +\infty$, and it is tied to 0 at time $t = 0$, i.e. $L_0 = 0$. The condition

$$\mathbb{E}\left[\Delta L_n\right] = 0$$

significantly facilitates the analysis of the forthcoming ECF estimations methods, in analogy with the analysis of the three-stage ECF method method, see Chapter 4. Although generally not satisfied by the Lévy processes presented in Chapter 2, it can be enforced by preprocessing our data, as is customary in classic time series analysis.

The Lévy-measure of L will be denoted by $\nu(dx) = \nu(dx, \eta^*)$, where η^* denotes an unknown parameter-vector with a known open range, say $D_\eta \subset \mathbb{C}^q$. The system dynamics depends on some unknown parameter-vector θ^*, taking its values from some known open set $D_\theta \subset \mathbb{C}^p$. Let D_θ^* and D_η^* be compact domains such that $\theta^* \in D_\theta^* \subset \text{int } D_\theta$ and $\eta^* \in D_\eta^* \subset \text{int } D_\eta$. We note that the parameters θ^* and η^* are real valued.

Condition 5.1. *The operator $A(\theta, q^{-1})$ is a stable, rational function of the backward-shift operator q^{-1} for all $\theta \in D_\theta$. Moreover $A(\theta, q^{-1})$ is three-times continuously differentiable w.r.t. θ for $\theta \in D_\theta$.*

Note that we did not assume the *inverse stability* of the operator $A(\theta, q^{-1})$, in contrast to standard identification methods such as PE, ML or our three-stage method. Contrarily, we assume the stability of $A(\theta, q^{-1})$. In particular, our method is suitable for the identification of moving average (MA) systems with unstable zeroes.

The smoothness of $A(\theta, q^{-1})$ w.r.t. θ should be interpreted as follows: there exists a state-space realization of $A(\theta, q^{-1})$ such that the state-matrices are three-times continuously differentiable w.r.t. θ for $\theta \in D_\theta$.

Condition 5.2. *We assume that for all* $q \geq 1$

$$\int_{|x|\geq 1} |x|^q \nu(dx) < +\infty. \tag{5.1.2}$$

Moreover, it is assumed that the driving noise (L_t) *is a zero mean process:*

$$\mathbb{E}\left[L_t\right] = 0.$$

Now we are in the position to apply the ECF method for dependent data, following the literature, in our special case. Consider the parametric family of systems (or equivalently time series)

$$\Delta y(\theta, \eta) = A(\theta)\Delta L(\eta), \tag{5.1.3}$$

with the time n taking its values in $-\infty < n < +\infty$. Note that for $(\theta, \eta) = (\theta^*, \eta^*)$ we recover our observed data in a statistical sense. $\Delta L(\eta)$ denotes the increments of a Lévy process with characteristics η, and it can be generated using the mechanism in (5.0.1). The ECF methods proposed in the literature, see [44],[27], are based on the computation of the joint characteristic function of blocks of unprocessed data, i.e. for blocks of (y_n). While this computation can indeed be carried out for special cases, such as for Gaussian or stable noise processes, the computation of the joint characteristic function is far from trivial in general. One of the main contributions of this study is to address this challenge.

For a start, fix a block length, say r, and define the r-dimensional blocks

$$\Delta Y_n^r(\theta, \eta) = (\Delta y_{n-1}(\theta, \eta), \ldots, \Delta y_{n-r}(\theta, \eta)).$$

Then the joint characteristic function of the block $\Delta Y_n^r(\theta, \eta)$, with $u = (u_1, \ldots, u_r)^T$ being an

arbitrary vector in $\mathbb{R}^r$, is given by

$$\varphi_n(u,\theta,\eta) = \mathbb{E}\left[e^{iu^T \Delta Y_n^r(\theta,\eta)}\right] = \mathbb{E}\left[e^{i\sum_{j=1}^r u_j \Delta Y_{n-j}(\theta,\eta)}\right].$$

Now, this can be explicitly computed, at least in theory. Letting $h_l(\theta),\ l = 0, 1, \ldots$ denote the impulse responses of the system $A(\theta)$, we can write

$$\begin{aligned}\varphi_n(u,\theta,\eta) = \mathbb{E}\left[\exp\left\{i\sum_{j=1}^r u_j \sum_{l=0}^{\infty} h_l(\theta)\Delta L_{n-j-l}(\eta)\right\}\right] = \\ \mathbb{E}\left[\exp\left\{i\sum_{k=1}^{\infty}\Delta L_{n-k}(\eta)\sum_{l\geq 0, j+l=k} u_j h_l(\theta)\right\}\right].\end{aligned} \tag{5.1.4}$$

Fix k and consider the last term. Set $l = k - j$ and introduce the notation

$$v_k(\theta,u) = \sum_{j=1}^r h_{k-j}(\theta)u_j,$$

with $h_l(\theta) = 0$ for $l < 0$. Then v is the convolution of h and u:

$$v = h * u.$$

Denoting the characteristic function of $\Delta L_n(\eta)$, for any n, by $\varphi_{\Delta L(\eta)}$, we get

$$\varphi_n(u,\theta,\eta) = \prod_{k=1}^{\infty}\varphi_{\Delta L(\eta)}(v_k(\theta,u)). \tag{5.1.5}$$

Now the ECF method could be defined by fitting this theoretical joint characteristic function to the empirical joint characteristic function. Without providing details we point out that it is not clear how to use such a procedure it in actual computations, since $\varphi_n(u,\theta,\eta)$ is given in terms of an infinite product.

To circumvent this difficulty let us return to the the definition of $\varphi_n(u,\theta,\eta)$. Note that a

simple unbiased estimation of $\varphi_n(u,\theta,\eta)$ is given by

$$e^{iu^T\Delta Y_n^r(\theta,\eta)} = e^{i\sum_{j=1}^r u_j \Delta Y_{n-j}(\theta,\eta)}.$$

We propose to fit this simulated value to the data, and introduce the scores

$$h_n(u,\theta,\eta) = e^{iu^T\Delta Y_n^r} - e^{iu^T\Delta Y_n^r(\theta,\eta)}. \tag{5.1.6}$$

Note that the score is essentially a kind of output error. Thus the proposed procedure will be a generalization of the output error identification method for the case when the actual input process generating the data is not observed, but only statistically known if η^* is known.

Note also that we can write the scores in the form

$$h_n(u,\theta,\eta) = e^{i\ (u*\Delta y)_n} - e^{i\ (u*\Delta y(\theta,\eta))_n}, \tag{5.1.7}$$

where u denotes the sequence $u_1,\ldots,u_r$, interpreted as an impulse response. The advantage of this representation is that, in theory, we can use infinite sequences of u-s representing the impulse responses of a finite dimensional stable linear filter.

A final note: in order to compute the above score functions one has to be able to generate the i.i.d. noise sequence $\Delta L_n(\eta)$ for any given η, having a prescribed c.f. $\varphi(u,\eta)$. This problem has been addressed and solved in [21].

To see the details of our procedure, suppose that we are given a sequence of observed data $\Delta y_1,\ldots,\Delta y_{N+r}$ being the outputs of (5.1.3) with $\theta=\theta^*,\eta=\eta^*$. Construct the blocks of observations $\Delta Y_n^r = (\Delta y_{n-1},\ldots,\Delta y_{n-r})$ for each n with $r<n\le N+r$. Take a set of vectors of dimension r, say $\mathbf{u}_1,\ldots,\mathbf{u}_M$, with $M>p+q$. Define the score functions as follows

$$h_{k,n}(\theta,\eta) = e^{i\mathbf{u}_k^T\Delta y_n} - e^{i\mathbf{u}_k^T\Delta y_n(\theta,\eta)} \tag{5.1.8}$$

for $k = 1, \ldots, M$ and $n = 1, \ldots, N$. Note that these are indeed appropriate score functions because

$$\mathbb{E}\left[h_{k,n}(\theta^*, \eta^*)\right] = 0.$$

For a fix n collecte the above scores over k we define the M-vector

$$h_n(\theta, \eta) = (h_{1,n}(\theta, \eta), \ldots, h_{M,n}(\theta, \eta))^T. \tag{5.1.9}$$

The sample average vector of the scores is defined as

$$\overline{h}_N(\theta, \eta) = \frac{1}{N} \sum_{n=r+1}^{N+r} h_n(\theta, \eta). \tag{5.1.10}$$

Let $g_N(\theta, \eta)$ denote the expected error, i.e. let

$$g_N(\theta, \eta) = \mathbb{E}\left[\overline{h}_N(\theta, \eta)\right]. \tag{5.1.11}$$

Also introduce its limit via

$$g(\theta, \eta) = \lim_{N \to \infty} g_N(\theta, \eta) = \mathbb{E}\left[\overline{h}_N^{(\mathrm{s})}(\theta, \eta)\right].$$

Clearly $\theta = \theta^*, \eta = \eta^*$ solves the over-determined system of M equations

$$g(\theta, \eta) = 0.$$

Since g is not computable we consider the alternative, approximating equation

$$\overline{h}_N(\theta, \eta) = 0,$$

which will typically has no solution since $M > p+q$. Denote by ρ the merged parameter vector, that is, $\rho = (\theta, \eta)$. Following the standard line of arguments for a fixed symmetric, positive definite $M \times M$ weighting matrix K we seek a solution of the 'half-gradient' equation

$$V_N(\rho) = \overline{h}^*_{\rho N}(\rho) K^{-1} \overline{h}_N(\rho) = 0 \tag{5.1.12}$$

to obtain the possibly complex valued $\hat{\rho}_N = (\hat{\theta}_N, \hat{\eta}_N)$. In the next section we concentrate on the identification of the system dynamics under the assumption that η^* is known. We construct the identification procedure along lines of the just presented procedure.

5.2 Estimating the system dynamics

Thus, suppose now that η^* is known and we are able to generate a sequence of i.i.d. random variables statistically equivalent to $\Delta L(\eta^*)$. With a slight abuse of notations we shall use the same notations for real and simulated noise sequences. Define the family of simulated time-series parameterized by θ as follows:

$$\Delta y_n(\theta) = A(\theta) \Delta L_n(\eta^*), \tag{5.2.1}$$

with $-\infty < n < +\infty$. Again, for $\theta = \theta^*$ we recover our observed data in a statistical sense. The score functions are defined as

$$h_{k,n}(\theta) = e^{iu_k^T \Delta y_n} - e^{iu_k^T \Delta y_n(\theta)}. \tag{5.2.2}$$

One could easily mimic the steps of the construction of $V_N(\theta, \eta)$ to define $V_N(\theta)$. Again, fix a finite set of u-s, say $(\mathbf{u}_1, \ldots, \mathbf{u}_M)$. Merge the above scores to define

$$h_n(\theta) = (h_{1,n}(\theta), \ldots, h_{M,n}(\theta))^T, \tag{5.2.3}$$

and its sample average as

$$\overline{h}_N(\theta) = \frac{1}{N} \sum_{n=r+1}^{N+r} h_n(\theta). \tag{5.2.4}$$

$g_N(\theta)$ stands for the expected value of $\overline{h}_N$:

$$g_N(\theta) = \mathbb{E}\left[\overline{h}_N(\theta)\right], \tag{5.2.5}$$

its limit is denoted by $g(\theta) = \lim_{N\to\infty} g_N(\theta)$. Clearly $\theta = \theta^*$ solves the over-determined system of equations

$$g(\theta) = 0.$$

By approximating g by $\overline{h}_N$ we define $\hat{\theta}_N$ as the possibly complex valued solution of

$$V_N(\theta) := \overline{h}^*_{\theta N}(\theta) K^{-1} \overline{h}_N(\theta) = 0. \tag{5.2.6}$$

With a little effort one can show that the asymptotic value of the l.h.s. is then defined as

$$W(\theta) = \lim_{N\to\infty} \mathbb{E}\left[V_N(\theta)\right] = g^*_\theta(\theta) K^{-1} g(\theta).$$

One can also write the recurring auxiliary problem

$$\bar{V}_N(\theta) = H^* K^{-1} \overline{h}_N(\theta) = 0,$$

with $H = g_\theta(\theta^*)$.

Condition 5.3. *θ^* is the unique solution of $W(\theta) = 0$ in D_θ^*.*

Following the arguments given in [31] we get the following result:

Theorem 5.1. *Under Conditions 5.1, 5.2 and 5.3 we have*

$$\hat{\theta}_N - \theta^* = W_\theta^{-1}(\theta^*)\bar{V}_N(\theta^*) + O_M(N^{-1}).$$

Now we are ready to calculate the asymptotic covariance of the estimator.

Theorem 5.2. *Under Conditions 5.1, 5.2 and 5.3 the optimal choice of the weighting matrix K is $K = 2C$ and the asymptotic covariance matrix of $\hat{\theta}_N$ exists and with this choice is given by*

$$\Sigma_{\theta\theta} = 2(H^* C^{-1} H)^{-1},$$

where the k^{th} row of H is given by $-\varphi_\theta^T(\mathbf{u}_k, \theta^*, \eta^*)$.

Proof. The Jacobian of $W(\theta)$ w.r.t. θ at θ^* is

$$R^* = g_\theta^*(\theta^*) K^{-1} g_\theta(\theta^*) = H^* K^{-1} H.$$

Note that since Δy_n and $\Delta y_n(\theta^*)$ are independent as they are generated using two independent copies of $\Delta L_n(\eta^*)$ sequences we have

$$\Lambda'_{k,l} := \mathbb{E}\left[h_{k,n}^{(\mathrm{s})}(\theta^*) h_{l,n}^{(\mathrm{s})*}(\theta^*) \right] = 2\left(\varphi(\mathbf{u}_k - \mathbf{u}_l, \theta^*, \eta^*) - \varphi(\mathbf{u}_k, \theta^*, \eta^*)\varphi(-\mathbf{u}_l, \theta^*, \eta^*)\right).$$

We note in passing that $\Lambda' = 2C$. Thus the asymptotic covariance of $\bar{V}_N$-s is

$$S = H^* K^{-1} \Lambda' K^{-1} H.$$

The asymptotic covariance of the estimator $\hat{\theta}_N$ is then

$$(H^*K^{-1}H)^{-1}\ H^*K^{-1}\Lambda' K^{-1}H\ (H^*K^{-1}H)^{-1}.$$

Using the same linear algebra arguments as for i.i.d. samples we get that the optimal choice of K is

$$K = \Lambda' = 2C$$

yielding the asymptotic covariance for $\hat{\theta}_N$

$$\Sigma_{\theta\theta} = (H^*\Lambda'^{-1}H)^{-1} = 2(H^*C^{-1}H)^{-1}.$$

□

Remark: Recall that $H = g_\theta(\theta^*)$, so that the k^{th} row of H is

$$\frac{\partial}{\partial\theta}\mathbb{E}\left[\overline{h}_k(\theta)\right]\Big|_{\theta=\theta^*} = -\varphi_\theta(\mathbf{u}_k, \theta^*, \eta^*),$$

and that the asymptotic covariance matrix of the noise estimate for i.i.d. sample is $\Sigma_{\eta\eta} = (G^*C^{-1}G)^{-1}$, using the notations of Chapter 3. The covariance matrices

$$\Sigma_{\theta\theta} = 2(H^*C^{-1}H)^{-1}$$

and

$$\Sigma_{\eta\eta} = (G^*C^{-1}G)^{-1}$$

have similar structure. The rows of H and G are derivatives of the characteristic function of the observed data with respect to the unknown parameters θ and η, respectively.

5.3 ECF for i.i.d. data revisited

In this section we give an extension of the ECF method for i.i.d. data under the assumption that the c.f. is not known explicitly, but we do have a computable random variable $\xi(\eta)$ such that

$$\varphi(u,\eta) = \mathbb{E}\left[e^{iu\xi(\eta)}\right].$$

More exactly, we assume that we have a mechanism to compute an i.i.d. sequence $\xi_n(\eta)$ given by

$$\xi_n(\eta) = F(\rho_n, \eta),$$

where ρ_n is an i.i.d. sequence that we can generate, and F is a known function of ρ and η, which is sufficiently smooth in η.

Let the true parameter be denoted by η^*, and let the observed sequence be

$$\xi_n^* = F(\rho_n^*, \eta^*),$$

where (ρ_n^*) is a realization of an i.i.d. sequence with given distribution. The problem is then to identify η^*. The purpose of this exercise is to apply the procedure presented in Section 5.1 for i.i.d. samples. An obvious candidate for a score function is now

$$h_n(u,\eta) = e^{iu\xi_n(\eta^*)} - e^{iu\xi(\eta)},$$

where $\xi_n(\eta^*)$ are real data and $\xi_n(\eta)$ are simulated data. Taking a finite set u-s, say $u_1, \ldots, u_M$, define

$$h_{k,n}(\eta) = e^{iu_k\xi_n(\eta^*)} - e^{iu_k\xi(\eta)}. \tag{5.3.1}$$

From here we may proceed like above to define the function $V_N(\eta)$ and the corresponding objects $\overline{h}_N(\eta)$, its expected value $g_N(\eta)$ and $g = \lim_{N\to\infty} g_N(\eta^*)$. The asymptotic covariance

matrix of this modified ECF estimator for i.i.d. data is given by the following result:

Theorem 5.3. *Denote the asymptotic covariance matrix of the estimator for i.i.d. data without a known characteristic function, but with a random variable that can be generated by (5.0.1), by $\Sigma'_{\eta\eta}$. Then we have*

$$\Sigma'_{\eta\eta} = 2(G^* C^{-1} G)^{-1}.$$

One can easily follow the line of reasoning presented in Chapter 3 and in the previous section to obtain the above asymptotic covariance matrix $\Sigma'_{\eta\eta}$. Note that $\Sigma'_{\eta\eta} = 2\Sigma_{\eta\eta}$.

5.4 Factor models

It has been a widely believed common hypothesis that the dynamics of several stocks, such as SP500, are determined by relatively small number of real or abstract economic factors. The number of these factors can be as low as three. This hypothesis led to the development of factor models. A factor model describes the dynamics of stock prices using only a small number of underlying factors, such as stock market average, gross national product, employment rate or abstract factors. Similar hypothesis can be formulated for the noise sources. Thus we are induced to consider the return process (Δy_t) defined by a linear stochastic systems:

$$\Delta y = F \Delta L, \tag{5.4.1}$$

where F is a transfer function of dimension $n \times m$, where n is the number of financial instruments y_i-s and m is the number of factors. Each Δy_i may correspond to the dynamic of the return of a single stock. The motivation of the problem suggests that n is typically much larger m. Factor models are particularly attractive when the driving factors are modeled by independent Lévy processes $(L_{i,t})$-s. By merging $(L_{i,t})$-s into a vector we define the m-dimensional vector valued Lévy process (L_t). The objective is then to develop some analytical tools for the study

of the random vector process $y_t = (y_{1,t}, \ldots, y_{n,t})$ with the above dynamics. Standard methods for identifying the systems dynamics apply the Kálmán filter with the PE method. The next lemma gives the c.f. of a single observation y_t.

Lemma 5.1. *Let us suppose that the n-dimensional process (Δy_t) is defined by (5.4.1), where (L_t) is a vector valued Lévy process. Then the characteristic function $\varphi_{\Delta y_t}(u)$ of Δy_t can be written as*

$$\exp\left\{\sum_{j=1}^{m}\sum_{s=0}^{\infty}\psi_j\left(u^T F_{\cdot,j,s}\right)\right\}. \tag{5.4.2}$$

where ψ_j is the characteristic exponent of ΔL_j, and $F_{\cdot,j,s}$ denotes the j^{th} column of F.

Proof. For the k^{th} component of y_t we have

$$\Delta y_{k,t} = \sum_{j=1}^{m}\sum_{s=0}^{\infty} f_{k,j,s}\Delta L_{j,t-s}. \tag{5.4.3}$$

Using that the coordinates of ΔL are independent write the c.f. of Δy_t as

$$\begin{aligned}\varphi_{\Delta y_t}(u) = \mathbb{E}\left[\exp\left\{i\sum_{k=1}^{n}u_k\Delta y_{k,t}\right\}\right] = \mathbb{E}\left[\exp\left\{i\sum_{k=1}^{n}\sum_{j=1}^{m}\sum_{s=0}^{\infty}u_k f_{k,j,s}\Delta L_{j,t-s}\right\}\right] = \\ \prod_{j=1}^{m}\prod_{s=0}^{\infty}\left(\exp\left\{\psi_j\left(\sum_{k=1}^{n}u_k f_{k,j,t-s}\right)\right\}\right) = \exp\left\{\sum_{j=1}^{m}\sum_{s=0}^{\infty}\psi_j\left(u^T F_{\cdot,j,s}\right)\right\},\end{aligned} \tag{5.4.4}$$

which concludes the proof. □

This lemma shows that the characteristic function of Δy_t is typically given in the form of an infinite product. Returning to the problem of estimating the dynamics, the key feature of the last observation is that this characteristic function is not known explicitly. The above result can be extended to blocks of data. Hence we face the same problem, the incomputability of an infinite product, like in the previous sections. A potential alternative is to follow the line of arguments presented there: replace the joint c.f. of blocks of unprocessed data by the simulated

empirical c.f. An interesting exception is when (L_j)-s are α-stable processes, in this case the above infinite product reduces to

$$\exp\left\{\sum_{j=1}^{m}|u^T F_{\cdot,j}|^{\alpha}\psi(1)\right\},$$

but generally (5.4.2) cannot be simplified.

Recall that the joint characteristic function of blocks of unprocessed data was also given by an infinite product, see (5.1.5). Thus, one may proceed like we did in the preceding sections to estimate the system dynamics of factor models.

Chapter 6

ECF estimation of the parameters of GARCH processes

6.1 Basic properties of GARCH processes

An important stylized fact of financial time series is that the conditional variance of the return process is not constant in time. This feature of financial data can be expressed by saying that it has a time varying volatility. Therefore the process cannot be modeled by linear systems. Thus, in particular, to analyze the dynamics of highly volatile financial instruments such as indices, foreign exchange rates and commodities, a more sophisticated model should be proposed that can reflect the dynamic volatility of past data resulting in the well-known phenomena of volatility clustering.

The first model that captured the above mentioned stylized fact was introduced by Engle [25]. His model, the so called autoregressive conditional heteroscedasticity (ARCH) model, was refined by Bollerslev [8]. Bollerslev's GARCH (generalized ARCH) model is one of the most widely accepted models recently in the area of financial modeling.

In this chapter we tackle one of key problems of the statistical analysis of GARCH models, the parameter estimation problem, see [37]. When it comes to the identification, the most principled method in the literature is the quasi-maximum likelihood method, see for example [7]. The main objective of this chapter is to study the possibility of adapting the ECF method to GARCH processes with i.i.d. driving noise having known characteristic function. This possibility has not been attracted many researchers until recently. In [43] a goodness of fit test is applied using the empirical characteristic function, while in [29] a Fourier type method is presented for power GARCH processes. Xu in [63] proposes to estimate the parameters of a GARCH model with normal driving noise using the ECF method and presents some empirical investigations.

Technically, the special type of a GARCH(r,s) model to be studied in this chapter is defined via the equations

$$y_n = \sigma_n \Delta L_n \tag{6.1.1}$$

$$\sigma_n^2 - \gamma^* = \sum_{i=1}^{r} \alpha_i^* (y_{n-i}^2 - \gamma^*) + \sum_{j=1}^{s} \beta_j^* (\sigma_{n-j}^2 - \gamma^*), \tag{6.1.2}$$

where $-\infty < n < +\infty$. The driving noise ΔL_n is obtained as the increment of a Lévy process (L_t) with $-\infty < t < +\infty$, and $L_0 = 0$, over an interval $[(n-1)h, nh)$, with $h > 0$ being a fixed sampling interval, and $-\infty < n < +\infty$. The noise characteristic will be denoted by η^*, i.e. the characteristic function of ΔL_n is $\varphi(u, \eta^*)$. We assume that L_t has zero mean and $\mathrm{Var}(\Delta L_n) = 1$. Let $\mathscr{F}^{\Delta L}$ denote the natural filtration, i.e. $\mathscr{F}_n^{\Delta L} = \sigma\{\Delta L_k : k \le n\}$. Under the above conditions γ^* is the conditional variance of y_n and σ_n given $\{y_i : i < n\}$. The unknown parameter vector θ^* is defined as $\theta^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_r^*, \beta_1^*, \ldots, \beta_r^*)^T$. The second order properties of a GARCH process was given by Bollerslev, see [8].

Theorem 6.1. *The GARCH(r,s) process defined by* $(6.1.2)$ *and* $(6.1.1)$ *is second-order stationary with*

$$\mathbb{E}[y_n] = 0,\ \mathrm{Cov}(y_n, y_m) = 0 \ for\ n \neq m$$

and

$$\mathbb{E}\left[y_n^2\right] = \mathbb{E}\left[\sigma_n^2\right] = \frac{\alpha_0^*}{1 - \sum_{i=1}^r \alpha_i^* - \sum_{j=1}^s \beta_j^*}$$

if and only if

$$\sum_{i=1}^r \alpha_i^* + \sum_{j=1}^s \beta_j^* < 1.$$

Definition 6.1. *We say that a Lévy process (L_t) satisfies the moment condition of order Q if*

$$\int_{\mathbb{R}} |x|^q \nu(dx) < \infty$$

holds for $1 \leq q \leq Q$, where the Lévy measure of (L_t) is denoted by $\nu(dx)$.

Define the polynomials

$$C^*(q^{-1}) = \sum_{i=1}^r \alpha_i^* q^{-1} \quad \text{and} \quad D^*(q^{-1}) = 1 - \sum_{j=1}^s \beta_j^* q^{-1}, \tag{6.1.3}$$

with q^{-1} being the backshift operator. In order to guarantee the invertibility of the sensitivity matrix we assume that C^* and D^* are relative prime. Using these polynomials $(6.1.2)$ can be written in the following compact form:

$$D^*(q^{-1})(\sigma_n^2 - \gamma^*) = C^*(q^{-1})(y_n^2 - \gamma^*). \tag{6.1.4}$$

Let us define the $(r+s)$-dimensional state vector

$$X_n^* = (y_n^2, \ldots, y_{n-r+1}^2, \sigma_n^2, \ldots, \sigma_{n-s+1}^2)^T. \tag{6.1.5}$$

It is easy to check that the dynamics of (X_n^*) is then

$$X_{n+1}^* = A_{n+1}^* X_n^* + u_{n+1}^*, \quad n \in \mathbb{Z}, \tag{6.1.6}$$

where $A_n^* \in \mathbb{R}^{(r+s)\times(r+s)}$ is defined in terms of (ΔL_n) as

$$A_n^* = \begin{pmatrix} A_{n;1,1}^* & A_{n;1,2}^* \\ A_{n;2,1}^* & A_{n;2,2}^* \end{pmatrix},$$

where

$$A_{n;1,1}^* = \begin{pmatrix} \alpha_1^*(\Delta L_n)^2 & \alpha_2^*(\Delta L_n)^2 & \cdots & \alpha_{r-1}^*(\Delta L_n)^2 & \alpha_r^*(\Delta L_n)^2 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{6.1.7}$$

$$A_{n;1,2}^* = \begin{pmatrix} \beta_1^*(\Delta L_n)^2 & \beta_2^*(\Delta L_n)^2 & \ldots & \beta_{s-1}^*(\Delta L_n)^2 & \beta_s^*(\Delta L_n)^2 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \tag{6.1.8}$$

$$A^*_{n;2,1} = \begin{pmatrix} \alpha^*_1 & \alpha^*_2 & \cdots & \alpha^*_{r-1} & \alpha^*_r \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \tag{6.1.9}$$

$$A^*_{n;2,2} = \begin{pmatrix} \beta^*_1 & \beta^*_2 & \dots & \beta^*_{s-1} & \beta^*_s \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{6.1.10}$$

and

$$u^*_n = (\alpha^*_0(\Delta L_n)^2, 0, \dots, 0, \alpha^*_0, 0, \dots, 0)^T,$$

for each n, with $\alpha^*_0 = \gamma^* \left(1 - \sum_{i=1}^r \alpha^*_i - \sum_{j=1}^s \beta^*_j\right)$. Note that $(A^*_n, u^*_n), n \in \mathbb{Z}$ is a sequence of i.i.d. random matrices. Moreover, (X^*_n) is a Markov process with unobservable components. The above given state space representation, which is the slight modification of the one introduced by Bougerol and Picard [9], will be useful for proving L-mixing properties of $(y_n), (\sigma_n)$ and related processes.

Define the $p = r + s + 1$-dimensional parameter vector

$$\theta = (\alpha_0, \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_r)^T$$

and the real domain

$$D = \left\{ \theta \left| \sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1 \right. \right\}.$$

Note that for $\theta \in D$ the corresponding GARCH process with parameter vector θ is well-defined.

We also define a corresponding complex domain D_ϵ by

$$D_\epsilon = \left\{ \theta \left| \mathrm{Re}\left(\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j \right) < 1,\ |\mathrm{Im}(\alpha_i)| < \epsilon,\ |\mathrm{Im}(\beta_j)| < \epsilon \right. \right\}$$

Let $D_\epsilon^* \subset \mathrm{int}\ D_\epsilon$ be a compact domain such that $\theta^* \in \mathrm{int}\ D_\epsilon^*$. For a fixed value $\theta \in D_\epsilon$ we invert the GARCH system to recover the driving noise. Define the process $(\sigma_n(\theta))$ in terms of y_n:

$$\sigma_n^2(\theta) - \gamma = \sum_{i=1}^{r} \alpha_i (y_{n-i}^2 - \gamma) + \sum_{j=1}^{s} \beta_j (\sigma_{n-j}^2(\theta) - \gamma), \tag{6.1.11}$$

with initial values $y_n = 0,\ \sigma_n^2(\theta) = \gamma,$ for all $n \leq 0$. Then the estimated driving noise is defined as

$$\varepsilon_n(\theta) = \frac{y_n}{\sigma_n(\theta)} \tag{6.1.12}$$

for $n \geq 0$. Note that for $\theta = \theta^*$ the stationary solution of the inverse is

$$\varepsilon_n^{(\mathrm{s})}(\theta^*) = \frac{y_n}{\sigma_n^{(\mathrm{s})}(\theta)} = \Delta L_n,$$

which is obtained by letting $-\infty < n < \infty$. Note that if $\theta = \theta^*$, then $\sigma_n(\theta)$ recovers σ_n at least in a statistical sense.

6.2 ML method for GARCH processes

In this section we develop and implement a ML estimator for GARCH models driven by Lévy processes. As in the subsequent sections we will present an ECF identification method corresponding to the third stage of three-stage identification method for linear systems that is essentially asymptotically efficient. Before proceeding to this in this section we compute the asymptotic covariance of the ML method itself, to be able to compare the two asymptotic covariance matrices.

It can be shown along the lines of the ML method for linear systems that for the joint density function f_Y of $(y_1, \ldots, y_n)$ and the joint density function $f_{\Delta L}$ of $(\Delta L_1, \ldots, \Delta L_n)$ we have

$$f_Y(y_1, \ldots, y_n) = \prod_{k=1}^{n} \sigma_k(\theta)^{-1} f_{\Delta L}(\Delta L_1, \ldots, \Delta L_n),$$

because the determinant of the Jacobian of the transformation $(\Delta L_1, \ldots, \Delta L_n) \to (y_1, \ldots, y_n)$ is $\prod_{k=1}^{n} \sigma_k(\theta)^{-1}$. We have that the asymptotic cost function of the ML estimator is given by

$$\begin{aligned} W(\theta) = \lim_{n\to\infty} \mathbb{E}\left[-\log\left(f\left(\varepsilon_n(\theta)\right)\sigma_n(\theta)^{-1}\right)\right] = \mathbb{E}\left[-\log\left(f\left(\varepsilon_n^{(\mathrm{s})}(\theta)\right)\sigma_n^{(\mathrm{s})}(\theta)^{-1}\right)\right] = \\ \mathbb{E}\left[-\log\left(f\left(\varepsilon_n^{(\mathrm{s})}(\theta)\right)\right) + \log\left(\sigma_n^{(\mathrm{s})}(\theta)\right)\right], \end{aligned} \tag{6.2.1}$$

where f denotes the density function of ΔL_n. It is easy to check that $W_\theta(\theta^*) = 0$ holds. For,

$$\begin{aligned} W_\theta(\theta^*) = \mathbb{E}\left[\frac{f'(\Delta L_n)}{f(\Delta L_n)}\Delta L_n \frac{\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)}{\sigma_n^{(\mathrm{s})}(\theta^*)} + \frac{\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)}{\sigma_n^{(\mathrm{s})}(\theta^*)}\right] = \\ \mathbb{E}\left[\mathbb{E}\left[\left(\frac{f'(\Delta L_n)}{f(\Delta L_n)}\Delta L_n + 1\right)\frac{\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)}{\sigma_n^{(\mathrm{s})}(\theta^*)}\middle| \mathscr{F}_{n-1}^{\Delta L}\right]\right] = \mathbb{E}\left[\frac{\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)}{\sigma_n^{(\mathrm{s})}(\theta^*)}\mathbb{E}\left[\frac{f'(\Delta L_n)}{f(\Delta L_n)}\Delta L_n + 1\right]\right], \end{aligned} \tag{6.2.2}$$

because $\varepsilon_{\theta n}(\theta^*) = -\Delta L_n \frac{\sigma_{\theta n}(\theta^*)}{\sigma_n(\theta^*)}$, and $\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)$ and $\sigma_n^{(\mathrm{s})}(\theta^*)$ are $\mathscr{F}_{n-1}^{\Delta L}$ measurable. Note that under appropriate regularity conditions on f we have

$$\mathbb{E}\left[\frac{f'(\Delta L_n)}{f(\Delta L_n)}\Delta L_n\right] = \int_{\mathbb{R}} f'(x)x dx = [xf(x)]_{-\infty}^{\infty} - \int_{\mathbb{R}} f(x)dx = -1, \tag{6.2.3}$$

which implies the claim. For the Hessian of W write

$$W_{\theta\theta}(\theta) = -\lim_{n\to\infty}\left(\mathbb{E}\left[\frac{f''(\varepsilon_n^{(\mathrm{s})}(\theta))}{f(\varepsilon_n^{(\mathrm{s})}(\theta))}\varepsilon_{\theta n}^{(\mathrm{s})}(\theta)\varepsilon_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta) - \frac{f'^2(\varepsilon_n^{(\mathrm{s})}(\theta))}{f^2(\varepsilon_n^{(\mathrm{s})}(\theta))}\varepsilon_{\theta n}^{(\mathrm{s})}(\theta)\varepsilon_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta) + \frac{f'(\varepsilon_n^{(\mathrm{s})}(\theta))}{f(\varepsilon_n^{(\mathrm{s})}(\theta))}\varepsilon_{\theta\theta n}^{(\mathrm{s})}(\theta)\right]\right.$$
$$\left. + \mathbb{E}\left[\frac{\sigma_{\theta\theta n}^{(\mathrm{s})}(\theta)\sigma_n^{(\mathrm{s})}(\theta) - \sigma_{\theta n}^{(\mathrm{s})}(\theta)\sigma_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta)}{\sigma_n^{(\mathrm{s})2}(\theta)}\right]\right). \tag{6.2.4}$$

Using again the fact that $\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)$ and $\sigma_n^{(\mathrm{s})}(\theta^*)$ are $\mathscr{F}_{n-1}^{\Delta L}$ measurable we get that $W_{\theta\theta}(\theta^*)$ equals to

$$\mathbb{E}\left[\frac{\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)\sigma_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta^*)}{\sigma_n^{(\mathrm{s})2}(\theta^*)}\right]\mathbb{E}\left[\frac{f'^2(\Delta L_n)}{f^2(\Delta L_n)}(\Delta L_n)^2 - \frac{f''(\Delta L_n)}{f(\Delta L_n)}(\Delta L_n)^2 - 2\frac{f'(\Delta L_n)}{f(\Delta L_n)}\Delta L_n - 1\right] +$$
$$\mathbb{E}\left[\frac{\sigma_{\theta\theta n}^{(\mathrm{s})}(\theta^*)}{\sigma_n^{(\mathrm{s})}(\theta^*)}\right]\mathbb{E}\left[\frac{f'(\Delta L_n)}{f(\Delta L_n)}\Delta L_n + 1\right]. \tag{6.2.5}$$

Using that under appropriate technical conditions we have

$$\int_{\mathbb{R}} f''(x)x^2dx = [f'(x)x^2]_{-\infty}^{\infty} - 2\int_{\mathbb{R}} f'(x)xdx = 2,$$

the previous formula can be written as

$$\mathbb{E}\left[\frac{\sigma_{\theta n}^{(\mathrm{s})}(\theta^*)\sigma_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta^*)}{\sigma_n^{(\mathrm{s})2}(\theta^*)}\right]\mathbb{E}\left[\frac{f'^2(\Delta L_n)}{f^2(\Delta L_n)}(\Delta L_n)^2 - 1\right]. \tag{6.2.6}$$

By almost identical calculation we get that the covariance of the gradient of log-likehood function

$$l_{\theta,n}(\theta) = \frac{f'(\Delta L_n)}{f(\Delta L_n)}\varepsilon_{\theta n}^{(\mathrm{s})}(\theta) + \frac{\sigma_{\theta n}^{(\mathrm{s})}(\theta)}{\sigma_n^{(\mathrm{s})}(\theta)}$$

at $\theta = \theta^*$ is given by

$$\mathrm{Cov}(l_{\theta,n}(\theta^*), l^T_{\theta,n}(\theta^*)) = \mathbb{E}\left[\frac{\sigma^{(\mathrm{s})}_{\theta n}(\theta^*)\sigma^{(\mathrm{s})\mathrm{T}}_{\theta n}(\theta^*)}{\sigma^{(\mathrm{s})2}_n(\theta^*)}\right]\mathbb{E}\left[\frac{f'^2(\Delta L_n)}{f^2(\Delta L_n)}(\Delta L_n)^2 - 1\right]. \tag{6.2.7}$$

Thus we have the following lemma.

Lemma 6.1. *Let $\hat{\theta}_N$ the ML estimate of the parameters of a GARCH process. Then the asymptotic covariance of $\sqrt{N}\left(\hat{\theta}_N - \theta^*\right)$ is*

$$\mu^{-1}(M^*)^{-1}, \tag{6.2.8}$$

with

$$M^* = \mathbb{E}\left[\frac{\sigma^{(\mathrm{s})}_{\theta N}(\theta^*)\sigma^{(\mathrm{s})\mathrm{T}}_{\theta N}(\theta^*)}{\sigma^{(\mathrm{s})2}_N(\theta^*)}\right],$$

and

$$\mu = \mathbb{E}\left[\frac{f'(\Delta L_N)^2}{f^2(\Delta L_N)}(\Delta L_N)^2 - 1\right]. \tag{6.2.9}$$

The precise definition of the random vector $\hat{\theta}_N$ is analogous with that of the solution of $(3.1.5)$. A very nice interpretation of this μ is that it can be also obtained as the Fisher information of a scale parameter estimation problem. Suppose that we are given an i.i.d. realization of the scaled random variable $\lambda\Delta L_1$, with the true value of λ being $\lambda^* = 1$. Then the λ-dependent density $f(x,\lambda)$ of $\lambda\Delta L_1$ is

$$f(x,\lambda) = f\left(\frac{x}{\lambda}\right)\frac{1}{\lambda},$$

where $f(\cdot)$ denotes the density function of ΔL_1. Write

$$\frac{\partial}{\partial\lambda}\log f(x,\lambda) = \frac{f'\left(\frac{\lambda}{x}\right)}{f\left(\frac{\lambda}{x}\right)}\left(-\frac{x}{\lambda^2}\right) - \frac{1}{\lambda}, \tag{6.2.10}$$

Hence taking into account (6.2.3) the Fisher information reads as

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\lambda}\log f(\Delta L_1,\lambda)\right)^2\Big|_{\lambda=1}\right]=\mathbb{E}\left[\left(-\frac{f'(\Delta L_1)}{f(\Delta L_1)}\Delta L_1-1\right)^2\right]= \\ \mathbb{E}\left[\frac{f'^2(\Delta L_1)}{f^2(\Delta L_1)}\left(\Delta L_1\right)^2+1+2\frac{f'(\Delta L_1)}{f(\Delta L_1)}\Delta L_1\right]=\mathbb{E}\left[\frac{f'^2(\Delta L_1)}{f^2(\Delta L_1)}\left(\Delta L_1\right)^2-1\right]. \tag{6.2.11}$$

Therefore we get the following lemma.

Lemma 6.2. *μ in (6.2.9) can be interpreted as a Fisher information of a scale parameter estimate.*

In analogy with the analysis of the efficiency of the three-stage method for linear systems this property of μ will have a key role in proving the essentially asymptotic efficiency of the ECF method for GARCH systems.

6.3 ECF method for GARCH processes

Now we turn to the problem of identifying the parameters of a GARCH process by adapting the approach of the ECF method. The ideas presented in this section show several similarities with those of Chapter 4, yet we will see that the different model structure poses numerous new problems. Despite the fact that the dynamics of a GARCH process can be described as a Markov process, the method presented in [14] does not solve this problem as it is not capable of dealing with unobservable components. For GARCH models only (y_n) is observable and (σ_n) is a latent process. The paper of Carrasco, Chernov, Florens and Ghysels [14] tackles the problem of estimating the parameters of an observable Markov process. Hereby we briefly summarize their findings. Let X_t be a Markov process that is generated with some unknown parameter vector θ_0. Let $\varphi(s|X_t;\theta)$ denote the conditional characteristic function

$$\mathbb{E}\left[e^{isX_{t+1}}|X_t\right].$$

The score functions used in the method are defined by

$$h(r, s, X_t, X_{t+1}; \theta) = e^{irX_t} \left(e^{isx_{t+1}} - \varphi \left(s | X_t; \theta_0 \right) \right).$$

They prove that under some conditions using continuum moment condition yields an estimator that reaches the Cramer-Rao bound.

While this is a very attractive result, it does not solves the problems we consider in this thesis. The process X_t is supposed to be observable, their proposed method cannot handle latent components. The presence of latent component is natural in GARCH processes, hence the method is not applicable for such processes. For such non-Markovian processes they propose to use the joint characteristic function instead of the conditional one. Define the joint characteristic function as

$$\varphi(\tau, t, n) = \mathbb{E}\left[e^{i\tau^T Y_t} \right],$$

where $\tau = (\tau_0, \ldots, \tau_n)^T$ and $Y_t = (X_t, \ldots, X_{t+n})$. This problem has been considered in [26] by Feuerverger. The unknown parameter θ is estimated by fitting the empirical joint characteristic function to the joint characteristic function using a weighting function. Feuerverger showed that this estimator is as efficient as the one that obtains $\hat{\theta}$ by solving

$$\frac{1}{N} \sum_{k=1}^{N} \frac{\partial}{\partial \theta} \ln f(X_{k+n} | X_{k+n-1}, \ldots, X_{k+n}; \theta) = 0,$$

and shows that the resulting estimator is not efficient for non-Markovian cases. In [26] it is claimed that the variance of the estimator can be arbitrarily close to the Cramer-Rao bound if n is chosen sufficiently large, but no proof is presented. Even if this claim were valid the implementation of the procedure for large L would problematic. Moreover, Carrasco et all. argues that for large n the available data provide only a few observation vectors of length n.

In this section as an alternative of the ML method we adapt the ECF method for GARCH

processes. The motivation behind the adaptation of the ECF method again lies in the fact that the density function of ΔL_n is typically unknown. Still our proposed procedure estimates θ^* as efficiently as the ML method. We suppose that the characteristic η^* of the noise is given and we are to identify the system parameters θ^*. Although the three-stage method can be applied for GARCH processes to identify the system and the noise characteristics, the results that we obtained for linear systems cannot be reproduced. The problem of identifying both the system parameters and the noise parameters will be briefly discussed at the end of the chapter. The following paragraphs present the identification method with known η^*.

First, for each θ we define the estimated volatility $\sigma_n^2(\theta)$ and the estimated driving noise $\varepsilon_n(\theta)$ for $\theta \in D_\epsilon$, see equations (6.1.11) and (6.1.12). Following the philosophy of the ECF method take a fix set u_i-s, $1 \leq i \leq M$. We define the $p \times 1$-dimensional modified primary score functions as

$$h_{k,n}(\theta) = \left(e^{iu_k \varepsilon_n(\theta)} - \varphi(u)\right) \frac{\sigma_{\theta n}(\theta)}{\sigma_n(\theta)}, \tag{6.3.1}$$

where the modification being the usage of the instrumental variable $\frac{\sigma_{\theta n}(\theta)}{\sigma_n(\theta)}$. The choice of the instrumental variable $\frac{\sigma_{\theta n}(\theta)}{\sigma_n(\theta)}$ is suggested by the construction of ECF method for linear systems. Namely, recall that for linear Lévy systems the modified primary score functions were defined via

$$\left(e^{iu_k \varepsilon_n(\theta)} - \varphi(u)\right) \varepsilon_{\theta n}(\theta),$$

where the instrumental variable $\varepsilon_{\theta n}(\theta)$ satisfies $\lim_{n\to\infty} \mathbb{E}\left[\varepsilon_{\theta n}(\theta^*)\varepsilon_{\theta n}^T(\theta^*)\right] = R_P^*$. By analogical thinking for GARCH processes we choose the instrumental variable $IV_n(\theta)$ such that $\lim_{n\to\infty} \mathbb{E}\left[IV_n(\theta^*)IV_n^T(\theta^*)\right] = M^*$, hence the choice of

$$\frac{\sigma_{\theta n}(\theta)}{\sigma_n(\theta)}.$$

Surprisingly we will see that this ad-hoc choice of instrumental variable yields an essentially

asymptotically efficient identification method.

Since $\sigma_n(\theta), \sigma_{\theta n}(\theta)$ are $\mathscr{F}_{n-1}^{\Delta L}$ measurable

$$\mathbb{E}\left[h_{k,n}^{(\mathrm{s})}(\theta^*)\right] = 0$$

holds. In analogy with the linear case merge the score functions $h_{k,n}(\theta)$-s into a $(r+s+1)M$-dimensional column vector

$$h_n(\theta) = \left(h_{1,n}^T(\theta), \ldots, h_{M,n}^T(\theta)\right)^T.$$

Define $\overline{h}_N(\theta) = \frac{1}{N}\sum_{n=1}^N h_n(\theta)$ the averaged score vector and

$$g_N(\theta) = \mathbb{E}\left[\overline{h}_N(\theta)\right] \text{ and } g(\theta) = \lim_{N\to\infty} g_N(\theta).$$

Note that the system of equations

$$g(\theta) = 0$$

is over-determined with solution $\theta = \theta^*$, hence following the idea presented in Chapter 4 we redefine the score function as follows. Fix a symmetric, positive definite, $pM \times pM$ weighting matrix K. Since g is not computable we approximate it by $\overline{h}_N$ and we seek a solution for the 'half-gradient' equation

$$V_N'(\theta) = \overline{h}_{\theta N}^*(\theta) K^{-1} \overline{h}_N(\theta) = 0 \tag{6.3.2}$$

to obtain $\hat{\theta}_N$. We note in passing that the system of equations in (6.3.2) is no longer over-determined because $\dim V_N = r+s+1$. We mimic the steps of Chapter 4 in defining

$$G = g_\theta(\theta^*),$$

and the auxiliary equation

$$\bar{V}'_N(\theta) = G^* K^{-1} \overline{h}_N(\theta) = 0. \tag{6.3.3}$$

The asymptotic cost function is then given by

$$\bar{W}'(\theta) = \lim_{N\to\infty} \mathbb{E}[V'_N(\theta)] = g^*_\theta(\theta) K^{-1} g(\theta),$$

and its Jacobian at $\theta = \theta^*$ is

$$R^*_G = G^* K^{-1} G.$$

Condition 6.1. *The equation $\bar{W}'(\theta) = 0$ has a unique solution $\theta = \theta^*$ in D^*_ϵ.*

We will use our recurring $M \times M$ auxiliary matrix C with elements

$$C_{k,l} = \varphi(u_k - u_l, \eta^*) - \varphi(u_k, \eta^*)\varphi(-u_l, \eta^*),$$

recall that C is the covariance matrix of the primary score functions used in the i.i.d. ECF method.

6.4 Analysis of the ECF method for GARCH processes

To analyze the process $\sigma^2_n(\theta)$ we expand the state vector

$$X^*_n = (y^2_n, \ldots, y^2_{n-r+1}, \sigma^2_n, \ldots, \sigma^2_{n-s+1})^T$$

to

$$\overline{X}_n(\theta) = (X^{*T}_n, \sigma^2_n(\theta), \ldots, \sigma^2_{n-s+1}(\theta))^T. \tag{6.4.1}$$

Then the dynamics of $\overline{X}_n(\theta)$ can be written as

$$\overline{X}_{n+1}(\theta) = \overline{A}_{n+1}(\theta)\overline{X}_n(\theta) + \overline{u}^*_{n+1}, \tag{6.4.2}$$

where

$$\overline{A}_n(\theta) = \begin{pmatrix} A^*_n & Z \\ M_{2,1}(\theta) & M_{2,2}(\theta) \end{pmatrix},$$

with Z being an $(r+s) \times s$ zero matrix,

$$M_{2,1}(\theta) = \left(\begin{array}{ccc|ccc} \alpha_1 & \cdots & \alpha_r & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right)$$

is an $s \times (r+s)$ dimensional matrix, and

$$M_{2,2}(\theta) = \left(\begin{array}{cccc|c} \beta_1 & \beta_2 & \cdots & \beta_{s-1} & \beta_s \\ \hline 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{array}\right)$$

is of dimension $s \times s$, and finally

$$\overline{u}^*_n = (u^{*T}_n, 0, \ldots, 0)^T.$$

First we state two theorems from the theory of block-triangular random matrices that we will use in the proofs, see [53]. $\rho(P)$ stands for the spectral radius of matrix P.

Theorem 6.2. *Let*

$$P = \begin{pmatrix} P_1 & 0 \\ B & P_2 \end{pmatrix}$$

be a random $(d_1+d_2)\times(d_1+d_2)$ *matrix in* $L^2(\Omega, \mathscr{F}, P)$, *with* P_1 *and* P_2 *being square matrices. Then*

$$\rho\left[\mathbb{E}\left[P \otimes P\right]\right] = \max\{\rho\left[\mathbb{E}\left[P_1 \otimes P_1\right]\right], \rho\left[\mathbb{E}\left[P_2 \otimes P_2\right]\right]\}$$

Similarly, let q *be a positive integer and let us assume that* $P \in L^q(\Omega, \mathscr{F}, P)$, *then*

$$\rho\left[\mathbb{E}\left[P^{\otimes q}\right]\right] = \max\{\rho\left[\mathbb{E}\left[P_1^{\otimes q}\right]\right], \rho\left[\mathbb{E}\left[P_2^{\otimes q}\right]\right]\}.$$

Theorem 6.3. *Let* (P_n) *be an i.i.d. sequence of random matrices such that* $||P_1|| \in L^q$. *Assume that for some even integer* $q \geq 2$

$$\rho\left[\mathbb{E}\left[P_1^{\otimes q}\right]\right] < 1$$

holds. Then

$$\lambda_q := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}||P_n \cdots P_1||^q < 0.$$

It follows that for any $\varepsilon > 0$ *we have*

$$\mathbb{E}||P_n \cdots P_1||^q \leq Ce^{(\lambda_q+\varepsilon)n}$$

with some $C = C(\varepsilon) > 0$.

The next lemma implies the L-mixing property of the state vector.

Lemma 6.3. *Let* $D(q^{-1})$ *be stable for all* $\theta \in D_\epsilon$ *and suppose that for some positive even* Q *we have*

$$\rho\left[\mathbb{E}\left[(A_0^*)^{\otimes Q}\right]\right] < 1.$$

Then the process $\left(\overline{X}_n(\theta)\right)$ *is* L*-mixing of order Q uniformly in* $\theta \in D_\epsilon$.

Proof. Fix a $\tau \in \mathbb{Z}^+$ and iterate the state space equation (6.4.2)

$$\begin{aligned}\overline{X}_n(\theta) = \overline{A}_n(\theta)\overline{X}_{n-1}(\theta) + \overline{u}_n^* = \overline{A}_n(\theta)\overline{A}_{n-1}(\theta)\overline{X}_{n-2}(\theta) + \overline{u}_n^* + \overline{A}_n(\theta)\overline{u}_{n-1}^* = \ldots = \\ \overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\overline{X}_{n-\tau}(\theta) + \overline{u}_n^* + \overline{A}_n(\theta)\overline{u}_{n-1}^* + \ldots + \overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\overline{u}_{n-\tau}^*\end{aligned} \tag{6.4.3}$$

Observe that

$$\overline{u}_n^* + \overline{A}_n(\theta)\overline{X}_{n-1}(\theta) + \ldots + \overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\overline{u}_{n-\tau}^*$$

is $\mathscr{F}_{n-\tau}^+ = \sigma\{\Delta L_i : i \geq n-\tau\}$ measurable, thus

$$\begin{aligned}&\mathbb{E}\left[\overline{X}_n(\theta)|\mathscr{F}_{n-\tau}^+\right] = \overline{u}_n^* + \overline{A}_n(\theta)\overline{X}_{n-1}(\theta) + \ldots + \overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\overline{u}_{n-\tau}^* + \\ &\mathbb{E}\left[\overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\overline{X}_{n-\tau}(\theta)|\mathscr{F}_{n-\tau}^+\right] = \\ &\overline{u}_n^* + \overline{A}_n(\theta)\overline{X}_{n-1}(\theta) + \ldots + \overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\overline{u}_{n-\tau}^* + \\ &\overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\mathbb{E}\left[\overline{X}_{n-\tau}(\theta)\right],\end{aligned}$$

because $\overline{X}_{n-\tau}(\theta)$ is independent of $\mathscr{F}_{n-\tau}^+$. It follows that

$$\overline{X}_n(\theta) - \mathbb{E}\left[\overline{X}_n(\theta)|\mathscr{F}_{n-\tau}^+\right] = \overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\left(\overline{X}_{n-\tau}(\theta) - \mathbb{E}\left[\overline{X}_{n-\tau}(\theta)\right]\right). \tag{6.4.4}$$

Since $\overline{X}_{n-\tau}(\theta)$ is independent of $\overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)$ and $||AB|| \leq ||A||\ ||B||$ for the L^q-norm of (6.4.4) we have

$$\begin{aligned}\mathbb{E}^{1/q}\left[||\overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)\left(\overline{X}_{n-\tau}(\theta) - \mathbb{E}\left[\overline{X}_{n-\tau}(\theta)\right]\right)||^q\right] \leq \\ \mathbb{E}^{1/q}\left[||\overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)||^q\right]\mathbb{E}^{1/q}\left[||\overline{X}_{n-\tau}(\theta) - \mathbb{E}\left[\overline{X}_{n-\tau}(\theta)\right]||^q\right].\end{aligned} \tag{6.4.5}$$

It is easy to see that $\overline{X}_n(\theta) - \mathbb{E}\left[\overline{X}_n(\theta)\right]$ is M-bounded of order Q, and for the first term of the two-term product on the l.h.s. using Theorem 6.2 with the choice $P_1 = A_n^*$ and $P_2 = M_{2,2}(\theta)$

yields $\rho\left[\mathbb{E}\left[\overline{A}_1(\theta)^{\otimes q}\right]\right] < 1$. Note that in this case the trivial version of Theorem 6.2 is used as P_2 is non-random. Hence, Theorem 6.3 implies that

$$\mathbb{E}^{1/q}||\overline{A}_n(\theta)\cdots\overline{A}_{n-\tau+1}(\theta)||^q \leq C^{1/q}e^{(\lambda_q+\varepsilon)\tau/q}.$$

Then choose $\varepsilon > 0$ such that $\lambda_q + \varepsilon < 0$. It follows that $\gamma_q(\tau, \overline{X}(\theta))$ is summable, which means by definition that $\left(\overline{X}(\theta)_n\right)$ is L-mixing or order Q uniformly in $\theta \in D_\epsilon$. □

Lemma 6.4. *The process* $\overline{X}_{e,n}(\theta) := \left(\overline{X}_n^T(\theta), \overline{X}_{\theta n}^T(\theta)\right)^T$ *is* L*-mixing of order* Q *uniformly in* $\theta \in D_\epsilon$.

Proof. In order to analyze the derivative process we first determine its dynamics. Suppose that we have a general parameter dependent recursion given by

$$\xi_{n+1}(\theta) = F_{n+1}(\theta)\xi_n(\theta) + v_{n+1}(\theta), \tag{6.4.6}$$

and we are interested in the dynamic of the derivative process $\xi_{\theta n}(\theta)$. For simplicity we assume that θ is a scalar parameter, differentiating (6.4.6) we obtain

$$\xi_{\theta,n+1}(\theta) = F_{\theta,n+1}(\theta)\xi_n(\theta) + F_{n+1}(\theta)\xi_{\theta,n}(\theta) + v_{\theta,n+1}(\theta). \tag{6.4.7}$$

Thus the dynamics of the extended state vector $\xi_{e,n} = (\xi_n^T(\theta), \xi_{\theta,n}^T(\theta))^T$ can be written in a compact form:

$$\xi_{e,n+1}(\theta) = F_{e,n+1}(\theta)\xi_{e,n}(\theta) + v_{e,n+1}(\theta), \tag{6.4.8}$$

with

$$F_{e,n}(\theta) = \begin{pmatrix} F_n(\theta) & 0 \\ F_{\theta,n}(\theta) & F_n(\theta) \end{pmatrix},$$

and $v_{e,n}(\theta) = \left(v_n^T(\theta), v_{\theta,n}^T(\theta)\right)^T$.

It follows that the state-transition matrix, say $\overline{A}_{e,n}(\theta)$, of the dynamics of $\overline{X}_{e,n}(\theta)$ has two identical blocks in the diagonal, namely $\overline{A}_n(\theta)$-s. Hence Theorem 6.3 implies that

$$\rho\left[\mathbb{E}\left[\overline{A}_{e,n}(\theta)^{\otimes q}\right]\right] = \rho\left[\mathbb{E}\left[\overline{A}_n(\theta)^{\otimes q}\right]\right] < 1.$$

Mimicking the steps of the proof of the previous lemma we obtain that $\overline{X}_{e,n}(\theta)$ is L-mixing of order Q uniformly in $\theta \in D_\epsilon$. Similarly, the same can be shown if we further expand $\overline{X}_{e,n}(\theta)$ with the higher order derivatives of $\sigma_n(\theta)$. □

As consequence we get that

$$h_{k,n}(\theta) = \left(e^{iu_k\varepsilon_n(\theta)} - \varphi(u)\right)\frac{\sigma_{\theta n}(\theta)}{\sigma_n(\theta)}$$

and their derivatives w.r.t. θ up to order three are L-mixing of order Q. From now on we may proceed as we did in Chapter 4. We get the following major result which is a precise characterization of the estimation error:

Theorem 6.4. *Assume that Condition 6.1 holds. Let $D(q^{-1})$ be stable for all $\theta \in D_\epsilon$ and suppose that for some positive even Q we have*

$$\rho\left[\mathbb{E}\left[(A_0^*)^{\otimes Q}\right]\right] < 1.$$

Then for the estimation error we have

$$\hat{\theta}_N - \theta^* = -(R_G^*)^{-1}\bar{V}_N'(\theta^*) + O_M^{Q/(2(r+s+1))}(N^{-1}).$$

The last formula equivalently can be written as

$$\hat{\theta}_N - \theta^* = -(R_G^*)^{-1}G^*K^{-1}\overline{h}_N(\theta^*) + O_M^{Q/(2(r+s+1))}(N^{-1}).$$

6.5 Efficiency of the ECF method for GARCH processes

In this section we show that the proposed ECF identification method gives an essentially asymptotically efficient estimate of the system characteristics of a GARCH process.

Theorem 6.5. *Choose $K = C \otimes M^*$, then for the estimate $\hat{\theta}_N$ obtained with the method presented in the previous section we have*

$$\mathbb{E}\left[N\left(\hat{\theta}_N - \theta^*\right)\left(\hat{\theta}_N - \theta^*\right)^*\right] = \Sigma_{\theta\theta} + O_M(N^{-1/2}),$$

where the asymptotic covariance matrix is given by

$$\Sigma_{\theta\theta} = \left(\phi^* C^{-1} \phi\right)^{-1} (M^*)^{-1},$$

with $\phi = (u_1\varphi'(u_1), \ldots, u_M\varphi'(u_M))^T$.

The proof is analogous with that of Theorem 4.6. Note that ϕ and ψ in Theorem 4.6 have similar structure, but now

$$\mathbb{E}\left[h^{(\mathrm{s})}_{\theta,k,n}(\theta^*)\right] = \mathbb{E}\left[e^{iu_k\Delta L_n} i u_k \varepsilon^{(\mathrm{s})}_{\theta n}(\theta^*) \frac{\sigma^{(\mathrm{s})}_{\theta n}(\theta^*)}{\sigma^{(\mathrm{s})}_{n}(\theta^*)}\right] = u_k\varphi'(u_k, \eta^*)M^*.$$

Now we will demonstrate that the above presented ECF method gives an essentially asymptotically efficient estimate $\hat{\theta}_N$. The line of reasoning is analogous with the one in the proof of Theorem 4.6. Suppose that we use the full continuum of moment conditions. Then the continuous version of (6.3.3) would read as

$$< K^{-1}G, \overline{h}_N >= 0,$$

where the inner product is defined on $H = L^2(\pi) = \left\{ f : \mathbb{R} \to \mathbb{C} \,\middle|\, \int |f(t)|^2 \pi(t) dt < \infty \right\}$ via

$$< f, g >= \int f(t) g^*(t) \pi(t) dt,$$

with π being a probability measure on $\mathbb{R}$.

Define the π-dependent covariance operator

$$(Cf)(s) = \int c(s,t) f(t) \pi(t) dt, \tag{6.5.1}$$

with

$$c(s,t) = \mathbb{E}\left[h_{s,n}(\theta^*, \eta^*) h^*_{t,n}(\theta^*, \eta^*) \right].$$

If the full continuum of u-s were defined via $u_s = s$ for all $s \in \mathbb{R}$, then the continuous version of Theorem 6.5 would give

$$\lim_{N\to\infty} \mathbb{E}\left[N \left(\hat{\theta}_N - \theta^* \right) \left(\hat{\theta}_N - \theta^* \right)^* \right] = \left(||u\varphi'(u, \eta^*)||^2_C \right)^{-1} (M^*)^{-1} \tag{6.5.2}$$

for the asymptotic covariance matrix of the estimate $\hat{\theta}_N$. Note that, like for linear system, in the above formula the asymptotic covariance matrix decouples, $||u\varphi'(u, \eta^*)||^2_C$ depends only on η^* and R^*_P depends on the parameters of the GARCH system.

Now we are ready to demonstrate that the proposed estimation method is essentially asymptotically efficient provided the full continuum of moment conditions is available.

Theorem 6.6. *Under the conditions of Theorem 4.6 the estimate $\hat{\theta}_N$ is essentially asymptotically efficient.*

Proof. Recall that the asymptotic covariance of the ML estimate of the parameters of GARCH processes is

$$\mu^{-1} (M^*)^{-1}, \tag{6.5.3}$$

with

$$M^* = \mathbb{E}\left[\frac{\sigma^{(\mathrm{s})}_{\theta N}(\theta^*)\sigma^{(\mathrm{s})\mathrm{T}}_{\theta N}(\theta^*)}{\sigma^{(\mathrm{s})2}_{N}(\theta^*)}\right],$$

and

$$\mu = \mathbb{E}\left[\frac{f'^2(\Delta L_n)}{f^2(\Delta L_n)}(\Delta L_n)^2 - 1\right]. \tag{6.5.4}$$

To complete the proof we only need to prove the following lemma.

Lemma 6.5. *Using the notations above we have*

$$\left(||u\varphi'(u,\eta^*)||^2_C\right)^{-1} = \mu^{-1}. \tag{6.5.5}$$

Again, we do not prove (6.5.5) using direct computation. Instead we show that $(||u\varphi'(u,\eta^*)||^2_C)^{-1}$ can be obtained as the asymptotic covariance of an efficient ECF method with the full continuum moment conditions for the problem of estimating the scale parameter λ^* of $\lambda^*\Delta L$, with $\lambda^* = 1$, given an i.i.d. realization of ΔL. The problem of efficiency, like the analogues problem in Theorem 4.2, is then reduced to the i.i.d. case.

To carry out the suggested argument solve the following identification problem: estimate the scale parameter λ given a sequence of i.i.d. realizations of the distribution $\lambda\Delta L$, where the true value of λ is $\lambda = \lambda^* = 1$. The characteristic function of ΔZ is denoted by φ, then the c.f. $\varphi_{\lambda\Delta L}(u,\lambda)$ of $\lambda\Delta L$ is given by $\varphi(u\lambda)$.

Recall that for an i.i.d. sample, which was generated by a random variable with a general characteristic function $\chi(u,\alpha^*)$, with α^* being an unknown parameter, the ECF method using the full continuum of u-s gives an asymptotically efficient estimate of α^* with asymptotic covariance

$$(||\chi_\alpha(u,\alpha^*)||^2_C)^{-1}.$$

Write the derivative of the c.f. of $\lambda\Delta L$ w.r.t. λ

$$\frac{\partial}{\partial\lambda}\mathbb{E}\left[e^{iu\lambda\Delta L}\right]=\mathbb{E}\left[e^{iu\lambda\Delta L}iu\Delta L\right],$$

choosing $\lambda=\lambda^*=1$ gives

$$\frac{\partial}{\partial\lambda}\varphi_{\lambda\Delta L}(u,\lambda)\bigg|_{\lambda=\lambda^*}=u\varphi'(u,\lambda^*).$$

Choosing $\chi=\varphi_{\lambda\Delta L}$ and $\alpha=\lambda$ we have $\chi_\alpha(u,\lambda^*)=u\varphi'(u,\eta^*)$. Hence for this identification problem the asymptotic covariance of the i.i.d. ECF method with full continuum u-s is

$$\left(||u\varphi'(u,\eta^*)||_C^2\right)^{-1}.$$

Since the ECF method with continuum u-s is exactly as efficient as the ML method we find that $(||u\varphi'(u,\eta^*)||_C^2)^{-1}$ equals to the inverse Fisher of the ML method, hence (6.5.5) follows. □

6.6 Discussion

The optimal choice of K is $C\otimes M^*$, but M^* is given by an expected value using the true value of parameters θ, so the optimal weighting matrix, like the optimal weighting matrix in the three-stage method for linear systems, is not computable. We follow the line of arguments presented in Section 4.7, we propose to approximate M^* is two steps. First, define the approximation $\hat{R}_P^*(\theta)$ by

$$\hat{M}^*(\theta)=\frac{1}{N}\sum_{n=1}^{N}\frac{\sigma_{\theta n}(\theta)\sigma_{\theta n}^T(\theta)}{\sigma_n^2(\theta)}$$

It would be convenient to use $\hat{M}^*(\theta^*)$, but since θ^* is unknown we approximate it by $\hat{\theta}_N^{(\mathrm{pre})}$, where $\hat{\theta}_N^{(\mathrm{pre})}$ is a preliminary estimate obtained by using the ECF method for GARCH systems

with the choice $K = I$. Thus, we apply the ECF method with the weighting matrix

$$K = C \otimes \hat{M}^* \left(\hat{\theta}_N^{(\mathrm{pre})} \right)$$

to get the approximation of $\hat{\theta}_N$. It is relatively easy to see that Theorem 6.4 and Theorem 6.5 are valid for this approximation of $\hat{\theta}_N$, too.

As we have already mentioned at the beginning of this chapter that although the three-stage method can be applied for GARCH processes to identify both the system and the noise characteristics, the results of Chapter 4 cannot be reproduced. In what follows we address this issue. Being aware of the steps of the three-stage method for linear Lévy systems a three-stage identification method for GARCH systems can be proposed in a natural manner. Suppose now that both θ^* and η^* are unknown. The steps of the proposed three-method can be summarized as follows:

1. Firstly estimate θ^* by applying the quasi-maximum likelihood method to obtain $\hat{\theta}_N$.

2. Secondly invert the GARCH system with $\theta = \hat{\theta}_N$ to generate the estimated noise process, then estimate η^* by pretending that these residuals are i.i.d., and apply the ECF method for i.i.d. data to obtain $\hat{\eta}_N$.

3. Finally re-estimate θ^* by applying the ECF method for system identification, pretending that $\hat{\eta}_N = \eta^*$, to obtain an estimate $\hat{\hat{\theta}}_N$ for the dynamics.

The problem with this three-stage method is that the tools presented for linear Lévy systems cannot be adapted for its analysis. For, in analogy with the three-stage method for linear Lévy systems the third step of the algorithm should give a consistent estimate of θ^* even if the noise characteristics η is misspecified. The η-dependent modified primary scores of the third step would be given by

$$h_{k,n}(\theta, \eta) = \left(e^{iu_k \varepsilon_n(\theta)} - \varphi(u, \eta) \right) \frac{\sigma_{\theta n}(\theta)}{\sigma_n(\theta)}.$$

Following the notations and the line of arguments of Section 6.3 in defining the η-dependent scores, 'half-gradient' equations and corresponding variables the asymptotic value of function $V'_N(\theta, \eta)$ would be given by

$$\bar{W}'(\theta, \eta) = \lim_{N \to \infty} \mathbb{E}[V'_N(\theta, \eta)] = g^*_\theta(\theta, \eta) K^{-1} g(\theta, \eta).$$

Observe that if we are given a misspecified η, then by solving $\bar{W}'(\theta, \eta) = 0$ for θ we typically have a solution $\theta^*(\eta)$ such that

$$\theta^*(\eta) \neq \theta^*.$$

The reason behind is that for the instrumental variable we typically have that

$$\mathbb{E}\left[\frac{\sigma_{\theta n}(\theta^*)}{\sigma_n(\theta^*)}\right] \neq 0.$$

The study of this interesting problem will be a subject of our further research.

Chapter 7

Recursive ECF estimation

In the literature the empirical characteristic function method is presented as an off-line identification method. While the results of the off-line methods are attractive, the proposed algorithms are ill-conditioned in many cases so that they requires special attention. As an alternative to the off-line method in this section we propose and analyze on-line empirical characteristic function methods. Such recursive methods enables us to carry out real-time statistical analysis as new data points are processed instantly. In constructing these algorithms we follow the general framework proposed by Djereveckii and Fradkov , see [22], and Ljung, see [45]. On-line methods are also used to complement a computationally expensive off-line identification method. Namely, it would be uneconomical to re-estimate θ^* using the off-line method when a new data point is received. Instead, we can argue that only a refinement of the estimate $\hat{\theta}_N$ should be computed using the newly received data point. This scenario not only shows a motivation behind the study of recursive algorithms but also suggests that it is reasonable to suppose that an initial guess of the parameter is close to θ^*.

7.1 General recursive estimation scheme

We present a recursive estimation scheme within a general setup first formulated and solved for dynamical systems by Djereveckii and Fradkov in [22] and Ljung in [45], hence the abbreviation DFL-scheme. Several recursive identification methods can be handled by this scheme, a nice summary of this can be found in [47]. The basic building block of the scheme is the following parameter-dependent state-space equation:

$$\overline{\xi}_{n+1}(x) = A(x)\overline{\xi}_n(x) + B(x)e_n,\ \overline{\xi}_0(x) = 0, \tag{7.1.1}$$

where the parameter x is an element of an open domain $D \subset \mathbb{R}^p$. In the above so-called frozen parameter system $\overline{\xi} \in \mathbb{R}^r$ is a state-vector with possibly unobservable components and $e \in \mathbb{R}^m$ is an exogenous noise. x will be allowed to be time-varying taking values (x_n) to be specified later. The next two conditions ensure the joint stability and the smoothness of the matrices $A(x)$ and $B(x)$.

Condition 7.1. *The family of $r \times r$ matrices $\{A(x), x \in D \subset \mathbb{R}^d\}$ is jointly stable, in the sense that there exists a positive-definite $n \times n$ matrix P, and a λ with $0 < \lambda < 1$ such that*

$$A^T(x)PA(x) \leq \lambda P,$$

holds for all $x \in D$.

Condition 7.2. *$A(x)$ and $B(x)$ are continuously differentiable up to third order in D.*

To analyze recursive algorithms we require the driving noise process (e_n) to be L-mixing, what is more we require that it is L^+-mixing in the sense defined below, defined in terms of the approximation error

$$\gamma_q(\tau, e) = \sup_{n \geq \tau} \mathbb{E}^{1/q}\left[\left|e_n - \mathbb{E}\left[e_n | \mathscr{F}^+_{n-\tau}\right]\right|^q\right],$$

see also the definition of L-mixing processes in Section 10.2.

Condition 7.3. (e_n) *is strictly stationary and it is also* L^+*-mixing with respect to a families of* σ*-algebras* $(\mathscr{F}_n, \mathscr{F}_n^+)$ *in the sense that for all integer* $\tau \geq 1$ *and* $q \geq 1$ *with some* $c > 0$ *we have*

$$\gamma_q(\tau, e) = O(\tau^{-1-c}).$$

A variety of methods that analyze recursive methods is based on the idea of approximating (x_n) using a trajectory of an ordinary differential equation (ODE). In the process of developing the ODE method an often used assumption is that e_n^2 has some finite positive exponential moments. This leads to the definition of class M^*.

Definition 7.1. *Let* $(u_n), n \geq 0$ *be a real-valued stochastic process. We say that* (u_n) *is in class* M^* *if for some* $\varepsilon > 0$

$$M^{\varepsilon}(u) := \sup_n \frac{1}{\varepsilon} \log \mathbb{E}\left[e^{\varepsilon u_n}\right] < \infty.$$

Condition 7.4. (e_n^2) *is in* M^*.

Let $Q : \mathbb{R}^r \times \mathbb{R}^d \to \mathbb{R}^d$ denote a function such that it is bounded by some polynomial of $\overline{\xi}$ and the same holds for the derivatives of Q up to order three. In many standard identification method Q is a quadratic-form in $\overline{\xi}$, but for the present application this will not hold. Define

$$F(x) = \lim_{n \to \infty} \mathbb{E}\left[Q(\overline{\xi}_n(x), x)\right].$$

Now we are ready to formulate the abstract estimation problem related to the DFL-scheme: solve for x the non-linear algebraic equation

$$F(x) = 0. \tag{7.1.2}$$

Without loss of generality we may assume that $x = x^* = 0$ is a solution. We may assume that $D_0 \subset D$ is a compact domain such that $x^* \in D_0$. Suppose that we are given an initial estimate of x^*, say x_0. Then the tentative recursion corresponding to the DFL scheme is given by

$$\xi_{n+1} = A(x_n)\xi_n + B(x_n)e_n, \quad \xi_0 = 0. \tag{7.1.3}$$

$$x_{n+1} = x_n + \frac{1}{n+1}Q(\xi_{n+1}, x_n) \quad x_0 \in D_0, \tag{7.1.4}$$

where (x_n) denotes the sequence of generated estimates. Typically the initial estimate x_0 is close to x^*. A controversial issue is the problem of keeping (x_n) in the domain D_0. In order to guarantee this a resetting mechanism is introduced. To make this modified recursion formal we denote the value of x computed at time $n+1$ using (7.1.4) by x_{n+1-} and define

$$x_{n+1} = \begin{cases} x_{n+1-} & \text{if } x_{n+1-} \in D_0 \\ x_0 & \text{if } x_{n+1-} \in D_0^c. \end{cases}$$

That is, if x_{n+1-} leaves the domain D_0 then a resetting is applied. This event is denoted by $B_{n+1} = \{\omega | x_{n+1-} \in D_0^c\}$. Hence $(7.1.4)$ is replaced by

$$x_{n+1} = x_n + (1 - \mathbf{1}_{B_{n+1}})\frac{1}{n+1}Q(\xi_{n+1}, x_n) + \mathbf{1}_{B_{n+1}}(x_0 - x_n), \tag{7.1.5}$$

where $\mathbf{1}_B$ is the indicator function of the event $B \subset \Omega$.

Now we define the differential equation, the solution trajectories of which reflect the pattern of behaviour of the sequence (x_n). This so-called associated ordinary differential equation (ODE) is defined by

$$\dot{y}_t = \frac{1}{t}F(y_t), \quad y_s = \zeta, \tag{7.1.6}$$

for $t \leq s \leq 1$. Alternatively, the associated ODE can be also defined as

$$\dot{y}_t = F(y_t),$$

we allow this ambiguity in the definition of the ODE. Note that since F is well defined in D and has continuous derivatives up to third order, the latter differential equation has a unique solution $y(t, s, \zeta)$ in some interval for t. A variety of convergence results is based on the stability of the above ODE (7.1.6), see [47], [17], [64] and [6]. For our application the stability of the associated ODE is specified by the next condition, which can be found in [32].

Condition 7.5. *Let $D_0 \subset D$ be the compact truncation domain such that $x^* \in \mathrm{int} D_0$. Assume that there exists a compact convex set D_0' such that $D_0 \subset D_0' \subset D$ and for all $t \geq s \geq 1$ we have*

$$y(t, s, \zeta) \in D_0' \quad \textit{for} \quad \zeta \in D_0 \quad \textit{and} \quad y(t, s, \zeta) \in D \quad \textit{for} \quad \zeta \in D_0'.$$

In addition $\lim_{t \to \infty} y(t, s, \zeta) = x^*$ for $\zeta \in D$ and

$$\left\| \frac{\partial}{\partial \zeta} y(t, s, \zeta) \right\| \leq C(s/t)^{\alpha}$$

with some $C > 1, \alpha > 0$ for all $\zeta \in D_0'$ and $t \geq s \geq 1$. We have an initial estimate $\zeta = y_1 = x_0$ such that for all $t \geq s \geq 1$ we have $y(t, s, \zeta) \in \mathrm{int} D_0$. Finally, for the star-like closure

$$D_0^* = \{ y \,|\, y = x^* + \lambda(x - x^*), 0 \leq \lambda \leq 1, x \in D_0 \}$$

of the set D_0 we have $D_0^* \subset D$.

The asymptotic covariance matrix of the estimates will be closely related to that of the

averaged correction terms. Hence define

$$H(n, x, \omega) := Q(\overline{\xi}_n^{(\mathrm{s})}(x), x)$$

and the matrix P^* in terms of H as

$$P^* = \sum_{m=-\infty}^{\infty} \mathbb{E}\left[H(m, x^*, \omega) H^*(0, x^*, \omega)\right]. \tag{7.1.7}$$

The following result from [32] states that the above recursion indeed defines a sequence of x_n-s that converges to the solution of the equation $F(x) = 0$ and the rate of convergence of the moments of the error is also given. What is more the result also gives it asymptotic covariance matrix of the estimate.

Theorem 7.1. *Assume that Conditions 7.1-7.4 are satisfied and further assume that the differential equation (7.1.6) satisfies Condition 7.5 with $\alpha > 1/2$, then we have $x_N = O_M(N^{-1/2})$. Moreover, the asymptotic covariance matrix of the error process $x_N - x^*$, defined by*

$$\Sigma_{xx} = \lim_{N \to \infty} N \mathbb{E}\left[(x_N - x^*)(x_N - x^*)^*\right],$$

exists and it satisfies the Lyapunov-equation

$$(A^* + I/2)\Sigma_{xx}^* + \Sigma_{xx}(A^* + I/2)^* + P^* = 0,$$

where $A^* = F_x(x^*)$.

An exciting special case is when the variable x can be split as $x = (x_1, x_2)$ so that the recursive estimation method is partially stochastic Newton w.r.t x_1, meaning that the Jacobian

matrix of the r.h.s. of the corresponding associated ODE at $x = x^*$ is of the form

$$\begin{pmatrix} -I & 0 \\ J_{2,1} & J_{2,2} \end{pmatrix}.$$

Then using simple linear algebra and Theorem 7.1 we conclude the following corollary.

Corollary 7.1. *Assume that the conditions of Theorem 7.1 hold. Assume further that we can split x as $x = (x^1, x^2)$ so that the recursive estimation method is a partially stochastic Newton method with respect to x^1. Then the asymptotic covariance matrix of the recursive estimate x_N^1 equals to $P_{1,1}^*$, which is the corresponding block of P^* defined in (7.1.7).*

7.2 Recursive ECF for i.i.d. sample

The DFL-scheme provides a solution for the problem of estimating the parameters of a distribution or a regression function using i.i.d. samples by simply choosing $A(x) = 0$ and $B(x) = I$ in (7.1.3). Although this is the subject of the classic paper Robbins-Monroe-scheme, see [57], to have a unified treatment we shall discuss this problem using the DFL-scheme. A possible motivation of this is the problem of identifying the noise characteristics of a Lévy process using i.i.d. samples $y_1, y_2, \ldots$ generated by the increments of the process.

We suppose that the characteristic function of y_i is known up to an unknown parameter η^*. Let the c.f. of y_i denoted by $\varphi(u, \eta^*)$. Fix a set of real u_i-s $1 \leq i \leq M$. In this case following the idea of the off-line ECF method our aim is to solve the non-linear equation $F(x) = 0$ in (7.1.2) with $x = \eta$ and

$$F(\eta) = \mathbb{E}\left[-\varphi_\eta^*(\eta) K^{-1} h_N(\eta)\right] = 0,$$

where

$$\varphi_\eta(\eta) = (\varphi_\eta(u_1, \eta), \ldots, \varphi_\eta(u_M, \eta))^T$$

and

$$h_N(\eta) = \left(e^{iu_1 y_N} - \varphi(u_1, \eta), \ldots, e^{iu_M y_N} - \varphi(u_M, \eta)\right)^T. \tag{7.2.1}$$

A stochastic Newton method corresponding to this equation would read as

$$\hat{\eta}_{N-} = \hat{\eta}_{N-1} - \frac{1}{N}(R_E^*)^{-1}\left(-\hat{\varphi}_{\eta,N}^* K^{-1}\hat{h}_N\right),$$

with $R_E^* = \varphi_\eta^*(\eta^*)K^{-1}\varphi_\eta(\eta^*)$. Since R^* is unknown we estimate it using the most current estimate of η^*. Hence we extend the parameter vector η to (η, R) and re-define the equation $F(\eta) = 0$ as

$$F(\eta, R) = \mathbb{E}\begin{bmatrix} -\varphi_\eta^*(\eta)K^{-1}h_N(\eta) \\ \varphi_\eta^*(\eta)K^{-1}\varphi_\eta(\eta) - R \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Let the variables $\hat{\varphi}_{\eta,N}$ and $\hat{h}_N$ be obtained by using the most current estimate of η^*, that is

$$\begin{aligned} \hat{\varphi}_{\eta,N} &= \varphi_{\eta,N}(\hat{\eta}_{N-1}) = (\varphi_\eta(u_1, \hat{\eta}_{N-1}), \ldots, \varphi_\eta(u_M, \hat{\eta}_{N-1}))^T \\ \hat{h}_N &= h_N(\hat{\eta}_{N-1}) = \left(e^{iu_1 y_n} - \varphi(u_1, \hat{\eta}_{N-1}), \ldots, e^{iu_M y_n} - \varphi(u_M, \hat{\eta}_{N-1})\right)^T. \end{aligned} \tag{7.2.2}$$

Let $\hat{\eta}_0$ and $\hat{R}_0$ be initial guesses and let $\hat{\eta}_N$ and $\hat{R}_N$ be computed using a partially stochastic Newton method as follows:

Algorithm 7.1 (Recursive i.i.d. ECF method).

$$\begin{aligned} \hat{\eta}_{N-} &= \hat{\eta}_{N-1} - \frac{1}{N}\hat{R}_{N-1}^{-1}\left(-\hat{\varphi}_{\eta,N}^* K^{-1}\hat{h}_N\right) \\ \hat{R}_{N-} &= \hat{R}_{N-1} + \frac{1}{N}\left(\hat{\varphi}_{\eta,N}^* K^{-1}\hat{\varphi}_{\eta,N} - \hat{R}_{N-1}\right). \end{aligned} \tag{7.2.3}$$

We note in passing that instead of applying a recursion for computing $\hat{R}_{N-}$ we could define

it by simple substitution:

$$\hat{R}_{N-} = \varphi_\eta^*(\hat{\eta}_N) K^{-1} \varphi_\eta(\hat{\eta}_N).$$

The reason behind our choice is that it fits into the general framework of DFL-scheme. Denote the expected value of $h_N(\eta)$ by $g(\eta)$:

$$g(\eta) = \mathbb{E}\left[h_N(\eta)\right] = \left(\varphi(u_1, \eta^*) - \varphi(u_1, \eta), \varphi(u_M, \eta^*) - \varphi(u_M, \eta)\right)^T.$$

The corresponding associated ODE with the extended variable is given by

$$\begin{aligned} \dot{\eta}_t &= -R_t^{-1}\left(-\varphi_\eta^*(\eta_t) K^{-1} g(\eta_t)\right) \\ \dot{R}_t &= \varphi_\eta^*(\eta_t) K^{-1} \varphi_\eta(\eta_t) - R_t \end{aligned} \tag{7.2.4}$$

for $t > 0$. We have seen that in case of an off-line identification method the optimal choice of K is $K = C$. Recall the notation C with entries

$$C_{k,l} = \varphi(u_k - u_l, \eta^*) - \varphi(u_k, \eta^*)\varphi(-u_l, \eta^*).$$

Obviously for i.i.d. samples most of the conditions of Theorem 7.1 are satisfied: in (7.1.3) we have $A = 0$ and $B = I$, furthermore we have that (y_n) is strictly stationary and L-mixing. Therefore Theorem 7.1 and Corollary 7.1 imply the following result:

Theorem 7.2. *Suppose that we are given an i.i.d. data generated by a random variable X such that X^2 is in M^*, suppose further that the differential equation (7.2.4) satisfies Condition 7.5. Then the Algorithm 7.1 equipped with resetting is convergent and we have*

$$\hat{\eta}_N - \eta^* = O_M(N^{-1/2}).$$

Furthermore, if $K = C$ then the asymptotic covariance matrix $\Sigma_{\eta\eta}^{(\mathrm{rec})}$ of $\hat{\eta}_N$ is given by

$$\Sigma_{\eta\eta}^{(\mathrm{rec})} = (\varphi_\eta^*(\eta^*) C^{-1} \varphi_\eta(\eta^*))^{-1}.$$

Hence, the estimate $\hat{\eta}_N$ is essentially asymptotically efficient.

Note that the local stability of the ODE with $\alpha > 1/2$ follows. For, the Jacobian of the ordinary differential equation at $\eta = \eta^*$ and $R = R^*$ is

$$\begin{pmatrix} -I & 0 \\ J_{2,1} & -I \end{pmatrix},$$

thus each eigenvalue of the Jacobian is -1, hence the top Lyapunov exponent can be chosen to be equal to $-1 + c$ with any $c > 0$, which implies that the ODE is locally stable with $\alpha > 1/2$. The structure of the above Jacobian also shows that the proposed method is a partially stochastic Newton method w.r.t. η.

7.3 Recursive ECF for linear Lévy systems with known noise characteristics

This section is devoted to the presentation of a recursive ECF method for linear Lévy systems with known noise characteristics. Recall the definition of linear Lévy systems from Chapter 4:

$$\Delta y = A(\theta^*, q^{-1}) \Delta L, \tag{7.3.1}$$

defined for the time range $-\infty < n < +\infty$, where ΔL_n is the increment of a Lévy process (L_t) with $-\infty < t < +\infty$, and $L_0 = 0$, over an interval $[(n-1)h, nh)$, with $h > 0$ being a fixed sampling interval, and $-\infty < n < +\infty$. Let us assume that a state space representation

in innovation form for this model is given by

$$\Delta X_{n+1} = H(\theta^*)\Delta X_n + K(\theta^*)\Delta L_n \tag{7.3.2}$$

$$\Delta Y_n = T(\theta^*)\Delta X_n + \Delta L_n. \tag{7.3.3}$$

Then $H(\theta) - K(\theta)T(\theta)$ is the state transition matrix of the inverse process. We will need the following stability conditions:

Condition 7.6. *It is assumed that the system matrix $H(\theta^*)$ is stable and $H(\theta) - K(\theta)T(\theta)$ are jointly stable for $\theta \in D_\theta$.*

The next condition guarantees the smooth dependence of the system matrices on θ :

Condition 7.7. *Assume that $H(\theta), K(\theta)$ and $T(\theta)$ are three-times continuously differentiable w.r.t. θ for $\theta \in D_\theta$.*

The novel problem of identifying the system parameters θ^*, using the ECF method, under the assumption that the noise characteristics η^* is known was presented in Chapter 4. We may wish to solve the same problem, but now with a recursive method.

Suppose that we are given the noise characteristics η^*. Fix a set of real u_i-s $1 \leq i \leq M$. Following the third step of the off-line estimation method we seek the solution of the non-linear equation $F(x) = 0$ with $x = \theta$ and

$$F(\theta) = \mathbb{E}\left[G^* K^{-1} h_N^{(\mathrm{s})}(\theta; \eta^*)\right],$$

where

$$h_N^{(\mathrm{s})}(\theta; \eta^*) = \left(\left(e^{iu_1 \varepsilon_N^{(\mathrm{s})}(\theta)} - \varphi(u_1, \eta^*)\right) \varepsilon_{\theta N}^{(\mathrm{s})\mathrm{T}}(\theta), \ldots, \left(e^{iu_M \varepsilon_N^{(\mathrm{s})}(\theta)} - \varphi(u_M, \eta^*)\right) \varepsilon_{\theta N}^{(\mathrm{s})\mathrm{T}}(\theta)\right)^T.$$

Clearly $\theta = \theta^*$ is the solution of this equation. Similarly to the i.i.d. case we would like to apply

a stochastic Newton method, which requires the introduction of the Jacobian R^*, the true value of which is known to be $R^* = G^*(\theta^*, \eta^*)K^{-1}G(\theta^*, \eta^*)$. In order to compute R^* we need the value of $G(\theta^*, \eta^*) = \psi \otimes R_P^*$, where $\psi = (iu_1\varphi(u_1, \eta^*), \ldots, iu_M\varphi(u_M, \eta^*))^T$, for $G(\theta^*, \eta^*)$ see the proof of Theorem 4.6. Since these are not computable without the knowledge of θ^*, and $G(\theta^*, \eta^*)$ can be computed only empirically, we approximate $G(\theta^*, \eta^*)$ using the most current estimate of θ^*. To this end we extend the parameter vector to (θ, G, R) and re-define equation $F(\theta) = 0$ by

$$F(\theta, G, R) = \mathbb{E}\begin{bmatrix} G^*K^{-1}h_N^{(\mathrm{s})}(\theta; \eta) \\ G - h_{\theta N}^{(\mathrm{s})}(\theta; \eta) \\ G^*K^{-1}G - R \end{bmatrix}, \tag{7.3.4}$$

where $h_{\theta N}^{(\mathrm{s})}(\theta)$ shows up in the derivative of $h_N^{(\mathrm{s})}(\theta)$ w.r.t. θ, and defined by

$$h_{\theta N}^{(\mathrm{s})}(\theta) := \left(\left(iu_1 e^{iu_1\varepsilon_N^{(\mathrm{s})}(\theta)}\varepsilon_{\theta,N}^{(\mathrm{s})}(\theta)\varepsilon_{\theta,N}^{(\mathrm{s})*}(\theta)\right)^T, \ldots, \left(iu_M e^{iu_M\varepsilon_N^{(\mathrm{s})}(\theta)}\varepsilon_{\theta,N}^{(\mathrm{s})}(\theta)\varepsilon_{\theta,N}^{(\mathrm{s})*}(\theta)\right)^T\right)^T, \tag{7.3.5}$$

which is obtained by dropping the term containing $-u_j^2 e^{iu_j\varepsilon_N^{(\mathrm{s})}(\theta)}\varepsilon_{\theta\theta,N}^{(\mathrm{s})}(\theta)$-s, which has zero expectation at $\theta = \theta^*$, from the derivative of $h_N^{(\mathrm{s})}(\theta)$.

Suppose we are given a set of initial values of the parameter: $\hat{\theta}_0 \in D_{0\theta}$ is the initial value of θ and $\hat{g}_{\theta,0}$ is the initial value of G. The set of initial values of the parameter is given by $\{\hat{\varepsilon}_0, \hat{\varepsilon}_{\theta,0}\}$. Here $\hat{\theta}_0$ might have been previously obtained by an off-line identification method. Likewise $\hat{\varepsilon}_0, \hat{\varepsilon}_{\theta,0}, \hat{g}_{\theta,0}$ might be obtained previously or we can set them to be equal to 0. The recursive algorithm at step N updates the estimates as follows: given the previous estimates

first compute the auxiliary variables using the most current estimate of θ^* according to

$$\begin{aligned}
\hat{\varepsilon}_N &= A^{-1}\left(\hat{\theta}_{N-1}\right)\Delta y_N \\
\hat{\varepsilon}_{\theta,N} &= A_\theta^{-1}\left(\hat{\theta}_{N-1}\right)\Delta y_N \\
\hat{h}_N(\eta^*) &= \left(\left(e^{iu_1\hat{\varepsilon}_N}-\varphi(u_1,\eta^*)\right)\hat{\varepsilon}_{\theta,N}^T,\ldots,\left(e^{iu_M\hat{\varepsilon}_N}-\varphi(u_M,\eta^*)\right)\hat{\varepsilon}_{\theta,N}^T\right)^T \\
\hat{h}_{\theta,N} &= \left(\left(iu_1e^{iu_1\hat{\varepsilon}_N}\hat{\varepsilon}_{\theta,N}\hat{\varepsilon}_{\theta,N}^*\right)^T,\ldots,\left(iu_Me^{iu_M\hat{\varepsilon}_N}\hat{\varepsilon}_{\theta,N}\hat{\varepsilon}_{\theta,N}^*\right)^T\right)^T.
\end{aligned} \tag{7.3.6}$$

Following the special form of the off-line estimation method presented in Chapter 4 we define a stochastic Newton method via the following algorithm:

Algorithm 7.2 (Re-estimating recursive ECF method).

$$\begin{aligned}
\hat{\theta}_{N-} &= \hat{\theta}_{N-1}-\frac{1}{N}\hat{R}_{S,N-1}^{-1}\left(\hat{g}_{\theta,N-1}^*K^{-1}\hat{h}_N(\eta^*)\right) \\
\hat{g}_{\theta,N-} &= \hat{g}_{\theta,N-1}+\frac{1}{N}\left(\hat{h}_{\theta,N}-\hat{g}_{\theta,N-1}\right),
\end{aligned} \tag{7.3.7}$$

where

$$\hat{R}_{S,N-1}=\hat{g}_{\theta,N-1}^*K^{-1}\hat{g}_{\theta,N-1}.$$

The Jacobian of the Note that the third component $G^*K^{-1}G-R$ in (7.3.4) is non-random, hence $\hat{R}_{S,N-1}$ is computed by simple substitution. These tentative values need to be modified with a suitable resetting mechanism as described in connection with the general DFL-scheme. In order to define the associated ODE first take the expectations of the frozen parameter correction terms $h_{\theta N}^{(\mathrm{s})}(\theta;\eta)$ and $h_N^{(\mathrm{s})}(\theta;\eta)$, showing up on the r.h.s. of 7.3.4:

$$\begin{aligned}
g(\theta;\eta^*) &= \mathbb{E}\left[\left(\left(e^{iu_1\varepsilon_n^{(\mathrm{s})}(\theta)}-\varphi(u_1,\eta^*)\right)\varepsilon_{\theta,n}^{(\mathrm{s})\mathrm{T}}(\theta),\ldots,\left(e^{iu_M\varepsilon_n^{(\mathrm{s})}(\theta)}-\varphi(u_M,\eta^*)\right)\varepsilon_{\theta,n}^{(\mathrm{s})\mathrm{T}}(\theta)\right)^T\right], \\
g_\theta(\theta) &= \mathbb{E}\left[\left(\left(iu_1e^{iu_1\varepsilon_n^{(\mathrm{s})}(\theta)}\varepsilon_{\theta,n}^{(\mathrm{s})}(\theta)\varepsilon_{\theta,n}^{(\mathrm{s})*}(\theta)\right)^T,\ldots,\left(iu_Me^{iu_M\varepsilon_n^{(\mathrm{s})}(\theta)}\varepsilon_{\theta,n}^{(\mathrm{s})}(\theta)\varepsilon_{\theta,n}^{(\mathrm{s})*}(\theta)\right)^T\right)^T\right].
\end{aligned}$$

Then the corresponding associated ODE reads as

$$\begin{aligned} \dot{\theta}_t &= -R_{S,t}^{-1}\left(g_{\theta,t}^* K^{-1} g(\theta_t;\eta^*)\right) \\ \dot{g}_{\theta,t} &= g_\theta(\theta_t) - g_{\theta,t}, \end{aligned} \tag{7.3.8}$$

where

$$R_{S,t} = g_{\theta,t}^* K^{-1} g_{\theta,t}.$$

It is easy to check that the Jacobian matrix of (7.3.8) at $(\theta^*, G(\theta^*,\eta^*))$ is a lower triangular matrix with $-I$ blocks in the diagonal, thus all eigenvalues are equal to -1. It follows that the solution of the ODE is locally stable with $\alpha = 1/2$, see Condition 7.5 for the definition of α. Moreover, the structure of the Jacobian also shoves that the on-line method is a partially stochastic Newton method w.r.t. θ.

Recall the notation of Chapter 4

$$R_P^* = \mathbb{E}\left[\varepsilon_{\theta n}^{(\mathrm{s})}(\theta^*)\varepsilon_{\theta n}^{(\mathrm{s})\mathrm{T}}(\theta^*)\right].$$

Theorem 7.1 and Corollary 7.1 together with Theorem 4.7, giving the asymptotic covariance matrix of the re-estimated system parameter, imply the following result:

Theorem 7.3. *Suppose that for the Lévy system Condition 7.6 and Condition 7.7 are satisfied. Suppose further that for the driving Lévy process we have that* $((\Delta L_n)^2)$ *is in* M^* *and that the differential equation (7.3.8) satisfies Condition 7.5.Then for the estimate* $\hat{\theta}_N$ *obtained by the above recursive method in Algorithm 7.2 modified by a suitable resetting mechanism we have*

$$\hat{\theta}_N - \theta^* = O_M(N^{-1/2}).$$

Moreover, if $K = C \otimes R_P^*$, then the asymptotic covariance matrix $\Sigma_{\theta\theta}^{(\mathrm{rec})}$ of $\hat{\theta}_N - \theta^*$ is given by

$$\Sigma_{\theta\theta}^{(\mathrm{rec})} = \left(\psi^* C^{-1} \psi\right)^{-1} (R_P^*)^{-1},$$

with $\psi = (iu_1\varphi(u_1, \eta^*), \ldots, iu_M\varphi(u_M, \eta^*))^T$. Hence, the proposed on-line method is essentially asymptotically efficient.

7.4 Recursive ECF for linear Lévy systems

Now we are ready to present and analyze a recursive identification method that estimates both the system and the noise characteristics by converting the three-stage method presented in Chapter 4 to a recursive method. Suppose that the dynamics of (y_n) follows (7.3.1). Fix a set of real u_i-s $1 \leq i \leq M$. In this case we define the non-linear equation $F(x) = 0$ in (7.1.2) by merging the asymptotic equations corresponding to the PE method, the ECF method for the noise characteristics η^* and the ECF method for the re-estimation of θ^*. Accordingly, x is defined as $x = (\theta_P, R_P, \eta, R_E, \theta_S, G, R_S)$. Observe that θ is duplicated, in the sense that $\hat{\theta}_P$ and $\hat{\theta}_S$ both are expected to converge to θ^*. This separation of the recursive PE estimate $\hat{\theta}_P$ and the ECF estimate $\hat{\theta}_S$ guarantees that the Jacobian matrix of the corresponding associated

ODE will be lower triangular. Now we are ready to define F by

$$F(\theta_P, R_P, \eta, R_E, \theta_S, G, R_S) = \mathbb{E}\left[\begin{array}{c} \varepsilon^{(\mathrm{s})}_{\theta_P N}(\theta_P)\varepsilon^{(\mathrm{s})}_{N}(\theta_P) \\ \varepsilon^{(\mathrm{s})}_{\theta_P N}(\theta_P)\varepsilon^{(\mathrm{s})\mathrm{T}}_{\theta_P N}(\theta_P) - R_P \\ \hline -\varphi^*_\eta(\eta)K_E^{-1}h^{(\mathrm{s})}_{E,N}(\theta_P,\eta) \\ \varphi^*_\eta(\eta)K_E^{-1}\varphi_\eta(\eta) - R_E \\ \hline G^*K_S^{-1}h^{(\mathrm{s})}_{S,N}(\theta_S,\eta) \\ G - h^{(\mathrm{s})}_{S,\theta N}(\theta_S,\eta) \\ G^*K_S^{-1}G - R_S \end{array}\right],$$

where the auxiliary variables are defined in analogy with the ones in the previous two sections as

$$\begin{aligned}
\varphi_\eta(\eta) &= (\varphi_\eta(u_1,\eta),\ldots,\varphi_\eta(u_M,\eta))^T \\
h^{(\mathrm{s})}_{E,N}(\theta,\eta) &= \left(e^{iu_1\varepsilon^{(\mathrm{s})}_N(\theta)} - \varphi(u_1,\eta),\ldots,e^{iu_M\varepsilon^{(\mathrm{s})}_N(\theta)} - \varphi(u_M,\eta)\right)^T \\
h^{(\mathrm{s})}_{S,N}(\theta,\eta) &= \left(\left(e^{iu_1\varepsilon^{(\mathrm{s})}_N(\theta)} - \varphi(u_1,\eta)\right)\varepsilon^{(\mathrm{s})\mathrm{T}}_{\theta N}(\theta),\ldots,\left(e^{iu_M\varepsilon^{(\mathrm{s})}_N(\theta)} - \varphi(u_M,\eta)\right)\varepsilon^{(\mathrm{s})\mathrm{T}}_{\theta N}(\theta)\right)^T \\
h^{(\mathrm{s})}_{S,\theta N}(\theta) &= \left(\left(iu_1 e^{iu_1\varepsilon^{(\mathrm{s})}_n(\theta)}\varepsilon^{(\mathrm{s})}_{\theta,n}(\theta)\varepsilon^{(\mathrm{s})*}_{\theta,n}(\theta)\right)^T,\ldots,\left(iu_M e^{iu_M\varepsilon^{(\mathrm{s})}_n(\theta)}\varepsilon^{(\mathrm{s})}_{\theta,n}(\theta)\varepsilon^{(\mathrm{s})*}_{\theta,n}(\theta)\right)^T\right)^T.
\end{aligned}$$

Note that to different ECF scores h_E and h_S are being used, one estimates η^* and another re-estimates θ^*. Let us suppose that we are given the initial values of the parameters: $\hat{\theta}_{P,0}, \hat{R}_{P,0}$ are the initial values of the recursive PE method, see [47], $\hat{\eta}_0, \hat{R}_{E,0}$ are the initial values of the recursive ECF method for the noise characteristics, see Section 7.2 and $\hat{\theta}_{S,0}, \hat{g}_{\theta,0}$ are the initial values of the recursive ECF re-estimation method, see Section 7.3. We assume that each of these initial values are the element of the corresponding truncation domain and $\hat{\theta}_{P,0} = \hat{\theta}_{S,0}$ is a reasonable choice. We are also given a set of initial values $\hat{\varepsilon}_{P,0}, \hat{\varepsilon}_{P,\theta,0}, \hat{\varepsilon}_{S,0}, \hat{\varepsilon}_{S,\theta,0}$. Clearly, these values might have been obtained by carrying out an off-line identification method, otherwise

we may set all of them to be equal to zero.

The recursive algorithm at step N updates the estimates as follows: given the previous estimates of the parameters first compute the estimated driving noise and its derivative using the most current values of the parameters as

$$\begin{aligned}
\hat{\varepsilon}_{P,N} &= A^{-1}\left(\hat{\theta}_{P,N-1}\right)\Delta y_N \\
\hat{\varepsilon}_{P,\theta,N} &= A_\theta^{-1}\left(\hat{\theta}_{P,\theta,N-1}\right)\Delta y_N \\
\hat{\varepsilon}_{S,N} &= A^{-1}\left(\hat{\theta}_{S,N-1}\right)\Delta y_N \\
\hat{\varepsilon}_{S,\theta,N} &= A_\theta^{-1}\left(\hat{\theta}_{S,\theta,N-1}\right)\Delta y_N.
\end{aligned} \tag{7.4.1}$$

While the first two equations correspond to the recursive PE method, the last two equations correspond to the re-estimating ECF method for linear systems. In analogy with the previous two sections we also define the auxiliary variables using the most current values of the parameters according to

$$\begin{aligned}
\hat{\varphi}_{\eta,N} &= \left(\varphi_\eta(u_1,\hat{\eta}_{N-1}),\ldots,\varphi_\eta(u_M,\hat{\eta}_{N-1})\right)^T \\
\hat{h}_{E,N} &= \left(e^{iu_1\hat{\varepsilon}_{P,N}}-\varphi(u_1,\hat{\eta}_{N-1}),\ldots,e^{iu_M\hat{\varepsilon}_{P,N}}-\varphi(u_M,\hat{\eta}_{N-1})\right)^T \\
\hat{h}_{S,N} &= \left(\left(e^{iu_1\hat{\varepsilon}_{S,N}}-\varphi(u_1,\hat{\eta}_{N-1})\right)\hat{\varepsilon}_{S,\theta,N}^T,\ldots,\left(e^{iu_M\hat{\varepsilon}_{S,N}}-\varphi(u_M,\hat{\eta}_{N-1})\right)\hat{\varepsilon}_{S,\theta,N}^T\right)^T \\
\hat{h}_{S,\theta,N} &= \left(\left(iu_1e^{iu_1\hat{\varepsilon}_{S,N}}\hat{\varepsilon}_{S,\theta,N}\hat{\varepsilon}_{S,\theta,N}^*\right)^T,\ldots,\left(iu_Me^{iu_M\hat{\varepsilon}_{S,N}}\hat{\varepsilon}_{S,\theta,N}\hat{\varepsilon}_{S,\theta,N}^*\right)^T\right)^T.
\end{aligned} \tag{7.4.2}$$

The recursive version of the three-stage method is then given as follows:

Algorithm 7.3 (Three-stage recursive ECF method). *First apply the recursive PE method defined as*

$$\begin{aligned}
\hat{\theta}_{P,N-} &= \hat{\theta}_{N-1} - \frac{1}{N}\hat{R}_{P,N-1}^{-1}\hat{\varepsilon}_{P,\theta N}\hat{\varepsilon}_{P,N}^T \\
\hat{R}_{P,N-} &= \hat{R}_{E,N-1} + \frac{1}{N}\left(\hat{\varepsilon}_{P,\theta N}\hat{\varepsilon}_{P,\theta N}^T - \hat{R}_{P,N-1}\right),
\end{aligned} \tag{7.4.3}$$

then apply the recursive ECF method for the noise characteristics defined via

$$\begin{aligned}\hat{\eta}_{N-} &= \hat{\eta}_{N-1} - \frac{1}{N}\hat{R}_{E,N-1}^{-1}\left(-\hat{\varphi}_{\eta,N}^{*}K_E^{-1}\hat{h}_{E,N}\right)\\ \hat{R}_{E,N-} &= \hat{R}_{E,N-1} + \frac{1}{N}\left(\hat{\varphi}_{\eta,N}^{*}K_E^{-1}\hat{\varphi}_{\eta,N} - \hat{R}_{P,N-1}\right),\end{aligned} \tag{7.4.4}$$

finally re-estimate θ^* using the recursive ECF method defined by

$$\begin{aligned}\hat{\theta}_{S,N-} &= \hat{\theta}_{N-1} - \frac{1}{N}\hat{R}_{S,N-1}^{-1}\left(\hat{g}_{\theta,N-1}^{*}K_S^{-1}\hat{h}_{S,N}\right)\\ \hat{g}_{\theta,N-} &= \hat{g}_{\theta,N-1} + \frac{1}{N}\left(\hat{h}_{S,\theta,N} - \hat{g}_{\theta,N-1}\right),\end{aligned} \tag{7.4.5}$$

where $\hat{R}_{S,N-1} = \hat{g}_{\theta,N-1}^{*}K^{-1}\hat{g}_{\theta,N-1}$.

These tentative values need to be modified using a suitable resetting mechanism as described in connection with the DFL-scheme. Write the expectations of the frozen parameters as

$$\begin{aligned}R_P(\theta_P) &= \mathbb{E}\left[\varepsilon_{\theta,n}^{(\mathrm{s})}(\theta_P)\varepsilon_{\theta,n}^{(\mathrm{s})\mathrm{T}}(\theta_P)\right]\\ h_E(\theta_P,\eta) &= \mathbb{E}\left[\left(e^{iu_1\varepsilon_n^{(\mathrm{s})}(\theta_P)} - \varphi(u_1,\eta),\ldots,e^{iu_M\varepsilon_n^{(\mathrm{s})}(\theta_P)} - \varphi(u_M,\eta)\right)^T\right]\\ h_S(\theta_S,\eta) &= \mathbb{E}\left[\left(\left(e^{iu_1\varepsilon_n^{(\mathrm{s})}(\theta_S)} - \varphi(u_1,\eta)\right)\varepsilon_{\theta,n}^{(\mathrm{s})\mathrm{T}}(\theta_S),\ldots,\left(e^{iu_M\varepsilon_n^{(\mathrm{s})}(\theta_S)} - \varphi(u_M,\eta)\right)\varepsilon_{\theta,n}^{(\mathrm{s})\mathrm{T}}(\theta_S)\right)^T\right]\\ g_\theta(\theta_S) &= \mathbb{E}\left[\left(\left(iu_1e^{iu_1\varepsilon_n^{(\mathrm{s})}(\theta_S)}\varepsilon_{\theta,n}^{(\mathrm{s})}(\theta_S)\varepsilon_{\theta,n}^{(\mathrm{s})*}(\theta_S)\right)^T,\ldots,\left(iu_Me^{iu_M\varepsilon_n^{(\mathrm{s})}(\theta_S)}\varepsilon_{\theta,n}^{(\mathrm{s})}(\theta_S)\varepsilon_{\theta,n}^{(\mathrm{s})*}(\theta_S)\right)^T\right)^T\right].\end{aligned}$$

Recall the notation

$$W_{P,\theta_P}(\theta_P) = \mathbb{E}\left[\varepsilon_{\theta n}^{(\mathrm{s})}(\theta_P)\varepsilon_n^{(\mathrm{s})}(\theta_P)\right].$$

In terms of the above expectations the ODE corresponding to the recursive PE method reads as

$$\begin{aligned}\dot{\theta}_{P,t} &= -R_{P,t}^{-1}W_{P,\theta_P}(\theta_{P,t})\\ \dot{R}_{P,t} &= R_P(\theta_{P,t}) - R_{P,t},\end{aligned} \tag{7.4.6}$$

while that of the recursive ECF method for noise characteristic is given by

$$\begin{aligned}\dot{\eta}_t &= -R_{E,t}^{-1}\left(-\varphi_\eta^*(\eta_t)K^{-1}h_E(\theta_{P,t},\eta_t)\right)\\ \dot{R}_{E,t} &= \varphi_\eta^*(\eta_t)K_E^{-1}\varphi_\eta(\eta_t) - R_{E,t}\end{aligned} \tag{7.4.7}$$

and finally the ODE of the ECF method system parameters can be written as

$$\begin{aligned}\dot{\theta}_{S,t} &= -R_{S,t}^{-1}\left(g_{\theta,t}^*K_S^{-1}h_S(\theta_{S,t},\eta_t)\right)\\ \dot{g}_{\theta,t} &= g_\theta(\theta_{S,t}) - g_{\theta,t},\end{aligned} \tag{7.4.8}$$

where $R_{S,t} = g_{\theta,t}^*K^{-1}g_{\theta,t}$. By merging the above three ODE-s we get the associated ODE of the recursive three-stage identification method:

$$\begin{aligned}\dot{\theta}_{P,t} &= -R_{P,t}^{-1}W_{P,\theta_P}(\theta_{P,t})\\ \dot{R}_{P,t} &= R(\theta_{P,t}) - R_{P,t}\\ \dot{\eta}_t &= -R_{E,t}^{-1}\left(-\varphi_\eta^*(\eta_t)K^{-1}h_E(\theta_{P,t},\eta_t)\right)\\ \dot{R}_{E,t} &= \varphi_\eta^*(\eta_t)K_E^{-1}\varphi_\eta(\eta_t) - R_{E,t}\\ \dot{\theta}_{S,t} &= -R_{S,t}^{-1}\left(g_{\theta,t}^*K_S^{-1}h_S(\theta_{S,t},\eta_t)\right)\\ \dot{g}_{\theta,t} &= g_\theta(\theta_{S,t}) - g_{\theta,t}.\end{aligned} \tag{7.4.9}$$

The Jacobian of the r.h.s. at

$$(\theta_P, R_P, \eta, R_E, \theta_S, G) = (\theta^*, R_P^*, \eta^*, R_E^*, \theta^*, G(\theta^*, \eta^*))$$

is given by

$$\begin{pmatrix} -I & 0 & 0 & 0 & 0 & 0 \\ J_{2,1} & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & -I & 0 & 0 & 0 \\ 0 & 0 & J_{4,3} & -I & 0 & 0 \\ 0 & 0 & 0 & 0 & -I & 0 \\ 0 & 0 & 0 & 0 & J_{6,5} & -I \end{pmatrix}.$$

Hence the solution of the ODE 7.4.9 is locally stable with $\alpha = 1/2$, see Condition 7.5 for the definition of α. The structure of the above Jacobian matrix and Theorem 7.1 together with Corollary 7.1 imply the next result.

Theorem 7.4. *Let $\hat{\theta}_{S,N}$ and $\hat{\eta}_N$ be the N^{th}-step estimate of the parameters obtained by the recursive estimation in Algorithm 7.3 using a suitable resetting mechanism. Suppose that for the Lévy system Condition 7.6 and Condition 7.7 are satisfied. Suppose further that for the driving Lévy process we have that $((\Delta L_n)^2)$ is in M^* and that the differential equation (7.4.9) satisfies Condition 7.5. Then we have*

$$\hat{\eta}_N - \eta^* = O_M(N^{-1/2}) \quad \text{and} \quad \hat{\theta}_{S,N} - \theta^* = O_M(N^{-1/2}).$$

Furthermore, if $K_E = C$ then the asymptotic covariance matrix $\Sigma_{\eta\eta}^{(\mathrm{rec})}$ of $\hat{\eta}_N$ is given by

$$\Sigma_{\eta\eta}^{(\mathrm{rec})} = (\varphi_\eta^*(\eta^*) C^{-1} \varphi_\eta(\eta^*))^{-1},$$

and if $K_S = C \otimes R_P^*$, then the asymptotic covariance matrix $\Sigma_{\theta\theta}^{(\mathrm{rec})}$ of $\hat{\theta}_{S,N} - \theta^*$ is given by

$$\Sigma_{\theta\theta}^{(\mathrm{rec})} = \left(\psi^* C^{-1} \psi\right)^{-1} (R_P^*)^{-1},$$

with $\psi = (iu_1\varphi(u_1, \eta^*), \ldots, iu_M\varphi(u_M, \eta^*))^T$. Hence, the proposed on-line method gives es-

sentially asymptotically efficient estimates of θ^* and η^*.

Remark:

Similarly to the off-line identification the optimal choice of K_E and K_S depend on the true values θ^*, η^*. Thus these values need to be approximated using the most recent estimates of the parameters, see Section 4.7. In any case it can be easily shown that the results of this Chapter remain valid even if these approximated weighting matrices are used.

Chapter 8

Stability of hybrid Lévy systems

8.1 Motivation

Continuous-time stochastic systems have attracted a lot of attention recently, due to their widespread use in finance for modelling price-dynamics. A widely used model for continuous-time returns has been, since the works of L. Bachelier, Gaussian white noise with drift. More recently models taking into accounts shocks have been developed by assuming that the return process is an infinitesimal Lévy process. For long term modelling a more suitable model is a stochastic system with poles close to 1 driven by a Lévy process, see [33].

Recall that stochastic processes driven or modulated by a Lévy process are called a Lévy system. Description of real data in terms of Lévy systems is far from being settled. In this chapter we focus on a particular technical problem that proved to be fundamental in the statistical analysis of continuous-time stochastic systems driven by Gaussian white noise, see [39]. Ultimately it is hoped that this technical result may contribute to the development of a continuous-time recursive maximum likelihood method for finite dimensional linear stochastic Lévy systems, along the lines of [38].

A possible motivation behind the study of time-varying systems is that the recursive identification method presented in Chapter 7 requires us to work with such systems. There we supposed the joint stability of the transition matrices, an alternative method for guaranteing the stability of the corresponding continuous-time system will be presented in this Chapter.

The problem is the stability analysis of time-varying stochastic systems driven or modulated by a Lévy process with discrete time interventions, such as parameter or state resetting. Such systems will be called hybrid Lévy systems. They are hybrid in the sense that jumps both in the dynamics and the state may occur. The peculiarity of our systems is that the jump-times are defined by a more or less arbitrary point process, but there exists an *asymmetry* in the system dynamics, inasmuch jumps can occur only one-way, after a period of slow variation, namely, back to a fixed point.

Stability of Markovian and switching system have been widely studied in the literature, [10],[66]. We note that the well-developed theory of switching stochastic systems, see [18], does not cover the problem that we consider. The novelty of our model relative to the theory of switching stochastic systems is two-fold. First, we allow slow time variation of the parameters, in a stochastic sense, without any statistical pattern, in the spirit of the classical stability result of Desoer, see [20]. Secondly, we allow certain jumps (resetting) in the system parameters almost without any a priori condition.

The structure of this chapter is as follows: in Section 8.2 we develop the basic technical tools, such as the geometric drift condition and the associated Lyapunov-function method, for the analysis of time-invariant Lévy systems, and provide estimates for higher order moments of the Lyapunov-function. In Section 8.3 we present the simplest version of an extension of Desoer's theorem. In Sections 8.3 we prove a stability result under parameter resetting.

8.2 Modified geometric drift condition for time-invariant systems

Consider the time-invariant linear stochastic system

$$dX_t = AX_{t-}dt + BdL_t + CdW_t \tag{8.2.1}$$

where L_t is Lévy process that has finite variation and has no continuous part and W_t is a standard Wiener process. Assume that A and B are time-independent constant matrices. $X_t \in \mathbb{R}^n, L_t \in \mathbb{R}^l, W_t \in \mathbb{R}^k A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times l}, C \in \mathbb{R}^{n\times k}$. We will denote the i^{th} component of a vector V with $V^{(i)}$. We will use the fact that if L_t is of finite variation then it can be written as $L_t = L_0 + bt + \sum_{s\le t} \Delta L_s$, and the quadratic covariance of two coordinates of such vector processes is of the form

$$\sum_{s\le t} \Delta L_s^{(i)} \Delta L_s^{(j)}.$$

Definition 8.1. *We say that a vector $L_t = (L_t^{(1)}, \dots, L_t^{(l)})$ with independent components that are Lévy processes satisfies the moment condition of order Q if $\int_{\mathbb{R}} |x|^q \nu^{(i)}(dx) < \infty$ holds for all $1 \le i \le l$, and for $1 \le q \le Q$, where the Lévy measure of $L_t^{(i)}$ is denoted by $\nu^{(i)}(x)$.*

The next definition is motivated by the geometric drift condition introduced in [39]. To estimate the moments of X_t we will use a quadratic Lyapunov function V_t.

Definition 8.2. *Let $L_t^{(i)}, 1 \le i \le l$ be independent Lévy processes with finite variation. Let f be a polynomial with coefficients bounded uniformly in t and $\deg f \le Q$. Given a process V_t satisfying with some $\varepsilon > 0$*

$$\begin{aligned} dV_t = V_{t-}\Big(u_t dt + dM_t\Big) + \\ V_{t-}^{1-\varepsilon} f(\Delta L_t^{(1)}, \dots, \Delta L_t^{(l)}), \end{aligned} \tag{8.2.2}$$

where $\Delta L_t^{(i)}$ *denotes the size of the jump of* $L^{(i)}$ *at* t. *We say that* V_t *satisfies the modified geometric drift condition of order* Q *if there exist* $\alpha, \gamma > 0$, *such that*

$$\begin{aligned} u_t &\leq -\alpha \\ \frac{d[M]_t}{dt} &\leq \gamma. \end{aligned} \tag{8.2.3}$$

Without loss generality may always assume that decomposition of L_t contains no drift term. Any possible drift term can be incorporated into $u_t dt$.

The next two lemmas show that our Lyapunov function V_t and its q^{th} power satisfy the modified geometric drift condition.

Lemma 8.1. *Let* X_t *be defined via* $(8.2.1)$, *and* P *given by Lemma* 8.4. *Define* $V_t = 1 + X_t^T P X_t$, *then* V_t *satisfies the modified geometric drift condition of order two.*

Lemma 8.2. *Let* X_t *be defined via* $(8.2.1)$. *Define* $V_t = 1 + X_t^T P X_t$, *then* V_t^q *satisfies the modified geometric drift condition of order* $2q$.

The result of the next Lemma will be used in the proof of Theorem 8.1.

Lemma 8.3. *Let us suppose that* V_t *satisfies the modified geometric drift condition of order* Q, *and suppose that* L_t *satisfies the moment condition of order* Q . *Then*

$$\mathbb{E}\left[V_t\right] < \infty$$

holds.

Proof. If a martingale M is present we may follow [39]. Thus, for the sake of simplicity, we may omit the martingale M_t from (8.2.2). So take a process of the form $(8.2.2)$ with $M_t = 0$. Then V_t satisfies

$$dV_t = V_{t-} Z_t, \tag{8.2.4}$$

with

$$Z_t = u_t dt + V_{t-}^{-\varepsilon} f(\Delta L_t^{(1)}, \ldots, \Delta L_t^{(l)}).$$

Using the Doleans-Dade exponential formula for processes with finite variation yields the solution for V_t :

$$V_t = e^{Z_t^{(c)} - Z_0^{(c)}} \prod_{s \le t} (1 + \Delta Z_s),$$

where $^{(c)}$ denotes the continuous part of a process. This V_t is also called as the stochastic exponential of U_t. Let c and M be uniform bounds for u_t and for the coefficients of f, respectively. Increasing both u_t and the coefficients of f and taking absolute value of the jumps we obtain a bound on the solution V_t :

$$\begin{aligned} V_t \le e^{ct} \prod_{s \le t} \left(1 + M \sum_{0 \le j_1 + \ldots + j_l \le Q} \prod_{i=1}^{l} \left| \Delta L_s^{(i)} \right|^{j_i} \right) \le \\ e^{ct} \prod_{i=1}^{l} \prod_{s \le t} \left(1 + M \sum_{j=1}^{Q} \left| \Delta L_s^{(i)} \right|^{j} \right) \end{aligned} \tag{8.2.5}$$

Since $L_t^{(i)}$-s are independent processes it is sufficient show that

$$\mathbb{E}\left[\prod_{s \le t} \left(1 + M \sum_{j=1}^{Q} \left| \Delta L_s^{(i)} \right|^{j} \right) \right] < \infty.$$

Since L_t satisfies the moment condition of order Q

$$\mathbb{E}\left[1 + M \sum_{j=1}^{Q} \left| \Delta L_s^{(i)} \right|^{j} \right] < \infty.$$

Hence applying Lemma 8.5, see Appendix, concludes the proof. □

The next theorem implies the stability of X_t defined in $(8.2.1)$.

Theorem 8.1. *Let us suppose that V_t satisfies the modified geometric drift condition of order*

Q, and suppose that L_t satisfies the moment condition of order Q . Then

$$\sup_{t\geq 0}\mathbb{E}\left[V_t\right] < \infty.$$

holds.

The proof will be given in the Appendix.

Choose $V_t = 1 + X_t^T P X_t$, we have seen that V_t satisfies the modified geometric drift condition, it follows that so does V_t^q. Thus Theorem 8.1 implies the next corollary

Corollary 8.1. *Let $V_t = 1 + X_t^T P X_t$, where X_t is defined in $(8.2.1)$ with L_t satisfying the moment condition of order Q. then*

$$\sup_{t\geq 0}\mathbb{E}\left[|X|_t^q\right] < \infty$$

for $1 \leq q \leq Q$.

8.3 A stochastic Desoer's Theorem

Consider now a parametric family of linear stochastic state-space systems given by the state space equations:

$$dX_t = A(\theta_t)X_t dt + B(\theta_t)dW_t + C(\theta_t)dL_t. \tag{8.3.1}$$

Condition 8.1. *$A(\theta)$ is stable for each $\theta \in D$, where $D \subset \mathbb{R}^p$ is an open set, and $A(\theta), B(\theta)$ and $C(\theta)$ are smooth in D.*

In this section we study slowly varying stochastic systems in the following sense:

Definition 8.3. *We say that θ_t is slowly varying in a stochastic sense if*

$$d\theta_t = \beta_t dt + \sigma_t dW_t + \rho_t dL_t, \tag{8.3.2}$$

with $|\beta|_t^2 + ||\sigma||_t^2 + ||\rho||_t^2 < \delta$, *for some* $\delta > 0$ *and all* t.

The following technical lemma will be used several times in this chapter.

Lemma 8.4. *Assume Condition 8.1, let* $D_0 \subset D$ *compact and* $\theta_0 \in D_0$. *Then, there exists a smooth function* $P(\theta), \theta \in D$ *and* $\alpha > 0$ *such that* $P(\theta) \geq P(\theta_0) \geq I$ *for all* $\theta \in D_0$ *and* $P(\theta)A(\theta) + A^T(\theta)P(\theta) \leq -\alpha P(\theta)$, *for all* $\theta \in D_0$.

The proof can be found in [39].

Theorem 8.2. *Let* $V_t = 1 + X_t^T P X_t$, *where* X_t *is defined in* $(8.3.1)$. *Assume that* L_t *satisfies the moment condition of order* Q *and* θ_t *is slowly varying in the stochastic sense above, furthermore assume that* $\theta_t \in \mathbb{R}^p$ *is an adapted process taking its values in a compact set* $D_0 \subset D$, *and that Condition 8.1 holds. Then for a sufficiently small* δ *we have*

$$\sup_{t \geq 0} \mathbb{E}\left[V_t^q\right] < \infty, \tag{8.3.3}$$

for $1 \leq q \leq Q$.

Proof. The case when no Lévy terms are present in the dynamics of x_t and θ_t has been settled in Theorem 1 of [39]. We may therefore assume that $B(\theta) = 0$ and $\sigma_t = 0$.

For a given θ, let $P(\theta) \in C^2$ be a symmetric, positive definite matrix that solves

$$P(\theta)A(\theta) + A(\theta)^T P(\theta) \leq -\alpha P(\theta), \tag{8.3.4}$$

with some $\alpha > 0$, and $P(\theta) \geq I$. Let $P_t = P(\theta_t)$, and consider $V_t = (1 + X_t^T P_t X_t)^{q/2}$. It is enough to prove that V_t satisfies the modified geometric drift condition. By Lemma 8.2 we only need to check that $1 + X_t^T P_t X_t$ satisfies the modified geometric drift condition. We can

write the dynamics of $1 + X_t^T P_t X_t = 1 + \mathrm{Tr}(P_t Z_t)$, with $Z_t = X_t X_t^T$ as

$$d\,\mathrm{Tr}(P_t Z_t) = \mathrm{Tr}(P_t dZ_t) + \mathrm{Tr}(dP_t Z_t) + \sum_{i,j} dP_{i,j} d[X^{(i)}, X^{(j)}]_t \tag{8.3.5}$$

The first term can be handled using Lemma 8.1. The dynamics of P_t is given by

$$dP_t = u_t dt + \Sigma_t dL_t, \tag{8.3.6}$$

with $||u||_t^2 + ||\Sigma||_t^2 < c\delta$, with some c. Thus, the second and the third term give drift terms that do not spoil the modified geometric drift condition. The typical form of the contribution of the second term up to a bounded constant multiplier is

$$X^{(i)} X^{(j)} dL^{(k)}, \tag{8.3.7}$$

and that of the third term is

$$dL_t^{(k)} d[L^{(i)}, L^{(j)}]_t. \tag{8.3.8}$$

Hence, $1 + X_t^T P_t X_t$ indeed satisfies the modified geometric drift condition. Thus, applying Theorem 8.1 concludes the proof. □

This result implies the stability of the parameter varying system defined by $(8.3.1)$.

Corollary 8.2. *Under conditions of the previous theorem*

$$\sup_{t \geq 0} \mathbb{E}\left[|X|_t^q\right] < \infty$$

holds for $1 \leq q \leq Q$.

8.4 Time-varying systems with parameter resettings

We now assume that the slowly parameter varying process θ_t resets at random times defined by an arbitrary point process with counting process N_t.

$$d\theta_t = \beta_t dt + \sigma_t dW_t + dL_t + (\theta_0 - \theta_{t-})dN_t, \tag{8.4.1}$$

where $|\beta_t|^2 + ||\sigma_t||^2 < \delta$.

Theorem 8.3. *Let X_t be defined via $(8.3.1)$, and the dynamics of θ_t via (8.4.1). Assume that Condition 8.1 holds, and that L_t satisfies the moment condition of order Q, then*

$$\sup_{t\geq 0} \mathbb{E}\left[|X|_t^q\right] < \infty$$

holds for $1 \leq q \leq Q$.

Proof. We may assume that there is no diffusion part in the dynamics of X_t and θ_t. Let $P(\theta)$ be defined by Lemma 8.4 so that it attains its minimum on D in θ_0. Define $V_t = (1 + X_t^T P(\theta_t) X_t)^{q/2}$. Let ξ_t be the size of the jump at t induced by the jump of θ, i.e.

$$\xi_t = (1 + X_t^T P(\theta_0) X_t)^{q/2} - (1 + X_t^T P(\theta_t) X_t)^{q/2}, \tag{8.4.2}$$

using this notation the dynamics of V_t can be written as

$$dV_t = V_{t-} U_t + \xi_t dN_t, \tag{8.4.3}$$

with

$$U_t = u_t dt + \sum_{0 \leq j_1 + \ldots + j_l \leq Q} c_{j_1,\ldots,j_l} \prod_{i=1}^{l} \left(\Delta L_t^{(i)}\right)^{j_i}.$$

By the minimality of $P(\theta_0)$, the jump term in $(8.4.3)$ causes a non-positive jump in V_t. Let ψ_t

be the stochastic exponential of U_t, then

$$V_t = \psi_t V_0 + \int_0^t \psi_t \psi_s^{-1} \xi_s dN_s \leq \psi_t V_0. \tag{8.4.4}$$

Since $\mathbb{E}[\psi_t V_0] < \infty$ is implied by Theorem 8.1, we conclude the proof. □

8.5 Discussion: state resetting for jump processes

Consider the hybrid linear system with state resettings

$$dX_t = AX_t + BdW_t + CdL_t + (X_0 - X_{t-})dN_t, \tag{8.5.1}$$

where W_t is a Wiener process, and N_t is a counting process.

Conjecture 8.1. *Suppose that L_t satisfies the moment condition of order Q, then for X_t defined by* $(8.5.1)$

$$\sup_{t \geq 0} \mathbb{E}[|X|_t^q] < \infty$$

holds for $1 \leq q \leq Q$.

The statement was proved for $C = 0$ in [39], but that proof cannot be adapted in case the presence of a Lévy process in the dynamic of X_t. A subject of our future research will be the study of this problem . Although the main ideas of the proof have been established some technical issues are still to be taken care of.

8.6 Proofs

Proof of Lemma 8.1:

This lemma is an extension of Lemma 8 in [39], where no Lévy processes are present in defining the dynamics of V_t. Thus, for the sake of simplicity, we may omit the martingale M_t from (8.2.2).

Write $V_t = 1 + X_t^T P X_t = 1 + \mathrm{Tr}(PZ_t)$, where $Z_t = X_t X_t^T$. The dynamic of Z_t can be written as

$$dZ_t = X_{t-} dX_t^T + dX_t X_{t-}^T + B d[L,L]_t B^T, \tag{8.6.1}$$

where $d[L,L]_t$ is an $l \times l$ matrix with entries representing quadratic covariances, that is $d[L,L]_t^{(i,j)} = d[L^{(i)}, L^{(j)}]$. Equation $(8.6.1)$ reads as

$$X_{t-}\left(X_{t-}^T A^T dt + dL_t^T B^T\right) + \left(AX_{t-}dt + BdL_t\right) X_t^T + B d[L,L]_t B^T \tag{8.6.2}$$

Thus the dynamics of $V_t = 1 + \mathrm{Tr}(PZ_t)$ can be written as

$$\left(PX_{t-}X_{t-}^T A^T + PAX_{t-}X_{t-}^T\right) dt + PX_{t-}dL_t B^T + PBdL_t X_{t-}^T + PBd[L,L]_t B^T. \tag{8.6.3}$$

So the dt terms in the dynamics of $V_t = 1 + \mathrm{Tr}(PZ_t)$ are given by

$$\mathrm{Tr}\left(X_{t-}^T A^T P X_{t-} + X_{t-}^T P A X_{t-}\right) \leq -\alpha X_{t-}^T P X_{t-} = -\alpha V_{t-} + \alpha, \tag{8.6.4}$$

for the terms having dL_t

$$\begin{aligned} &\mathrm{Tr}\left(P(X_{t-}dL_t B^T + BdL_t X_{t-}^T)\right) = \\ &\mathrm{Tr}\left((P + P^T)BdL_t X_{t-}^T\right) = \\ &2X_{t-}^T(P+P^T)BdL_t = \psi_t^T dL_t, \end{aligned} \tag{8.6.5}$$

with $|\psi_t|^2 = 4X_{t-}^T(P+P^T)BB^T(P+P^T)X_{t-} \leq 4KV_{t-}$, with some fixed K. Finally for the

term with $d[L,L]_t$

$$\mathrm{Tr}\left(PBd[L,L]_tB^T\right)=\sum_{i,j=1}^{l}c_{i,j}d[L^{(i)},L^{(j)}]_t=\sum_{i,j=1}^{l}c_{i,j}\Delta L_t^{(i)}\Delta L_t^{(j)}, \tag{8.6.6}$$

with some $c_{i,j}, 1\leq i,j\leq l$ constants. It follows that the dynamic of V_t can be written as

$$dV_t=V_{t-}u_tdt+V_{t-}^{1/2}\left(\sum_{i=1}^{l}\frac{\psi_t^{(i)}}{V_{t-}^{1/2}}\Delta L_t^{(i)}+\sum_{i,j=1}^{l}\frac{c_{i,j}}{V_{t-}^{1/2}}\Delta L_t^{(i)}\Delta L_t^{(j)}\right), \tag{8.6.7}$$

with uniformly bounded $u_t, \frac{\psi_t^{(i)}}{V_{t-}^{1/2}}, \frac{c_{i,j}}{V_{t-}^{1/2}}$ for any $1\leq i,j\leq l$, which concludes the proof. □

Proof of Lemma 8.2:

The dynamics of V_t^q can be written as

$$\begin{aligned}
&dV_t^q=qV_{t-}^{q-1}dV_{t,(c)}+V_t^q-V_{t-}^q=\\
&qV_{t-}^{q-1}u_tdt+(V_{t-}+\Delta V_t)^q-V_{t-}^q=\\
&qV_{t-}^qu_t/V_{t-}dt+\sum_{k=1}^{q}\binom{q}{k}(\Delta V_t)^kV_{t-}^{q-k},
\end{aligned} \tag{8.6.8}$$

with $u_t/V_{t-}<\alpha$. Using that

$$\Delta V_t=V_{t-}^{1-\varepsilon}f(\Delta L_t^{(1)},\ldots,\Delta L_t^{(l)})$$

we obtain that a typical jump term in (8.6.8) reads as up to constant multiplier

$$V_{t-}^{q-k\varepsilon}f(\Delta L_t^{(1)},\ldots,\Delta L_t^{(l)})^k. \tag{8.6.9}$$

This implies that V_t^q satisfies the modified geometric drift condition of order $2q$. □

The next two technical Lemmas will be used in the proof of Theorem 8.1.

Lemma 8.5. *Let L_t be a Lévy process with Lévy measure ν. Suppose that a function f satisfies*

$$\int_{\mathbb{R}} f(x)\nu(dx) < \infty,$$

then

$$\mathbb{E}\left[\prod_{s\leq t}(1+f(\Delta L_s))\right] = e^{t\int_{\mathbb{R}} f(x)\nu(dx)}$$

holds for any t.

Proof. First suppose that L_t is a compound Poisson process with rate λ, then the expected value of

$$\psi_t = \prod_{s\leq t} f(\Delta L_s)$$

can be estimated by conditioning on the number of jumps of L_t. Let N_t, J_t denote the number of jumps of L_t on $[0,t]$, and the set of time indices when L jumps on $[0,t]$, respectively. Define $D_t^n = \{(t_1,\ldots,t_n) : 0 \leq t_i \leq t, \text{ for all } 1 \leq i \leq n\}$.

$$\begin{aligned}
&\mathbb{E}\left[\Psi_t\right] = \sum_{n=0}^{\infty}\mathbb{E}\left[\Psi_t | N_t = n\right]\mathbb{P}\left(N_t = n\right) = \\
&\int_{D_t^n}\mathbb{E}[\Psi_t | N_t = n, J_t = \{t_1,\ldots,t_n\}]P(N_t = n)dt_1\ldots dt_n = \\
&\sum_{n=0}^{\infty}(m+1)^n e^{-\lambda t}\frac{(\lambda t)^n}{n!} = e^{\lambda m t},
\end{aligned}$$

where $m = \mathbb{E}\left[f(\Delta L_t) | L \text{ jumps at } t\right]$, and P is the joint probability density of the jump times. For the general case define the truncated Lévy measure

$$\nu^{\varepsilon}(x) = \frac{\mathbf{1}_{|x|>\varepsilon}\nu(x)}{\int_{|x|>\varepsilon}\nu(dx)},$$

and let L_t^{ε} be the Lévy process with Lévy measure ν^{ε}. Then L_t is the weak limit of L_t^{ε} as ε

tends to zero.

$$m_f^\varepsilon = \int f(x)\nu^\varepsilon(dx)$$
$$m^\varepsilon = \mathbb{E}\left[f(\Delta L_t^\varepsilon)|L^\varepsilon \text{ jumps at } t\right]$$
$$\lambda^\varepsilon = \int_{|x|>\varepsilon} \nu(dx)$$

writing (8.6) for L^ε yields

$$\mathbb{E}\left[\prod_{s\le t}(1+f(\Delta L_s^\varepsilon))\right] = e^{\lambda^\varepsilon(m^\varepsilon-1)t}$$

Note that $\lambda^\varepsilon m_f^\varepsilon = \int_{|x|>\varepsilon} f(x)\nu(dx)$, it follows that $e^{\lambda^\varepsilon m^\varepsilon t}$ has finite limit as $\varepsilon \to 0^+$ provided $\int_{\mathbb{R}} f(x)\nu(dx) < \infty$ which is the case. Hence, $\mathbb{E}\left[\psi_t\right] = e^{t\int_{\mathbb{R}} f(x)\nu(dx)}$ follows. □

Lemma 8.6. *Let the one dimensional process L_t with Lévy measure ν satisfy the moment condition of order Q. Let f be a polynomial with $\deg f \le Q$, and $f(0) = 0$. Then*

$$\mathbb{E}\left[\int_0^t e^{-\alpha(t-s)} f(\Delta L_s)\right] = \frac{1-e^{-\alpha t}}{\alpha}\int_{\mathbb{R}} f(x)\nu(dx).$$

Proof. First consider the case when L_t is a compound Poisson process with intensity λ. Let N_t, J_t denote the number of jumps of L_t on $[0,t]$, and the set of time indices when L jumps on $[0,t]$, respectively. Define $D_t^n = \{(t_1,\ldots,t_n) : 0 \le t_i \le t, \text{ for all } 1 \le i \le n\}$.

$$\mathbb{E}\left[\int_0^t e^{-\alpha(t-s)} f(\Delta L_s)\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[\int_0^t e^{-\alpha(t-s)} f(\Delta L_s)|N_t = n\right] P(N_t = n). \qquad (8.6.10)$$

Calculating one term in the above sum

$$\int_{D_t^n} \mathbb{E}\left[\int_0^t e^{-\alpha(t-s)} f(\Delta L_s)|N_t = n, J_t = \{t_1, \ldots, t_n\}\right] P(N_t = n)\frac{dt_1 \ldots dt_n}{t^n} =$$
$$\int_{D_t^n} \sum_{i=1}^{n} e^{-\alpha(t-t_i)} \mathbb{E}\left[f(\Delta L_s)|N_t = n, J_t = \{t_1, \ldots, t_n\}\right] e^{-\lambda t}\frac{(\lambda t)^n}{n!}\frac{dt_1 \ldots dt_n}{t^n} =$$
$$n\int_{D_t^{n-1}} \int_0^t e^{-\alpha(t-t_n)} \mathbb{E}\left[f(\Delta L_s)|t_1 \in J_t\right] e^{-\lambda t}\frac{(\lambda t)^n}{n!}\frac{dt_n}{t}\frac{dt_1 \ldots dt_{n-1}}{t^{n-1}} =$$
$$ne^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathbb{E}\left[f(\Delta L_s)|t_1 \in J_t\right]\frac{1-e^{-\alpha t}}{\alpha t}.$$

Now using this result compute further (8.6.10) to get

$$\frac{1-e^{-\alpha t}}{\alpha t}\mathbb{E}\left[f(\Delta L_s)|t_1 \in J_t\right]\sum_{n=0}^{\infty} ne^{-\lambda t}\frac{(\lambda t)^n}{n!} =$$
$$\frac{1-e^{-\alpha t}}{\alpha}\lambda\mathbb{E}\left[f(\Delta L_s)|t_1 \in J_t\right]. \tag{8.6.11}$$

For the general case define like in the proof of Lemma 8.5 process L_t^ε and its Lévy measure $\nu^\varepsilon(dx)$, and $m^\varepsilon = \mathbb{E}\left[f(\Delta L_s)|t_1 \in J_t\right]$. Writing (8.6.11) for L_t^ε we obtain

$$\mathbb{E}\left[\int_0^t e^{-\alpha(t-s)} f(\Delta L_s^\varepsilon)\right] =$$
$$\frac{1-e^{-\alpha t}}{\alpha}\lambda^\varepsilon\mathbb{E}\left[f(\Delta L_s^\varepsilon)|t_1 \in J_t\right] = \tag{8.6.12}$$
$$\frac{1-e^{-\alpha t}}{\alpha}\int_{|x|>\varepsilon} f(x)\nu(dx).$$

Since L_t is the weak limit of L_t^ε as ε tends to zero, allowing $\varepsilon \to 0^+$ concludes the proof. □

Proof of Theorem 8.1:

Let V_t satisfy the modified geometric drift condition, that is

$$dV_t = u_t dt + V_{t-}^{1-\varepsilon} f(\Delta L_t^{(1)}, \ldots, \Delta L_t^{(l)}). \tag{8.6.13}$$

In the presence of a martingale in the dynamics of V_t one can apply methods shown in [39]. Applying Cauchy formula gives

$$V_t = \int_0^t e^{-\alpha(t-s)} V_{s-}^{1-\varepsilon} f(\Delta L_s^{(1)}, \dots, \Delta L_s^{(l)}) \tag{8.6.14}$$

Now we estimate the expected values of V_t using $V_T^* := \sup_{0\le s\le T} \mathbb{E}[V_s]$.

$$\mathbb{E}[V_t] = \int_0^t \mathbb{E}\left[e^{-\alpha(t-s)} V_{s-}^{1-\varepsilon} f(\Delta L_s^{(1)}, \dots, \Delta L_s^{(l)})\right] =$$

$$\int_0^t e^{-\alpha(t-s)} \mathbb{E}[V_{s-}^{1-\varepsilon}] \mathbb{E}\left[f(\Delta L_s^{(1)}, \dots, \Delta L_s^{(l)})\right] \le \tag{8.6.15}$$

$$\int_0^t e^{-\alpha(t-s)} \mathbb{E}[V_{s-}]^{1-\varepsilon} \mathbb{E}\left[f(\Delta L_s^{(1)}, \dots, \Delta L_s^{(l)})\right] \le \tag{8.6.16}$$

$$(V_T^*)^{1-\varepsilon} \int_0^t e^{-\alpha(t-s)} \mathbb{E}\left[f(\Delta L_s^{(1)}, \dots, \Delta L_s^{(l)})\right] \le \tag{8.6.17}$$

$$(V_T^*)^{1-\varepsilon} \prod_{i=1}^{l} \int_0^t e^{-\alpha(t-s)} \mathbb{E}\left[g_i(\Delta L_s^{(i)})\right] \tag{8.6.18}$$

First we used Fubini's theorem and the independency of V_{s-} and $\Delta L_s^{(i)}$. In (8.6.15) Hölder inequality was applied after that in (8.6.16) we used the definition of V_T^*. Finally, in (8.6.18) we estimate f with products of g_i polynomials as we did in (8.2.5), clearly $\deg g_i \le Q$ holds for all i. Applying Lemma 8.6 gives for (8.6.18)

$$(V_T^*)^{\varepsilon} \le \prod_{i=1}^{l} \frac{1-e^{-\alpha t}}{\alpha} \int_{\mathbb{R}} g_i(x) \nu_i(dx),$$

where ν_i is the Lévy measure of $L_t^{(i)}$. Since $1 - e^{-\alpha t} < 1$ we obtained a bound on V_T^* that do not depend on T. Hence, $\sup_{0\le t} \mathbb{E}[V_t] < \infty$, which concludes the proof. □

Chapter 9

Simulation results

9.1 Simulating CGMY processes

In this chapter we test the performance of our algorithms using simulated data. In each simulation the driving noise is a CGMY process. A possible simulation of the tempered stable distribution is obtained by using the series representation of such processes, following [58]: we will first generate the positive jumps of the process with Lévy measure

$$\nu_+(dx) = \frac{Ce^{-Mx}}{|x|^{1+Y}} \text{ for } x > 0,$$

and $\nu_+(dx) = 0$ otherwise. Let $(E_j), (U_j)$ and (P_j) be sequences of independently distributed random variables such that each E_j has exponential distribution with rate M, each U_j is uniformly distributed on $[0, 1]$ and P_j-s are the arrival times of a Poisson process with rate 1. Then the series representation of the process (X_t) corresponding to the Lévy measure $\nu_+(dx)$ is given by

$$X_t = \sum_{j=1}^{\infty} \min\left(\left(\frac{2Ct}{MP_j} \right)^{1/Y}, E_j U_j^{1/Y} \right). \tag{9.1.1}$$

So if N is large enough one can take

$$X_t \approx X_t^{(N)} = \sum_{j=1}^{N} \min\left(\left(\frac{2Ct}{MP_j}\right)^{1/Y}, E_j U_j^{1/Y}\right) \tag{9.1.2}$$

to simulate sample paths of the process with Lévy measure $\nu_+(dx)$. Similarly, one can simulate the process with Lévy measure

$$\nu_-(dx) = \frac{Ce^{-G|x|}}{|x|^{1+Y}}, \text{for } x < 0,$$

and $\nu_-(dx) = 0$ otherwise, to get the process generated by the negative jumps. By adding up the two processes one gets an approximation of the tempered stable process with parameters C, G, M, Y. In Figure 9.1 an approximation of the sample path of a CGMY process is plotted using the above formula with $N = 10$. In Figure 9.2 a similar approximation can be seen, but generated with $N = 10$.

9.2 Re-estimation of ARMA system paramters by ECF method

In this section we present some simulation results for ARMA processes. As the numerical aspects of the estimation of the noise characteristics for i.i.d. data is extensively studied in the literature, see for example [16], here we confine our study to the estimation of the system parameters. More precisely, we test the third step of our method presented in Chapter 4. To this end we simulated ARMA processes defined by

$$A(\theta^*)\Delta y_n = C(\theta^*)\Delta L_n \quad \text{for } n > 0, \tag{9.2.1}$$

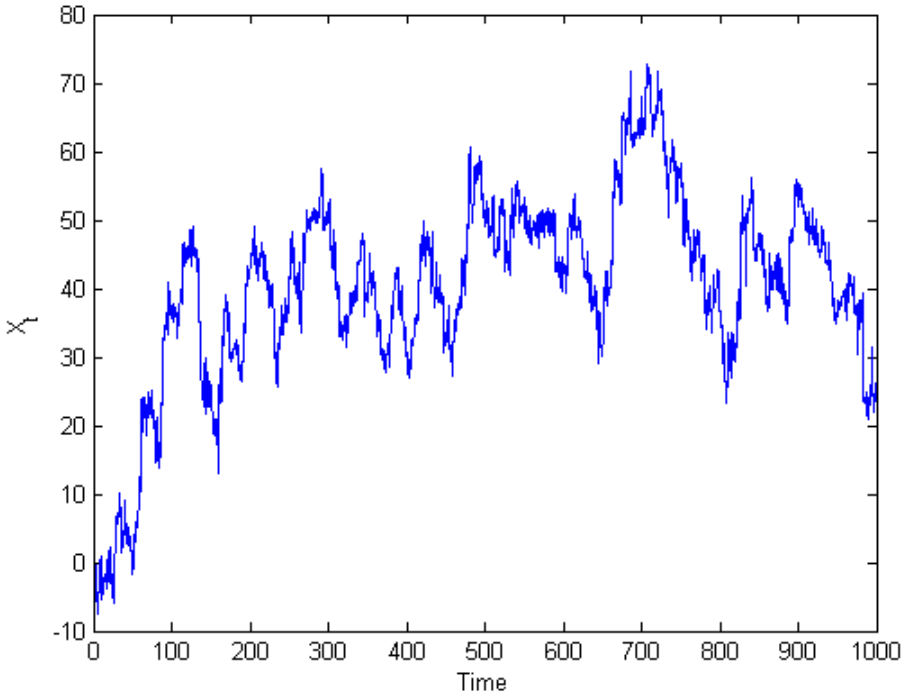

Figure 9.1: Approximation of a CGMY process

and $\Delta y_n = 0$, for $n \leq 0$, where ΔL_n is the increment of a CGMY process with parameters $\eta^* = [C^*, G^*, M^*, Y^*]$, to be specified later. Let A and C be polynomials of the back-shift operator q^{-1}:

$$A(\theta^*) = 1 + \theta_1^{(A)} q^{-1} + \ldots + \theta_{p_1}^{(A)} q^{-p_1} \tag{9.2.2}$$

$$C(\theta^*) = 1 + \theta_1^{(C)} q^{-1} + \ldots + \theta_{p_2}^{(C)} q^{-p_2}, \tag{9.2.3}$$

and θ^* be the unknown system parameter with components $\theta^* = (\theta_1^{(A)}, \ldots, \theta_{p_1}^{(A)}; \theta_1^{(C)}, \ldots, \theta_{p_2}^{(C)})$. We test our method using different ARMA processes, in each case 10.000 simulated observations were used. Table 8.1 and 8.2 present the type of the process the true parameters and the estimated parameters. Our computational experience suggested that we should limit the number of u-s, hence the ECF identification method was applied with $u_k = k/10, \quad k = 1, \ldots, 10$ and

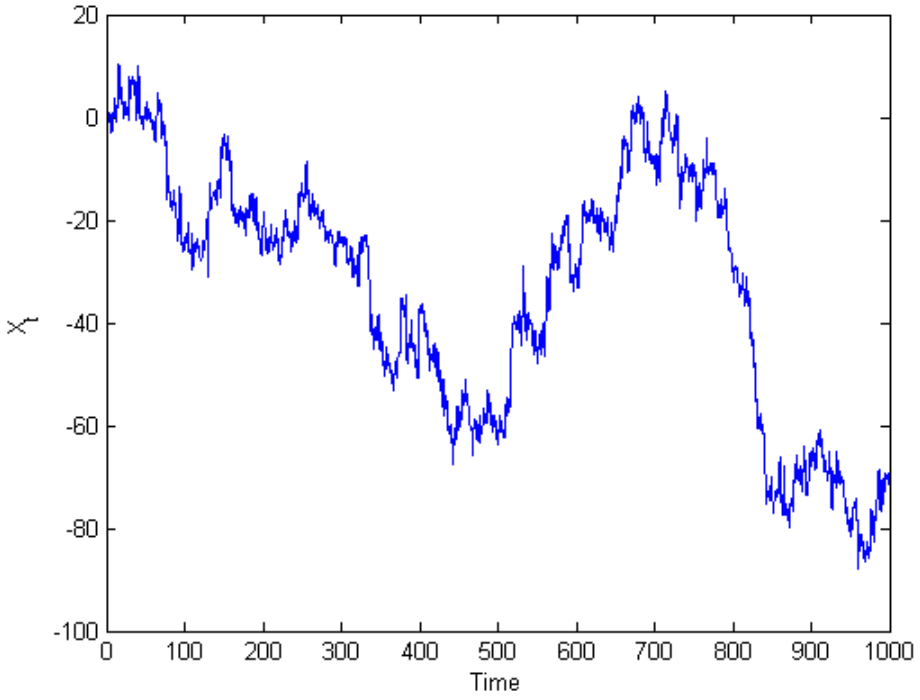

Figure 9.2: Approximation of a CGMY process

K being the identity matrix of appropriate dimension.

Table 9.1: ARMA simulation results

Process	$\theta^{(A)*}$	$\theta^{(B)*}$	η^*	$\hat{\theta}_N^{(A)}$	$\hat{\theta}_N^{(B)}$
AR(1)	0.4	n/a	(1,0.5,0.5,0.3)	0.4130 + 0.0003i	n/a
AR(2)	(0.4,0.1)	n/a	(1,0.5,0.5,0.3)	(0.4009 - 0.0018i,0.1040 + 0.0031i)	n/a

Table 9.2: ARMA simulation results

Process	$\theta^{(A)*}$	$\theta^{(B)*}$	η^*	$\hat{\theta}_N^{(A)}$	$\hat{\theta}_N^{(B)}$
MA(1)	n/a	0.6	(1,0.5,0.5,0.3)	n/a	0.6035 + 0.0129i
ARMA(1,1)	0.25	0.75	(1,0.5,0.5,0.3)	0.2749 - 0.0221i	0.7645 + 0.0197i

9.3 Re-estimation of ARMA system parameters by recursive ECF method

This section is devoted to the presentation of the simulation results for the recursive algorithm proposed in Section 7.3. We use the notations of the last section. Recall that the recursive estimation requires the definition of domains for each variable. In case a variable leaves its domain it is reset to its initial value. Here, we reset all variables if the absolute value of the system parameter θ is greater than 1. Again, we restrict our attention to recursive identification of the system parameters, i.e. we suppose that η^* is given. The set of u-s and K are defined like in the previous section. The estimated parameters after 5.000 steps can be seen in Table 8.3. In Figure 9.3 the norm of the estimation error, i.e. $\left|\hat{\theta}_N - \theta^*\right|$, is plotted for a MA(1) process with parameters $\theta^* = 0.3$ and noise characteristics $\eta^* = (1, 0.5, 0.5, 0.3)$, the initial value of the parameter θ was defined by $\hat{\theta}_0 = 0.6$.

Table 9.3: ARMA on-line simulation results

Process	$\theta^{(A)*}$	$\theta^{(B)*}$	η^*	$\hat{\theta}_N^{(A)}$	$\hat{\theta}_N^{(B)}$
AR(1)	0.4	n/a	(1,0.5,0.5,0.3)	0.3991 - 0.0037i	n/a
MA(1)	n/a	0.3	(1,0.5,0.5,0.3)	n/a	0.2897 + 0.0044i
ARMA(1,1)	0.25	0.75	(1,0.5,0.5,0.3)	0.2701 + 0.0113i	0.7419 - 0.0201i

9.4 ECF estimate of GARCH processes

Finally, we test the ECF method for GARCH processes presented in Chapter 6. Again, we suppose that the driving noise process is given by the increments of a tempered stable process with parameter vector η^*. Table 8.4 presents the results of simulation for different GARCH processes. In each case 30.000 simulated observations were used. The ECF method was applied with $u_k = k/10, \quad k = 1, \ldots, 10$ and K being the identity matrix of appropriate dimension.

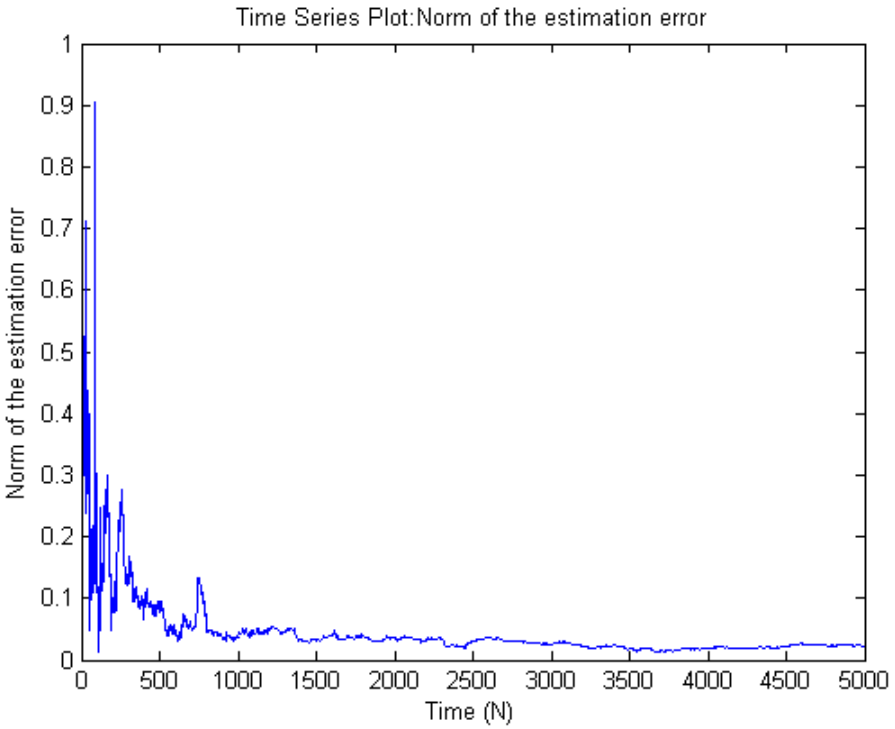

Figure 9.3: Norm of $\hat{\theta}_N - \theta^*$

For simplicity we assumed that $\alpha_0^* = 0.1$ is known.

Table 9.4: GARCH simulation results

Process	θ^*	η^*	$\hat{\theta}_N$
GARCH(1,1)	(0.2,0.5)	(1,0.5,0.5,0.3)	(0.2169 + 0.0068i, 0.4895 - 0.0109i)
GARCH(1,1)	(0.4,0.3)	(1,0.5,0.5,0.3)	(0.3966 - 0.0005i, 0.2980 - 0.0003i)
GARCH(1,1)	(0.099,0.8)	(1,0.5,0.5,0.3)	(0.0950 - 0.0014i, 0.8041 + 0.0017i)

Remark:

The empirical calculation of the asymptotic covariance matrix is beyond the scope of this work. The simulation results presented above show that these methods are computationally feasible. Although the algorithms were implemented with a naive choice of u-s and choosing the weighting matrix $K = I$, the results of the simulations are convincing.

Chapter 10

Appendix

10.1 Gamma processes

In this section we provide a brief introduction to gamma-processes that play a major role in VG modelling. To construct a gamma process let ξ_n be an i.i.d. sequence of random variables with exponential distribution having density $\lambda e^{-\lambda x}$ for $x > 0$. Let $s_k = \xi_1 + \ldots + \xi_k$. The probability density function and characteristic function of s_k are given by

$$\lambda^k x^{k-1} e^{-\lambda x}/(k-1)! \tag{10.1.1}$$

and

$$\mathbb{E}\left[\exp(ius_k)\right] = \left(\frac{\lambda}{\lambda - iu}\right)^k = \left(\frac{1}{1 - \frac{iu}{\lambda}}\right)^k. \tag{10.1.2}$$

For the means and the variances of s_k we have:

$$\mathbb{E}\left[s_k\right] = k/\lambda \quad \text{and} \quad \sigma^2(s_k) = k/\lambda^2.$$

We can look upon (s_k) as a stochastic process defined over the positive integers, with independent and identically distributed increments. Let us re-parametrize the density above by introducing the new variables

$$\nu = 1/\lambda \quad \text{and} \quad t = k/\lambda = k\nu. \tag{10.1.3}$$

Here $\nu = 1/\lambda$ is the *mean life-time*. Then define

$$\gamma_t = s_k = s_{t/\nu}.$$

Remember that t/ν is the number of exponential terms. The probability density function of γ_t can be written as

$$f_t(x) = \left(\frac{1}{\nu}\right)^{t/\nu} \frac{x^{t/\nu-1}e^{-x/\nu}}{\Gamma(t/\nu)}. \tag{10.1.4}$$

The characteristic function of $\gamma_t(\mu, \nu)$ is given by

$$\varphi_t(u; \mu, \nu) = \left(\frac{1}{1 - i\frac{\nu}{\mu}u}\right)^{\mu^2 \frac{t}{\nu}}. \tag{10.1.5}$$

Finally, the means and the variances of γ_t are

$$\mathbb{E}[\gamma_t] = t \quad \text{and} \quad \sigma^2(\gamma_t) = t\nu. \tag{10.1.6}$$

Now it can be shown that $f_t(x)$ as defined above, is a density function for *any* real $t \geq 0$. This is called a gamma-density. The corresponding characteristic function is given by (10.1.5) for *any* real $t \geq 0$. Obviously, thus set of characteristic functions is closed under multiplication. Thus gamma-densities are closed under convolution. Consequently, we can construct a stochastic process (γ_t), with $t \geq 0$ real, with stationary independent increments, so that the the density function of $\gamma_{t+h} - \gamma_t$ is f_h. This is called a gamma-process. Obviously, the means and variances

of γ_t are obtained as in (10.1.6) for *any* real t. Therefore we say that the *mean rate* of (γ_t) is 1, and its *variance rate* is ν.

Finally, we can re-scale the process by setting, with some $\mu' > 0$,

$$t' = t/\mu' \quad \text{and} \quad \gamma'_{t'} = \gamma_t. \tag{10.1.7}$$

Then

$$\mathbb{E}\left[\gamma'_{t'}\right] = \mu' t' \quad \text{and} \quad \sigma^2(\gamma'_{t'}) = (\mu' t')\nu = (\mu'\nu)t'. \tag{10.1.8}$$

Correspondingly, we say that the *mean-rate* of the re-scaled process is μ', and the *variance rate* of the re-scaled process is $\nu' = \mu'\nu$. We can express the old variables in terms of the new variables by the the following scaling equations:

$$t = \mu' t' \quad \text{and} \quad \nu = \nu'/\mu'.$$

Expressing the density function of $\gamma'_{t'} = \gamma_t$ in terms of these parameters, and changing the roles of parameters with and without superscripts, and correspondingly making the replacements

$$t \to \mu t \quad \text{and} \quad \nu \to \nu/\mu$$

we get

$$f_t(x;\mu,\nu) = \left(\frac{\mu}{\nu}\right)^{\mu^2\frac{t}{\nu}} \frac{x^{\mu^2\frac{t}{\nu}-1}e^{-\mu\frac{x}{\nu}}}{\Gamma(\mu^2\frac{t}{\nu})}. \tag{10.1.9}$$

Note that the following scaling property holds: for any $c > 0$

$$f_t(x;\mu,\nu) = f_{ct}(x;\mu/c,\nu/c). \tag{10.1.10}$$

A random variable with this distribution will be denoted by $\gamma_t = \gamma_t(\mu,\nu)$, with μ denoting the

mean rate and ν denoting the variance rate. Its characteristic function is given by

$$\phi_t(u;\mu,\nu) = \left(\frac{1}{1-i\frac{\nu}{\mu}u}\right)^{\mu^2\frac{t}{\nu}}.$$

Similarly, a stochastic process (γ_t) with stationary independent increments, so that the the density function of $\gamma_{t+h}-\gamma_t$ is $f_h(x;\mu,\nu)$ will be denoted by $\gamma_t(\mu,\nu)$. This is then a Lévy process, the Lévy density of which can be explicitly determined, see [56].

$$\nu(dx;\mu,\nu) = \frac{\mu^2}{\nu}\frac{e^{-\mu\frac{x}{\nu}}}{x}\mathbf{1}_{x>0}dx. \tag{10.1.11}$$

Note that the integral of $\nu(dx)$ is infinite, hence the gamma process has an infinite number of jumps in any finite interval. It is also said that the gamma process is an infinite activity process. Clearly, most of these jumps are very small as the Lévy measure is concentrated at the origin.

10.2 L-mixing processes

In this section we summarize the most important definitions and theorems in the area of L-mixing processes. While we follow [30] the concept of L-mixing is presented with a minor modification. Let θ be a d-dimensional parameter vector. The results can be easily extended for complex valued parameters.

Definition 10.1. *We say that $x_n(\theta)$ is M-bounded of order Q if for all $1 \le q \le Q$,*

$$M_q^Q(x) = \sup_{n>0,\theta\in D} \mathbb{E}^{1/q}\left[|x_n(\theta)|^q\right] < \infty$$

Define $\mathscr{F}_n = \sigma\{e_i : i \le n\}$ and $\mathscr{F}_n^+ = \sigma\{e_i : i > n\}$ where e_i-s are i.i.d. random variables.

Definition 10.2. *We say that a stochastic process $(x_n(\theta))$ is L-mixing of order Q with respect to $(\mathscr{F}_n, \mathscr{F}_n^+)$ uniformly in θ if it is $\mathscr{F}_n$ progressively measurable, M-bounded of order Q with*

any positive r and

$$\gamma_q(\tau, x) = \sup_{n \geq \tau, \theta \in D} \mathbb{E}^{1/q} \left[\left| x_n(\theta) - \mathbb{E}\left[x_n(\theta) | \mathscr{F}_{n-\tau}^+ \right] \right|^q \right],$$

we have for any $1 \leq q \leq Q$,

$$\Gamma_q(x) = \sum_{\tau=1}^{\infty} \gamma_q(\tau, x) < \infty.$$

The definition can be extended to continuous time processes as well. We note in passing that with $Q = \infty$ we get L-mixing processes proposed by Gerencsér in [30]. Thus, usage of L-mixing meaning that L-mixing with order $Q = \infty$ should not cause any confusion.

Definition 10.3. *We say that a stochastic process $(x_t(\theta))$ is L-mixing of order Q with respect to $\left(\mathscr{F}_t, \mathscr{F}_t^+\right)$ uniformly in θ if it is $\mathscr{F}_n$ progressively measurable, M-bounded of order Q with any positive τ and*

$$\gamma_q(\tau, x) = \sup_{t \geq \tau, \theta \in D} \mathbb{E}^{1/q} \left[\left| x_t(\theta) - \mathbb{E}\left[x_t(\theta) | \mathscr{F}_{t-\tau}^+ \right] \right|^q \right],$$

we have for any $1 \leq q \leq Q$,

$$\Gamma_q(x) = \int_{0=1}^{\infty} \gamma_q(\tau, x) d\tau < \infty.$$

Example:

Let the process X_t given by

$$dX_t = AX_t dt + BdW_t, X_0 = 0, \tag{10.2.1}$$

where W_t is an m-dimensional Brownian motion, $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}$ and A is stable. Then

X_t is L-mixing.

$$X_t = e^{A\tau} X_{t-\tau} + \int_{t-\tau}^{t} e^{A(t-r)} B dW_r, \tag{10.2.2}$$

hence $\mathbb{E}\left[X_t | \mathscr{F}_{t-\tau}^{+}\right] = \int_{t-\tau}^{t} e^{A(t-r)} B dW_r$, it follow that

$$X_t - \mathbb{E}\left[X_t | \mathscr{F}_{t-\tau}^{+}\right] = e^{A\tau} X_{t-\tau}.$$

One can easily check that X_t is M-bounded, thus the stability of A implies $\Gamma_q(x) < \infty$.

The following well-known lemma is used in proving several properties of L-mixing processes.

Lemma 10.1. *Let ξ be an M-bounded random variable and let $\mathscr{F}' \subset \mathscr{F}$ be some σ-algebra. Then for all $1 \leq q < \infty$ and for any $\mathscr{F}'$ measurable η we have*

$$\mathbb{E}^{1/q}\left[|\xi - \mathbb{E}\left[\xi | \mathscr{F}'\right]|^q\right] \leq 2\mathbb{E}^{1/q}\left[|\xi - \eta|^q\right]$$

Lemma 10.2. *Let X_n and Y_n be L-mixing process with respect to the same filtration, then the process $(XY)_n$ is L-mixing as well.*

Theorem 10.1. *Let $(u_n), n \geq 0$ be an L-mixing process of order Q with $\mathbb{E}[u_n] = 0$ for all n, and let (f_n) be a deterministic sequence. Then we have for all $1 \leq m \leq Q/2$,*

$$\mathbb{E}^{1/(2m)}\left[\left|\sum_{n=1}^{N} f_n u_n\right|^{2m}\right] \leq C_m \left(\sum_{n=1}^{N} f_n^2\right)^{1/2} M_{2m}^{1/2}(u) \Gamma_{2m}^{1/2}(u)$$

where $C_m = 2(2m-1)^{1/2}$.

Define

$$\Delta x / \Delta^{\alpha} \theta = |x_n(\theta + h) - x_n(\theta)| \,/\, |h|^{\alpha}$$

for $n \geq 0, \theta \neq \theta + h \in D$ with $0 < \alpha \leq 1$.

Definition 10.4. *We say that $x_n(\theta)$ is M-Hölder continuous of order Q in θ with exponent α if the process $\Delta x/\Delta^\alpha\theta$ is M-bounded of order Q.*

Now let us suppose that $(x_n(\theta))$ is measurable, separable, M-bounded of order Q and M-Hölder of order Q in θ with exponent α for $\theta \in D$. The realizations of $(x_n(\theta))$ are continuous in θ almost surely hence

$$x_n^* = \max_{\theta \in D_0} |x_n(\theta)|$$

is well defined for almost all ω, where $D_0 \subset \text{int } D$ is a compact domain. Since the realizations of $(x_n(\theta))$ are continuous, x_n^* is measurable with respect to $\mathscr{F}$.

Theorem 10.2. *Assume that $(x_n(\theta))$ is measurable, separable, M-bounded of order Q and M-Hölder of order Q in θ with exponent α for $\theta \in D$. Then we have for all positive $q \leq Q\alpha/s$ and $p/\alpha < s \leq Q/q$,*

$$M_q(x^*) \leq C\left(M_{qs}(x) + M_{qs}(\Delta x/\Delta^\alpha\theta)\right)$$

where C depends only on p, q, s, α and D_0, D.

Choosing $f_n = 1$ and $\alpha = 1$ and using Theorem 10.1 and 10.2 we obtain

Theorem 10.3. *Let $(u_n(\theta))$ be an L-mixing of order Q uniformly in $\theta \in D$ such that $\mathbb{E}[u_n(\theta)] = 0$ for all $n \geq 0, \theta \in D$, and assume that $\Delta u/\Delta\theta$ is also L-mixing of order Q, uniformly in $\theta, \theta + h \in D$. Then*

$$\sup_{\theta \in D_0} \left| \frac{1}{N} \sum_{n=1}^{N} u_n(\theta) \right| = O_M^{Q/p}(N^{-1/2}) \tag{10.2.3}$$

Theorem 10.4. *Let D_0 and D be as above and let $W_\theta(\theta), \delta W_\theta(\theta),\ \theta \in D \subset \mathbb{R}^p$ be $\mathbb{R}^p$-valued continuously differentiable functions, let for some $\theta^* \in D_0, W_\theta(\theta^*) = 0$, and let $W_{\theta\theta}(\theta^*)$ be nonsingular. Then for any $d > 0$ there exists positive*

numbers d', d'' *such that*

$$|\delta W_\theta(\theta)| < d' \text{ and } \|\delta W_{\theta\theta}(\theta)\| < d'' \tag{10.2.4}$$

for all $\theta \in D_0$ *implies that the equation* $W_\theta(\theta) + \delta W_\theta(\theta) = 0$ *has exactly one solution in a neighborhood of radius* d *of* θ^*.

Chapter 11

Conclusions

In this thesis we investigated the problem of identifying the parameters of finite dimensional stochastic systems driven by the increments of a Lévy process. This technical problem is motivated by the statistical analysis of financial time series. The common special feature of the presented problems is that the characterization of the driving noise is given by the characteristic function of the noise instead of the density function. We have seen that by using the combination of standard methods of system identification, such as the prediction error method and the ECF method we get essentially equivalent alternatives of the ML method in terms of efficiency.

By adapting the ECF method we proposed a three-stage identification method that estimates the system and noise parameters of linear Lévy systems essentially asymptotically efficiently. The problem of estimating the system parameters of Lévy driven GARCH processes can be solved using the idea of the third step of our three-stage identification method. We have proved that the resulting procedure is essentially as efficient as the ML method.

We have also demonstrated that a properly adapted ECF method using blocks of simulated data can be applied to identify finite dimensional linear Lévy systems with possibly unstable zeros. The method can be best described as an output error method and we provided the

analysis of the procedure along the lines of the classical ECF method.

The general framework of recursive estimation method, namely the DFL-scheme, enabled us to convert the three-stage off-line method for linear Lévy systems into a recursive identification method in a natural manner. Using advanced ODE techniques we proved that the resulting on-line estimation procedure gives essentially efficient estimates of the system and noise parameters of the process.

Motivated by our interest in the analysis of continuous time systems we proved a particular technical result: the stability of time-varying stochastic systems driven by a Lévy process with arbitrary discrete time parameter resettings. In connection with this we developed a novel Lyapunov function method to prove the stability of linear Lévy systems.

Bibliography

[1] A. Almendral and C. W. Oosterlee. Accurate evaluation of european and American options under the CGMY process. *SIAM Journal on Scientific Computing*, 29(1):93–117, 2007.

[2] S. Asmussen, D. Madan, and M. Pistorius. Pricing equity default swaps under an approximation to the CGMY Lévy model. *Journal of Computational Finance*, 11(2):79, 2007.

[3] K. Åström and T. Söderström. Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model. *IEEE Transactions on Automatic Control*, 19(6):769–773, 1974.

[4] L. Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.

[5] F. E. Benth and S. J. Benth. Dynamic pricing of wind futures. *Energy Economics*, 31(1):16–24, 2009.

[6] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*. Springer, 2012.

[7] I. Berkes, L. Horváth, and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9(2):201–227, 2003.

[8] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[9] P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52(1):115–127, 1992.

[10] E. Boukas, P. Shi, and K. Benjelloun. On stabilization of uncertain linear systems with jump parameters. *International Journal of Control*, 72(9):842–850, 1999.

[11] P. J. Brockwell and E. Schlemm. Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. *Journal of Multivariate Analysis*, 115(C):217–251, 2013.

[12] M. Calder and R. Davis. Inference for linear processes with stable noise. In *A practical guide to heavy tails*, pages 159–176. Birkhauser Boston Inc., 1998.

[13] P. Carr, H. Geman, D. B. Madan, and M. Yor. The fine structure of asset returns: an empirical investigation. *The Journal of Business*, 75(2):305–333, 2002.

[14] M. Carrasco, M. Chernov, J.-P. Florens, and E. Ghysels. Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2):529–573, 2007.

[15] M. Carrasco and J.-P. Florens. Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16(06):797–834, 2000.

[16] M. Carrasco and J.-P. Florens. Efficient GMM estimation using the empirical characteristic function. Working Paper; Department of Economics: University of Rochester. `http://neeo.univ-tlse1.fr/919/1/efficientgmm.pdf`, 2002. [Online; accessed 22-February-2014].

[17] H.-F. Chen and L. Guo. *Identification and stochastic adaptive control*. Springer, 1991.

[18] P. Colaneri. Dwell time analysis of deterministic and stochastic switched systems. *European Journal of Control*, 15(3):228–248, 2009.

[19] R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Boca Raton, 2004.

[20] C. Desoer. Slowly varying system $\dot{x} = a(t)x$. *IEEE Transactions on Automatic Control*, 14(6):780–781, 1969.

[21] L. Devroye. An automatic method for generating random variates with a given characteristic function. *SIAM Journal on Applied Mathematics*, 46(4):698–719, 1986.

[22] D. Djereveckii and A. Fradkov. Applied theory of discrete adaptive control systems. *Nauka, Moscow*, 1981.

[23] T. V. Do. Multi-server Markov queueing models: computational algorithms and ICT applications. Dissertation for the Doctor degree of the Hungarian Academy of Sciences. `http://real-d.mtak.hu/270/4/DoVanTien_5_Mu.pdf`, 2011. [Online; accessed 22-February-2014].

[24] E. Eberlein and E. A. v. Hammerstein. Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes. In *Seminar on Stochastic Analysis, Random Fields and Applications IV*, pages 221–264. Springer, 2004.

[25] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[26] A. Feuerverger. An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics*, 18(2):155–161, 1990.

[27] A. Feuerverger and P. McDunnough. On some Fourier methods for inference. *Journal of the American Statistical Association*, 76(374):379–387, 1981.

[28] A. Feuerverger and P. McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society. Series B*, 43:20–27, 1981.

[29] C. Francq and S. Meintanis. Fourier–type estimation of the power GARCH model with stable–Paretian innovations. `http://mpra.ub.uni-muenchen.de/41667/1/MPRA_paper_41667.pdf`, 2012. [Online; accessed 22-February-2014].

[30] L. Gerencsér. On a class of mixing processes. *Stochastics*, 26(3):165–191, 1989.

[31] L. Gerencsér. On the martingale approximation of the estimation error of ARMA parameters. *Systems & Control Letters*, 15(5):417–423, 1990.

[32] L. Gerencsér. A representation theorem for the error of recursive estimators. *SIAM Journal on Control and Optimization*, 44(6):2123–2188, 2006.

[33] L. Gerencsér and M. Mánfay. Identification of finite dimensional linear stochastic systems driven by Lévy processes. In *Proceedings of 12th European Control Conference, ECC 2013*, pages 2415–2420, 2013.

[34] L. Gerencsér and M. Mánfay. Empirical characteristic function identification of linear stochastic systems with possibly unstable zeros. *Submitted to the 13th European Control Conference, ECC 2014, Strasbourg, France*, 2014.

[35] L. Gerencsér and M. Mánfay. Identification of finite dimensional linear systems driven by Lévy processes. *Submitted to Automatica*, 2014.

[36] L. Gerencsér, G. Michaletzky, and Z. Reppa. A two-step maximum-likelihood identification of non-Gaussian systems. In *Proceedings of the 15th IFAC World Congress, Barcelona*, volume 15, pages 157–157, 2002.

[37] L. Gerencsér and Z. Orlovits. Real time estimation of stochastic volatility processes. *Annals of Operations Research*, 200(1):223–246, 2012.

[38] L. Gerencsér and V. Prokaj. Recursive identification of continuous-time linear stochastic systems-convergence w.p. 1 and in L_q. In *Proceedings of the 10th Europen Control Conference, ECC 2009*, pages 1209–1214, 2009.

[39] L. Gerencsér and V. Prokaj. Stability of a class of hybrid linear stochastic systems. *IEEE Transactions on Automatic Control*, 55(5):1233–1238, 2010.

[40] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, pages 1029–1054, 1982.

[41] A. Hirsa and D. B. Madan. Pricing american options under variance gamma. *Journal of Computational Finance*, 7(2):63–80, 2004.

[42] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. Springer-Verlag Berlin, 1987.

[43] B. Klar, F. Lindner, and S. G. Meintanis. Specification tests for the error distribution in GARCH models. *Computational Statistics & Data Analysis*, 56(11):3587–3598, 2012.

[44] J. L. Knight and J. Yu. Empirical characteristic function in time series estimation. *Econometric Theory*, 18(03):691–721, 2002.

[45] L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 22(4):551–575, 1977.

[46] L. Ljung. *System identification: theory for the user*. Prentice Hall, 1999.

[47] L. Ljung and T. Söderström. *Theory and practice of recursive identification*. The MIT Press, 1983.

[48] D. B. Madan, P. P. Carr, and E. C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2(1):79–105, 1998.

[49] D. B. Madan and E. Seneta. The variance gamma (VG) model for share market returns. *Journal of Business*, 63(4):511–524, 1990.

[50] B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[51] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144, 1976.

[52] Y. Miyahara and A. A. Novikov. Geometric Lévy process pricing model. *Proceedings of the Steklov Mathematical Institute*, 237:185–200, 2002.

[53] Z. Orlovits. *Staistical Analysis of Stochastic Volatility Models*. PhD thesis, Eötvös Lóránd University, 2011.

[54] E. Parzen. Statistical inference on time series by RKHS methods. In *12th Biennial Seminar Canadian Mathematical Congress Proceedings. Canadian Mathematical Society, Montreal*, 1970.

[55] J. Poirot and P. Tankov. Monte Carlo option pricing for tempered stable CGMY processes. *Asia-Pacific Financial Markets*, 13(4):327–344, 2006.

[56] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer, 1999.

[57] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.

[58] F. Ruggeri, R. Kenett, and F. W. Faltin. *Encyclopedia of statistics in quality and reliability*. John Wiley, 2007.

[59] G. Saridis. Intelligent robotic control. *IEEE Transactions on Automatic Control*, 28(5):547–557, 1983.

[60] K.-I. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

[61] E. Schlemm and R. Stelzer. Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electronic Journal of Statistics*, 6:2185–2234, 2012.

[62] G. Terdik and W. A. Woyczyński. Rosiński measures for tempered stable and related Ornstein-Uhlenbeck processes. *Probability and Mathematical Statistics*, 26(2):213, 2006.

[63] D. Xu. Continuous empirical characteristic function estimation of GARCH models. `http://economics.uwaterloo.ca/documents/12-004DX.pdf`, 2012. [Online; accessed 22-February-2014].

[64] G. G. Yin and H. J. Kushner. *Stochastic approximation and recursive algorithms and applications*. Springer, 2003.

[65] J. Yu. Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23(2):93–123, 2004.

[66] L. Zhang. H_∞ estimation for discrete-time piecewise homogeneous Markov jump linear systems. *Automatica*, 45(11):2570–2576, 2009.